



Article

# Identification and Analysis of Financial Technology Risk Factors Based on Textual Risk Disclosures

Lu Wei, Yuqi Deng, Jie Huang, Chen Han and Zhongbo Jing \*

School of Management Science and Engineering, Central University of Finance and Economics,  
Beijing 102206, China; weilu@cufe.edu.cn (L.W.); 2019311520@email.cufe.edu.cn (Y.D.);  
2019310446@email.cufe.edu.cn (J.H.); 2019311579@email.cufe.edu.cn (C.H.)

\* Correspondence: zbjing@cufe.edu.cn

**Abstract:** With the development of financial technology (referred to as fintech), the risks faced by fintech companies have received increasing attention. This paper uses the Sentence Latent Dirichlet Allocation (Sent-LDA) topic model to comprehensively identify risk factors in the fintech industry based on textual risk factors disclosed in Form 10-K. Furthermore, this paper analyzes the importance of risk factors and the similarities of the risk factors for the whole fintech industry and different fintech sub-sectors from the perspectives of risk factor types and risk factor contents. In the empirical analysis, 53,452 risk factor headings of 34 fintech companies included in the KBW Nasdaq Financial Technology Index (KFTX) over the period 2015–2019 are collected. The empirical results show that 20 risk factors of the fintech industry are identified. However, the important risk factors vary differently among different fintech sub-sectors. For the analysis of risk factor similarity, mean values of similarity of risk factor types and the similarity of risk factor contents both increased from 2015 to 2019, which indicates that the risks faced by fintech companies are becoming increasingly similar. The mean value of similarity of risk factor contents is 42.13%, while the mean value of similarity of risk factor types is 80.93%. Thus, although the types of risk factors faced by different fintech companies are similar, the contents of risk factors disclosed by different companies are still quite different. The comprehensive identification of fintech risk factors lays an important foundation for the further measurement and management of risks in the fintech industry. In the future, we will further make effective risk estimations of the fintech industry based on the identified fintech risk factors.

**Keywords:** fintech; textual analysis; risk factor; Sent-LDA; risk importance; risk similarity



**Citation:** Wei, L.; Deng, Y.; Huang, J.; Han, C.; Jing, Z. Identification and Analysis of Financial Technology Risk Factors Based on Textual Risk Disclosures. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 590–612.  
<https://doi.org/10.3390/jtaer17020031>

Academic Editor: Eduardo  
Álvarez-Miranda

Received: 10 March 2022

Accepted: 25 April 2022

Published: 26 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

At present, the development of global fintech is unstoppable [1–3]. A large amount of capital has been poured into the fintech field, which has caused the rapid growth of the fintech industry and has a huge impact on the high-quality development of the world economy [4]. According to the report “The Pulse of Fintech H1 2021” released by Klynveld Peat Marwick Goerdeler (KPMG), one of the four major international accounting firms, the global financial technology investment reached 98 billion USD in the first half of 2021, and financial technology has become the engine of the development of the global financial industry. However, with the rapid development of fintech, it also faces many potential risks [5,6], such as information system vulnerabilities, customer data, and privacy leaks [7–9]. In the process of applying advanced technology, if criminals take advantage of its loopholes to make huge profits, it will cause incalculable losses to society. Thus, studying the risks faced by the fintech industry is significant for the industry's subsequent risk measurement and management.

Generally, corporations are exposed to various risks [10]. Understanding these risks is essential for measuring corporate risks [11]. Many studies have engaged in identifying the risks faced by corporations. For instance, Mirakur [12] manually classified 29 risk

types for more than 100 companies. Grundke [13] and Bellini [14] made contributions to identifying bank risks. Thus, risk identification has become an important basis for corporate risk management, and the comprehensive identification of risk sources is of the utmost importance for explaining fintech corporate risks. The accuracy of risk assessment and the management of fintech corporations depend on comprehensively identifying the correct risk factors [15].

The current research works about fintech risks are mainly divided into two categories: one type of literature is that in which researchers summarize the risks or challenges faced by fintech by analyzing the characteristics of fintech [16–20]; another type of literature is based on the specific business of fintech companies making risk judgments. Specifically, Lee and Shin [21] analyzed that fintech start-ups had to deal with both financial and regulatory risks. Vives [22] found that fintech has the potential to disrupt established financial intermediaries as new business models based on the use of big data emerge. Systemic issues arising from operational risk and cyber risk will intensify as fintech activities are carried out. Fintech credit is growing rapidly while the relevant regulatory development is insufficient, which can lead to regulatory risks [23].

Thus, some qualitative studies on the risks and challenges of fintech have appeared in recent years. However, there is no unified understanding of the risks faced by fintech companies in the academic circle. Furthermore, the identification of risk sources of corporations is difficult and complex [11,24]. Previous studies identifying fintech risks mainly depended on the researchers' adjustments. However, as the fintech industry becomes increasingly complex, depending on researchers' adjustments is infeasible for comprehensively and accurately identifying fintech corporate risks. Therefore, a comprehensive discovery of fintech risk factors is essential for managing the risks faced by fintech corporations.

Since 2005, the US Securities and Exchange Commission (SEC) has required listed companies to add a separate part 1A to their Form 10-K to disclose "the most important factors that make the offering speculative or risky" [25]. The disclosure text of the listed company provided us with a detailed description of the risks faced by the company, and some kinds of literature [26–28] have also shown that the risk factor disclosures in the Form 10-K were highly related to the basic situation of the company. Therefore, there have been a great number of studies into identifying the risks faced by enterprises from the risk factor part of financial statements [29].

Specifically, Wei et al. [30] used a semi-supervised text mining algorithm to identify bank risk factors from financial statements. Some studies identified the risk factors of energy companies by analyzing risk disclosure texts from Form 10-K [31], and Li et al. [32] further measured the risk dependence between energy companies. In addition, many researchers have analyzed the risk disclosure texts of corporations to discover the risks faced by corporations [33–35]. Among them, Bao and Datta [34] especially proposed an unsupervised Sentence Latent Dirichlet Allocation (Sent-LDA) method to comprehensively discover and quantify the risk types from text risk disclosure. Thus, analyzing the text risk disclosure of the fintech industry by using the text mining method is a feasible way to identify the risk factors of fintech companies. However, to the best of our knowledge, there is no work that applies text analysis to the textual risk disclosures of the fintech industry currently.

The development of fintech in the US has always been at the leading level in the world. At present, the business contents of the fintech industry in the US have been expanded to many aspects, such as payment, big data analysis, trading platforms, rating agencies, software service, online loan platforms, online banks, and information consulting [36]. As the first stock market in the world to adopt electronic trading, the Nasdaq has now become one of the world's largest securities trading markets, and its indexes have extensive influence. KBW Nasdaq Financial Technology Index (KFTX) is the fintech index announced by KBW Investment Bank, Stifel Financial Corporation, and Nasdaq jointly in 2016, which included influential fintech companies—i.e., visa, lending club, etc.—mainly engaged in big data, exchange, transaction, and payment. The announcement of KFTX aims to accurately

track the performance of companies that use high technology to issue financial products and services. According to KBW, these companies account for 18% of the U.S. financial sector and have a market value of \$785 billion. Thus, it provides a reliable research sample for us to study the risk sources of the fintech industry and the key risk factors in different sub-industries.

Therefore, this paper proposes a new perspective for comprehensively identifying the risk factors of fintech corporations by first introducing the textual mining approach of Sent-LDA to analyse textual risk disclosures of the fintech industry. Based on the typical fintech companies included in the KFTX, this paper comprehensively identifies the risks faced by fintech companies from the textual risk factor part of financial statements based on 53,452 sentences in 169 Form 10-K filings of 34 fintech companies over the period 2015–2019. Furthermore, this paper analyzes the importance of fintech risk factors, which can help bank managers and regulators to focus on these important risks. Besides, this paper also studies the difference in risk factors among fintech companies of the whole fintech industry and different fintech sub-industries by analyzing the similarity of risk factor types and risk factor contents. Based on the identification results of fintech risk factors in this paper, researchers can comprehensively and effectively select risk factors when measuring the risks of the entire fintech industry or fintech subsectors. Hence, our identified fintech risk factors constitute fundamental support for further fintech corporate risk measurement.

The structure of this paper is as follows. In Section 2, we introduce our approach. Section 3 explains the data collection and description. In Section 4, we give the empirical results based on sample companies. Finally, we provide some concluding remarks in Section 5.

## 2. Methods

In this paper, we use Sent-LDA to comprehensively identify the risk factors of fintech companies. Then, we use a commonly used indicator to measure risk factor importance [30,31] and two widely used methods to measure the similarity of risk factor types and similarity of risk factor contents to analyze the similarity of the risks faced by the fintech industry [32,37]. Next, we introduce these three methods in detail.

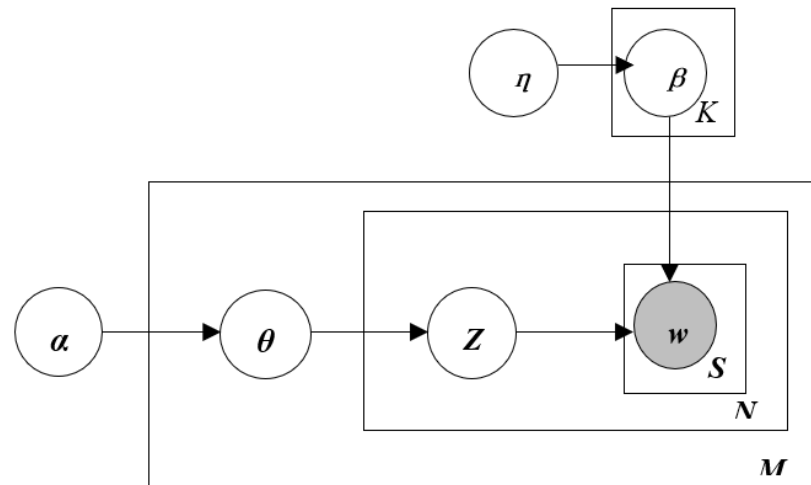
### 2.1. Sent-LDA Topic Model

This paper uses the topic model [34] to identify risk factors faced by fintech companies from the textual risk factor disclosed in Form 10-K reports. As a commonly used topic model, the main idea of LDA is that each document is a mixture of multiple topics, and each word has its corresponding topic. However, for short texts, a sentence may only express one topic. The assumption of Sent-LDA is that a sentence in the document comes from a topic. The risk disclosure part of the fintech company's 10-K report is only a short text, and one sentence expresses only one topic. Thus, previous studies proposed and used the Sent-LDA model to identify risk factors from textual risk disclosures [30,31,34].

Specifically, let  $K$ ,  $M$ ,  $S$ ,  $N$ , and  $V$  denote the number of topics, the number of documents, the number of sentences in the document, the number of words in the document, and the size of the vocabulary in a corpus, respectively.  $\text{Dirichlet}(\cdot)$  and  $\text{Multinomial}(\cdot)$ , respectively, represent the Dirichlet distribution and multinomial distribution with parameters.  $\beta_k$  is the  $v$ -dimensional word distribution of topic  $k$ , while  $\theta_d$  is the  $K$ -dimensional topic ratio of document  $d$ . In addition,  $\alpha$  and  $\eta$  represent the hyperparameters of the corresponding Dirichlet distribution, and  $w$  is a list of words in a sentence  $s$ . The graphical representation of the Sent-LDA model is shown in Figure 1, and the specific generation process is as follows:

- (1) For each topic  $k \in \{1, 2, \dots, K\}$ , draw a Dirichlet distribution on the vocabulary words  $\beta_k \sim \text{Dirichlet}(\eta)$ ;
- (2) For each document  $d$ , draw a Dirichlet distribution over topics  $\theta_d \sim \text{Dirichlet}(\alpha)$ ;
- (3) For each sentence  $s$  in document  $d$ , draw a topic distribution  $z_{d,s} \sim \text{Multinomial}(\theta_d)$ ;

- (4) For each word  $w_{d,s,n}$  in sentence  $s$ , draw a word multinomial distribution  $w_{d,s,n} \sim \text{Multinomial}(\beta_{z,d,s})$ .



**Figure 1.** The graphical representation of the Sent-LDA model.

According to Bao and Datta [34], who proposed the Sent-LDA model, when using the model, the key problem we should solve is to calculate the posterior distribution of the hidden variables  $\theta$  (topic proportion) and  $z$  (topic distribution) given the model parameters and the set of words  $w$  observed from the sentence:

$$p(\theta, z | w, \alpha, \beta) = \frac{p(\theta, z, w | \alpha, \beta)}{p(w)} \quad (1)$$

Based on the above Sent-LDA model, we can comprehensively identify risk factor topics and the assignment of sentences in these identified topics. These topics clustered through the Sent-LDA model represent the risk factors of fintech companies.

## 2.2. Method of Measuring the Similarity between Risk Factor Types

In recent years, cosine similarity has often been used to measure text similarity [38,39]. Specifically, it uses the cosine value of the angle between two vectors to reflect the similarity between the two vectors [38]. Since it can be applied to any number of dimensions, this similarity measure has been applied in many studies [39]. This paper uses the cosine similarity between risk factor vectors to measure the similarity of risk factor types between two companies.

Based on the risk factors identified by Sent-LDA and the risk factor distribution of the sentences in the text disclosure, we can construct a risk factor vector for each corporation [32]. Formally, let  $A$  represent the total number of identified risk factors.  $I$  is the total number of companies, while  $T$  is the number of all sample years. For company  $i \in \{1, 2, \dots, I\}$  in year  $t \in \{1, 2, \dots, T\}$ , we construct an  $A$  vector, denoted as  $W_{i,t}$ . The value of the vector  $W_{i,t}$  only includes 1 or 0, which indicates whether the risk factors disclosed by the company in year  $t$  include the particular risk factors  $a \in \{1, 2, \dots, A\}$ , as in the following Expression (2):

$$W_{i,t,a} = \begin{cases} 1, & \text{if the risk factor disclosure of corporation } i \text{ in year } t \text{ contains risk factor } a; \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

where  $W_{i,t,a}$  stands for the  $a$ th element in the risk factor vector  $W_{i,t}$ . As a result, the company's textual risk factors have been converted into digital vectors, which could be used for further risk similarity measures.

Let  $RS_{i,j,t}$  represent the risk similarity of enterprise  $i$  and  $j$  in year  $t$ . This can be calculated as Formula (3):

$$RS_{i,j,t} = \cos \omega_{i,j,t} = \frac{W_{i,t}^T W_{j,t}}{|W_{i,t}| |W_{j,t}|} \quad (3)$$

where  $\omega_{i,j,t}$  represents the angle between these two vectors, and  $\cos \omega_{i,j,t}$  is the cosine of this angle.

According to the above method, the more common risk factors appear in the two vectors, the greater the degree of risk similarity between the two companies, and the higher the value of  $RS_{i,j,t}$ . If the risk factor types of two companies are the same, their risk similarity is 1. In contrast, if the risk factor types of the two companies are completely different, the risk similarity value of the two companies is 0.

### 2.3. Method of Measuring the Similarity between the Risk Factor Contents

The commonly used method to measure the similarity between textual contents is to calculate the vocabulary similarity score of two texts [37]. In this paper, we also calculate the vocabulary similarity score of risk factor contents disclosed in Form 10-K of two companies to respect the similarity of the two companies' risk factor contents. To be specific, firstly we extract the vocabulary included in the risk disclosure section of each company's Form 10-K reports to summarize the vocabulary used in the risk disclosure part of all companies in the sample and construct a total phrase vector  $P$ . The length of the vector  $P$  is equal to the number of unique words used in the risk disclosure section of all companies' Form 10-K reports. To focus on the risk, we remove common words including articles, conjunctions, personal pronouns, and abbreviations from the stop words list. For the risk description vocabulary, for a given company  $i$ , a vector  $X_i$  can be constructed. The constructed vector is a binary vector. Let the length of each vector be  $L$  (that is, there are  $L$  components). Use each component of the vector  $X_i$  to compare each component of the total vector  $P$  in turn. If a company  $i$  uses the word given in  $P$  in its risk disclosure to describe the risk it faces, then fill 1 for this component in  $X_i$ ; if not used, fill this value with 0. Then, unitize each vector to get the following expression:

$$V_i = \frac{X_i}{\sqrt{X_i \cdot X_i}} \quad \forall i, j \quad (4)$$

To obtain the risk similarity between the two companies in the industry, we use the vectors  $V_i$  and  $V_j$  to represent a pair of companies  $i$  and  $j$  and calculate the cosine similarity of the two companies' risk factor contents (or the company's pairwise similarity), as follows:

$$\text{Content Cosine Similarity}_{i,j} = (V_i \cdot V_j) \quad (5)$$

Since fintech companies use a large number of words to describe the risks they face in Form 10-K, and  $V_i, V_j$  are unitized vectors, the value of  $\text{Content Cosine Similarity}_{i,j}$  is an unrestricted real number in the interval  $[0, 1]$ . Intuitively, when companies  $i$  and  $j$  use more of the same words to describe the risks they face, the calculated cosine similarity is higher and closer to 1.

## 3. Data

To analyze the risks faced by the fintech industry, we select the fintech companies included in KFTX. KFTX is the fintech index announced by KBW Investment Bank, Stifel Financial Corporation, and Nasdaq jointly in 2016, which includes 49 fintech companies—i.e., visa, lending club, etc.—mainly engaged in big data, exchange, transaction, and payment [40]. According to KBW, these companies account for 18% of the U.S. financial sector and have a market value of \$785 billion. Thus, the announcement of KFTX aims to



accurately track the performance of companies that use high technology to issue financial products and services.

Therefore, to analyze the fintech industry, we collect the textual risk factor disclosures reported in item 1A of the Form 10-K reports of all 49 fintech companies included in the KFTX. Each textual risk disclosure includes headings and the following descriptions. According to the EDGAR database on the US Securities and Exchange Commission website, some fintech companies did not disclose risk factors in their 10-K form. By removing these companies, we finally obtain a total of 34 companies as sample companies. Since the Form 10-K filings as of 2019 are publicly disclosed, the risk disclosure sentences collected in this paper are all from the disclosures of sample companies in the EDGAR database from 2015 to 2019, and the description of the sample risk factor sentences for the corporations is shown in Table A1 of Appendix A. Our data set contains a total of 169 Form 10-K documents and 53,452 sentences describing risk factors from 34 fintech companies from 2015 to 2019. Table 1 shows five examples of sentences describing risk factors in our data set.

**Table 1.** The examples of risk disclosure sentences in Form 10-K.

No.	Examples
1	Rising interest rates, credit market dislocations, decisions and actions by credit rating agencies can affect the availability and cost of our funding and adversely affect our net income.
2	Our success is dependent, in part, upon our executive officers and other key personnel, and the loss of key personnel could materially adversely affect our business.
3	The European Union continues to face great economic uncertainty which could impact the overall world economy or various other regional economies.
4	These tax authorities may challenge our positions or apply existing laws and regulations more broadly, which may potentially result in a significant increase in liabilities for taxes and interest in excess of accrued liabilities.
5	If we are not able to protect our intellectual property, our revenue and profitability could be negatively affected.

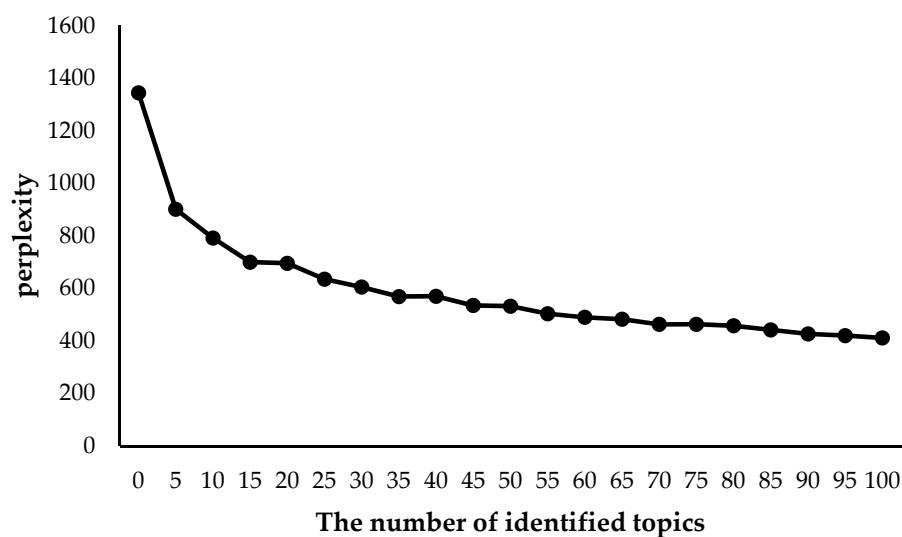
## 4. Empirical Results

### 4.1. Identification Results of Fintech Risk Factors

#### 4.1.1. The Identified Risk Factors of the Whole Fintech Industry

After collecting textual risk factor headings, we perform data preprocessing on the original texts. This process includes word segmentation and the removal of stop words [10,14]. First, we segment risk factor headings into words. Then, we use stop words list matching to remove the names of companies, some special symbols, punctuation marks, modal verbs, and some common words such as “risk” and “company” that are not representative of fintech risks.

After completing data preprocessing, the remaining words contained in the textual risk disclosures can be classified into different topics of fintech risk factors by using the Sent-LDA model. The number of topics is determined based on the accuracy of the classification, which is called perplexity [34]. The lower the perplexity, the higher the accuracy of the model. Blei and Lafferty [41] pointed out that the perplexities of the models were obtained by 10-fold cross-validation. Figure 2 shows the perplexities of the topic number in the range of [0, 100]. We can see from Figure 2 that when the number of topics grows to more than 30, the perplexity tends to stabilize. Then, by manually comparing the classification results of Sent-LDA with topic numbers in the range of [0, 100], we find that Sent-LDA provides the clearest classification when the number of topics is 45. Thus, the topic number of the risk factor types of this paper is set at 45.



**Figure 2.** The perplexity of the Sent-LDA model with a topic number in the range [0, 100].

Bao and Datta [34] used the manual labeling procedure designed by Huang and Li [33] to label different topics of risk factors. Thus, we also use the manual labeling procedure to label these 45 risk factor topics. By manually labeling the clustered 45 topics, it is found that several topics obtained from the Sent-LDA model are very similar, such as “transaction” and “trading”, “merger risk” and “acquisition risk”, etc. After merging these similar topics, 20 unique risk factors are finally obtained. For textual topic classification, each topic is usually visualized using word clouds [30,31,34]. The word cloud diagram of each topic intuitively shows which words are clustered under the topic and the frequency of these words in the topic. A larger size of the word in the word cloud indicates that the word appears many times under the topic and has a high frequency. Thus, we here visualize these 20 risk factors with the word clouds in Appendix B, where the font size of the word in Figure A1 corresponds to the probability of the word occurring in the risk factor topic.

#### 4.1.2. The Identified Risk Factors of Fintech Sub-Industry

The fintech industry is divided into eight sub-industries according to the classification of fintech companies released by the Nasdaq Index by Wangdaizhijia and Yingcan Consultation. The eight fintech sub-industries include payment companies, big data analysis companies, trading platforms, rating agencies, software service companies, online loan platforms, online banks, and information consulting companies. Thus, according to the business description of the 34 companies in the sample by the Nasdaq Stock Exchange and the New York Stock Exchange, the 34 fintech companies are classified into these 8 fintech sub-industries.

Based on the risk factors identified by Sent-LDA and the risk factor distribution of the sentences in the text disclosure, we trace back to the risk factors disclosed in each sentence in each company’s risk disclosure contents, and then we can obtain the corresponding risk factors of fintech sub-sectors from 2015 to 2019.

We find that the six sub-sectors of payment, big data analysis, trading platforms, rating agencies, software services, and information consulting companies all had these 20 identified risk factors from 2015 to 2019. However, for the online bank sub-sector, there was no infringement risk from 2015 to 2019, and there was no global financial market risk, credit rating risk, and information disclosure risk from 2015 to 2017. The credit rating risk and information disclosure risk of online banks only appeared in 2018, and the global financial market risk only appeared in 2019. As for the sub-sector of the online loan platform, there was no data security risk from 2015 to 2016, which began to appear in 2017. The global financial market risk appeared from 2016 to 2018, while there was no global financial market risk in 2015 and 2019. The absence of these risks in these two sub-sectors may be

due to the fact that, among 34 fintech companies, there is only one company in each of the two sub-sectors, which leads to the sample size of a company being too small to represent the whole sub-industry.

#### 4.2. The Results of Fintech Risk Factor Importance

After discovering 20 fintech risk factors, we further analyze which risk factors are more important for the fintech industry. In this way, when measuring the risks faced by the fintech industry, it is more effective to select several important risk factors instead of putting all risk factors into the model. Thus, this paper calculates the importance of risk factors faced by fintech companies and the annual changes of risk factor importance from two aspects of the fintech industry and sub-industries.

##### 4.2.1. The Risk Factor Importance of the Whole Fintech Industry

This paper uses the method of calculating the importance of risks based on prior research [30] to calculate the importance of the risks of fintech companies. The more frequently a risk factor is disclosed, the more risk-disclosure sentences are clustered into this risk factor, and the more attention is paid by fintech companies to the risk factor, which further shows that the risk factor is more important to fintech companies. The annual disclosure frequency of a risk factor is the number of sentences clustered into the risk factor in each sample year. Thus, the annual disclosure frequency of each risk factor can reflect the importance of the risk factor, and we calculate the importance ratio by dividing the number of sentences that are clustered to the risk factor by the total number of textual risk factor sentences. Having calculated the importance ratio of each risk factor for every year, we can also calculate the total importance ratio of a risk factor in the sample period.

We use the importance ratio to measure the importance of each risk factor. The larger the importance ratio value, the higher the importance level of the risk factor to the company. The total importance ratios of all 20 fintech risk factors during 2015–2019 are summarized in Table 2, in which the 20 identified fintech risk factors are ranked from high to low values of the importance ratio. To select the most important risk factors of fintech companies, we use the cumulative importance ratio of 80% as the boundary. The cumulative importance ratio of 12 fintech risk factors—information system security risk (13.22%), product risk (9.39%), investment risk (8.93%), business risk (7.92%), legal risk (6.66%), compliance risk (6.54%), transaction payment security risk (5.98%), infringement risk (5.74%), economic and market condition risk (5.12%), capital risk (4.47%), acquisition risk (4.13%), and tax risk (3.72%)—accounts for more than 80% (81.82%). This shows that these 12 fintech risk factors are very important to the risk measurement of fintech companies, and the cumulative importance of the remaining 8 fintech risk factors only accounts for less than 20%.

**Table 2.** The importance ratios of 20 fintech risk factors from 2015 to 2019.

Risk Factor	Importance Ratio (%)	Cumulative Importance Ratio (%)
Information system security risk	13.22	13.22
Product risk	9.39	22.61
Investment risk	8.93	31.54
Business risk	7.92	39.46
Legal risk	6.66	46.12
Compliance risk	6.54	52.66
Transaction payment security risk	5.98	58.64
Infringement risk	5.74	64.38
Economic and market condition risk	5.12	69.50
Capital risk	4.47	73.97
Acquisition risk	4.13	78.10
Tax risk	3.72	81.82
Personnel risk	3.20	85.02
Data security risk	3.05	88.07



**Table 2.** *Cont.*

Risk Factor	Importance Ratio (%)	Cumulative Importance Ratio (%)
Foreign exchange risk	2.88	90.95
Credit risk	2.28	93.23
Regulatory risk in the international market	1.89	95.12
Global financial market risk	1.76	96.88
Credit rating risk	1.65	98.53
Information disclosure risk	1.47	100

In addition, we want to study whether the important risk factors of fintech companies are the same during the sample year. Thus, we calculate the annual importance ratio of 20 fintech risk factors to analyze the annual change in the importance of risk factors from 2015 to 2019. We still use the cumulative importance ratio of 80% as the boundary to select the most important annual risk factors of fintech companies. The importances of 20 risk factors from 2015 to 2019 are summarized in Table 3 for comparison.

It can be seen from Table 3 that the top 12 risk factors identified above were still the annual important risk factors for fintech companies from 2015 to 2019. Among them, the importance of information system security risk in the sample year ranked first and experienced an upward trend from 2015 (12.99%) to 2019 (14.25%). Thus, for the fintech industry, the information system security risk is the most important and needs to be paid more attention to. Besides, the importance of legal risk has increased year by year, indicating that, with the development of the fintech industry, the supervision and law systems enacted by governments for fintech companies have become more complete, resulting in fintech companies possibly facing more legal risks. The importance of product risk and investment risk shows a downward trend year by year. In addition, although the importances of foreign exchange risk, international market regulatory risk, and global financial market risk are not ranked in the top 12, these three types of risks have been increasing in recent years. The importance ratios of other risks are relatively stable within 5 years.

**Table 3.** The annual importance ranking of 20 fintech risk factors from 2015 to 2019.

	2015	2016	2017	2018	2019
Information system security risk (%)	1(12.99)	1(12.38)	1(13.31)	1(13.09)	1(14.25)
Product risk (%)	2(10.31)	2(10.09)	2(9.78)	4(8.75)	2(8.61)
Investment risk (%)	3(9.64)	4(9.03)	3(8.96)	3(8.88)	3(8.42)
Business risk (%)	4(9.19)	3(9.18)	4(8.51)	2(8.94)	4(8.06)
Compliance risk (%)	5(6.25)	5(6.39)	6(6.39)	6(6.81)	6(6.74)
Infringement risk (%)	6(5.95)	8(6.02)	8(5.63)	8(5.54)	8(5.56)
Legal risk (%)	7(5.62)	6(6.21)	5(6.63)	5(6.89)	5(7.44)
Transaction payment security risk (%)	8(5.56)	7(6.12)	7(5.85)	7(5.60)	7(5.85)
Economic and market condition risk (%)	9(5.45)	9(5.04)	9(4.78)	9(4.87)	9(5.13)
Capital risk (%)	10(4.91)	10(4.46)	10(4.30)	10(4.55)	10(4.36)
Acquisition risk (%)	11(4.40)	12(3.65)	11(4.16)	11(4.09)	11(4.00)
Tax risk (%)	12(3.52)	11(3.78)	12(4.02)	12(3.82)	12(3.53)
Personnel risk (%)	13(3.14)	13(3.03)	14(3.01)	13(3.16)	13(3.10)
Data security risk (%)	14(2.86)	14(2.95)	13(3.24)	14(2.97)	15(2.88)
Foreign exchange risk (%)	15(2.68)	15(2.94)	15(2.76)	15(2.87)	14(3.03)
Credit risk (%)	16(2.28)	16(2.32)	16(2.22)	16(2.21)	17(2.00)
Regulatory risk in the international market (%)	17(1.62)	18(1.70)	18(1.69)	18(1.75)	18(1.84)
Credit rating risk (%)	18(1.42)	19(1.56)	20(1.45)	19(1.65)	19(1.56)
Information disclosure risk (%)	19(1.34)	20(1.33)	19(1.48)	20(1.38)	20(1.34)
Global financial market risk (%)	20(0.86)	17(1.81)	17(1.84)	17(2.20)	16(2.30)

Furthermore, according to the business characteristics of fintech companies, we divide the 20 risk factors into technological risks and financial risks. The technological risks include information system security risk, data security risk, and transaction payment security risk. Then, the remaining 17 risks are classified as financial risks. As a result, the changes in the importance ratios of technological and financial risks faced by sample fintech companies from 2015 to 2019 are shown in Table 4.

**Table 4.** The annual changes in the importance ratios of technical risk and financial risk.

	2015	2016	2017	2018	2019
Technological risk (%)	21.41	21.45	22.40	21.66	22.98
Financial risk (%)	78.59	78.55	77.60	78.34	77.02

From Table 4, we can observe that the importance of the two types of risks faced by fintech companies in the sample has remained relatively stable. Financial attributes have continued to account for about 80% and technical attributes account for about 20%. From the overall perspective of the 5 years, technical risks show a slight upward trend.

#### 4.2.2. The Risk Factor Importance of Fintech Sub-Industry

In order to analyze whether there are differences in important risk factors among different sub-industries, we further calculate the importance of risk factors corresponding to each sub-industry in the sample of 5 years. The risk factor importance results of each fintech subsector are shown in Tables 5–12.

**Table 5.** The importance ratios of risks faced by the payment industry from 2015 to 2019.

Payment	Importance Ratio (%)	Cumulative Importance Ratio (%)
Information system security risk	13.29	13.29
Transaction payment security risk	12.03	25.32
Compliance risk	9.35	34.67
Business risk	8.21	42.88
Legal risk	8.15	51.03
Investment risk	7.36	58.39
Product risk	7.30	65.69
Tax risk	4.72	70.41
Infringement risk	3.97	74.38
Capital risk	3.75	78.13
Data security risk	3.63	81.76
Foreign exchange risk	2.82	84.58
Economic and market condition risk	2.64	87.22
Acquisition risk	2.45	89.67
Regulatory risk in the international market	2.29	91.96
Personnel risk	2.27	94.23
Credit risk	2.04	96.27
Global financial market risk	1.49	97.76
Credit rating risk	1.23	98.99
Information disclosure risk	1.01	100.00

For the payment industry, the five most important risks are ranked in order of importance: information system security risk, transaction payment security risk, compliance risk, business risk, and legal risk. The cumulative importance ratio of these five risks exceeds 50%. We have observed that transaction payment security risk ranks second, which is closely related to payment companies mainly conducting transaction payment business.

**Table 6.** The importance ratios of risks faced by the information consulting industry from 2015 to 2019.

Information Consulting	Importance Ratio (%)	Cumulative Importance Ratio (%)
Information system security risk	16.88	16.88
Product risk	11.37	28.25
Business risk	10.06	38.31
Investment risk	9.20	47.51
Infringement risk	7.15	54.66
Economic and market condition risk	5.78	60.44
Tax risk	5.49	65.93
Compliance risk	4.98	70.91
Legal risk	4.35	75.26
Personnel risk	4.23	79.48
Capital risk	3.88	83.36
Transaction payment security risk	3.49	86.85
Acquisition risk	3.28	90.13
Global financial market risk	2.79	92.92
Data security risk	1.99	94.91
Foreign exchange risk	1.83	96.74
Information disclosure risk	1.08	97.82
Credit risk	0.98	98.80
Regulatory risk in the international market	0.77	99.57
Credit rating risk	0.43	100.00

For information consulting companies, the most important risks are information system security risk, product risk, and business risk. The importance of information system security risk (16.88%) is much higher than other risks. This situation may be because the information consulting services provided by the industry are greatly affected by the information system, and damage to the information system may result in an inability to carry out most of the business.

**Table 7.** The importance ratios of risks faced by the big data analysis industry from 2015 to 2019.

Big Data Analysis	Importance Ratio (%)	Cumulative Importance Ratio (%)
Product risk	16.15	16.15
Information system security risk	11.14	27.29
Investment risk	9.36	36.65
Business risk	7.50	44.15
Infringement risk	7.05	51.20
Legal risk	6.01	57.21
Capital risk	5.38	62.59
Data security risk	4.84	67.43
Compliance risk	4.79	72.22
Tax risk	4.34	76.56
Acquisition risk	3.91	80.47
Personnel risk	3.35	83.82
Transaction payment security risk	3.11	86.93
Credit risk	2.49	89.42
Regulatory risk in the international market	2.31	91.73
Foreign exchange risk	2.23	93.96
Economic and market condition risk	1.88	95.84
Credit rating risk	1.69	97.53
Global financial market risk	1.64	99.17
Information disclosure risk	0.83	100.00

The most important risks faced by the big data analysis industry are product risk, information system security risk, and business risks. Big data analysis companies mainly provide data analysis services and research data analysis-related application products. Thus, the industry should focus on the above three risks in risk management.

**Table 8.** The importance ratios of risks faced by the trading platform industry from 2015 to 2019.

Trading Platform	Importance Ratio (%)	Cumulative Importance Ratio (%)
Economic and market condition risk	14.42	14.42
Information system security risk	11.60	26.02
Business risk	10.41	36.43
Acquisition risk	10.37	46.80
Investment risk	8.90	55.70
Capital risk	6.90	62.60
Product risk	6.85	69.45
Infringement risk	5.80	75.25
Legal risk	4.26	79.51
Compliance risk	3.94	83.45
Personnel risk	2.42	85.87
Data security risk	2.32	88.19
Credit rating risk	2.31	90.50
Global financial market risk	1.99	92.49
Foreign exchange risk	1.75	94.24
Transaction payment security risk	1.28	95.52
Regulatory risk in the international market	1.27	96.79
Credit risk	1.10	97.89
Information disclosure risk	1.09	98.98
Tax risk	1.02	100.00

From Table 8, we can observe that the main risks faced by the trading platform sub-industry are economic and market condition risk, information system security risk, and business risk. The trading platform is a third-party trading security assurance platform, which includes B2B trading platforms, foreign exchange trading platforms, cloud trading platforms, and so on. For B2B trading platforms, the important risk they face is information system security risk. Once the information system is damaged, the company will suffer a major blow. Foreign exchange trading platforms are greatly affected by the economic environment. Therefore, they face larger economic and market condition risks.

**Table 9.** The importance ratios of risks faced by the rating agency industry from 2015 to 2019.

Rating Agency	Importance Ratio (%)	Cumulative Importance Ratio (%)
Information system security risk	18.58	18.58
Compliance risk	14.29	32.87
Legal risk	12.92	45.79
Product risk	8.51	54.30
Infringement risk	7.44	61.74
Business risk	5.48	67.22
Credit rating risk	5.27	72.49
Personnel risk	4.14	76.63
Transaction payment security risk	3.36	79.99
Economic and market condition risk	3.36	83.35
Data security risk	3.13	86.48
Tax risk	2.50	88.98
Global financial market risk	2.41	91.39
Foreign exchange risk	2.38	93.77
Regulatory risk in the international market	1.96	95.73

**Table 9.** *Cont.*

Rating Agency	Importance Ratio (%)	Cumulative Importance Ratio (%)
Capital risk	1.94	97.67
Acquisition risk	0.89	98.56
Credit risk	0.51	99.07
Information disclosure risk	0.51	99.58
Investment risk	0.42	100.00

It can be seen from Table 9 that the most important risks for rating agencies are information system security risk, compliance risk, and legal risk. The importance ratios of these three risks are close to 50%. The rating agency is an organization that is composed of specialized economic, legal, and financial experts who conduct rating ratings of securities issuers and securities credit, and their operations are subject to relatively large restrictions by laws and regulations. Therefore, rating agencies face greater compliance risk and legal risk.

**Table 10.** The importance ratios of risks faced by the software service industry from 2015 to 2019.

Software Service	Importance Ratio (%)	Cumulative Importance Ratio (%)
Business risk	13.82	13.82
Information system security risk	12.95	26.77
Product risk	11.14	37.91
Investment risk	10.27	48.18
Infringement risk	9.14	57.32
Personnel risk	6.08	63.40
Legal risk	5.29	68.69
Economic and market condition risk	4.11	72.80
Credit risk	4.11	76.91
Tax risk	3.40	80.31
Global financial market risk	3.25	83.56
Transaction payment security risk	3.21	86.77
Compliance risk	2.91	89.68
Capital risk	2.86	92.54
Acquisition risk	2.27	94.81
Regulatory risk in the international market	1.65	96.46
Foreign exchange risk	1.43	97.89
Data security risk	1.17	99.06
Credit rating risk	0.64	99.70
Information disclosure risk	0.30	100.00

For the software service industry, the main risks it faces are business risk, information system security risk, product risk, investment risk, and infringement risk. Software service companies mainly provide mature software products, implementation training services, and subsequent continuous improvement services. Thus, based on the launch of its main business, the industry will face greater business risk, information system security risk, and product risk.

**Table 11.** The importance ratios of risks faced by the online bank industry from 2015 to 2019.

Online Bank	Importance Ratio (%)	Cumulative Importance Ratio (%)
Foreign exchange risk	19.54	19.54
Investment risk	17.92	37.46
Information system security risk	16.13	53.59
Business risk	10.41	64.00
Legal risk	5.12	69.12
Compliance risk	4.52	73.64

**Table 11.** *Cont.*

Online Bank	Importance Ratio (%)	Cumulative Importance Ratio (%)
Capital risk	4.52	78.16
Personnel risk	4.35	82.51
Product risk	3.58	86.09
Transaction payment security risk	2.73	88.82
Credit risk	2.56	91.38
Data security risk	2.30	93.68
Tax risk	1.78	95.46
Acquisition risk	1.71	97.17
Economic and market condition risk	1.62	98.79
Regulatory risk in the international market	0.60	99.39
Credit rating risk	0.26	99.65
Information disclosure risk	0.26	99.91
Global financial market risk	0.09	100.00
Infringement risk	0.00	100.00

The main risks faced by the online bank are foreign exchange risk, investment risk, and information system security risk. Among them, the importance of foreign exchange risk accounts for close to 20%, ranking first. This is also in line with the industry characteristics of the online bank. The main businesses of the online bank are deposit and withdrawal businesses, so its operations are greatly affected by exchange rates.

From Table 12, we observe that the main risks faced by online loan platforms are investment risk, information disclosure risk, credit risk, and product risk. Among them, investment risk accounts for 23.80%, ranking first. Online loan platforms mainly provide two types of products: investments and loans. Therefore, this sub-sector will suffer greater losses due to improper investment and financial management or breach of the agreement by both borrowers and lenders.

**Table 12.** The importance of risks faced by the online loan platform industry from 2015 to 2019.

Online Loan Platform	Importance Ratio (%)	Cumulative Importance Ratio (%)
Investment risk	23.80	23.80
Information disclosure risk	10.45	34.25
Credit risk	10.18	44.43
Product risk	7.75	52.18
Legal risk	7.36	59.54
Foreign exchange risk	6.97	66.51
Information system security risk	5.83	72.34
Infringement risk	5.60	77.94
Capital risk	5.48	83.42
Business risk	4.58	88.00
Acquisition risk	3.13	91.13
Tax risk	1.60	92.73
Transaction payment security risk	1.56	94.29
Economic and market condition risk	1.37	95.66
Personnel risk	1.25	96.91
Compliance risk	1.09	98.00
Regulatory risk in the international market	0.94	98.94
Credit rating risk	0.59	99.53
Data security risk	0.35	98.88
Global financial market risk	0.12	100.00

#### 4.3. The Risks Similarities of Fintech Corporations

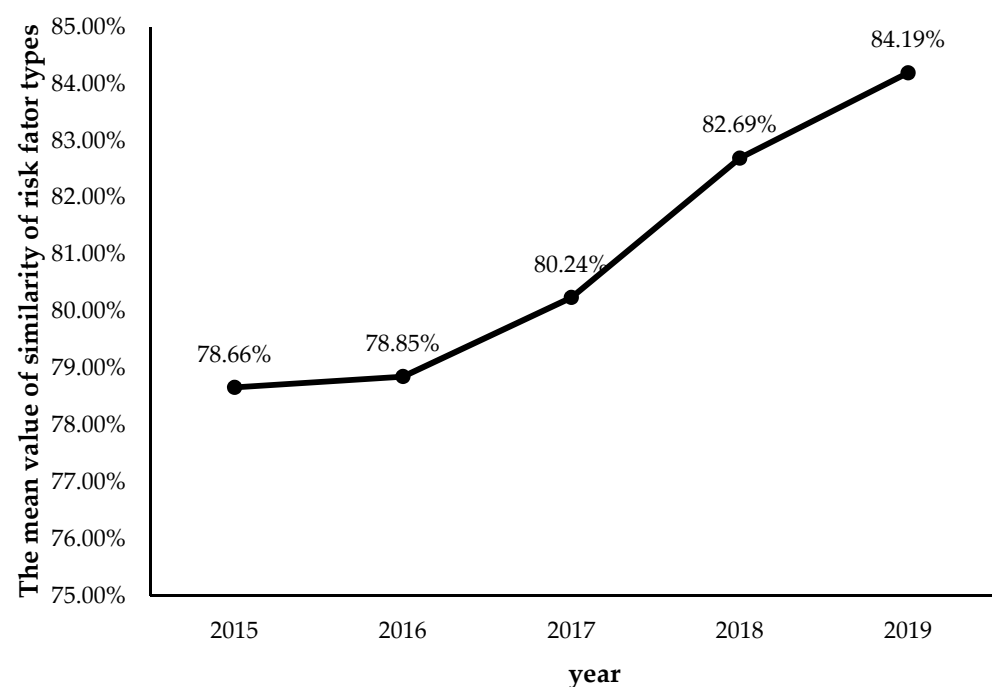
##### 4.3.1. The Similarities of Risk Factor Types of Fintech Corporations

In this subsection, we present the similarity results of risk factor types and risk factor contents of the 34 fintech corporations from 2015 to 2019. Furthermore, since there is only



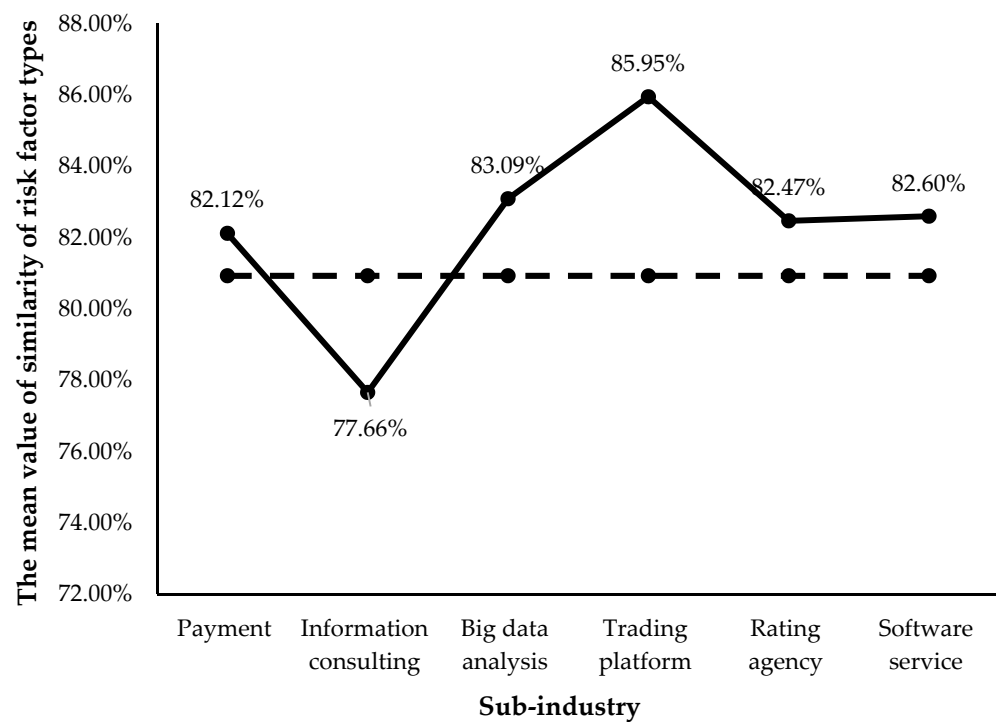
one company for each subsector of online banks and online loan platforms, we further analyze the risk similarity of the other six fintech sub-sectors except for online banks and online loan platforms.

Regarding the similarity of risk factor types, the more common the risk factors appear in two companies, the higher the similarity of risk factor types between the two companies. By calculating the similarity of risk factor types between every two fintech companies, we obtain the change of the mean of similarity of risk factor types across the fintech industry from 2015 to 2019, as shown in Figure 3. From Figure 3, the mean value of similarity of risk factor type of the fintech industry has increased year by year during the sample years, from 2015 (78.66%) to 2019 (84.19%). The five-year mean value of the similarity of fintech risk factor types is 80.93%, indicating that the similarities of the risk factor types of fintech companies are relatively high, which further reflects that the types of risk factors faced by fintech companies are relatively similar.



**Figure 3.** Changes in the mean value of the similarity of risk factor types of the fintech industry from 2015 to 2019.

Furthermore, we calculate the mean value of similarity of risk factor types between companies within each fintech sub-industry during the 5 sample years. Since there is only one company in the online bank and online loan platform categories in the sample, the similarity of these two sub-sectors is not calculated. The results are compared with the mean value of similarity of risk factor types of the whole fintech industry, which is represented by a dotted line in Figure 4.



**Figure 4.** The mean value of similarity of risk factor types within different fintech subsectors.

From Figure 4, we observe that, except for information consulting companies, the average values of similarity of risk factor types for the other five sub-sectors are higher than that of the whole fintech industry. This indicates that for most fintech sec-sectors, companies belonging to the same sub-industry have a higher similarity of risk factor type. In other words, compared with the whole fintech industry, the types of common risks faced by companies in sub-industries are more consistent. Particularly, the trading platform has the highest average value of similarity of risk factor types (85.95%) while the lowest is for the subsector of information consulting (77.66%). The possible reason is that, as a knowledge-based industry, an information consulting company uses various information processing technologies to collect, process, organize, and analyze many types of information and provide customers with information products such as solutions and strategies to problems. Companies in this industry have different sources of information, a large amount of data, and complex analysis methods and processes. Thus, the types of risks faced by individual information consulting companies may vary.

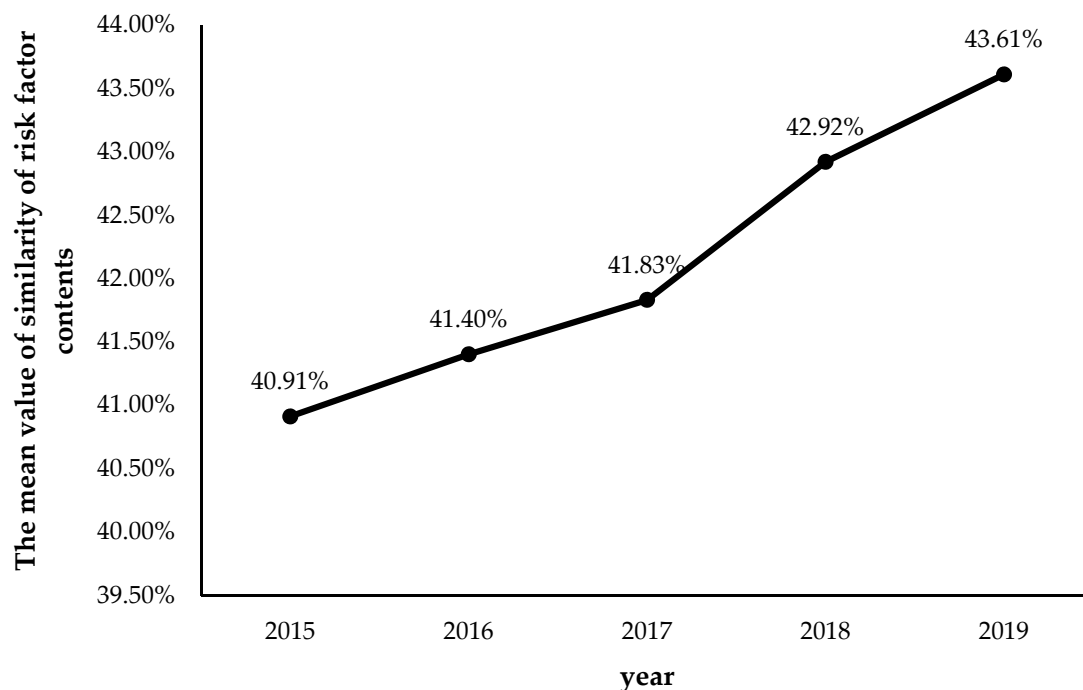
Furthermore, we calculate the mean value of the similarity of risk factor types of each sub-industry from 2015 to 2019, which is shown in Table 13. Then, we observe the annual changes in the similarity of risk factor types of the fintech sub-industry. From Table 13, it can be seen that the mean value of similarity of risk factor types for all the six fintech subsectors experienced an upward tendency from 2015 to 2019, indicating that the similarity of risk factor types of these sub-sectors has become higher in recent years, and the risk factors faced by companies in these sub-sectors have become increasingly similar.

**Table 13.** The annual changes in the similarity of risk factors in the fintech sub-industries.

	Payment (%)	Information Consulting (%)	Big Data Analysis (%)	Trading Platform (%)	Rating Agency (%)	Software Service (%)
2015	82.33	74.62	81.05	79.83	79.80	79.84
2016	79.79	78.84	83.54	81.29	81.32	78.94
2017	82.57	69.64	83.55	89.31	82.68	76.74
2018	81.99	81.22	82.92	89.71	83.09	88.79
2019	83.91	83.99	84.37	89.59	85.46	88.67

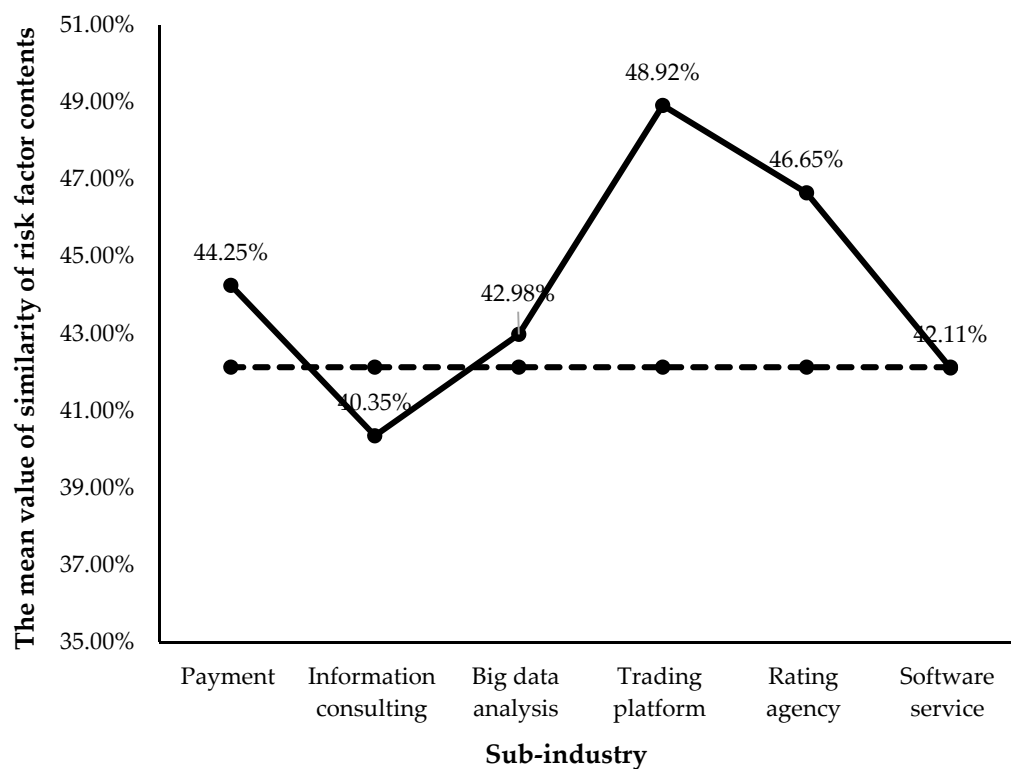
#### 4.3.2. The Similarities of Risk Factor Contents of Fintech Corporations

Fintech companies use a wide vocabulary to describe the risks they face in the 10-K form; when two companies use much of the same words to describe the risks they face, the calculated cosine similarity will be higher, which indicates that the two companies have a higher similarity of risk factor contents. Figure 5 shows the change of the average value of similarity of risk factor contents of the fintech industry from 2015 to 2019.

**Figure 5.** The change of the mean value of the similarity of risk factor contents from 2015 to 2019.

From Figure 5, it is clear that the mean value of the similarity of risk factor content across the fintech industry has also increased year by year from 2015 (40.91%) to 2019 (43.61%), with the five-year mean value being 42.13%. Thus, in general, there are not many common words used by different fintech companies to describe the risks they faced. Although the types of risk factors faced by different fintech companies are of high consistency, with the mean value of similarity of risk factor types being 80.93%, the words used to describe the common risk factor types are not very similar. There are still great differences in the specific description content on risk factor types among different fintech companies.

In addition, we also use the method of content similarity to measure the similarity of the five-year risk factor contents of six fintech sub-sectors and calculate the average value. The mean value of similarity of risk factor contents among different fintech subsectors over the period 2015 to 2019 is shown in Figure 6, in which the dotted line represents the average value of the similarity of risk factor content of the whole fintech industry.



**Figure 6.** The mean value of similarity of risk factor contents among different fintech subsectors over the period 2015 to 2019.

It can be seen from Figure 6 that the sub-sectors of the payment companies (44.25%), big data analysis companies (42.98%), trading platforms (48.92%), and rating agencies (46.65%) have higher average values for the similarity of risk factor contents than the average value of the similarity of the whole fintech industry (42.13%). However, the mean values of the similarity of risk factor contents of the subsectors of the information consulting companies (40.35%) and software service companies (42.11%) are lower than that of the whole fintech industry. Thus, for most of the fintech sub-industries, the companies that belong to the same fintech subsector have a higher similarity of risk factor contents.

Furthermore, the results of risk content similarity within different fintech subsectors are consistent with the results of risk factor type similarity within different fintech subsectors, as shown in Figure 4. Figure 4 also shows that the subsector of the trading platform has the highest mean value of similarity of risk factor types, while the mean value of the similarity of risk factor types is the lowest for the subsector of the information consulting company. Therefore, in the fintech industry, the risks faced by companies in the trading platform sub-industry are more similar, while the risks faced by companies in the information consulting sub-industry are less similar.

Moreover, we calculate the average similarities of the risk disclosure contents of each sub-industry from 2015 to 2019 and observe the annual changes in the similarity of the risk disclosure contents in the fintech sub-industry. The results are shown in Table 14.

**Table 14.** The annual changes in the similarity of risk factor contents in different fintech sub-industries.

	Payment (%)	Information Consulting (%)	Big Data Analysis (%)	Trading Platform (%)	Rating Agency (%)	Software Service (%)
2015	43.91	38.18	42.35	47.46	45.38	40.59
2016	43.82	39.38	43.56	48.20	45.74	40.44
2017	43.95	39.44	42.53	49.49	46.34	40.34
2018	44.66	41.62	43.39	49.54	47.84	43.53
2019	44.93	42.40	43.07	49.91	47.95	45.63

It can be seen from Table 14 that all the six fintech subsectors—payment, information consulting, trading platforms, rating agencies, big data, and software service —have also shown an upward trend from 2015 to 2019. Thus, the similarities of risk factor contents of these six sub-industries are increasing.

## 5. Conclusions

This paper comprehensively identifies the risk factors of the fintech industry for the first time. Furthermore, this paper analyzes the fintech risk factors from the perspective of risk factor importance and risk factor similarity of types and contents. Besides analyzing the risk factors of the whole industry, this paper also studies the risk factors of different fintech sub-sectors, including payment, information consulting, trading platforms, rating agencies, big data, software service, online loan platforms, and online banks. The identification of fintech risk factors can provide suggestions for the effective selection of risk factors, laying foundational support for further fintech corporate risk estimation.

In theory, through the empirical analysis based on 53,452 textual risk factor sentences disclosed in 169 Form 10-K filings of 34 fintech companies from 2015 to 2019, we identify 20 fintech risk factors. According to the order of importance from high to low, the identified 20 fintech risk factors are information system security risk, product risk, investment risk, business risk, legal risk, compliance risk, transaction payment security risk, infringement risk, economic and market condition risk, capital risk, acquisition risk, tax risk, personnel risk, data security risk, foreign exchange risk, credit risk, regulatory risk in the international market, global financial market risk, credit rating risk, and information disclosure risk.

For the similarity analysis, the similarities of risk factor types and the risk factor contents of disclosed risks both increased from 2015 to 2019, which shows that the risks faced by fintech companies are becoming increasingly similar. The mean of similarity of risk factors is higher, with an average value of 80.93%, while the average value of risk disclosure contents' similarity is only 42.13%, indicating that although the risk factor types faced by fintech companies are very similar, their descriptions of risks are still very different. Furthermore, the similarity results of different fintech subsectors show that, in general, companies belonging to the same fintech industry have a higher similarity in both risk factor type and risk factor contents.

From the perspective of practical management, with regard to the risks faced by the fintech industry that have a huge impact on the high-quality development of the world economy, this paper comprehensively identifies fintech risk factors based on textual risk disclosures, which solves the important problem of selecting risk factors for fintech corporate risk measurement. The identification results are highly significant in terms of practical applications. The identified fintech risk factors can support financial regulators and managers of fintech companies to better measure and manage risks, which has practical significance for the robust operation of the fintech industry.

This study is not without limitations. A comprehensive selection of risk factors is of the utmost importance for explaining corporate risks. The comprehensive identification of fintech risk factors lays the foundation for making effective risk estimations. Therefore, one limitation of this paper is that we have not analyzed how to use the identified fintech risk factors to measure the risks faced by fintech companies. Therefore, in future research, we

will further measure the risks of the fintech industry based on the identified fintech risk factors.

**Author Contributions:** Conceptualization, L.W. and Z.J.; methodology, Y.D.; software, J.H.; validation, C.H.; formal analysis, L.W. and Y.D.; investigation, Y.D.; resources, L.W.; data curation, Y.D. and C.H.; writing—original draft preparation, L.W. and Y.D.; writing—review and editing, L.W. and Z.J.; visualization, J.H.; supervision, Z.J.; project administration, L.W.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by grants from the National Natural Science Foundation of China (72001223).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** The description of risk factor sentences from fintech corporations' Form 10-K filings during 2015–2019.

CIK Number	Corporations' Name	Number of Sentences
0000004962	American Express Co. (AXP)	1916
0000033185	Equifax, Inc. (EFX)	1214
0000036047	CoreLogic, Inc. (CLGX)	890
0000064040	S&P Global, Inc. (SPGI)	1141
0000350894	SEI Investments Co. (SEIC)	846
0000798354	Fiserv, Inc. (FISV)	955
0000814547	Fair Isaac Corp. (FICO)	1198
0000935036	ACI Worldwide, Inc. (ACIW)	1193
0001013237	FactSet Research Systems Inc. (FDS)	620
0001059556	Moody's Corp. (MCO)	1004
0001101215	Alliance Data Systems Corp. (ADS)	1349
0001120193	Nasdaq, Inc. (NDAQ)	1942
0001136893	Fidelity National Information Services, Inc. (FIS)	1372
0001141391	MasterCard, Inc. (MA)	1746
0001156375	CME Group, Inc. (CME)	1421
0001175454	FleetCor Technologies, Inc. (FLT)	1967
0001278021	MarketAxess Holdings, Inc. (MKTX)	2270
0001299709	BofI Holding, Inc. (BOFI)	1172
0001309108	WEX, Inc. (WEX)	1974
0001337619	Envestnet, Inc. (ENV)	2439
0001365135	The Western Union Co. (WU)	2075
0001374310	CBOE Holdings, Inc. (CBOE)	1428
0001383312	Broadridge Financial Solutions, Inc. (BR)	905
0001386278	Green Dot Corp. (GDOT)	1758
0001402436	SS&C Technologies Holdings, Inc. (SSNC)	1694
0001403161	Visa, Inc. (V)	1321
0001409970	LendingClub Corp. (LC)	2555
0001442145	Verisk Analytics, Inc. (VRSK)	1339
0001559865	EVERTEC, Inc. (EVTC)	2030
0001571949	Intercontinental Exchange, Inc. (ICE)	1915
0001592386	Virtu Financial, Inc. (VIRT)	1755
0001627014	Black Knight Financial Services, Inc. (BKFS)	1677
0001633917	PayPal Holdings, Inc. (PYPL)	2324
0001598014	IHS Markit Ltd. (INFO)	2047



## Appendix B



Figure A1. The word clouds of the identified 20 fintech risk factors.

## References

1. Weichert, M. The future of payments: How FinTech players are accelerating customer-driven innovation in financial services. *J. Paym. Strategy Syst.* **2017**, *11*, 23–33.
2. Barbu, C.M.; Florea, D.L.; Dabija, D.C.; Barbu, M.C.R. Customer Experience in Fintech. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 80. [CrossRef]

3. Alt, R.; Beck, R.; Smits, M.T. FinTech and the transformation of the financial industry. *Electron. Mark. Int. J. Netw. Bus.* **2018**, *28*, 235–243. [CrossRef]
4. KPMG. Pulse of Fintech H1 2021—Global. 2021. Available online: <https://home.kpmg/xx/en/home/insights/2021/08/pulse-of-fintech-h1-2021-global.html> (accessed on 1 March 2022).
5. Sloboda, L.Y.; Demianyk, O.M. Prospects and risks of the fintech initiatives in a global banking industry. *Probl. Econ.* **2020**, *1*, 275–282. [CrossRef]
6. Chen, X.; Hu, X.; Ben, S. How individual investors react to negative events in the fintech era? Evidence from China's peer-to-peer lending market. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 52–70. [CrossRef]
7. Financial Stability Board. Financial Stability Implications from FinTech. 2017. Available online: <http://www.fsb.org/2017/06/financial-stability-implications-from-fintech/> (accessed on 10 November 2018).
8. Ng, A.W.; Kwok, B.K.B. Emergence of Fintech and cybersecurity in a global financial center: Strategic approach by a regulator. *Financ. Regul. Compliance* **2017**, *25*, 422–434. [CrossRef]
9. Gomber, P.; Koch, J.A.; Siering, M. Digital Finance and FinTech: Current research and future research directions. *J. Bus. Econ.* **2017**, *87*, 537–580. [CrossRef]
10. Sanusi, M.S.; Ahmad, F. Modelling oil and gas stock returns using multi factor asset pricing model including oil price exposure. *Financ.* **2016**, *18*, 89–99. [CrossRef]
11. Bianconi, M.; Yoshino, J.A. Risk factors and value at risk in publicly traded companies of the nonrenewable energy sector. *Energy Econ.* **2014**, *45*, 19–32. [CrossRef]
12. Mirakur, Y. *Risk Disclosure in SEC Corporate Filings*; Working Paper; University of Pennsylvania: Philadelphia, PA, USA, 2011; Available online: [http://repository.upenn.edu/wharton\\_research\\_scholars/85](http://repository.upenn.edu/wharton_research_scholars/85) (accessed on 1 March 2022).
13. Grundke, P. Top-down approaches for integrated risk management: How accurate are they? *Eur. J. Oper. Res.* **2010**, *203*, 662–672. [CrossRef]
14. Bellini, T. Integrated bank risk modeling: A bottom-up statistical framework. *Eur. J. Oper. Res.* **2013**, *230*, 385–398. [CrossRef]
15. Boyer, M.M.; Fillion, D. Common and fundamental factors in stock returns of Canadian oil and gas companies. *Energy Econ.* **2007**, *29*, 428–453. [CrossRef]
16. Chuen, D.K.; Teo, E.G.S. Emergence of FinTech and the LASIC principles. *J. Financ. Perspect.* **2015**, *3*, 1.
17. Philippon, T. The FinTech Opportunity. NBER Working Papers. 2016. Available online: [https://www.nber.org/system/files/working\\_papers/w22476/w22476.pdf](https://www.nber.org/system/files/working_papers/w22476/w22476.pdf) (accessed on 1 March 2022).
18. Gai, K.; Qiu, M.; Sun, X. A survey on Fintech. *J. Netw. Comput. Appl.* **2018**, *103*, 262–273. [CrossRef]
19. Gomber, P.; Kauffman, R.J.; Parker, C.; Weber, B. On the Fintech Revolution: Interpreting the Forces of Innovation, Disruption and Transformation in Financial Services. *J. Manag. Inf. Syst.* **2018**, *35*, 220–265. [CrossRef]
20. Milian, E.Z.; Spinola, M.M.; Carvalho, M.M. Fintechs: A literature review and research agenda. *Electron. Commer. Res. Appl.* **2019**, *34*, 100833. [CrossRef]
21. Lee, I.; Shin, Y.J. FinTech: Ecosystem, business models, investment decisions, and challenges. *Bus. Horiz.* **2018**, *61*, 35–46. [CrossRef]
22. Vives, X. The impact of FinTech on banking. *Eur. Econ.* **2017**, *2*, 97–105.
23. Claessens, S.; Frost, J.; Turner, G.; Feng, Z. Fintech Credit Markets around the World: Size, Drivers and Policy Issues. BIS Quarterly Review 2018. Available online: [http://www.bis.org/publ/qtrpdf/r\\_qt1809e.pdf](http://www.bis.org/publ/qtrpdf/r_qt1809e.pdf) (accessed on 1 March 2022).
24. Gupta, K. Oil price shocks, competition, and oil & gas stock returns-global evidence. *Energy Econ.* **2016**, *57*, 140–153. [CrossRef]
25. Securities and Exchange Commission (SEC). Securities and Exchange Commission Final Rule, Release No. 33-8591(FR-75). 2005. Available online: <http://www.sec.gov/rules/final/33-8591.pdf> (accessed on 1 March 2022).
26. Securities and Exchange Commission (SEC). *Report on Review of Disclosure Requirements in Regulation S-K*; SEC Offices: Washington, DC, USA, 2013. Available online: <http://www.sec.gov/news/studies/2013/reg-sk-disclosure-requirements-review.pdf> (accessed on 1 March 2022).
27. Monga, V.; Chasan, E. The 109,894-word annual report; as regulators require more disclosures, 10-Ks reach epic lengths; how much is too much? *Wall Str. J.* **2015**. (Online). Published on 2 June 2015.
28. Dyer, T.; Lang, M.; Stice-Lawrence, L. The evolution of 10-K textual disclosure: Evidence from latent dirichlet allocation. *J. Account. Econ.* **2017**, *64*, 221–245. [CrossRef]
29. Campbell, J.L.; Chen, H.C.; Dhaliwal, D.S.; Lu, H.M.; Steele, L.B. The information content of mandatory risk factor disclosures in corporate filings. *Rev. Account. Stud.* **2014**, *19*, 396–455. [CrossRef]
30. Wei, L.; Li, G.; Zhu, X.; Li, J. Discovering bank risk factors from financial statements based on a new semi-supervised text mining algorithm. *Account. Financ.* **2019**, *59*, 1519–1552. [CrossRef]
31. Wei, L.; Li, G.; Zhu, X.; Sun, X.; Li, J. Developing a hierarchical system for energy corporate risk factors based on textual risk disclosures. *Energy Econ.* **2019**, *80*, 452–460. [CrossRef]
32. Li, J.; Li, J.; Zhu, X. Risk dependence between energy corporations: A text-based measurement approach. *Int. Rev. Econ. Financ.* **2020**, *68*, 33–46. [CrossRef]
33. Huang, K.W.; Li, Z.L. A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. *ACM Trans. Inf. Syst.* **2011**, *2*, 1–19. [CrossRef]

34. Bao, Y.; Datta, A. Simultaneously discovering and quantifying risk types from textual risk disclosures. *Manag. Sci.* **2014**, *60*, 1371–1391. [[CrossRef](#)]
35. Miller, G.S. Discussion of “The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation”. *J. Account. Econ.* **2017**, *64*, 246–252. [[CrossRef](#)]
36. U.S. Fintech Investigation Report. 2017. Available online: <https://www.idf.pku.edu.cn/attachments/d64ed11be7224f9d8efa8eaa69e7d09f.pdf> (accessed on 1 March 2022).
37. Hoberg, G.; Phillips, G. Text-Based Network Industries and Endogenous Product Differentiation. *J. Political Econ.* **2016**, *124*, 1423–1465. [[CrossRef](#)]
38. Wu, J.; Zhu, S.; Liu, H.; Xia, G. Cosine interesting pattern discovery. *Inf. Sci.* **2012**, *184*, 176–195. [[CrossRef](#)]
39. Nguyen, H.T.; Duong, P.H.; Cambria, E. Learning short-text semantic similarity with word embeddings and external knowledge sources. *Knowl. Based Syst.* **2019**, *182*, 104842. [[CrossRef](#)]
40. KBW Nasdaq Financial Technology Index (KFTX). 2016. Available online: <https://indexes.nasdaqomx.com/Index/Overview/KFTX> (accessed on 1 March 2022).
41. Blei, D.M.; Lafferty, J.D. A correlated topic model of science. *Ann. Appl. Stat.* **2007**, *1*, 17–35. [[CrossRef](#)]