Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 214 (2022) 280–287

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Information Technology and Quantitative Management

Bank risk aggregation based on dual perspectives of bank managers and credit rating agencies

Haozhe Jing[a], Xiyuan Miao[a], Lu Wei[a]*

[a]School of Management Science and Engineering, Central University of Finance and Economics, No. 39 Xueyuan South Road, Haidian District, Beijing 100081, China

Abstract

In this paper, the risk disclosure text information from the dual perspectives of bank managers and credit rating agencies is innovatively incorporated into the bank risk aggregation, which to some extent avoids the deviation of bank risk assessment caused by the risk perception from a single perspective, and measures the overall bank risk more comprehensively. In the empirical analysis part, this paper extracted 342 risk profit and loss data, 511 text risk information in 10-K forms and 356 available bank rating reports from 134 listed commercial banks in the United States. The Sent-LDA model was used to identify and compare risk factors from different perspectives. In addition, by constructing a total risk-adjusted index, this paper studies the effectiveness of textual risk disclosure from the dual perspectives of bank management and credit rating agencies in the overall risk measurement of commercial banks. The empirical results show that the text contains the incremental information of risk assessment, and the inclusion of double-perspective text risk information into the risk aggregation will amplify or reduce the overall risk of banks calculated only based on financial numerical data. In practice, it will be more effective to measure the overall risk of banks by considering the textual risk disclosure from dual perspectives.

Keywords: bank risk aggregation; credit rating report; text mining; risk factors; financial statements

* Corresponding author. Tel.: +86 18510628807.
E-mail address: weilu@cufe.edu.cn.

1877-0509
10.1016/j.procs.2022.11.176

1. Introduction

One of the keys to the stable operation of the financial system is bank risk management [1]. Improving the level of bank risk management is crucial for preventing financial sector crisis [2]. The main risks faced by commercial banks are credit risk, market risk and operational risk, and bank risk integration is the core of bank risk management [3].

In the research of bank risk aggregation, the method based on financial statements has been widely used because financial numerical data are intuitive and easy to obtain. Many scholars choose financial indicators as the variables of risk assessment, and link the income statement and balance sheet items with various risks [4-6] to measure the overall risk of banks through the risk profit and loss data.

At present, many studies have identified risks by analyzing textual information disclosed in financial statements [7-9]. Some studies interpret section 1A of the 10-K form of American listed companies through text mining and incorporate it into bank risk aggregation [10], which can avoid the limitation of simply using financial numerical data for risk aggregation.

However, identifying and analyzing bank risks only from the unilateral and internal perspective of managers may lead to bias in risk measurement results [11]. Credit rating agencies selectively use public information of issuers to provide ratings [12], which conveys informed signals from a neutral agency about the overall condition of the company [13]. Credit rating reports issued by credit rating agencies are another important source of textual risk disclosure of banks which can become an effective strategic complement with the textual disclosure in the annual financial statements [14].

Therefore, this paper innovatively expands the source of textual risk information from the single perspective of bank management to the dual perspectives of bank management and credit rating agencies. In the empirical analysis part, this paper extracted 342 risk profit and loss data, 511 text risk information in 10-K forms and 356 available bank rating reports from 134 listed commercial banks in the United States. The Sent-LDA model was used to identify and compare risk factors from different perspectives. In addition, by constructing a total risk- adjusted index, this paper studies the effectiveness of textual risk disclosure from the dual perspectives of bank management and credit rating agencies in the overall risk measurement of commercial banks. The experimental results shows that the aggregation of financial statements and credit rating reports into bank risk aggregation is conducive to improving the rationality and comprehensiveness of the results of risk aggregation.

The remainder of this paper is organized as follows. Section 2 mainly introduces the Sent-LDA model, the construction of integrated risk adjustment index and the bank risk aggregation method. Section 3 introduces the data used in this paper. Section 4 is the empirical analysis. Section 5 concludes this paper.

2. Methodology

2.1. Sent-LDA model

Based on the assumption of "one topic per sentence", Sent-LDA model can effectively extract the potential risk topics in the text information, and has more accurate clustering effect compared with the original LDA model [7]. In order to better mine the prominent risks faced by the banking industry from the text information, this paper uses the Sent-LDA model to find the potential topics in a series of text content, so as to identify the key risk factors. The implementation of Sent-LDA model in this paper is as follows:

- Collect the text contents of risk disclosure in 10-K statements and credit rating reports of commercial banks and summarize them into two text sets respectively.
- The text content in the text set is preprocessed (word segmentation, eliminating stopping words, etc.) and converted into text format that can be processed by the model.
- The Sent-LDA model is run to calculate the complexity to determine the optimal number of topics.
- Sent-LDA model was used for training, and sentences reflecting the same risk point were clustered into a topic, and word cloud was generated according to the word frequency in each topic.
- According to the frequency of important words and the meaning of the subject sentence, manually annotate the risk factors reflected in each topic, and the types of bank risks covered in the text content are obtained.

2.2. Construction of total risk-adjusted index

According to the study of Wei et al, the importance ratio and the negative tone value of risk factor i in year t was defined as follows[9]:

$$IR_{i,t} = \frac{F_{i,t}}{\sum_{t=1} F_{i,t}} \tag{1}$$

$$NTONE_{i,t} = \frac{NEG_{i,t}}{NEG_{i,t} + POS_{i,t}} \tag{2}$$

Where $F_{i,t}$ represents the disclosure times of risk factor i in year t, $POS_{i,t}$ and $NEG_{i,t}$ represent the word frequencies of positive and negative emotional tone words describing risk factor i in year t, respectively.

When a certain type of risk is disclosed more frequently and the tone of the text is more negative, it indicates that the sender pays more attention to this type of risk and has a more pessimistic view, which further indicates the severity of this type of risk. Therefore, define the importance tone value of bank risk type j in year t as:

$$\overline{IT_{j,t}} = \frac{Rating\_agency\_IT_{j,t} + Bank\_IT_{j,t}}{2} \tag{3}$$

Where, $IT_{j,t}$ is denoted as the mean importance of the bank risk type j in year t under the text information from the two perspectives. $Rating\_agency\_IT_{j,t}$ refers to the importance tone value of this type of risk j in year t from the perspective of credit rating agencies. $Bank\_IT_{j,t}$ refers to the importance tone value of this type of risk j in year t from the perspective of bank management.

From the dual perspectives, the total risk-adjusted index (*TRAI*) of risk type j in year t is defined as follows:

$$TRAI_{j,t} = \frac{\overline{IT_{j,t}}}{M_j} \tag{4}$$

Where, M_j represents the mean value of the importance tone value of risk type j in the sample period T under the dual perspective, which is defined as follows:

$$M_j = \frac{\sum_{t=1}^{T} \overline{IT_{j,t}}}{T} \tag{5}$$

2.3. Bank risk aggregation approach

In this paper, referring to the research methods of Zhu et al.[15] and Wei et al. [16], three types of major risks of banks are mapped into the financial statements to collect the historical profit and loss data of three types of risks. After obtaining the risk profit and loss data, the risk profit and loss is divided by risk-weighted assets (*RWA*) and converted into risk rate of return [5]. The risk rate of return is calculated as follows:

$$r_{i,j,t} = \frac{R_{i,j,t}}{RWA_{i,t}} \tag{6}$$

Where $r_{i,j,t}$ and $R_{i,j,t}$ respectively represent the risk rate of return of bank i with respect to risk j in period t and the risk profit and loss data, which are collected from the income statement by mapping different risk types into the income statement. $RWA_{i,t}$ denotes the risk-weighted assets of bank i in period t.

For reference, Li et al. [17] and Wei et al. [16] used numerical data to measure bank risk, adjusted bank risk rate of return to the sum of risk expected rate of return and random volatility, and recalculated the risk rate of return of each bank after data processing. The method is as follows.

First, the risk-expected return $\overline{r_{i,j}}$ of bank i with respect to risk j in period T is defined as:

$$\overline{r_{i,j}} = \frac{\sum_{t=1}^{T_i} r_{i,j,t}}{T_i} \tag{7}$$

The deviation of the expected rate of return on risk with respect to the original rate of return on risk $\Delta_{i,j,t}$:

$$\Delta_{i,j,t} = r_{i,j,t} - \overline{r_{i,j}} \tag{8}$$

Therefore, for a given bank k , the risk rate of return obtained by the bank is expressed as:

$$r_{k,j,t} = \overline{r_{k,j}} + \sum_i \Delta_{i,j,t} \tag{9}$$

After the bank risk rate data is obtained through the above data processing, the value at risk (VaR) is used to measure the bank risk.

This paper uses simple addition method to integrate bank risk which refers to the simple combination of different types of risk values of banks to determine the overall risk of a bank. This method assumes that the worst case of each risk type occurs at the same time, and obtains the upper limit of the total loss faced by the bank [18]. It focuses on the impact of the comprehensive text information from the dual perspective on the overall risk under the extreme case that all risk types occur at the same time.

3. Data

The data in this paper are selected from American Commercial Bank (SIC code: 6021, 6022, 6029), during the period of 2013 to 2020. After treating different branches as the same company, according to data availability, 342 risk profit and loss data, 511 text risk information in 10-K forms and 356 available bank rating reports were finally extracted from 134 listed commercial banks in the United States. In order to empirically analyze the overall risk of American commercial banks without focusing on a specific real bank, Rosenberg and Schuermann [4] and Kretzschmar et al. [6] set the research object as an imaginary American commercial bank constructed using the median method.

4. The empirical analysis

4.1. Identification results of bank risk factors

Based on the Sent-LDA model, a popular used topic model to identify risk factors [19, 20], this paper conducts text mining on the 1A part of the text risk information disclosure of financial statements of banks and the Moody's bank rating report, and identifies the risk factors faced by the banking industry from two perspectives. Among them, 28 risk factors are identified from the perspective of bank management : laws and regulations, loan loss, technology, acquisition, loss of key personnel, tax, economic conditions, operational risk, interest rate, mortgage, accounting standard, external risk event, dividends from subsidiaries, damage to reputation, stock price, information security, competition, claim and legal action, internal control, credit rating, new products and services, funding, liquidity, environmental liability, goodwill and other intangible assets, capital requirement, other institutions interaction, merger.

There are 27 kinds of risk factors identified from the perspective of credit rating agencies, which are as follows:

acquisition, asset quality, capital ratio, capitalization, core deposits, coronavirus, credit loss, debt load, deposit, earning, economic downturn, franchise, funding, government support, interest rate, liquidity, loan growth, loan loss, management, merger, mortgage, originator ability, product, real estate, service, total asset, unemployment.

With 70% as the dividing line of cumulative importance ratio, Table 1 summarizes the top 10 risk factors of cumulative importance ratio from the different perspectives from 2013 to 2020:

Table 1. Top 10 risk factors of overall importance from 2013 to 2020 from a dual perspective

Bank's Perspective		Credit rating agency perspective	
Interest rate	13.63%	Asset quality	16.65%
Laws and regulations	13.52%	Acquisition,	11.10%
Loan loss	9.12%	Capital ratio	6.77%
Information security	8.89%	Loan growth	6.58%
Economic conditions	6.63%	Deposit	6.11%
Stock price	5.49%	Interest rate	4.98%
Mortgage	4.50%	Management	4.64%
Accounting standard	3.94%	Capitalization	4.55%
Loss of key personnel	3.65%	Core deposits	4.36%
Internal control	3.56%	Loan loss	4.36%
Total	72.93%	Total	70.10%

Wei et al. [16], when studying the relationship between textual risk disclosure and bank risk aggregation, mapped risk factors into different risk types based on the definition of risk factors and bank risk types, and pointed out its feasibility and rationality. Therefore, this article uses the same mapping method, according to the above definition of credit risk, market risk and operational risk, and get the definition of the risk factors of text mining, the bank management, credit rating agencies under the two perspectives of bank risk factor and credit risk, market risk and operational risk to establish a one-to-one mapping relationship, as shown in Table 2:

Table 2. Mapping between bank risk factors and risk types

Risk types	Risk factors from the perspective of bank management	Risk factors from the perspective of credit rating agencies
Credit risk	Loan loss	Loan loss
	Mortgage	Mortgage
		Credit loss
		Loan growth
		Unemployment
Market risk	Economic conditions	Economic downturn
	Interest rate	Interest rate
	Other institutions interaction	Real estate
	Credit rating	Asset quality
	Stock price	Capitalization
Operational risk	Internal control	Management
	New products and services	Service
	External risk event	Coronavirus
	Funding	Funding
	Acquisition	Acquisition
	Merger	Merger
	Liquidity	Liquidity
	Capital requirement	Capital ratio

	Laws and regulations	Total asset
	Technology	Deposit
	Loss of key personnel	Originator ability
	Accounting standard	Government support
	Damage to reputation	Franchise
	Tax	Core deposits
	Information security	Earning
	Claim and legal action	Debt load
	Environmental liability	
	Goodwill and other intangible assets	
	Competition	
	Dividends from subsidiaries	
	Operational risk	

4.2. Bank risk aggregation results from dual perspective

Table 3 shows the historical VaR values of credit risk, market risk and operational risk from 2013 to 2020 and the overall risk (%) obtained by simple addition method under 99% confidence intervals:

Table 3. VaR value of Credit, Market, Operational Risk and Overall Risk, 2013-2020 (%)

Bank risk factors	2013	2014	2015	2016	2017	2018	2019	2020
Credit risk	-1.70	-1.93	-2.27	-1.34	-1.32	-1.27	-1.79	-1.52
Market risk	-0.028	-0.047	-0.028	-0.0022	-0.0056	-0.0059	-0.0024	-0.0035
Operational risk	-11.77	-10.54	-9.74	-11.17	-9.94	-7.13	-12.30	-9.85
Overall risk	-13.50	-12.52	-12.04	-12.51	-11.27	-8.41	-14.09	-11.37

Through the relevant formulas constructed by the above indicators, the risk adjustment coefficients of credit risk, market risk and operational risk in each year can be calculated. Table 4 presents the adjusted indices for the three risk types from 2013 to 2020:

Table 4. Total risk-adjusted index of credit risk, market risk and operational risk

Year	TRAI of credit risk	TRAI of market risk	TRAI of operational risk
2013	1.00	0.98	1.11
2014	1.08	1.03	0.89
2015	1.01	0.94	0.90
2016	1.02	0.91	0.92
2017	1.16	0.96	0.91
2018	1.13	0.97	0.88
2019	0.83	1.49	1.27
2020	0.96	1.08	1.30

The total risk-adjusted index (*TRAI*) calculated above was used to adjust the overall risk of the bank calculated based on financial numerical data from 2013 to 2020, and the overall risk of the bank after incorporating the text risk information from the dual perspective was obtained, as shown in Table 4 and figure 1:

Table 5. Overall risk and Overall risk from a dual perspective

Types	2013	2014	2015	2016	2017	2018	2019	2020
Overall risk	-13.50%	-12.52%	-12.04%	-12.51%	-11.27%	-8.41%	-14.09%	-11.37%
Overall risk from dual perspective	-14.83%	-11.48%	-11.06%	-11.67%	-10.56%	-7.71%	-17.13%	-14.26%

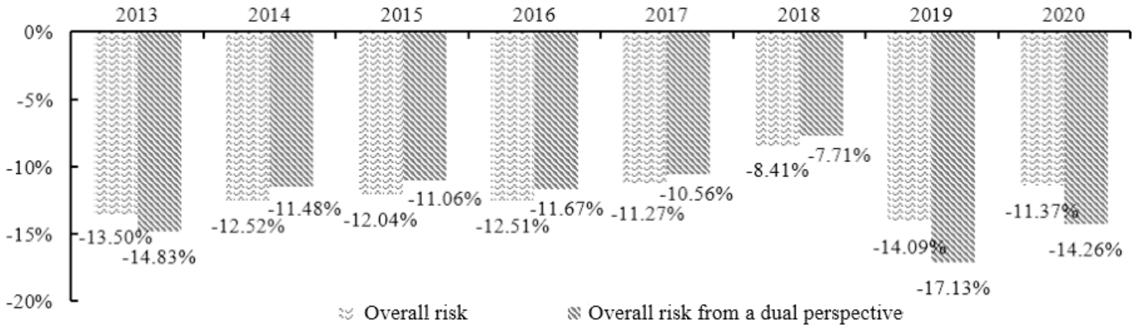

Fig. 1. Overall risk and Overall risk from a dual perspective between 2013 and 2020

From 2014 to 2018, the overall risk from a dual perspective showed a trend of decreasing volatility and reached the minimum value of -7.71% in 2018. Compared with the overall risk calculated only based on financial numerical data, the overall bank risk decreased by 1.04%, 0.98%, 0.84%, 0.71% and 0.70% after the inclusion of dual perspective text risk information from 2014 to 2018. The reason may be that the global economic recovery after the end of the financial crisis has been maintained. Banks have recovered from the severe damage of the financial crisis, replenished their capital stock and performed significantly in cost management. Although there are still some challenges in terms of revenue and profit growth, textual risk disclosures from the perspective of bank management and credit rating agencies still include positive signals about the future as interest rates recover and other new winds emerge, so incorporating them into the bank risk mix will reduce the overall risk.

However, in 2019 and 2020, the overall risk of banks showed a trend of steep rise from the dual perspective and reached the maximum value of -17.13% in 2019, which increased by 3.04% and 2.89% respectively compared with the overall risk calculated only based on financial numerical data. The possible reasons are as follows. On the one hand, flexible consumption and the increase of M&A activities in the American banking industry make the business growth of commercial banks and the stock price performance increase, so the risk aggregation based on financial numerical data underestimates the risk of banks to a certain extent. On the other hand, as a result of the turmoil the constantly changing international political and economic situation such as trade friction between China and America, global debt increase caused by the uncertainty of the external economy, including a new outbreak of the effects of extreme events, makes bankers and credit rating agencies foresee a pessimistic, negative attitudes about the future, they believe that Banks could face a more serious loss, Therefore, the inclusion of text risk disclosure containing negative signals in the bank risk aggregation will increase the overall risk of the bank.

To sum up, since text contains incremental information of risk assessment, the aggregation of double-perspective text risk information into risk aggregation will expand or reduce the overall risk of the bank. Considering the disclosure of double-perspective text risk can measure the overall risk of the bank more comprehensively and comprehensively.

5. Conclusion

In this paper, the Sent-LDA model is used to conduct text mining on the risk disclosure texts reported by banks' financial reports and credit rating agencies. Then, use the bank's risk factor ratio and negative tone value to build total risk-adjusted index to adjust the bank risk calculated based on numerical data, so as to incorporate the text risk information from both the internal perspective of bank management and the external perspective of credit rating agencies into the bank risk aggregation. The empirical analysis shows that since the text contains the incremental information of risk assessment, the aggregation of dual-perspective text risk information into the risk aggregation will amplify or reduce the overall risk of the bank calculated based on financial numerical data. In practice, aggregating textual risk information and financial numerical data from the dual perspective into the risk aggregation method can measure the overall risk of a bank more effectively. This article has certain reference significance for improving bank risk aggregation management and preventing bank crisis in the era of big data.

Acknowledgements

This research was funded by grants from the National Natural Science Foundation of China (NSFC Grant Numbers: 72001223) and the Engineering Research Center of National Financial Security, Ministry of Education.

References

[1] Baker H K, Filbeck G. (2015)"Investment risk management", in Oxford University Press
[2] Zhu X, Xie Y, Li J, Wu D. (2015) "Change point detection for subprime crisis in American banking: From the perspective of risk dependence." *International Review of Economics & Finance* **38**: 18-28.
[3] Basel Committee on Banking Supervision.(2006) " International convergence of capital measurement and capital standards: a revised framework" , in *Bank for International Settlements*
[4] Rosenberg J V, Schuermann T A. (2006) "General approach to integrated risk management with skewed, fat-tailed risks." *Journal of Financial Economics* **79(3)**: 569-614.
[5] Kuritzkes A, Schuermann T.(2007) "The known, the unknown and the unknowable in financial risk management", in Princeton University Press
[6] Kretzschmar G, McNeil A J, Kirchner A. (2010) "Integrated models of capital adequacy-why banks are undercapitalized." *Journal of Banking & Finance* **34(12)**: 2838-2850.
[7] Bao Y, Datta A. (2014) "Simultaneously discovering and quantifying risk types from textual risk disclosures." *Management Science* **60(6)**: 1371-1391.
[8] Zhu X, Yang S Y, Moazeni S.(2016) "Firm risk identification through topic analysis of textual financial disclosures." *Symposium Series on Computational Intelligence* 1-8.
[9] Wei L, Li G, Zhu X, Li J.(2019a) "Discovering bank risk factors from financial statements based on a new semi-supervised text mining algorithm." *Accounting & Finance* **59(3)**: 1519-1552.
[10] Wei L, Li G, Zhu X, Sun X, Li J.(2019b) "Developing a hierarchical system for energy corporate risk factors based on textual risk disclosures." *Energy Economics* **80**: 452-460.
[11] Wei L, Han C, Yao Y.(2022) "The Bias Analysis of Oil and Gas Companies' Credit Ratings Based on Textual Risk Disclosures." *Energies* **15(7)**: 2390.
[12] Farkas M.(2021) "Competition, communication and rating bias." *Journal of Economic Behavior and Organization* **189**: 637-656.
[13] Ryan P A, Villupuram S V, Zygo J G.(2017) "The value of credit rating changes across economic cycles." *Journal of Economics and Business* **92**: 1-9.
[14] Lee K K, Schantl S F.(2019) "Financial Reporting and Credit Ratings: On the Effects of Competition in the Rating Industry and Rating Agencies' Gatekeeper Role." *Journal of Accounting Research* **57(2)**: 545-600.
[15] Zhu X, Wei L, Wu D, Li J.(2018) "A general framework for constructing bank risk data sets." *Journal of risk* **21(1)**: 37-59.
[16] Wei L, Li G, Li J, Zhu X.(2019c) "Bank risk aggregation with forward-looking textual risk disclosures." *The North American Journal of Economics and Finance* **50**: 185-205.
[17] Li J, Wei L, Lee C F, Zhu X, Wu D.(2018) "Financial statements based bank risk aggregation." *Review of Quantitative Finance & Accounting* **50(3)**: 673-694.
[18] Dimakos X K, Aas K.(2004) "Integrated risk modelling." *Statistical modelling* **4(4)**: 265-277.
[19] Li J, Li G, Liu M, Zhu X, Wei L.(2022) A novel text-based framework for forecasting agricultural futures using massive online news headlines. *International Journal of Forecasting* **38(1)**: 35-50.
[20] Zhu X, Wang Y, Li J.(2022) What drives reputational risk? Evidence from textual risk disclosures in financial statements. *Humanities and Soccial Sciences Communications* **9:** 318.