American Journal of Industrial and Business Management, 2023, 13, 930-948
https://www.scirp.org/journal/ajibm
ISSN Online: 2164-5175
ISSN Print: 2164-5167

Bibliometric Analysis of Credit Risk Based on the Web of Science (WOS)

Jian Xue, Yixue Fan*

School of Economics and Management, Shaanxi University of Science and Technology, Xi'an, China
Email: *210711017@sust.edu.cn

How to cite this paper: Xue, J., & Fan, Y. X. (2023). Bibliometric Analysis of Credit Risk Based on the Web of Science (WOS). *American Journal of Industrial and Business Management, 13,* 930-948.
https://doi.org/10.4236/ajibm.2023.139052

Received: August 20, 2023
Accepted: September 11, 2023
Published: September 14, 2023

Abstract

To clarify the evolution of credit risk research, the article searches the literature data on credit risk research from 2007 to 2022 through the Web of Science (WOS) database and adopts the scientometric method to analyze the number of publications, the distribution of countries/regions in the field of credit risk research with the help of Co-Occurrence (Cooc) metrology software, respectively, distribution of important journals, and analysis of highly cited literature, which elaborates the current status and research hotspots of credit risk research; secondly, we analyze the distribution pattern and evolution trend by drawing the relevant knowledge maps of keyword co-occurrence network, keyword clustering analysis, author-keyword clustering, and research hotspots. It aims to summarize the achievements and shortcomings of the existing literature, clarify the future research direction, and promote the development of corporate credit research.

Keywords

Credit Risk, Bibliometric Study, COOC Analysis, Visualization Analysis

1. Introduction

Credit risk has been a research area of great interest in financial risk management, especially when we face a globalized economy and changing financial markets (Augustin et al., 2009; Liu et al., 2022a). With the rapid development of financial innovations and the interdependence of the global economy, financial institutions and investors are generally exposed to credit risks, such as debt defaults and bond defaults (Puliga et al., 2014), the existence of which may lead to financial market turmoil, the bankruptcy of financial institutions and an overall economic recession (Yfanti et al., 2022). Identifying, assessing, and managing credit risk is essential to ensure the stability and sustainability of financial institutions and the economic system as a whole. The study of credit risk has achieved extensive research and study over the past few decades. Scholars have proposed different definitions and assessment models and explored various types of risk management methods from different disciplinary perspectives, such as economics, finance, and risk management. The purpose of this paper is to summarize and analyze the current state of research as well as the future direction of development through a review of credit risk-related literature to provide financial institutions and practitioners with an in-depth understanding of credit risk and more effective credit risk management methods.

In terms of definition, credit risk can be viewed as the potential risk that a debtor will not be able to meet its payment obligations as contractually required (Abedin et al., 2022). From an economic perspective, credit risk can affect prices and liquidity in bond and loan markets (Merton, 1973). Meanwhile, the definition of credit risk also encompasses a macroeconomics perspective focusing on financial stability to understand the potential impact of credit risk on the economic system as a whole (Freixas & Rochet, 1997). Understanding the definition and characterization of credit risk helps us to better understand its nature and influencing factors.

Scholars have proposed various assessment models and methods to assess and manage credit risk. Credit risk assessment models are important tools widely used in finance to help financial institutions accurately assess the creditworthiness of borrowers to make decisions and manage risks. Traditional assessment models include discriminant analysis (Mahmoudi & Duman, 2015), logistic regression (Wang et al., 2015), and decision trees (Liu et al., 2022b), among others. For example, Silva et al. (2020) applied a logistic regression model to credit score data from a Portuguese financial institution to assess the risk of default on consumer loans. It was found that the model correctly predicted 89.79% of defaults. Jagric et al. (2012) developed a credit assessment model using a learning vector quantization (LVQ) neural network and compared the model with a logistic regression model. Finally, an empirical study was conducted using data from Slovenian banks. The results show that the LVQ neural network model is accurate: it outperforms the logistic regression model and leads to more accurate assessment results. These traditional models typically rely on human-defined features and rules to build predictive models to assess credit risk by training and learning from sample data. However, traditional models are limited by factors such as feature selection and model complexity and often fail to capture complex patterns and non-linear relationships in the data. With the rapid development of technology and data, emerging models based on machine learning and artificial intelligence have emerged. Models such as support vector machine (SVM) (Harris, 2013), random forest (Wang, 2022a), deep learning (Zhang et al., 2020), and neural networks (Li & Fu, 2023; Wang, 2022b) have been shown to have higher predictive ability and flexibility in credit risk assessment. For example, Li & Fu (2023) proposed a credit risk prediction model based on PCA-GA-SVM for supply chain finance. The running results show that the model has good generalization ability and can provide a reference for commercial banks to improve the credit risk management ability of Supply Chain Finance. Researchers have tried to combine and integrate traditional and deep learning models to improve the accuracy and robustness of credit risk assessment. For example, Zhong & Wang (2023) propose a new deep learning credit scoring model based on deep forest (DF) and random under-sampling. The results show that the RUS-DF model obtains better classification performance and stability than other models and is suitable for solving the credit scoring problem with imbalanced data. In conclusion, credit risk assessment models play a key role in financial risk management. From traditional models based on discriminant analysis to models based on neural networks, emerging new methods, and technologies provide new ideas and tools to improve the accuracy and efficiency of credit assessment. With the increase of data and the improvement of computing power, credit risk assessment models will be further developed and improved to better meet the needs of financial institutions.

Credit risk management is an important area in business and financial institutions, which have adopted a variety of approaches to reduce and control risk. Traditional approaches typically focus on the use of historical data and statistical models to assess and manage credit risk, including credit rating systems (Moon et al., 2011), diversified portfolios (Consiglio et al., 2018), and risk diversification (Uchiyama et al., 2019). Traditional credit risk management methods play an important role in assessing and managing credit risk but have limitations in adapting to the challenges of an increasingly complex financial environment and new types of data. With the development of technology and the application of new types of data, new types of credit risk management methods are emerging, such as big data analytics (Niu et al., 2020), blockchain technology (Wang, 2021), and artificial intelligence (Xu et al., 2019), which also provide new opportunities for risk management. The introduction of new methods can assess credit risk more comprehensively and accurately and provide more effective risk management strategies. As technology continues to advance, credit risk management will continue to see innovation and development.

To summarize, credit risk research has achieved many results, but there are still challenges and problems to be solved. This paper will bring together and integrate research results from various academic fields and provide an in-depth analysis and assessment of credit risk to identify the research hotspots and future trends in the field. To this end, this paper employs bibliometrics to quantitatively as well as qualitatively analyze credit risk. By comprehensively analyzing and synthesizing the field of credit risk research, our goal is to provide references for future research, as well as to provide financial institutions and investors with feasible credit risk management frameworks and strategies to further enhance the stability and development of the financial market.

The main contribution of this paper is to analyze the field of credit risk research using bibliometric methods, through systematically sorting out the development history and trends of credit risk research, research concerns and hotspots in different periods and regions, identifying key journals and keywords, and discovering research hotspots and future research directions. This will help to promote the further development of credit risk research and improve the academic level and practical effect in related fields.

Bibliometric studies on credit risk have improved the in-depth understanding of credit risk by financial institutions and practitioners from a multidimensional, systems-based perspective. The rest of the paper is structured as follows: in Section 2, we describe the data sources and analytical methods used in our study. In Section 3, we sort out the literature in the field of credit risk and map the corresponding knowledge map using COOC and VOSviewer software, including the number of journal articles, the number of issuing countries/regions, core journal publications, and highly cited literature. In Section 4, we analyze the research hotspots, including keyword co-occurrence, keyword clustering, author-keyword clustering, and evolutionary path analysis. In Section 5, we present conclusions and outlook.

2. Data Sources and Analytical Methods

2.1. Data Sources

The data for this study was obtained from the Web of Science (WOS) database, which is one of the most comprehensive and authoritative databases containing more than 12,000 high-quality journals (Jiang et al., 2022; Liu et al., 2022c). In this paper, data will be obtained by searching the following databases included in WOS: Science Citation Index Expanded (SCI-E), Social Science Citation Index (SSCI), and Conference Proceedings Citation Index (CPCI). In this paper, "credit risk" was used as the subject search term, and the search period was from 2000 to 2022. To ensure the relevance and authority of the literature, we screened according to the language (English) and the type of literature (Article, Conference papers & Review), and a total of 2407 articles were retrieved.

To ensure the accuracy of the data and the reliability of the results, it is necessary to carry out a series of pre-processing of the raw data to carry out the following analysis work (Wang et al., 2023). In this paper, we use the "Data Extraction Module" of the new bibliometric software Co-Occurrence14.3 (COOC14.3) to merge and save the retrieved 2407 documents (Academic Dots & Bibliometric, 2023), and then apply the "Data Cleaning Module" to remove weight (title duplicates) and merge synonyms (e.g., keywords with similar meanings, "default probability" is merged into "probability of default"), In the final data checking, to avoid data duplication, we again merged synonyms, eliminated (not related to credit risk research, such as climate change and other literature) and other standardized screening and finishing, resulting in a total of 2080 pieces of literature on credit risk.

2.2. Analyzing Methods

This paper uses bibliometrics and scientific knowledge mapping methods to analyze the progress of credit risk research based on searching the literature. Bibliometrics includes structured representation, dynamic description, evaluation, prediction, and scientometrics, so the analytical methods used need to be systematized. Available software includes COOC, VOSviewer, CiteSpace, Bibexcel, and Bicomb. This paper focuses on COOC and VOSviewer, each with different features. COOC is the most complete and relatively simple-to-operate software in the bibliometrics field. COOC allows for the rapid construction of relational matrices and the immediate export of matrix results such as co-occurrence and dissimilarity matrices. More importantly, it can also pre-process the data, such as batch merging synonyms and removing unnecessary keywords. Currently, COOC does not allow citation analysis, so this paper combines it with VOSviewer, developed in collaboration with Nees Jan van Eck and Ludo Waltman of Leiden University in the Netherlands, which is a bibliometric scientometrics and visualization and analysis software for constructing and viewing bibliometric mappings, which is based on the principle of co-citation and co-citation of literature, and performs the keyword co-occurrence analysis of the bibliometric data to get the clustering mappings, and it is an important tool for efficiently viewing, mining and analyzing the issues in the research field (van Eck & Waltman, 2010).

This paper adopts the research method of bibliometric analysis to analyze the keyword co-occurrence of credit risk-related literature, draw keyword co-occurrence mapping, keyword clustering mapping, and keyword evolution paths, to explore the hot spots and development trends of credit risk research, provide the basis for credit risk management for enterprises, and provide a strong guarantee for the stable operation of the market economy.

3. Literature Characterization

3.1. Trend Analysis of Publications

The number of published papers is an important indicator of the development of an academic field. Changes in the number of publications directly reflect changes in the amount of scientific knowledge in the field, as the growth of knowledge is closely related to the number of journal articles (Hao et al., 2021; Zeng et al., 2021; Fahimnia et al., 2015). **Figure 1** shows a graph of annual and cumulative trends in the number of publications for 2080 documents related to credit risk. As can be seen in **Figure 1**, the trend of the number of publications in the field of credit risk is increasing, with only 33 papers published in 2007, and the number of papers increasing to 245 in 2022, with an average annual growth rate of 22.66%. During this period, the big event subprime mortgage crisis and the new crown pneumonia epidemic occurred, credit risk has gradually been paid attention to, especially the evaluation and control of corporate credit risk on the international financial stability of the importance of the more and more obvious. Using statistical principles to fit the power function of the cumulative number of publications on credit risk, the $R^2 = 0.9942$ of the fitted curve is obtained, which indicates that the cumulative number of publications in the field of credit risk is growing in a power function type. This indicates that research on credit risk has received widespread attention from scholars and the heat continues to rise.

3.2. Country/Region Analysis

The article uses COOC software, combined with relevant econometric statistical methods to visualize and analyze the top 10 countries with the highest number of publications in the field of credit risk research from 2007 to 2022, as shown in Figure 2. The international financial crisis hit the world economy in 2008, which triggered a wide range of impacts, and scholars deeply realized that credit risk, once out of control, would lead to a serious crisis, so the academic research on

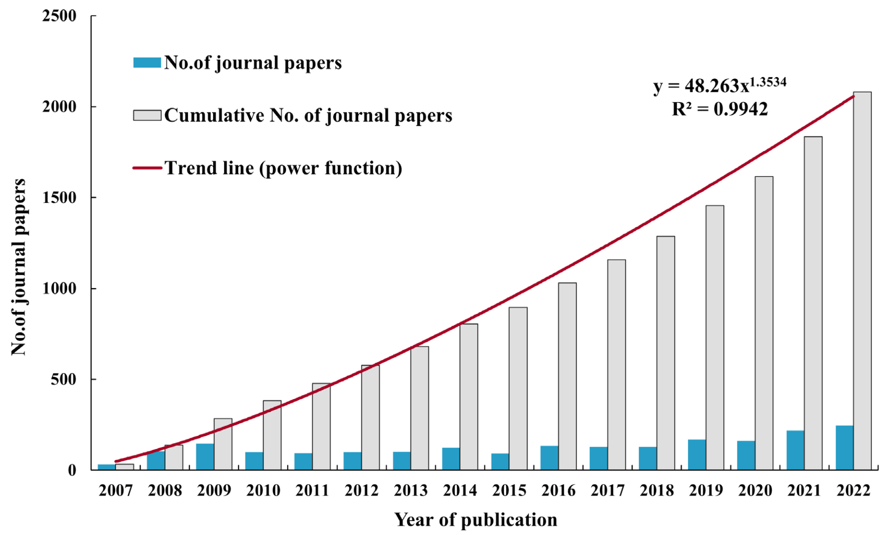


Figure 1. The number of papers per year and the cumulative number of papers on credit risk (Excel).

Figure 2. Top 10 countries with the highest number of publications from 2007 to 2022 (COOC).

credit risk has gradually increased. The articles published in 2007-2010 are mainly from China, the USA and Germany, etc. Romania and Spain are also involved in the publication of papers in this field, which shows that there is a global concern about this issue. With the abnormal stock market shock in 2015 and the outbreak and spread of the new Crown pneumonia epidemic in 2020, related literature continued to emerge in 2021, with a total of 2080 articles published as of December 31, 2022, mainly from China, USA, England, and Italy. At the same time, there is also cooperation between different countries on this topic, which is most likely due to the economic downturn, epidemic impact, and other effects, the international bond market also frequently defaults (Morelli et al., 2022), and scholars in various countries have a deeper knowledge of credit risk, which has caused great concern in the scientific community about credit risk in different countries.

3.3. Journal Publication Analysis

After counting 2080 articles, it was found that 857 journals published credit risk, of which 726 journals published 2 or fewer. According to Bradford's law, journals are ranked in decreasing order according to the number of papers they publish in a particular subject specialization, and they can be classified into core, relevant, and fringe zones (Bradford, 1985; Vickery, 1948). The core zone can be calculated based on Bradford's law: $N = 2\ln\left(e^{E} \cdot P\right)$, where N is the number of core zones, E is the Euler coefficient (0.5772), and P is the number of articles carried by the highest-volume journal. In this paper, $N = 2\ln\left(1.781 \times 97\right) = 10.30$, that is, the top 10 journals are in the core area, as shown in **Figure 3**.

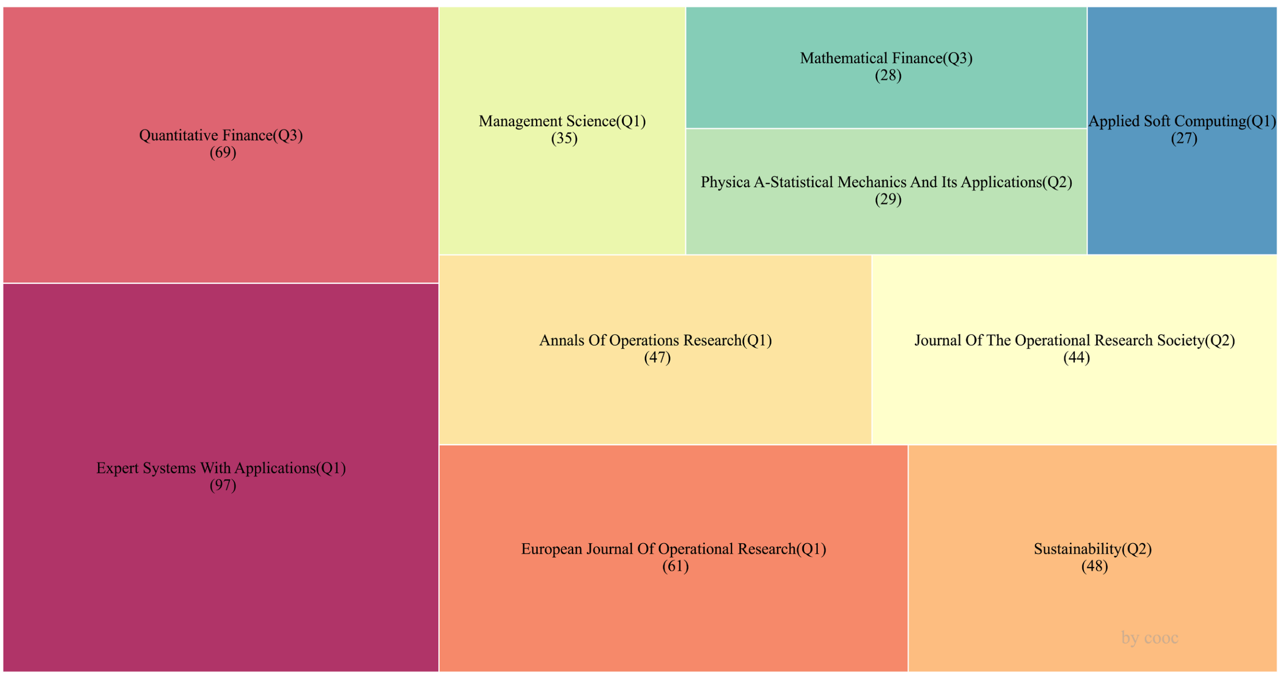


Figure 3. Treemap of the core area published journals (COOC).

Figure 3 shows the 485 journal papers included in the core journals, re-presenting 23.3% of the total 2080 journal papers retrieved. Most of the JCR partitions of these journal papers are Q1 and Q2, which represent the top journals and high-level journals in each field, respectively. It can be proved that scholars' research on credit risk has a certain depth, and the research in this field has reached the stage of prosperous development. Research on credit risk focuses on five areas: Artificial Intelligence, Risk Analysis, Business & Economics, Environmental Sciences & Ecology, Operations Research & Management Science. Among the core journals, Expert Systems With Applications is ranked first in terms of volume, with a total of 97 articles published in the field of credit risk from 2007-2022, accounting for 4.7% of the total number of papers in the journal.

3.4. Analysis of Highly Cited Literature

By analyzing the highly cited literature in the field of credit risk, it is possible to see the academic papers and major research scholars that have made a significant impact in this field and then to deeply analyze the seminal literature and theoretical foundations of the field. According to Table 1, the most cited article is Diebold & Yilmaz's "On the Network Topology of Variance Decompositions:

Table 1. Analysis of highly cited literature on credit risk.

Rank	Year	Title	Journal	First Author	Citations
1	2014	On the network topology of variance decompositions: measuring the connectedness of financial firms	Journal of econometrics	Francis X. Diebold	1530
2	2014	Evaluation of clustering algorithms for financial risk analysis using MCDM methods	Information sciences	GangKou	589
3	2014	Hazardous times for monetary policy: what do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking?	Econometrica	Gabriel Jimenez	437
4	2017	Machine learning models and bankruptcy prediction	Expert systems with applications	Flavio Barboza	288
5	2014	Genetic algorithm-based heuristic for feature selection in credit risk assessment	Expert systems with applications	Stjepan Oreski	263
6	2007	Recent developments in consumer credit risk assessment	European journal of operational research	Jonathan N. Crook	245
7	2018	Trade credit, risk sharing, and inventory financing portfolios	Management science	S. Alex Yang	237
8	2016	A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification	Expert systems with applications	Aytug Onan	214
9	2008	Credit risk assessment with a multistage neural network ensemble learning approach	Expert systems with applications	Lean Yu	201
10	2010	Affine point processes and portfolio credit risk	Siam journal on financial mathematics	Eymen Errais	193

Measuring The Connectedness Of Financial Firms" published in the Journal of Econometrics. They developed and apply a unified framework for conceptualizing and empirically measuring connectedness at a variety of levels, from pairwise through systemwide, using variance decompositions from the approximating model (Diebold & Yilmaz, 2014). The second is "Evaluation Of Clustering Algorithms For Financial Risk Analysis Using MCDM Methods" by Kou et al. published in Information Sciences in 2014. They proposed an MCDM-based approach for clustering algorithms evaluation in the domain of financial risk analysis (Kou et al., 2014). The third most cited is "Hazardous Times for monetary policy: what do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking?" by Jimenez et al. published in Econometrica in 2014. They identify the effects of monetary policy on credit risk-taking with an exhaustive credit register of loan applications and contracts. They find that a lower overnight interest rate induces lowly capitalized banks to grant more loan applications to ex-ante risky firms and to commit larger loan volumes with fewer collateral requirements to these firms, yet with a higher ex-post likelihood of default (Jimenez et al., 2014).

4. Research Hot Spots

4.1. Visual Analysis of Keyword Co-Occurrence

Keyword co-occurrence analysis allows us to describe the core content and structure of the field of credit risk, and at the same time indicates the research frontiers of the field. The visualization function of VOSviewer can do an in-depth analysis of keywords in a large number of kinds of literature, and through the interpretation of the keyword co-occurrence map, we can learn about the history and the development of the current situation of the field of credit risk research. In this paper, with the help of VOSviewer software, keywords with a frequency of 10 or more are visualized to identify the research hotspots in the field. **Figure 4** shows the keyword co-occurrence mapping of journal papers from 2007-2022.

With the turbulent international situation in recent years, the world economy is also characterized by uncertainty and instability. Enterprises in the development process will have different risks, but the credit risk is especially obvious and prominent, and the impact on the enterprise is also substantial and most serious. The size of the displayed nodes in **Figure 4** indicates the frequency of the keyword's occurrence in the article, with more frequency indicating greater importance of the research topic characterized by the keyword. In the research on credit risk, the keywords with the highest frequency are credit risk (672), credit scoring (140), machine learning (126), credit risk assessment (116), and probability of default (64), reflecting the fact that the topic of evaluating credit risk is studied more in research in the field of credit risk. While the frequency of keywords such as model, empirical, and data is also relatively high, it can be further speculated that empirical research on credit evaluation is favored in this research area.

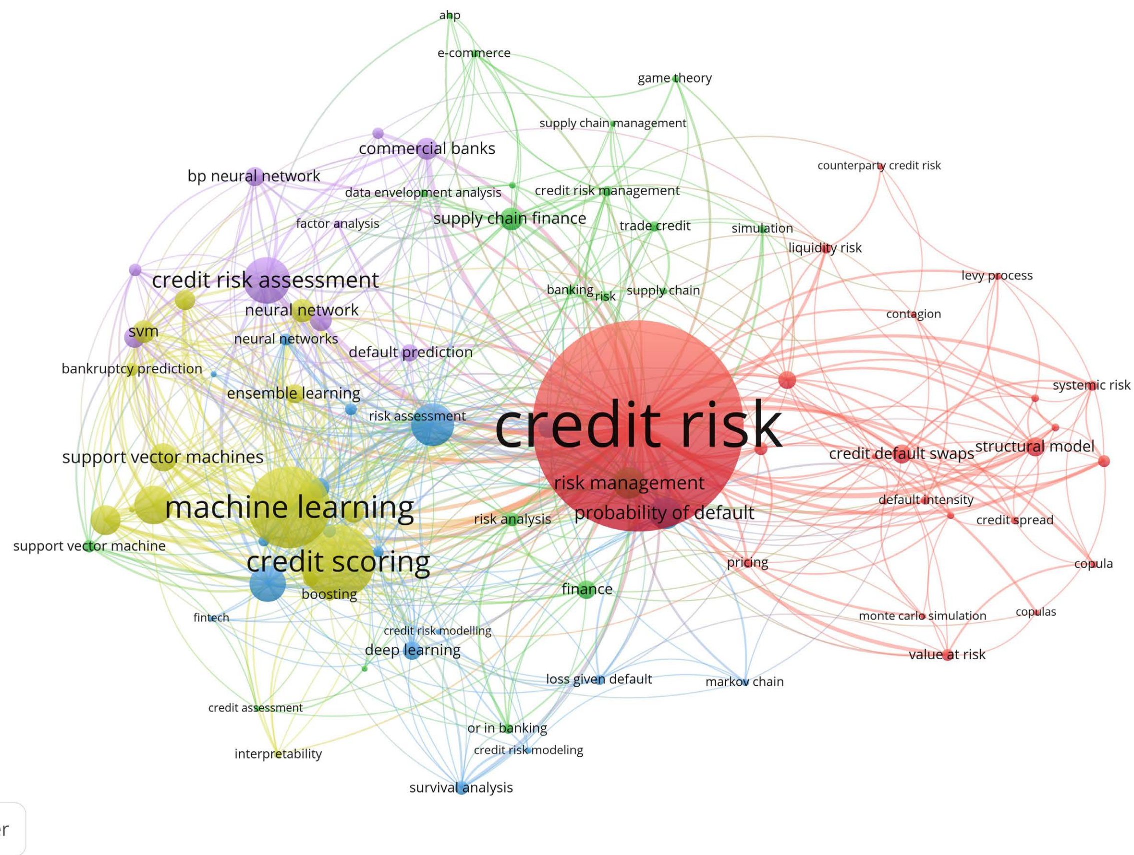


Figure 4. Co-occurrence mapping of keywords in the field of credit risk research (VOSviewer).

Overall, the research in this field is mainly concerned with the quantification and evaluation of credit risk, which, as one of the important components of financial market risk, is measured in a way that is always emphasized by market participants. Investors, financial institutions, and regulators tend to be more precise and dynamic in the measurement of credit risk due to the demand for risk control. Credit risk measurement models play the most critical role in this regard, and their scientificity and accuracy have become the fundamental factors of risk measurement results. Scholars are also trying to improve the limitations and deficiencies of traditional models to produce more accurate quantitative forecasts.

4.2. Keyword Clustering Analysis

Keyword co-occurrence mapping (**Figure 4**) can only show the keywords with the highest frequency in the papers in the field of credit risk research and the co-linear relationship between the keywords, which can represent the current hotspots of credit risk research to a certain extent, but it can't comprehensively reflect the relationship between them, and then it is necessary to carry out a co-word analysis. Now we choose 20 high-frequency keywords with word frequency greater than 30 as the basis of cluster analysis. The 20 high-frequency keywords in the research field of Credit risk are counted two by two to construct a 20 × 20 keyword co-occurrence matrix. To truly reveal the co-occurrence relationship between keywords, the co-occurrence matrix is first transformed into the correlation matrix by the Ochiia coefficient, to eliminate the influence of the co-occurrence matrix due to the too-large difference in frequency. The formula of the Ochiia coefficient is:

$$I_{mn} = \frac{C_{mn}}{\sqrt{C_m} \times \sqrt{C_n}},$$

where I_{mn} represents the Ochiia coefficient between keywords m and n, and the total number of occurrences of keywords m and n are C_m and C_n, respectively. Since there are too many 0 values in the correlation matrix, it is easy to cause too large an error when counting, to facilitate further processing, each data in the correlation matrix is subtracted by "1" to obtain the dissimilarity matrix that indicates the degree of dissimilarity between two words (e.g., **Table 2**). Different from the correlation matrix, the data in the dissimilarity matrix are dissimilar, and the larger the value is, the more distant the keywords are from each other and the worse the degree of similarity is.

The clustering results reflect the affinity between these keywords, further reflecting the research hotspot of credit risk. The principle of keyword clustering analysis is to take the frequency of keywords appearing two by two in the same article as the object of analysis and use the statistical method of clustering to gather closely related keywords together to form a cluster. In this paper, COOC software is used to cluster the high-frequency keyword dissimilarity matrix, and the clustering results obtained are shown in **Figure 5**.

Based on the results of the cluster analysis we classified them into 5 Clustering. Clustering 1 is the study of corporate credit risk in the financial sector, which consists of 2 subclassifications, of which 1 subclassification is the risk management of commercial banks (Sui et al., 2022), which consists of three keywords: commercial banks, risk management, and finance; and 2 subclasses are the quantification of credit risk, which consists of two keywords: probability of default and credit risk two keywords. Clustering 2 is the study of a quantitative assessment of credit risk in P2P lending (Wu et al., 2021), which consists of two subcategories, where subcategory 1 is the rating of default risk, which consists of the keywords credit rating and default risk, and subcategory 2 is the study of default factors in P2P lending, which consists of the keywords P2P lending and logistic regression two keywords. Clustering 3 is research on credit risk management of small and medium enterprises (Li et al., 2020), consisting of two keywords: credit risk management and small and medium enterprises. Clustering 4 is research on the credit risk assessment model for supply chain finance (Huang et al., 2021) and consists of six keywords: supply chain finance, machine learning, credit scoring, feature selection, classification, and data mining. Clustering 5 consists of assessing credit risk with neural network model (Li & Sun, 2020) and consists of two keywords: neural network and credit risk assessment.

Table 2. High-frequency keyword dissimilarity matrix (partial).

	Credit Risk	Credit Scoring	Machine Learning	Credit Risk Assessment	Probability Of Default	Risk Management
Credit Risk	0.0000	0.9218	0.8866	0.9893	0.8746	0.9125
Credit Scoring	0.9218	0.0000	0.8118	0.9922	0.9683	0.9787
Machine Learning	0.8866	0.8118	0.0000	0.9669	0.9777	0.9439
Credit Risk Assessment	0.9893	0.9922	0.9669	0.0000	0.9884	0.9883
Probability Of Default	0.8746	0.9683	0.9777	0.9884	0.0000	0.9843
Risk Management	0.9125	0.9787	0.9439	0.9883	0.9843	0.0000

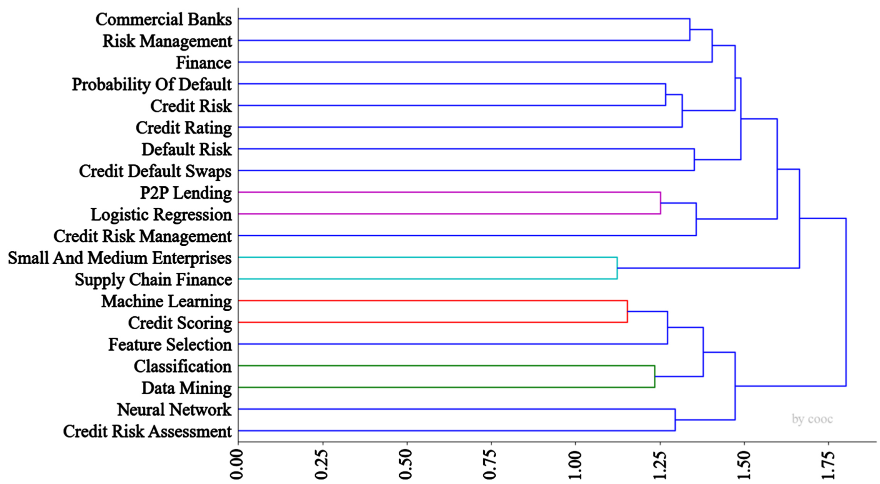
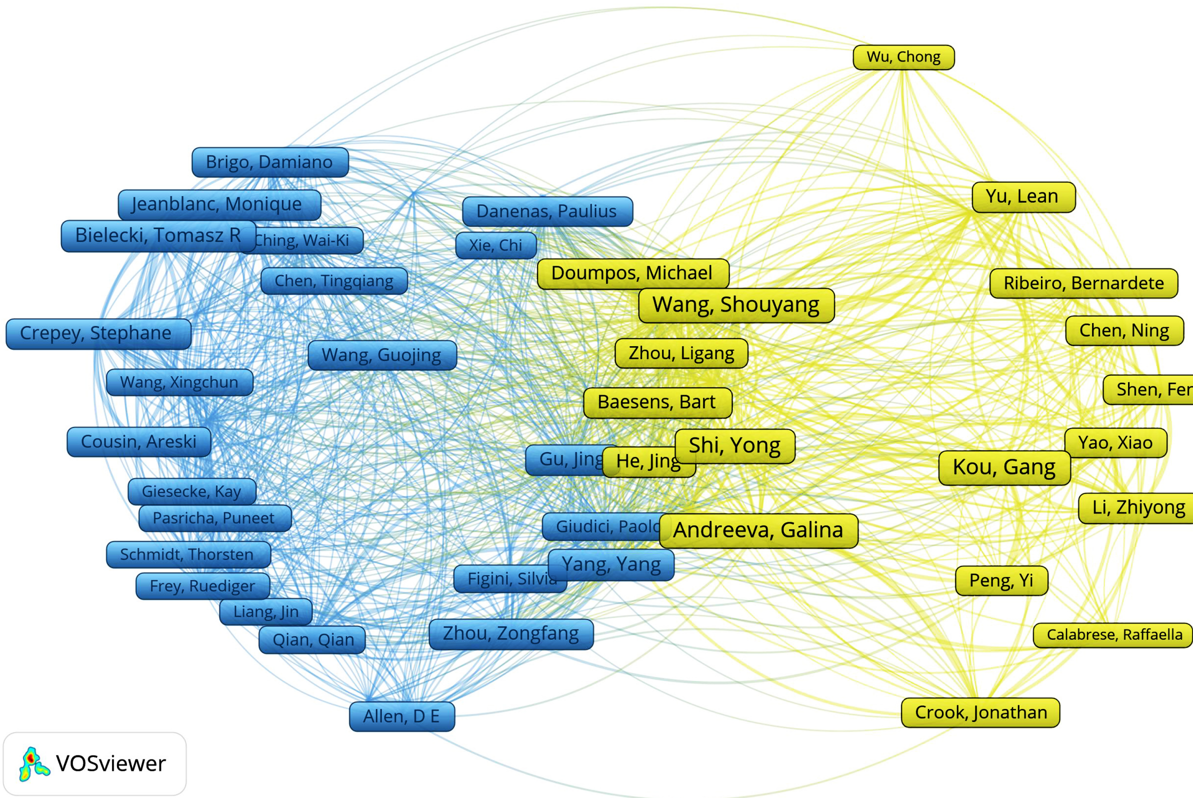


Figure 5. Credit risk research area keyword clustering map (COOC).

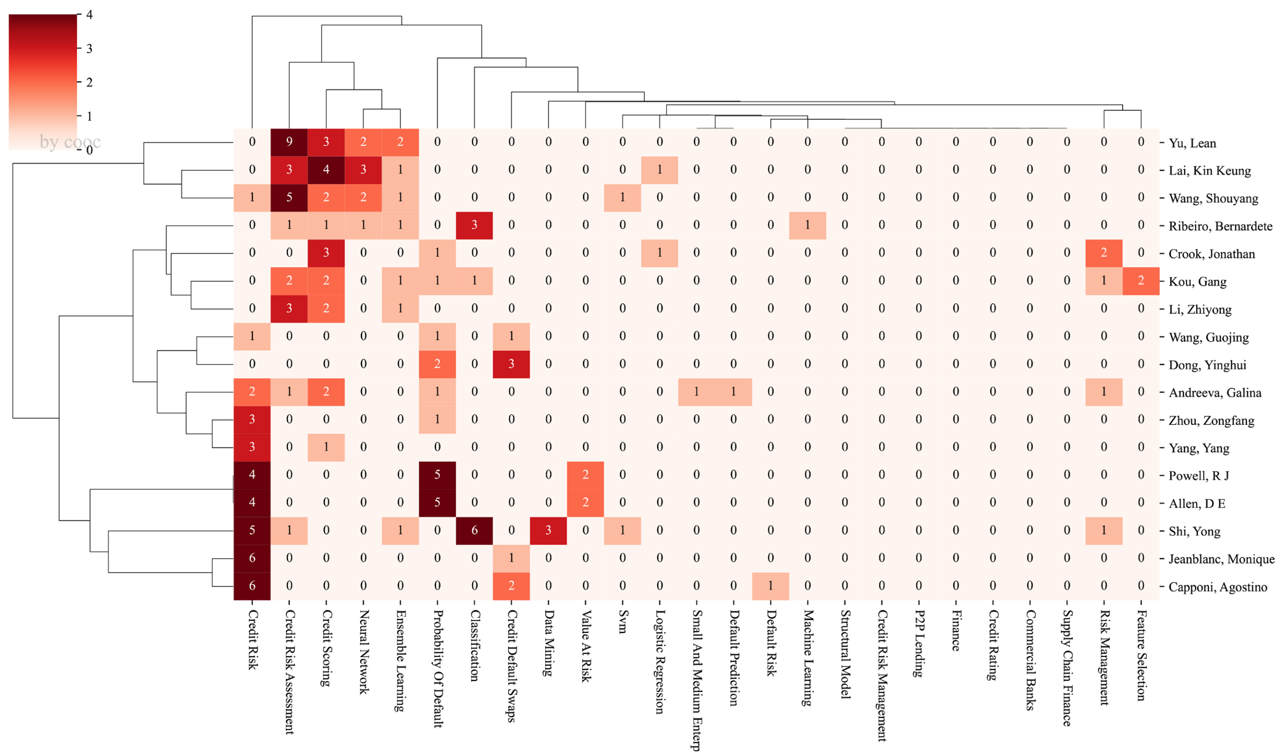
4.3. Author-Keyword Clustering Analysis

The research of many scholars has proved that the coupling analysis method can be used to reveal the hot topics and knowledge structure of research in subject areas (Zhao & Strotmann, 2008; Song et al., 2022). In this paper, we utilize the coupling analysis method to construct the keyword-based coupling relationship of journal authors and further excavate the implied potential correlation relationship between authors, the more the authors have the same keywords, the greater the similarity of the research direction between the two. Based on this similarity, this paper carries out similarity analysis and clustering research on authors based on constructing an author-keyword bimodal network, to achieve the revelation of the tacit knowledge of the subject area characterized by scholars.

The research direction of the authors can be more intuitively demonstrated through Figure 6(a) and Figure 6(b). Figure 6(a) shows the author-keyword coupling network diagram, which can clearly show the main potential collaborators in the field of credit risk research, in the coupling network of Figure 6(a), the nodes represent the authors of the journals, and the connecting lines between the nodes represent the coupling relationship between the authors of the



(a)



(b)

Figure 6. Keywords and authors analysis. (a) Keywords and authors coupling matrix (VOSviewer), (b) Keywords and author two-modular matrix (COOC).

journals, the thicker the connecting lines, the greater the coupling strength between the authors of the journals, and the stronger the correlation relationship between the two, i.e., the similarity in the research direction between the authors The thicker the line, the stronger the coupling strength between journal authors, the stronger the association between them, i.e., the stronger the similarity of research directions between authors. For example, Yang, Y., Bielecki, T.R., Zhou, Z.F., and other 34 authors have similar research areas. Meanwhile, we can also find more detailed research areas of the authors from the biclustering diagram of the constructed author-keyword bimodal matrices. For example, Prof. Yu researched credit risk assessment, credit scoring, neural network, and ensemble learning.

4.4. Analysis of the Evolutionary Path of Research Themes

The network analysis presented earlier in this paper can provide a preliminary understanding of the current status and research hotspots of credit risk. To elucidate the trend of research in the field of credit risk over time, this paper utilizes COOC software to draw the mutation detection map of 20 high-frequency keywords (**Figure 7**), which can better grasp the annual hotspot issues, and provide references for the future research and development of the industry through the mutated words over the course of this year.

Figure 7 shows the top 20 exploding keywords and the duration of the explosion. Fintech, interpretability, deep learning, machine learning, and ensemble learning were the research hotspots from 2007-2010. The emerging credit risk assessment, machine learning, lasted only one year in 2012, while P2P learning and big data lasted until 2014 and 2016, respectively, during which most of the scholars utilized big data to establish a credit risk assessment system for their research (Sang, 2021), and Du et al. (2020) used a large number of data samples to establish a network credit risk early warning model, and the constructed model was trained and tested. Finally, the neural network is optimized using a genetic

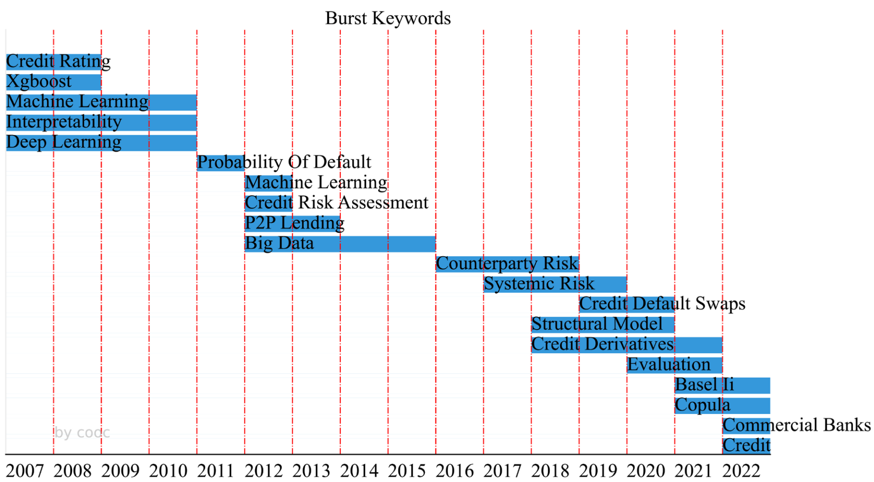


Figure 7. Top 20 keywords temporal emergence (COOC).

algorithm (GA) to improve the accuracy of early warning. Starting from 2018, research in the field of credit derivatives continues until 2021, during which research on credit default swaps, a derivative, also appears in 2019, e.g. Sun et al. (2021) examine whether and how payout policy affects credit risk using evidence from the credit default swap (CDS) market. It is worth noting that starting in 2021, research in the emerging area of copula and operational risk will continue through 2023, joining the two directions of credit and commercial banks as the focus of a new area that is likely to be an important topic in the future of credit risk.

5. Conclusion and Future Research Prospects

In this paper, we systematically review the research literature on credit risk. The number of journal papers, issuing countries/regions, core journal publications, keyword clustering, and research hotspots of credit risk are analyzed and visualized using COOC and VOSviewer software. The main results are as follows: 1) In terms of the number of journal papers, the cumulative number of publications in the field of credit risk has increased in a power function type, and the core research in the field began with the subprime mortgage crisis in 2008. 2) Currently, credit risk research has not yet formed a stable core group of authors, but it has already formed a core of issuing countries (China and the United States) as well as a core of research journals. 3) Credit risk assessment has become a key topic in the field, and research hotspots include supply chain finance, machine learning, credit scoring, probability of default, and commercial banks.

The future development of credit risk research should consider big data and artificial intelligence applications. With the development of big data and artificial intelligence technology, researchers are increasingly focusing on how to utilize large-scale data and intelligent algorithms to improve the prediction and assessment of credit risk. However, academics have several shortcomings in their research on credit risk. For example, machine learning and deep learning models are often considered black box models that lack interpretability and explainability. In credit risk assessment, the interpretability of models is important for policymakers and regulators. In addition, credit risk research is limited by available data, and traditional financial data may not provide the information needed for a comprehensive and accurate credit risk assessment.

Therefore, future research should consider the following aspects: 1) Traditional credit risk assessment relies on the experience and judgment of domain experts. Future research could combine AI techniques with domain expertise to create more accurate and interpretable credit risk models. 2) Scholars could explore ways to improve the explanatory and interpretable nature of machine learning and deep learning models so that policymakers and regulators can better understand the models' predictions and decision-making rationale. 3) Scholars could explore in future research how to integrate multiple data sources, including traditional financial data, nontraditional data, and external data to improve the accuracy and comprehensiveness of credit risk assessments.

Acknowledgements

Decision Consultation Project of Shaanxi Provincial Association for Science and Technology [grant number 23JCZX009R2]. We thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

Credit Authorship Contribution Statement

Jian Xue: Conceptualization, Methodology, Supervision. **Yixue Fan:** Methodology, Software, Writing—Original draft, Writing—Review & Editing.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data were obtained from the publication statistics of the Web of Science platform.

References

Abedin, M. Z., Chi, G. T., Hajek, P., & Zhang, T. (2022). Combining Weighted SMOTE with Ensemble Learning for the Class-İmbalanced Prediction of Small Business Credit Risk. *Complex & Intelligent Systems, 9,* 3559-3579. https://doi.org/10.1007/s40747-021-00614-4

Academic Dots, Bibliometric, COOC A Software for Bibliometric and Knowledge Mapping [CP/OL]. https://github.com/2088904822

Augustin, L., David, S., & David, T. (2009). Financial Risk Management: When Does Independence Fail? *American Economic Review, 99,* 454-458.

Bradford, S. C. (1985). Sources of Information on Specific Subjects. *Journal of Information Science, 10,* 173-180. https://doi.org/10.1177/016555158501000406

Consiglio, A., Lotfi, S., & Zenios, S. A. (2018). Portfolio Diversification in the Sovereign Credit Swap Markets. *Annals of Operations Research, 266,* 5-33. https://doi.org/10.1007/s10479-017-2565-5

Diebold, F. X., & Yilmaz, K. (2014). On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *Journal of Econometrics, 182,* 119-134. https://doi.org/10.1016/j.jeconom.2014.04.012

Du, G. S., Liu, Z. X., & Lu, H. F. (2020). Application of Innovative Risk Early Warning Mode under Big Data Technology in Internet Credit Financial Risk Assessment. *Journal of Computational and Applied Mathematics, 386,* Article ID: 113260. https://doi.org/10.1016/j.cam.2020.113260

Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green Supply Chain Management: A Review and Bibliometric Analysis. *International Journal of Production Economics, 162,* 101-114. https://doi.org/10.1016/j.ijpe.2015.01.003

Freixas, X., & Rochet, J. C. (1997). *The Microeconomics of Banking*. The MIT Press.

Hao, Z., Rallings, A. M., Espinoza, V., Luo, P. P., Duan, W. L. et al. (2021). Flowing from

East to West: A Bibliometric Analysis of Recent Advances in Environmental Flow Science in China. *Ecological Indicators, 125,* Article ID: 107358. https://doi.org/10.1016/j.ecolind.2021.107358

Harris, T. (2013). Quantitative Credit Risk Assessment Using Support Vector Machines: Broad versus Narrow Default Definitions. *Expert Systems with Applications, 40,* 4404-4413. https://doi.org/10.1016/j.eswa.2013.01.044

Huang, X. H., Sun, J. H., & Zhao, X. Y. (2021). Credit Risk Assessment of Supply Chain Financing with a Grey Correlation Model: An Empirical Study on China's Home Appliance Industry. *Complexity, 2021,* Article ID: 9981019. https://doi.org/10.1155/2021/9981019

Jagric, T., Jagric, V., & Kracun, D. (2012). Does Non-Linearity Matter in Retail Credit Risk Modeling? *Finance a Uver, 61,* 384-402.

Jiang, J. Y., Lyu, W., & Chen, N. (2022). A Bibliometric Analysis of Diffuse Large B-Cell Lymphoma Research from 2001 to 2020. *Computers in Biology and Medicine, 146,* Article ID: 105565. https://doi.org/10.1016/j.compbiomed.2022.105565

Jimenez, G., Ongena, S., Peydro, J. L., & Saurina, J. (2014). Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk-Taking? *Econometrica, 82,* 463-505. https://doi.org/10.3982/ECTA10104

Kou, G., Peng, Y., & Wang, G. X. (2014). Evaluation of Clustering Algorithms for Financial Risk Analysis Using MCDM Methods. *Information Sciences, 275,* 1-12. https://doi.org/10.1016/j.ins.2014.02.137

Li, G. R., Yang, J. X., & Huang, Y. (2020). Supply Chain Finance Credit Risk Evolving Intelligent Analysis System Based on System Dynamic Model. *Journal of Intelligent & Fuzzy Systems, 38,* 7837-7847. https://doi.org/10.3233/JIFS-179853

Li, M. Y., & Fu, Y. J. (2023). Prediction of Supply Chain Financial Credit Risk Based on PCA-GA-SVM Model. *Sustainability, 14,* Article No. 16376. https://doi.org/10.3390/su142416376

Li, X. T., & Sun, Y. (2020). Application of RBF Neural Network Optimal Segmentation Algorithm in Credit Rating. *Neural Computing & Applications, 33,* 8227-8235. https://doi.org/10.1007/s00521-020-04958-9

Liu, J., Liu, S. M., Li, J., & Li, J. Z. (2022a). Financial Credit Risk Assessment of Online Supply Chain in Construction Industry with a Hybrid Model Chain. *International Journal of Intelligent Systems, 37,* 8790-8813. https://doi.org/10.1002/int.22968

Liu, W. A., Fan, H., & Xia, M. (2022b). Multi-Grained and Multi-Layered Gradient Boosting Decision Tree for Credit Scoring. *Applied Intelligence, 52,* 5325-5341. https://doi.org/10.1007/s10489-021-02715-6

Liu, X., Zhao, S., Tan, L., Tan, Y., Wang, Y., Ye, Z. Y. et al. (2022c). Frontier and Hot Topics in Electrochemiluminescence Sensing Technology Based on CiteSpace Bibliometric Analysis. *Biosensors & Bioelectronics, 201,* Article ID: 113932. https://doi.org/10.1016/j.bios.2021.113932

Mahmoudi, N., & Duman, E. (2015). Detecting Credit Card Fraud by Modified Fisher Discriminant Analysis. *Expert Systems with Applications, 42,* 2510-2516. https://doi.org/10.1016/j.eswa.2014.10.037

Merton, R. C. (1973). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance, 29,* 449-470. https://doi.org/10.1111/j.1540-6261.1974.tb03058.x

Moon, T. H., Kim, Y., & Sohn, S. Y. (2011). Technology Credit Rating System for Funding SMEs. *Journal of the Operational Research Society, 62,* 608-615. https://doi.org/10.1057/jors.2010.15

Morelli, J. M., Ottonello, P., & Perez, D. J. (2022). Global Banks and Systemic Debt Crises. *Econometric, 90,* 749-798. https://doi.org/10.3982/ECTA17433

Niu, A. W., Cai, B. Q., & Cai, S. S. (2020). Big Data Analytics for Complex Credit Risk Assessment of Network Lending Based on SMOTE Algorithm. *Complexity, 2020,* Article ID: 8563030. https://doi.org/10.1155/2020/8563030

Puliga, M., Caldarelli, G., & Battiston, S. (2014). Credit Default Swaps Networks and Systemic Risk. *Scientific Reports, 4,* Article No. 6822. https://doi.org/10.1038/srep06822

Sang, B. (2021). Application of Genetic Algorithm and BP Neural Network in Supply Chain Finance under Information Sharing. *Journal of Computational and Applied Mathematics, 384,* Article ID: 113170. https://doi.org/10.1016/j.cam.2020.113170

Silva, E. C. E., Lopes, I. C., Correia, A., & Faria, S. (2020). A Logistic Regression Model for Consumer Default Risk. *Journal of Applied Statistics, 47,* 2879-2894. https://doi.org/10.1080/02664763.2020.1759030

Song, Y. H., Lei, L. X., Wu, L. J., & Chen, S. J. (2022). Studying Domain Structure: A Comparative Analysis of Bibliographic Coupling Analysis and Co-Citation Analysis Considering All Authors. *Online Information Review, 47,* 123-137. https://doi.org/10.1108/OIR-12-2020-0540

Sui, X., Wen, H. M., Gao, J., & Lu, S. P. (2022). Research on Contagion and the Influencing Factors of Personal Credit Risk Based on a Complex Network. *Discrete Dynamics in Nature and Society, 2022,* Article ID: 4730479. https://doi.org/10.1155/2022/4730479

Sun, C. Z., Wang, S. J., & Zhang, C. (2021). Corporate Payout Policy and Credit Risk: Evidence from Credit Default Swap Markets. *Management Science, 67,* 5755-5775. https://doi.org/10.1287/mnsc.2020.3753

Uchiyama, Y., Kadoya, T., & Nakagawa, K. (2019). Complex Valued Risk Diversification. *Entropy, 21,* 119. https://doi.org/10.3390/e21020119

Van Eck, N. J., & Waltman, L. (2010). Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics, 84,* 523-538. https://doi.org/10.1007/s11192-009-0146-3

Vickery, B. C. (1948). Bradford's Law of Scattering. *Journal of Documentation, 4,* 198-203. https://doi.org/10.1108/eb026133

Wang, H., Xu, Q. S., & Zhou, L. F. (2015). Large Unbalanced Credit Scoring Using Lasso-Logistic Regression Ensemble. *PLOS ONE, 10,* e0117844. https://doi.org/10.1371/journal.pone.0117844

Wang, X. G. (2022b). Analysis of Bank Credit Risk Evaluation Model Based on BP Neural Network. *Computational Intelligence and Neuroscience, 2022,* Article ID: 2724842. https://doi.org/10.1155/2022/2724842

Wang, Y. J. (2021). Research on Supply Chain Financial Risk Assessment Based on Blockchain and Fuzzy Neural Networks. *Wireless Communications & Mobile Computing, 2021,* Article ID: 5565980. https://doi.org/10.1155/2021/5565980

Wang, Y. S. (2022a). Credit Risk Evaluation of Asset Securitization of PPP Project of Sports Public Service Venues Based on Random Forest Algorithm. *Computational Intelligence and Neuroscience, 2022,* Article ID: 5177015. https://doi.org/10.1155/2022/5177015

Wang, Z., Sun, H., & Yang, L. (2023). A Bibliometric Analysis of Research on Historical Buildings and Digitization. *Buildings, 13,* Article No. 1607. https://doi.org/10.3390/buildings13071607

Wu, F., Su, X., Ock, Y. S., & Wang, Z. Y. (2021). Personal Credit Risk Evaluation Model of P2P Online Lending Based on AHP. *Symmetry-Basel, 13,* Article No. 83. https://doi.org/10.3390/sym13010083

Xu, Y. Z., Zhang, J. L., Hua, Y., & Wang, L. Y. (2019). Dynamic Credit Risk Evaluation Method for E-Commerce Sellers Based on a Hybrid Artificial Intelligence Model. *Sustainability, 11,* Article No. 5521. https://doi.org/10.3390/su11195521

Yfanti, S., Karanasos, M., Zopounidis, C., & Christopoulos, A. (2022). Corporate Credit Risk Counter-Cyclical Interdependence: A Systematic Analysis of Cross-Border and Cross-Sector Correlation Dynamics. *European Journal of Operational Research, 304,* 813-831. https://doi.org/10.1016/j.ejor.2022.04.017

Zeng, J. J., Qu, J. S., & Ma, H. Q. (2021). Characteristics and Trends of Household Carbon Emissions Research from 1993 to 2019: A Bibliometric Analysis and Its Implications. *Journal of Cleaner Production, 295,* Article ID: 126468. https://doi.org/10.1016/j.jclepro.2021.126468

Zhang, Z. M., Niu, K., & Liu, Y. (2020). A Deep Learning Based Online Credit Scoring Model for P2P Lending. *IEEE Access, 8,* 177307-177317. https://doi.org/10.1109/ACCESS.2020.3027337

Zhao, D. Z., & Strotmann, A. (2008). Evolution of Research Activities and Intellectual Influences in Information Science 1996-2005: Introducing Author Bibliographic-Coupling Analysis. *Journal of the American Society for Information Science and Technology, 59,* 2070-2086. https://doi.org/10.1002/asi.20910

Zhong, Y., & Wang, H. L. (2023). Internet Financial Credit Scoring Models Based on Deep Forest and Resampling Methods. *IEEE Access, 11,* 8689-8700. https://doi.org/10.1109/ACCESS.2023.3239889