Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/74/20240448

Advancing financial crisis prediction: Big data, AI, and economic theories

Shirui Li

University College London, London, UK

2482516799@qq.com

Abstract. This paper explores the theoretical foundations of financial crisis early warning systems, focusing on the integration of big data, ensemble learning, and various financial crisis theories. It begins with defining key concepts such as big data, financial crisis, and ensemble learning, highlighting the evolution and significance of these terms in the context of financial crisis management. The literature review covers extensive research on financial crisis early warning indicators and models, tracing the development from traditional statistical methods to advanced artificial intelligence approaches, including neural networks, decision trees, and machine learning algorithms. The paper critically assesses the current methodologies and emphasizes the necessity of incorporating big data and machine learning for more accurate and comprehensive early warning systems. Theoretical foundations related to financial crises, such as information asymmetry, behavioral economics, economic cycle theory, and contingency theory, are discussed to understand their impact on financial crisis prediction and management. The conclusion synthesizes the findings and suggests future research directions, emphasizing the integration of diverse methodologies and interdisciplinary approaches to enhance the early warning capabilities of enterprises.

Keywords: Big Data, Financial Crisis, Machine Learning, Artificial Intelligence, Behavioral Economics, Economic Cycle Theory

1. Introduction

This paper delves into the complex domain of financial crisis early warning systems, a critical area of study given the volatile nature of global markets and the profound impact of financial crises on economies and businesses. The investigation begins by defining and dissecting key concepts such as big data, financial crises, and ensemble learning, laying the groundwork for a comprehensive understanding of the topic. The focus then shifts to an extensive literature review, which examines a wide range of methodologies and models developed over the years for predicting financial crises. This review spans from early statistical models to sophisticated artificial intelligence techniques, including neural networks and machine learning algorithms. The paper aims to bridge the gap between traditional financial crisis prediction methods and the innovative potential of big data and AI-driven techniques. It also explores relevant theoretical foundations, including information asymmetry, behavioral economics, and economic cycle theory, to provide a multifaceted understanding of the factors influencing financial crises [1]. By offering an in-depth analysis of these diverse elements, the paper sets the stage for a nuanced discussion on the evolution and future trajectory of financial crisis early warning systems.

DOI: 10.54254/2755-2721/74/20240448

2. Theoretical Foundations in Financial Crisis Management

2.1. Relevant Concept Definitions

2.1.1. Big Data

In the mid-1990s, the term "big data" emerged and, after two decades of relative obscurity, regained prominence with recent technological advancements. Given the diversity of research perspectives and fields, academia has not unified its definition of big data. From a resources perspective, big data reflects the dynamic state of the material and intellectual world. Farboodi et al. define big data, in the context of its integration with physical businesses, as data generated alongside economic activities, enabling companies to transform decision-making paradigms and enhance operational efficiency [2]. Big data, as a derivative of computer technology and system integration, is valued for its practical application post-analysis. This definition distinguishes big data's core attributes from traditional data, encapsulating its "4V" characteristics: Volume, Variety, Velocity, and Value. Drawing from these definitions, this paper considers big data for corporate crisis warning as massive, high-speed, and diverse data from social media. The method involves collecting text from the Dongfang Fortune Stock Bar and quantifying it using an emotion dictionary to extend the definition of big data indicators.

2.1.2. Financial Crisis

Foreign scholars often define financial crises or distress around the concept of bankruptcy. Altman identified bankruptcy filing as a financial crisis, legally defining the term. Beaver included overdue preferred stock dividends, bond defaults, and bank overdrafts in his definition of a financial crisis. Altman later summarized business failure, inability to pay debts, default, and bankruptcy as four manifestations of a financial crisis. Chinese scholars, apart from bankruptcy, have diverse definitions including payment difficulties, ST standards, and negative net profits. In 1998, the China Securities Regulatory Commission introduced special treatment (ST) for listed companies with abnormal financial or other conditions [3]. Later, Chinese scholars shifted the definition of financial crisis from bankruptcy signs to whether a company is ST-listed. Quantitatively, scholars differ in their definitions. Rafie considered a company in financial crisis if its retained assets were less than 50%. Platt and Platt saw negative net profit as a crisis indicator. Song Peng and Li Tingting refined this to include negative net profits at the end of each quarter. Some scholars define a financial crisis as occurring when a company's cash flow cannot meet its daily operational expenses. Effective crisis prediction focuses on early warning rather than measurement, identifying significant financial risks in advance. Therefore, this paper selects Chinese A-share companies specially treated (ST) as research samples based on the above definitions.

2.1.3. Ensemble Learning

Ensemble learning, widely applied in practical domains, enhances machine learning through the combination of weak learners. It applies to regression ensembles, multiclass classification, feature selection, and anomaly detection. Its philosophy, akin to "many hands make light work," involves aggregating individual judgments to analyze complex problems, typically outperforming any single judgment. Core ideas in ensemble algorithms include Bagging (Bootstrap aggregating), Boosting, and Stacking. Bagging relies on multiple bootstrap samples to generate diverse sub-samples and decide outcomes through voting. Boosting transforms weak learners into stronger classifiers. Stacking involves secondary and meta-classifiers in a layered, heterogeneous model. This paper's model is based on the Random Forest algorithm, a representative of the Bagging approach [4].

DOI: 10.54254/2755-2721/74/20240448

2.2. Literature Review

2.2.1. Research on Financial Crisis Early Warning

In the 1930s, with the plummeting of the U.S. financial market and the onset of the "Great Depression", the management of financial crisis early warning emerged. Scholars dedicated themselves to developing more accurate financial crisis early warning models to help businesses predict crises. The effectiveness of a financial crisis early warning system primarily hinges on comprehensive warning indicators and cutting-edge models, prompting extensive research in these areas globally.

Research on Financial Crisis Early Warning Indicators

Due to technological disparities, scholars have chosen different types of indicators and constructed varied combinations. The selection of financial crisis prediction indicators can be divided into two categories: one derived from corporate financial reports (financial indicators), and the other from external information (non-financial indicators).

Considering the initial definition of a financial crisis linked to negative profits and inability to repay debts on time, the earliest financial indicators used for prediction mainly measured a company's debt-paying ability and profit return. A notable example is Fitzpatrick, who used net profit/equity and equity/debt in financial crisis prediction. Subsequently, scholars like Altman and Zmijewski began incorporating diverse financial indicators like retained earnings ratio, asset utilization, and working capital ratio into the early warning system, enhancing the accuracy of bankruptcy prediction. Further research delved into various aspects like a company's capabilities, value, and equity structure, improving the prediction accuracy [5].

The role of financial indicators in early warning is undeniable, but the inclusion of non-financial indicators offers deeper insights into the reasons for financial crises. For instance, Kuehn and Schmid utilized credit risk indicators in predicting financial crises, finding a higher occurrence during economic downturns. Studies also focused on the relationship between capital market reactions and financial crises, with researchers like Charalambakis and Garrett [6] enhancing prediction capabilities by including stock market information in their models. There's growing interest in exploring models outside the financial domain, linking executive characteristics and behavior to financial crises. For example, overconfident managers tend to overinvest and prefer value-destructive diversification, leading to higher stock crash risks [7]. Research on the emotional tone of Management Discussion and Analysis (MD&A) in financial reports also influences crisis prediction. These diverse perspectives enrich the financial crisis early warning indicator system.

Research on Financial Crisis Early Warning Models

Global scholars have been dedicated to developing more precise financial crisis early warning models. From the initial univariate analysis to complex multivariate models, their evolution reflects advancements in statistical theories and information technology. The integration of artificial intelligence represents a new direction in current research.

(1) Traditional Statistical Model Warning Methods

a) Univariate Analysis and Multivariate Discriminant Analysis (MDA)

Researchers like Fitzpatrick and Beaver initially used univariate analysis in their models, noting significant differences in predictive effectiveness based on indicators. Altman integrated MDA with financial crisis early warning, marking a significant milestone. Subsequent improvements by Altman et al. led to the ZETA model, which incorporated capitalization and scale indices, enhancing predictive accuracy. Gupta and Chaudhry represent recent use of MDA in predicting financial crises.

b) Logistic and Probit Models

Despite improvements, MDA's stringent assumptions limit its application. Logistic and Probit models, as multivariate conditional probability models, overcome these limitations. Martin and Ohlson were among the first to apply Logistic regression in crisis prediction. Chinese researchers have also extensively used Logistic models, with Wu Shinong and Jia et al. [8] achieving strong predictive performance. Probit models, though computationally intensive compared to Logistic models, have also been used effectively.

(2) Artificial Intelligence Model Warning Methods

a) Artificial Neural Networks (ANN)

ANNs, particularly BP neural networks, have been widely used due to their ability to identify and utilize non-linear relationships between input variables. Odom and Shard demonstrated ANN's superior predictive accuracy over traditional MDA.

b) Decision Trees (DT)

Decision trees effectively solve classification and regression problems, as evidenced by Frydman et al., Sung et al., and Olson et al. Their use in financial crisis prediction has been validated, though they are less capable than ensemble learning methods.

c) Genetic Algorithms (GA)

Initially used by Varetto for financial crisis early warning, GA's predictive accuracy was less than MDA. Subsequent studies combined GA with other models, like Ye Huanzhuo et al. combining it with adaptive Bayesian networks, offering new approaches to intelligent financial crisis early warning and diagnosis.

d) Rough Set Theory (RST)

RST, effective in classifying incomplete information, has been integrated with BP neural networks and SVM models for improved financial crisis early warning.

(3) Other Warning Methods

In addition to the above models, researchers continue to explore new models, such as Cox survival models and Cumulative Sum (CUSUM) models, considering their complexity and interdisciplinary nature. Honjo was among the first to apply Cox survival models to financial crisis research. The incorporation of option pricing models into KMV models has also filled gaps in research linking capital market stock price fluctuations with financial crises [9].

Financial Crisis Early Warning Models	Traditional Statistical Models	Univariate Analysis
		Multivariate Discriminant Analysis
		Logistic Regression
		Probit Model
	Artificial Intelligence Models	Artificial Neural Network
		Decision Tree
		Rough Set Theory
		Support Vector Machine
		Case-Based Reasoning
	Other Warning Methods	Accurate Pricing Model
		KMV Model
		Cox Survival Analysis

Figure 1. Summary of financial crisis early warning models.

Finally, due to the cross-disciplinary nature of the latest developments in artificial intelligence, specifically machine learning algorithms, their learning and usage present certain challenges and barriers. This has resulted in their application in the field of corporate financial crisis research being significantly less than that of traditional parametric models such as Logit and Multivariate Discriminant Analysis. However, the superiority of machine learning algorithms over traditional methods like Logit in the performance of financial crisis early warning systems is gradually being validated by scholars both domestically and internationally. Appropriately incorporating artificial intelligence and machine learning algorithms to predict financial crises, thereby enhancing the early warning capabilities of enterprises, and providing more objective data support and reference for regulatory bodies and investors, represents a new direction in the research of financial crisis early warning systems.

DOI: 10.54254/2755-2721/74/20240448

2.3. Theoretical Foundations Related to Financial Crises

2.3.1. Theory of Information Asymmetry

Information asymmetry refers to the disparities in the channels and methods of information collection among investors, businesses, and external agencies in the real-world capital market. This disparity leads to variations in the quantity and quality of information each party receives. The party with more comprehensive information may exploit this advantage for personal gain, often to the detriment of the less informed, negatively impacting the efficient functioning of the capital market.

Listed companies primarily disclose financial and relevant non-financial information through financial reports. The quality of this disclosed information is crucial for accurate early warning of financial crises. Low transparency in corporate information is one of the reasons for information asymmetry in the capital market. In an asymmetric information environment, the informationally disadvantaged parties actively seek information through various channels, such as online searches and social media forums like stock bars, selectively assimilating information beneficial to them.

2.3.2. Behavioral Economics Theory

Behavioral economics, a blend of economics and psychology, highlights that human economic behavior is often irrational. This revelation significantly impacts managerial decision-making. Managers are influenced by overconfidence, cognitive biases, etc., leading to the neglect of low-probability events or the opinions of stakeholders in their decisions, thus diminishing decision quality. Irrational behavior by managers increases the likelihood and severity of crises faced by businesses. Managers need to be aware of their biases and tendencies, taking appropriate measures to mitigate decision-making risks.

2.3.3. Economic Cycle Theory

The economic cycle, also known as the macroeconomic environment, refers to the regular expansion and contraction of the broader economy. It can be detailed into four stages: prosperity, recession, depression, and recovery, and broadly divided into periods of economic upswing and downturn. The development and operation of any enterprise are influenced by these economic cycles. During upswings, with robust market demand, businesses operate in a relaxed economic environment. During downturns, with sluggish markets and falling consumer demand, intense competition among businesses can trigger various latent risks, especially financial ones. Therefore, businesses can develop strategies for different periods by analyzing the patterns of economic cycles.

Table 1. Summary of Key Theories Related to Financial Crisis Management

Theory	Brief Description
Theory of Information Asymmetry	Addresses disparities in information access in capital markets, leading to unequal advantages and impacting market efficiency. Focuses on the importance of transparent information disclosure by companies.
Behavioral Economics Theory	Combines economics and psychology to explain irrational human economic behaviors and their impact on managerial decision-making and business risks.
Economic Cycle Theory	Describes the economy's regular expansion and contraction phases, emphasizing how these cycles affect business strategies and operations.

3. Conclusion

In summary, this paper underscores the critical need for more advanced and integrated approaches in developing early warning systems for financial crises. While traditional methods have provided a foundation, the dynamic nature of global economies and the complex interplay of various factors demand a more holistic and flexible approach. The integration of big data, artificial intelligence, and machine learning represents a significant step forward, offering the potential to capture a broader range of indicators and predict crises with greater accuracy. However, this integration must be carefully balanced with considerations of behavioral economics, information asymmetry, and the overarching economic cycle. Future research should focus on developing models that are not only technologically

DOI: 10.54254/2755-2721/74/20240448

advanced but also deeply rooted in a comprehensive understanding of economic and behavioral theories. Such interdisciplinary and multifaceted approaches are essential for creating robust and effective early warning systems capable of navigating the complexities of the modern financial landscape.

References

[1] Hancock, John T., and Taghi M. Khoshgoftaar. "CatBoost for big data: an interdisciplinary review." Journal of big data 7.1 (2020): 1-45.
[2] Deepa, Natarajan, et al. "A survey on blockchain for big data: Approaches, opportunities, and future directions." Future Generation Computer Systems 131 (2022): 209-226.
[3] Caplar R and Kulisic P 1973 Proc. Int. Conf. on Nuclear Physics (Munich) vol 1 (Amsterdam:North-Holland/American Elsevier) p 517
[4] Sestino, Andrea, et al. "Internet of Things and Big Data as enablers for business digitalization strategies." Technovation 98 (2020): 102173.
[5] Fujimoto, Yuka, and Jasim Uddin. "Inclusive leadership for reduced inequality: economic–social–economic cycle of inclusion." Journal of business ethics (2021): 1-20.
[6] Karmalita, Viacheslav. Stochastic dynamics of economic cycles. Walter de Gruyter GmbH & Co KG, 2020.
[7] Nadiya, Dekhtyar. "THEORY OF ECONOMIC CYCLES IN FORECASTING THE ACTIVITY OF THE TOURISM INDUSTRY IN UKRAINE." The 10th International scientific and practical conference "Dynamics of the development of world science"(June 10-12, 2020) Perfect Publishing, Vancouver, Canada. 2020. 848 p.. 2020.
[8] Greener, Joe G., et al. "A guide to machine learning for biologists." Nature Reviews Molecular Cell Biology 23.1 (2022): 40-55.
[9] Mahesh, Batta. "Machine learning algorithms-a review." International Journal of Science and Research (IJSR).[Internet] 9.1 (2020): 381-386.