Academic Education Publishing House
Philosophy and Social Science Vol. 2 No. 8, 2025

Research on the Application of Big Data Technology in Financial Risk Management

Meifang Jiang
Suzhou Agricultural Institute, Suzhou, JiangSu, China

Abstract: This article focuses on the empowering value of big data technology in financial risk management, aiming to break through the limitations of traditional risk control data coverage being narrow, response lagging, and low correlation. Firstly, from a theoretical perspective, this paper analyzes the three major shortcomings of traditional financial risk control that rely on structured data and static models. It elaborates on the adaptability of big data technology in alleviating information asymmetry (such as increasing the credit evaluation coverage of small and micro enterprises from 35% to 82%), achieving comprehensive risk collaborative monitoring (reducing cross risk warning response to 12 minutes), and improving risk pricing accuracy (reducing bad debt rates by 18%) through multi-source data integration, real-time processing, and intelligent analysis.

Secondly, based on practical cases, analyze the application effectiveness of big data in four types of risk control: in the field of credit risk, the optimization evaluation and approval of the "Puyin Point Loan" model of Shanghai Pudong Development Bank and the "310" model of online commercial banks; In the field of market risk, platforms such as CITIC Securities and Bank of China have implemented forward-looking warnings; In the field of operational risk, China Merchants Bank's anti fraud system and Industrial and Commercial Bank of China's monitoring platform achieve pre prevention and control; In the field of systemic risk, the financial risk correlation graph helps with global monitoring.

Furthermore, it points out three major challenges: data (quality defects, compliance pressure), models (poor interpretability, insufficient robustness), and resources (high investment, talent shortage). Finally, it proposes the trends of technology integration (federated learning, knowledge graph), scenario diversification (ESG risk control, cross-border risk control), and regulatory collaboration (intelligent compliance, regulatory sandbox), as well as suggestions for financial institutions to build in stages and regulatory departments to improve frameworks, providing reference for promoting the transformation of risk control from "experience driven" to "data-driven".

Keywords:Big Data, Financial Risk, Management

1. Theoretical Foundations of Big Data Technology and Financial Risk Management

The core objective of financial risk management is to identify, measure, and control various types of risks. Traditional approaches rely on structured financial data (such as corporate balance sheets and personal credit reports) and static statistical models (like logistic regression and VaR historical simulation). However, in the context of digital finance, three major limitations have become increasingly apparent: First, the narrow scope of data coverage makes it difficult to capture non-financial risk signals (e.g., corporate supply chain stability and individual consumption behavior preferences). Second, model responsiveness lags behind, as traditional batch processing requires T+1 or even longer cycles to update risk parameters, failing to address dynamic scenarios such as high-frequency trading and real-time cross-border payments. Third, low risk correlation means isolated business data struggles to identify cross-market and cross-institutional risk transmission pathways (e.g., the risk linkage between shadow banking and traditional banks).

The adaptability of big data technology to financial risk control essentially represents a technological empowerment of "risk management theory." From the perspective of information asymmetry theory, the integration of multi-source data reduces the information gap between financial institutions and clients. For instance, small and micro enterprises can compensate for incomplete financial statements by providing e-commerce transaction records and logistics information, thereby increasing credit assessment coverage from the traditional 35% to 82%. From the standpoint of Enterprise Risk Management (ERM) theory, real-time data processing capabilities enable the coordinated monitoring of credit risk, market risk, and operational risk. A global investment bank, by building an integrated risk management platform, reduced cross-risk-type alert response time from 4 hours to 12 minutes. Regarding risk pricing theory, intelligent analytical models can dynamically adjust risk premiums. For example, consumer finance companies modify loan interest rates based on clients' real-time repayment behavior, improving risk pricing accuracy by 30% and reducing default rates by 18%.

2. Core Application Scenarios and Practical Cases of Big Data Technology in Financial Risk Control

Big data technology has deeply penetrated the core field of financial risk control, achieving a dual improvement in risk management efficiency and accuracy through specific scenarios. The following analysis is based on practical cases of domestic and foreign financial institutions:

2.1 Credit Risk Management: From "Single Assessment" to "Dynamic Profile"

Credit risk is the most core type of risk for financial institutions, and big data technology reconstructs the credit evaluation system through multi-source data integration and real-time monitoring. In the personal credit scenario, the "Puyin Diandai" product launched by Shanghai Pudong Development Bank integrates customer credit card transaction data (consumption frequency, merchant type, repayment record in the past 12 months), mobile banking APP behavior data (login duration, function usage preference, geographical location trajectory), third-party payment data (Alipay/WeChat transfer frequency, bill amount), builds a credit scoring model using LightGBM algorithm, compresses the approval process from the traditional three working days to one hour, and controls the non-performing loan rate at 1.1%, lower than the industry average (1.8%). In the credit scenario of small and micro enterprises, the "online loan" of online merchant bank broke through the traditional mode of relying on financial statements, accessed the transaction flow of Taobao/Tmall stores (average daily order volume, customer unit price, return rate), supply chain data (cooperation duration of upstream suppliers, stability of downstream customers), logistics information (express shipment volume, distribution of receiving addresses), combined with federal learning technology to achieve data sharing with Alipay (not disclosing the original data), and built a "310" credit model (3-minute application, 1-second approval, zero manual intervention). By 2024, this product serves more than 30 million small and micro enterprises, with a non-performing loan rate of 0.45%, significantly lower than 2.5% of small and micro loans of traditional banks%. In the post loan management scenario, China Merchants Bank has established a "real-time risk warning system" that monitors the borrower's account dynamics (such as abnormal changes in transaction locations, large transfers in the early morning, frequent transactions with high-risk merchants), credit inquiry records (number of inquiries in the past 3 months, type of inquiry institution), social behavior (negative comments on social media, credit deterioration of related parties), calculates the probability of default (PD) in real time, identifies 75% of potential defaulting customers 3 months in advance, and takes intervention measures such as SMS reminders and adjusting repayment plans to improve the efficiency of post loan risk disposal by 45%.

2.2 Market Risk Management: From "Historical Retrospective" to "Forward-Looking Prediction"

Market risks are influenced by multiple factors such as macroeconomic conditions, policy changes, and market sentiment. Big data technology achieves forward-looking risk control through high-frequency data processing and multidimensional analysis. In the risk scenario of the stock market, CITIC Securities has built an "intelligent risk control platform" that integrates Level-2 high-frequency market data from the Shanghai and Shenzhen Stock Exchanges (10 orders per second, transaction data), internal trading data from securities firms (customer position changes, trading volume distribution), and financial public opinion data (news texts and analyst research reports from Sina Finance and Caixin). It uses LSTM deep learning models to predict stock price fluctuations and combines Monte Carlo simulation to calculate VaR (Value at Risk), increasing VaR prediction accuracy by 25%. During the volatile period of the A-share market in 2024, the platform issued a position risk warning to institutional clients 20 minutes in advance, helping clients reduce their holdings of high-risk stocks and minimize potential losses by 32%. In the risk scenario of the foreign exchange market, Bank of China's "Real time Foreign Exchange Risk Monitoring System" integrates real-time market data from 12 major foreign exchange markets worldwide (exchange rates of major currency pairs such as USD/EUR and USD/JPY), cross-border trade settlement data (import and export amounts of enterprises, settlement currency types), macroeconomic indicators (GDP growth rate, inflation rate, interest rate policy), predicts exchange rate fluctuations through ARIMA-GARCH model, and monitors the risk exposure of customers' foreign exchange positions in real time. Through this system, a multinational enterprise adjusted its foreign exchange hedging strategy in advance to reduce exchange losses of 18 million yuan in 2024 when the Federal Reserve's interest rate increase led to the appreciation of the U.S. dollar. In the risk scenario of the commodity market, the "Commodity Risk Warning System" developed by Nanhua Futures utilizes satellite remote sensing image data (such as planting area and rainfall in soybean producing areas, operating rate of extraction facilities in crude oil producing areas), logistics and transportation data (port inventory, freight volume), and industry chain data (upstream raw material prices, downstream processing enterprise operating rate) to construct a commodity price prediction model, helping futures investors identify price fluctuation risks in advance. In 2023, the investment return rate of users of the system is 15% higher than the market average level.

2.3 Operational Risk Management: From "Post Accountability" to "Prevention and Control"

Operational risks include internal fraud, external fraud, process vulnerabilities, and other types. Big data technology achieves pre interception of risks through behavior characteristic analysis and anomaly recognition. In external fraud prevention and control scenarios, China Merchants Bank's "Intelligent Anti Fraud System" integrates customer transaction characteristics (amount, time, merchant type), device characteristics (phone model, IP address, device fingerprint), biometric features (fingerprint recognition, facial recognition, gesture code trajectory), and constructs a 200+dimensional risk feature library, using the Isolation Forest algorithm to identify abnormal transactions. For example, when the system detects that "a customer frequently logs in with their mobile phone in Shanghai and another unfamiliar device attempts a large transfer in Shenzhen 10 minutes later," it will trigger a second verification in real time (SMS verification code+facial recognition). By 2024, the system's accuracy in intercepting fraudulent transactions will reach 99.3%, and it will recover more than 500 million yuan in losses for customers every year. In the internal operational risk scenario, Industrial and Commercial Bank of China's "Employee Operational Risk Monitoring Platform" analyzes teller business processing data (transaction frequency, authorization times, business duration), system access data (login time, operation logs, sensitive permission calls), compliance inspection data (past violation records, training and assessment scores), and constructs an operational risk heatmap. When the system detects that a teller has been calling customer credit data during non working hours for three consecutive days without any compliance approval records, it will issue a warning to the risk control department. By 2023, the platform will reduce the incidence of internal fraud incidents by 70%. In the scenario of process vulnerability identification, Shanghai Pudong Development Bank's "Process Risk Intelligent Diagnosis System" identifies weak links in the process by mining historical business data, such as reasons for rejection in loan approval processes and process issues in customer complaints. For example, the system found that "there is a lack of verification steps for the actual controller's affiliated enterprises in the approval of loans for small and micro enterprises", and promoted process optimization, resulting in a 22% decrease in non-performing loan rates caused by affiliated risks.

2.4 Systemic Risk Management: From "Local Monitoring" to "Global Prevention and Control"

Systemic risk involves cross institutional and cross market risk transmission, and big data technology achieves global risk monitoring through knowledge graph construction. A state-owned bank, in collaboration with the People's Bank of China, the China Securities Regulatory Commission, and the China Banking and Insurance Regulatory Commission, has developed a "Financial Risk Correlation Map" that integrates business data from banks, securities, insurance, trusts, and other institutions (such as interbank lending amounts, investment portfolio overlap, and risk exposure distribution) to identify the interrelationships between financial institutions. For example, through graph analysis, it was found that "a certain city commercial bank has large interbank cooperation with three high-risk trust companies, and these trust companies hold a large amount of non-performing real estate assets". Risk isolation measures were taken in advance (reducing interbank credit and requiring mortgage guarantees) to avoid risk transmission to the banking system. In 2024, this map will help regulatory authorities identify 5 financial institutions with systemic risk hazards and prevent the spread of risks through early intervention.

3. The Practical Challenges of Big Data Technology in Financial Risk Control Applications

Although big data technology has achieved significant results in financial risk control, it still faces challenges in the three dimensions of data, models, and resources in practical applications, which restrict the full release of technological value

3.1 At the Data Level: The Dual Pressure of Quality Defects and Compliance Constraints

Data is the foundation of big data risk control, and financial institutions currently face two core issues in data applications: firstly, uneven data quality. In the data collection process, the stability of external data interfaces is insufficient (such as delayed data updates from third-party credit reporting companies and fraudulent transaction data on e-commerce platforms), resulting in a loss of data timeliness and authenticity. According to statistics from a consumer finance company, about 15% of the external data it accesses has update delays (exceeding 24 hours), and 8% has data inconsistencies (information conflicts between the same customer on different platforms). In the data cleaning process, unstructured data processing is difficult, such as semantic ambiguity in customer consultation texts and professional terminology in research reports, resulting in a data standardization rate of only 65%, which is lower than the 98% of structured data; The handling of missing values relies on manual imputation (such as mean imputation and median imputation), which can easily lead to data bias. Due to improper handling of missing values, the risk control model of a certain bank expanded the credit score error by 20%. Secondly, data compliance risks are highlighted. With the implementation of Personal Information Protection Law, Data Security Law, GDPR and other laws and regulations, the data application of financial institutions faces strict constraints: in the data collection process, it is necessary to obtain clear informed consent from customers. An Internet bank was fined 20 million yuan by the regulatory authorities for not fully informing customers of the purpose of data; In the cross-border flow of data, GDPR requires that data must meet the principle of "full protection" when leaving the country. A foreign bank was fined 48 million euros by EU regulatory authorities for transferring European customer data to China for processing. In addition, the data sharing mechanism is not perfect, and financial institutions are unwilling to share risk information due to concerns about data leakage, resulting in the problem of "data silos". The data dimension of small and medium-sized banks is only 40% of that of large banks, making it difficult to build accurate risk control models.

3.2 Model Level: Double Bottleneck of Insufficient Interpretability and Robustness

Big data risk control models mostly rely on complex algorithms such as machine learning and deep learning, and their "black box" characteristics and robustness problems restrict the implementation of applications. First, the model's interpretability is difficult to meet regulatory and business needs. At the regulatory level, the China Banking and Insurance Regulatory Commission requires financial institutions to provide clear explanations of the decision-making logic of risk models. However, the decision-making process of deep learning models (such as LSTM and Transformer) is difficult to break down, and the credit rating model of a city commercial bank has been suspended by regulatory authorities due to its inability to explain why customer A's rating is lower than customer B. At the business level, risk control specialists have difficulty understanding the risk triggering logic of the model. For example, when the model rejects a loan application from a small and micro enterprise, the specialist cannot determine whether it is due to "insufficient transaction flow" or "related enterprise risks", which affects the efficiency of risk disposal. At the customer level, borrowers have doubts about trust in the "black box" model, and a consumer finance company's inability to explain the pricing basis of loan interest rates to customers has led to a 35% increase in customer complaint rates. Secondly, the robustness of the model is insufficient and vulnerable to abnormal data. In the scenario of data disturbance, a small amount of abnormal data (such as large transactions caused by customer misoperation and noise in third-party data) can lead to model decision bias. A bank's anti fraud model mistakenly identified "large transfers from other places during the Spring Festival period" as fraudulent transactions, resulting in a 12% increase in normal transaction interception rate. In the model adversarial scenario, criminals evade model monitoring by falsifying data (such as brushing orders, forging e-commerce transaction records, and PS false financial statements). In 2024, a certain gang fraudulently obtained a bank loan of 120 million yuan by falsifying transaction data of small and micro enterprises, exposing the problem of insufficient model adversarial capability. In addition, the model iteration lags behind and is difficult to adapt to market changes. For example, when the macroeconomic downturn occurs, the risk parameters of traditional models are not adjusted in a timely manner, resulting in a 25% increase in non-performing loan ratios.

3.3 Resource Level: Dual Constraints of High Technological Investment and Talent Shortage

The implementation of big data risk control requires a large amount of resource support, and small and medium-sized financial institutions in particular face resource bottlenecks: firstly, high technology investment costs. In terms of hardware investment, the average annual procurement cost of distributed storage servers and high-performance computing nodes exceeds 10 million yuan, and the initial investment in the construction of a big data platform by a city commercial bank reached 150 million yuan. In terms of software investment, the licensing fees and customized development costs for big data platforms such as Hadoop and Flink exceed 5 million yuan per year, and the monthly cost of cloud computing resources (such as GPU servers) required for model training exceeds 1 million yuan. In terms of maintenance costs, the average annual expenses for data interface maintenance, system upgrades, and security protection reach 3 million yuan, far exceeding the risk control budget of small and medium-sized banks (approximately 8 million yuan per year). Secondly, there is a severe shortage of versatile talents. Big data risk control requires composite talents with "financial business+data technology+algorithm modeling". Currently, only 15% of the total number of fintech talents in the market possess these three types of abilities simultaneously. Specifically, data engineers need to master data collection, cleaning, and storage technologies, as well as understand financial business logic (such as credit approval processes and anti money laundering rules). The salary level for such talents is 80% higher than that of traditional financial positions; Algorithm engineers need to be familiar with machine learning algorithms and be able to optimize models for financial scenarios (such as adjusting the feature weights of credit scoring models). A top bank offers algorithm engineers an annual salary of 600000-800000 yuan, which is difficult for small and medium-sized banks to afford; Risk control specialists need to have data analysis skills and be able to interpret model output results. Currently, only 30% of risk control specialists have Python data analysis skills. In addition, the talent cultivation cycle is long, and graduates from the interdisciplinary field of "finance+data science" in universities only account for 10% of graduates in finance majors, which is difficult to meet market demand.

4. Future Trends and Practical Suggestions for Risk Control in Big Data Finance

In response to current challenges, combined with technological development and industry demand, the future of big data financial risk control will develop towards the direction of "technology integration, scenario diversification, and regulatory collaboration". At the same time, multiple stakeholders need to collaborate to promote practical implementation

4.1 Technological Integration: Breaking Through the Limitations of a Single Technology and Achieving Collaborative Empowerment

Federated learning technology will become the core tool for data sharing, solving the problem of data silos between financial institutions through a "data available but invisible" model. For example, horizontal federated learning can achieve the sharing of customer behavior data among institutions of the same type (such as multiple city commercial banks), improving the accuracy of credit evaluation for small and micro enterprises; Vertical federated learning can enable the sharing of customer characteristic data among different types of institutions (such as banks and insurance companies), optimizing risk assessment for cross selling. A provincial cooperative, in collaboration with 10 city commercial banks, adopted vertical federated learning to construct a regional small and micro enterprise credit model, which increased the credit evaluation coverage from 55% to 90%. Knowledge graph technology will deepen systematic risk monitoring by constructing a multi-layered network of "enterprise individual institution industry" to track risk transmission paths in real-time. For example, when a real estate company defaults on its debt, the knowledge graph can quickly identify its associated banks (loan amount), trust companies (financing scale), and suppliers (accounts receivable), helping regulatory authorities develop risk isolation measures. In addition, explainable AI (XAI) technology will solve the "black box" problem of models by SHAP （SHapley Additive exPlanations）、LIME（Local Interpretable Model-agnostic Explanations） Using tools to break down the decision-making logic of the model, such as quantifying the impact of each feature on credit scores using SHAP values, and clearly explaining to regulators and customers that 'Customer A's low score is due to three overdue records in the past six months'.

4.2 Diversified Scenarios: Extending from Traditional Risk Control to Emerging Fields

ESG (Environmental, Social, Governance) risk assessment will become a new focus of risk control. Financial institutions will integrate ESG data (carbon emissions and energy intensity at the environmental level; employee compensation and customer satisfaction at the social level; board independence and information disclosure quality at the governance level), construct ESG risk scoring models, and incorporate them into credit approval and investment decision-making processes. For example, the "Green Credit" product launched by Industrial Bank offers a 10% -15% interest rate discount to companies with an ESG rating of A. In 2024, the non-performing loan ratio of this product will only be 0.3%. Dynamic stress testing will achieve accurate risk prediction, combined with macroeconomic simulations (such as a 2% decrease in GDP growth and a 3% increase in CPI) and big data predictions, to automatically generate risk assessment results for multiple scenarios. For example, a state-owned bank, through dynamic stress test, predicted that "the real estate price drops by 20%" would lead to its non-performing loan rate rising by 1.2 percentage points, and adjusted its credit policy in advance (reducing the proportion of real estate loans and increasing the mortgage rate). Cross border risk control will become a key focus for international financial institutions. By integrating cross-border transaction data (such as SWIFT messages and foreign exchange settlement records) and geopolitical data (such as trade frictions and exchange rate policy changes), a cross-border risk monitoring model will be constructed. Through this model, a multinational bank adjusted its cross-border fund allocation three days in advance during a country's exchange rate crisis in 2024, reducing losses by 230 million yuan.

4.3 Regulatory Synergy: Regtech Promotes Seamless Integration Between Risk Control and Regulation

The intelligent compliance system will enhance regulatory response efficiency by automatically parsing regulatory policies (such as anti money laundering notices from the China Banking and Insurance Regulatory Commission and KYC guidelines from the central bank) through NLP technology, transforming policy requirements into risk control rules, and embedding them into business processes in real-time. For example, Minsheng Bank's "Intelligent Anti Money Laundering System" automatically identifies the "high-risk country/region list" in regulatory policies, intercepts suspicious transactions with these regions in real time, and reduces anti money laundering compliance costs by 40%. The regulatory sandbox mechanism will accelerate the implementation of technology. Regulatory authorities will set up "pilot areas" to allow financial institutions to test big data risk control technologies (such as federated learning and knowledge graphs) in a controllable environment. For example, the People's Bank of China has launched a fintech sandbox in Shenzhen, where 12 financial institutions have tested big data credit risk control models, reducing product launch time by an average of 60%. Blockchain technology will enable the sharing of regulatory data, build a distributed data ledger between regulatory agencies and financial institutions, and synchronize risk control data in real-time (such as loan approval records and risk warning information). A certain provincial banking and insurance regulatory bureau has shortened the risk investigation time from 15 days to 2 days . through a blockchain regulatory platform, improving regulatory efficiency.

References

[1] Cui Dongyu The Application of Big Data in Bank Risk Management [J]. Financial Electronics, 2014 (7): 50-51

[2] Li Jikun The Application of Big Data Analysis Technology in Financial Investment Risk Management in the Information Age [J]. China Science and Technology Investment, 2016 (34). DOI: 10.3969/j.issn.1673-5811.2016.34.348

[3] Gong Linan, Party Jie, Sun Bo The Application of Big Data Technology in Financial Information Risk Management [J]. 2024 (8): 416-418

[4] Tao Zhigang Research on P2P Online Loan Risk Management in the Era of Big Data: A Case Study of Company A [D]. Nanjing Normal University [2014-10-09]

[5] Li Huaxiong, Zhou Xianzhong Research progress on cost sensitive credit evaluation under data-driven conditions [J]. 2016