Research on the Application of Alternative Data in Credit Risk Management

Gai Yan *

Department of Southwestern University of Finance and Economics, Chengdu, China

* Corresponding Author Email: shuaigai@smail.swufe.edu.cn

Abstract. With the development of financial technology, traditional credit assessment models have gradually shown their limitations. Especially in assessing borrowers with no credit history or weak credit records. The rise of alternative data provides a new dimension for credit risk prediction, including but not limited to social media behavior, online transaction records, geographic location data, etc. This paper explores the current application status, challenges, and future development trends of alternative data in personal credit risk assessment, and explores the application and effects of various forms of alternative data through different classifications. This paper refers to the relevant literature on alternative data and credit risk management and finds that the application of alternative data can not only supplement part of the information reference to enhance the risk management model but also further provide certain credit credentials for groups that cannot obtain credit services with traditional credit data. It has potential contributions to improving credit risk management and promoting the development of inclusive finance.

Keywords: Alternative data, personal credit, risk management.

1. Introduction

The problem of information asymmetry has always been an important issue in the financial industry. Lenders may not fully understand the borrower's true financial situation and willingness to repay debts, which may lead to improper allocation of credit resources. Borrowers may hide negative information or exaggerate positive information to obtain loans, increasing the risk of lenders. This information difference may cause high-quality borrowers to be overwhelmed by high-risk ratings, thereby increasing the borrowing costs and risks of the entire market, becoming an important issue in credit risk management.

Morse reviewed the relevant literature on fintech loans. She believed that capturing relevant soft information can better analyze loans and applicants can also improve their access to credit or the price of credit, thereby solving the problem of information asymmetry [1]. With the development of digital technology, a large amount of data has been generated from individuals' online activities and transaction behaviors, which is believed to provide additional credit information for borrowers. For example, alternative data such as social media behavior, online shopping habits, and geographic location information are used to supplement traditional credit assessment methods. In contrast, traditional credit scoring model data cannot fully cover the borrower information needs, which leads to a large part of the credit record missing and the limitation of credit scoring services. In the United States, Jennings estimated using the FICO data set that 53 million people cannot obtain credit scores due to insufficient credit records or no records at all [2].

The rapid development of the personal credit market requires financial institutions to accurately assess borrowers' credit risk. Against this backdrop, the use of alternative data has garnered increasing attention. The application of alternative data in personal credit risk management involves leveraging non-traditional information sources, such as social media activities, online shopping records, and geographic location information, to supplement traditional credit scoring models. These data assist financial institutions in comprehensively assessing borrowers' credit risk, particularly for those lacking sufficient traditional credit records. By analyzing alternative data, banks and lenders can predict borrowers' repayment behavior more accurately and make more informed loan decisions.

However, analyzing alternative data also faces some challenges, including data privacy protection, the complexity of analysis, and the reliability of data sources.

This article reviews previous literature research and analyzes the advantages and disadvantages of alternative data in the field of credit risk and its application cases. By deeply exploring the importance of alternative data in personal credit risk management, this article hopes to contribute to the development of the financial technology field.

2. Overview of Alternative Data

2.1. Definition and Types of Alternative Data

Alternative data refers to data sources that are not fully utilized in the traditional credit assessment process. These data sources include but are not limited to social media activities, browsing history, geolocation information, mobile phone usage data, and online shopping records. Compared with traditional credit assessment data (such as credit history, income, and liabilities), alternative data provides more insights into borrowers' behavior and debt repayment ability, helping financial institutions to more accurately assess credit risk. Social media activities reflect borrowers' social behavior and personal habits. Financial institutions can obtain indirect information about the borrower's credit attitude and debt repayment ability. To date, a large number of commercial organizations have developed and used scoring models based on non-traditional data. For example, Experian uses. For rental data, FICO uses utility data, evictions, property values, and other variables, Ant Financial uses individual consumer behavior data, etc.

Alternative data can be divided into multiple types according to their source and nature, including but not limited to social media data (such as user activities and interactions on social networks), consumer behavior data (such as online shopping history and payment behavior), geolocation data (including GPS data and geolocation tags), device usage data (such as mobile app usage and log data), and web browsing history. In the academic literature, more and more researchers are incorporating unconventional variables into credit scoring.

2.2. Differences between Alternative Data and Traditional Credit Scoring Data

Traditional credit data, primarily based on a borrower's financial history—such as credit score, income, liabilities, and assets—has long been the mainstay for banks and financial institutions in assessing credit risk. In contrast, alternative data encompasses a broader range of non-financial information, including but not limited to social media behavior, online shopping habits, geographic location information, and device usage data. These alternative data sources can reveal a borrower's living habits and preferences, offering assessment opportunities for individuals lacking traditional credit history, particularly young people and small and micro business owners. Research by Jagtiani and Lemieux provides relevant evidence that fintech lenders use alternative data to bridge the credit gap left by traditional banks, often serving borrowers who are not typically served by banks but have good creditworthiness [3].

Although some alternative data can play a key role in enabling lenders to fully understand the credit quality of potential borrowers and allow certain consumers to obtain credit that would otherwise be unavailable, the collection and analysis of alternative data also face challenges in privacy protection, data accuracy, and analytical methods. Nevertheless, the application of alternative data has shown great potential in improving the comprehensiveness and accuracy of credit decisions, leading to a revolution in the field of financial technology. By combining alternative data with traditional data, financial institutions can more accurately identify and manage credit risks, promote the development of the credit market, and improve financial inclusion.

3. Application of Alternative Data in Personal Credit Risk Management

3.1. Alternative Data are Used to Enhance Credit Risk Models

Alternative data is used to enhance credit risk models by supplementing information beyond traditional assessment indicators. This data provides insights into borrowers' living habits and preferences, helping credit institutions assess borrowers' debt repayment ability and credit risk. Cheney's research demonstrates that incorporating consumer information, including complete files, as an alternative data item in various scoring models enhances predictive capability [4]. Simultaneously, integrating alternative data into risk models can generate new scores for previously unscorable groups, addressing gaps in traditional credit risk models.

3.2. Constraints of Alternative Data on Moral Hazard

Alternative data can be used not only to analyze consumers' credit quality before lending and to supplement information for credit risk assessment but also to track consumers' economic behavior through various devices. However, the acquisition of such information must be legal and authorized by the user. Analyzing individual economic behavior can significantly contribute to assessing default risk, enabling comprehensive credit assessment and risk prediction throughout the entire life cycle, provided there is proper permission. As Rosamond noted, "all data is credit data" [5]. This "complete picture of credit applicants" obtained by using all available data (such as social networks and Internet search data) can achieve a vision, that is, from top to bottom, from data surveillance to self-observation overlap [6]. "Data subjects" can know where they are suitable to be monitored and tracked, and each of your potential data-generated behaviors may be submitting a "report card" for your credit behavior, whether it is a plus or minus item [7].

3.3. Application Cases and Effect Analysis of Different Types of Alternative Data.

The application of different types of alternative data in personal credit risk management covers multiple dimensions, from social media data, browsing history, and geographic location information to online payment records. Social media data helps assess borrowers' social behavior and credit attitudes; browsing history reflects consumption tendencies and preferences, and implies the level of economic activity; geographic location information reveals the stability of life and work; online payment records provide direct evidence of consumption capacity and financial behavior. Through the analysis of these data, financial institutions can have a more comprehensive understanding of borrowers' credit risks, and thus make more accurate loan decisions.

Based on personal social media data, can help evaluate the borrower's social behavior and credit attitude. Roa et al. first tried to classify users based on tweets on Twitter, using some keywords as factors, such as gender, age, region, work, and political stance [8]. The behavioral data on users' social networks can also provide rich information for judging personal credit. Feida Zhu et al. used personal social media data on Weibo to analyze personal credit and divided potential users into four dimensions, including positive weight dimensions and negative weight dimensions for credit. Users who use "lottery" and "money" and mostly forward tweets have negative weights. The characteristics of this type of people can be inferred that these people are often not rich and usually hope to forward advertising posts to win prizes [9]. Tweets contain words such as "country" and "society", and the posting time is all during the day. There is rarely any late-night behavior. This type of dimension is in the positive weight. This part of the people care about public affairs, are often responsible adults, and may have good credit, and so on. Even some emoticons, sentence patterns, and tones can infer the user's mental state and credit status, which makes a certain contribution to personal credit judgment.

An individual's browsing history can reflect his or her consumption and preferences, and imply his or her level of economic activity. In the era of big data, the storage and collection capabilities of data have significantly improved, encompassing credit score data as well as usage records and digital footprints from users' social networks [10]. The most representative of an individual's browsing history is the user's online purchase and product browsing history, which is also the easiest type of data to collect and obtain by retailers and online shopping platforms. Betty et al. added data from a US retailer to the model to test the user's default rate [11]. The variables introduced were the number of visits to the customer account, the number of visits to the company's website, the number of terms and conditions checked, and the number of mobile devices used, plus two application variables (customer age and brand type) and macroeconomic variables. The trained ROC models all meet commercial standards, that is, the corresponding weight contribution to the default rate prediction model is given in both the forward and reverse directions.

In addition, personal online payment records have become one of the most important data for mobile users at present. Personal online payment records can more intuitively reflect consumption ability and financial behavior. Tobias et al. collectively refer to information about users' consumption behavior as "digital footprints", which include consumption goods, consumption time, age, gender, region, etc., not just the amount of consumption flow, and test the default rate displayed after adding "digital footprints". For example, the default rate of Android users is higher than that of the benchmark category, which is consistent with the fact that consumers who buy iPhones are generally wealthier than consumers who buy other smartphones. At present, there are more and more credit scoring, loan, or insurance pricing companies known to use online payment records, such as Klarna, one of the largest payment service providers in Europe, covering 90,000 merchants and 60 million end customers; LenddoEFL, a company that provides credit scoring, was founded at the Harvard University Entrepreneurship Finance Laboratory; Zhixin, the largest credit scoring provider in China.

4. Opportunities and Challenges

4.1. Opportunities

Advancements in machine learning and artificial intelligence have significantly facilitated the utilization of alternative data in credit risk assessment. These advanced technologies enable financial institutions to process and analyze large amounts of unstructured alternative data, such as social media posts, text messages, and online behavioral records. Machine learning algorithms can identify patterns and correlations from this data to help predict the credit performance of borrowers. Advances in artificial intelligence, such as natural language processing, further improve the ability to extract useful information from alternative data, providing new dimensions and depth to credit risk management.

Alternative data has great long-term potential in the field of credit risk assessment, and its development direction may include deeper application of data analysis technology, integration of cross-domain data, and advancement of privacy protection technology. With the continuous development of big data analysis, artificial intelligence, and machine learning technology, the analysis of alternative data will become more accurate and efficient. In addition, with the increasing emphasis on data privacy and security, new technologies will be developed to protect borrowers' information while allowing financial institutions to use this data to improve the quality of credit decisions. New types of alternative data and their applications may include data generated by Internet of Things (IoT) devices, such as the use of smart home devices, which can provide in-depth insights into borrowers' living habits and consumption behaviors. It may also include biometric data, such as heart rate and activity monitoring, which can reflect an individual's health status and life stability. With the advancement of technology, the analysis of these new types of data will provide more dimensional information for credit risk assessment and help financial institutions make more accurate credit decisions.

4.2. Challenges

The challenges and limitations of the application of alternative data in credit risk management include data privacy and compliance issues, data quality and accuracy, and technical and cost considerations. Privacy concerns are particularly crucial as the collection and analysis of individuals' non-financial data may give rise to legal and ethical considerations. In addition, the heterogeneous and unstructured characteristics of alternative data require the use of advanced data processing and analysis techniques, which may lead to increased resource requirements. Finally, ensuring the accuracy and representativeness of data is another major challenge. Erroneous or biased data may lead to inaccurate risk assessment.

5. Conclusion

This article first introduces the definition and classification of alternative data and summarizes previous literature. It is known that the use of alternative data has become a trend in the development of the big data era. For instance, financial companies have increasingly started to collect and utilize alternative data sources such as social media behavior, online transaction records, and geographic location data. Although they cannot replace traditional information, they can serve as a good supplement. Secondly, based on the classification of different alternative data, we analyzed the different application scenarios and application benefits of alternative data in credit risk management and found that the reference methods of different alternative data are also different.

The application of alternative data brings an innovative perspective to credit risk management and can provide information on borrower behavior and preferences that traditional credit data cannot cover. Despite the challenges posed by data privacy, processing complexity, and data quality, the potential of alternative data in enhancing credit decision-making accuracy and risk prediction capabilities cannot be overlooked. In the future, as technology advances and privacy protection measures strengthen, the application of alternative data in credit risk management is expected to become more widespread and profound.

References

[1] Morse A. Peer-to-peer crowdfunding: information and the potential for disruption in consumer lending, NBER Working Paper, 2015.

[2] Jagtiani J, Lemieux C. Do fintech lenders penetrate areas that are underserved by traditional banks? Federal Reserve Bank of Philadelphia Working Paper, 2018, 18-13.

[3] Jagtiani J, Lemieux C. The roles of alternative data and machine learning in fintech lending: evidence from the lendingclub consumer platform. Federal Reserve Bank of Philadelphia Working Paper, 2018, 18-15.

[4] Julia S. C. Alternative data and its use in credit scoring thin- and no-file consumers. Discussion Paper of Payment Cards Center, 2008.

[5] Rosamond E. All data is credit data: Reputation, regulation and character in the entrepreneurial imaginary. Paragrana, 2016, 25(2), 112–124.

[6] Carney M. Flush with $20M from Peter Thiel, ZestFinance is measuring credit risk through non-traditional big data. Pando, 2013.

[7] Lupton D. The diverse domains of quantified selves: Self-tracking modes and dataveillance. Economy and Society, 2016, 45 (1), 101–122.

[8] Rao D, Yarowsky D, Shreevats A, Gupta M. Classifying latent user attributes in twitter. In SMUC, 2010, 37–44.

[9] Guangming G, Feida Z. Personal credit profiling via latent user behavior dimensions on social media. Institutional Knowledge at Singapore Management University, 2016, 4.

[10] Campbell-Verduyn M, Goguen M, Porter T. Big data and algorithmic governance: the case of financial practices. New Political Economy, 2017, 22(2): 219-236.

[11] Betty J, Garzon R, Jonathan C, Galina A. The role of web browsing in credit risk prediction. Decision Support Systems, 2023, 164.