



Effect Comparison and Ethical Issues of Alternative Data in Credit Scoring

Leyi Shi

School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, China
leyi.shi@outlook.com

Abstract. The traditional credit reporting system has obvious limitations in covering people and updating data. A large number of people who lack traditional credit records find it difficult to obtain financial services and the rise of alternative data provides a new way for credit scoring. This paper summarizes the differences between the application scenarios and effects of personal data and business process data in credit scoring. Through literature comparison, it is found that the two types of data present a complementary relationship in credit evaluation. Personal data is good at capturing individual behavior characteristics, especially suitable for the credit evaluation of borrowers who lack traditional financial records. Business process data reflects the actual operating conditions of enterprises. It is authentic and objective and has unique advantages in enterprise credit evaluation and monitoring. The combination of the two can build a more comprehensive credit scoring system. However, when applying alternative data, people should also pay attention to the ethical issues caused by it, such as privacy protection, algorithm fairness and data governance. In the future, people need to find a balance between technology and ethics, promote data fusion and improve the data governance framework.

Keywords: Alternative Data, Credit Score, FinTech.

1 Introduction

In recent years, with the rapid development of FinTech and big data, the credit scoring system has undergone a diversified transformation. The traditional credit scoring system mainly relies on traditional data such as the central bank's credit reporting and financial statements, which have great limitations in terms of coverage and data update speed. It results in a large number of groups without traditional credit records having difficulty in obtaining financial services. In this context, the rapid rise of alternative data such as digital footprint, social network data and transaction records provides a new dimension for credit scoring and promotes the development of Inclusive Finance.

Existing studies pointed out that social network data can significantly improve the prediction ability of the credit scoring model [1], and mobile app list information is also confirmed to be significantly associated with the borrower's credit risk [2]. At the

same time, digital footprints such as e-commerce credit scores and online shopping behavior can improve the accuracy of default prediction by about 50% [3]. Transaction flow and open bank data also show an effective supplement to the traditional model [4]. However, most of the existing studies focus on the credit prediction ability of a single data type and lack a horizontal comparison of the prediction effect of different alternative data.

This paper aims to systematically sort out the application of personal data and business process data in credit scoring and discuss it from the two dimensions of effect comparison and ethical analysis. When comparing the effects of the two types of alternative data, this study pays attention to the difference in their prediction ability in different application scenarios. When analyzing ethical issues, it focuses on the core challenges involved in the application of alternative data, such as privacy protection, algorithm fairness, and data governance. Through the systematic review and comparative analysis of the existing literature, this paper hopes to provide a theoretical reference for financial institutions to optimize the credit scoring model and regulators to improve the alternative data governance framework. Also, it lays the foundation for further research on the integration mechanism and synergy effect of the two types of data.

2 Application Scenarios of Alternative Data

Alternative data is increasingly widely used in credit scoring and its value is mainly reflected in three aspects: identifying risks, improving credit availability and monitoring post-loan behavior. Different studies from their respective perspectives show the unique role of alternative data in the whole process of credit.

Alternative data is the most widely used in risk identification and default prediction, showing the prediction ability beyond traditional financial data. Hu Jun et al. found that two types of digital footprints, e-commerce credit score and online shopping behavior, can improve the accuracy of default prediction by about 50% [3]. Among them, the e-commerce credit score has the strongest prediction ability. At the same time, the mobile app lists information and has a significant ability to explain the default risk. Adding app information to the model significantly improves the distinguishing ability of the credit risk evaluation model [2]. Combining call data with traditional data can also find that the model based on call behavior characteristics performs well in both AUC and profit [5]. The use of open bank data can significantly improve the accuracy of credit assessment of new customers, thus helping banks make better credit decisions when recruiting new customers [4]. The above studies show that alternative data can capture the behavior characteristics of borrowers from different aspects to provide a supplement to the vacancy of traditional data.

Alternative data also plays an equally important role in improving credit availability and financial inclusiveness. Razavi and Elbahnasawy [6] used the mobile usage data of more than 1500 people. They found that although the mobile usage attribute cannot completely replace the FICO score, it can significantly improve the accuracy of the classification model and provide a way for people without bank

accounts in emerging markets to obtain credit. The research of Muñoz-Cancino et al. further showed that the dynamic combination of social network characteristics and credit history can continuously improve the performance of credit evaluation [7]. Social network information can effectively fill the information gap, especially when the borrower lacks a credit history. For those thin-file borrowers who are rejected by traditional financial institutions due to a lack of credit history, the methods of graph representation learning offer an effective solution. These methods can effectively integrate social network information, significantly improve the evaluation performance of individuals without credit history, and help thin-file borrowers smoothly enter the financial system [8]. It can be seen that the wide application of alternative data gives more people access to credit opportunities and vigorously promotes the development of Inclusive Finance.

In addition, business process data also plays a unique role in post-loan behavior monitoring. The supply chain stability index based on dynamic transaction data can monitor the changes in the operating conditions of small and micro enterprises in real time and provide banks with post-loan warning signals [9]. The comprehensive credit risk index system includes behavior information, regulatory information and policy information, in which indicators such as the number of abnormal water and electricity charges and the number of days overdue for loans can dynamically reflect the operational stability of enterprises [10]. Fayyaz et al. Based on the network science method, people extracted the network characteristics from the relationship between participants in the supply chain finance network [11]. This method improved the accuracy of credit risk prediction and provided decision support for financial institutions' post-loan management in supply chain finance. These studies expand the dimension of credit evaluation from static financial indicators to dynamic business behavior and provide a more timely tool for post-loan monitoring.

The application scenario of alternative data has covered the whole process of credit, from risk identification and decision-making before a loan to behavior monitoring after a loan. Different data types have different emphases, which together constitute a powerful supplement to the traditional credit reporting system.

3 Effect Comparison of Different Alternative Data

3.1 Evaluation of the Effect of Personal Data

The evaluation advantage of personal data is mainly reflected in the capture of borrower behavior and social interaction. Hu Jun et al. [3] found that among the four types of information, including self-report, central bank credit reporting, e-commerce credit score and online shopping behavior, e-commerce credit score has the strongest predictive ability. Their study also showed that online shopping behavior has the same predictive effect as self-report information and still has additional information content relative to e-commerce credit score. This finding shows that even though the e-commerce credit score has integrated different data within the platform, the behavior data from other platforms can still provide additional information, reflecting the complementarity between personal data from different sources. At the same time, the

recognition ability of bookkeeping, takeout, stock and house buying apps to default risk is particularly outstanding. After adding the app list information to the model, the AUC value is significantly improved, indicating that the use of mobile apps can effectively reflect the borrower's consumption habits and asset status [2]. The research of Óskarsdóttir et al.[5] points out that the value of social network information. The model based only on the call behavior characteristics performs best at the profit level and the call behavior characteristics are also the most important predictive variable in other models. It can be seen that personal data provides a unique information dimension for credit assessment by reflecting individual behavior patterns and social influence, which cannot be covered by traditional data. Its value has been verified in multiple scenarios.

However, the application of personal data also has obvious limitations. On the one hand, the acquisition of personal data often involves user privacy, which requires a strict authorization mechanism and privacy protection measures [5]. This limitation becomes particularly prominent after the introduction of relevant laws and regulations. Therefore, the severe challenge facing the industry is how to mine valuable data while ensuring compliance. On the other hand, there may be a problem of timeliness in the evaluation of personal data. The impact of social network characteristics on personal score tends to stabilize after about 6 months [7], which means the evaluation model based on social relations needs to be updated regularly. In addition, personal data may also have the problem of the digital divide. Some vulnerable groups have less access to networks and electronic devices, resulting in an insufficient digital footprint and making it difficult to benefit from it.

3.2 Evaluation of the Effect of Business Process Data

The evaluation advantage of business process data lies in the accurate identification of enterprise operation stability and performance ability. Hu Chen et al. [9] constructed the supply chain stability index based on the dynamic transaction data of commercial banks, revealing the differential impact of customer stability and supplier stability on default risk. Customer stability has a stronger inhibitory effect on the default of small and micro enterprises in the manufacturing industry, while the U-shaped effect of supplier stability suggests that excessive supplier concentration may lead to encroachment risk. This finding verifies the importance of the supply chain relationship in enterprise credit evaluation. Wang and Zhang found that business process indicators such as quick ratio, operating cash flow, abnormal times of water and electricity charges, and overdue days of corporate loans are the most influential credit risk factors through the nuclear feature selection method [10]. The research verified the key role of non-financial information in the evaluation of small and micro enterprises. The research of Fayyaz et al. further shows that integrating the characteristics of the supply chain network into the prediction model can significantly improve the accuracy of credit risk prediction [11]. The characteristics of the network level and the organization level can complement each other, providing a new analysis angle for the credit evaluation of supply chain finance. Compared with personal data, business process data is closer to the actual operating conditions of enterprises and

can reflect their repayment ability and stability in real time. Moreover, the data comes from real transactions, which are highly objective and difficult to tamper with.

However, the threshold for obtaining business process data is high, requiring banks to establish long-term business relationships with enterprises. Some small and micro enterprises lack standardized financial records, which may affect the integrity and accuracy of business process data.

In short, personal data and business process data show a complementary relationship in credit scoring. Personal data is good at capturing individual behavior preferences and social impact, especially for personal borrowers who lack traditional credit reporting data, while business process data tends to reflect the actual operating conditions of enterprises, showing unique advantages in enterprise credit evaluation. The combination of these two types of data can promote a more comprehensive and accurate credit evaluation system. Future research can further explore the fusion mechanism of the two types of data, as well as the optimized application under the premise of privacy protection and algorithm fairness.

4 Ethical Issues of Alternative Data in Credit Scoring

While alternative data enhances the ability to assess credit scores, it also raises a series of ethical issues. These problems mainly involve privacy protection, algorithm fairness and data governance.

Privacy protection is the core ethical issue in alternative data applications. The digitized personal information is often collected and used indefinitely after a single authorization, which makes it difficult to fully protect the legitimate rights and interests of credit reporting objects [12]. When building a credit scoring model using call data, people must face strict privacy regulations and seek a balance between data sharing and privacy protection [5]. Therefore, establishing an effective institutional boundary between data use and privacy protection has become the primary ethical challenge for alternative data applications.

The issues of algorithmic discrimination and fairness cannot be ignored either. Although alternative data can help to expand the coverage of credit, the algorithm mechanism behind it may copy the existing discrimination and reflect the long-standing prejudice in social life [12]. Yang Fan found that the credit evaluation algorithm carries the economic benefits of the platform and the opacity of the algorithm makes it difficult for the evaluated person to obtain effective relief for algorithm bias [13].

Data governance and regulatory difficulties also exacerbate the ethical risks of alternative data. According to the research of Shang Bowen [12], there are internal contradictions in China's credit reporting business management measures and the collection standard of alternative data is not clear. While the platform integrates information providers, users, market regulators and other roles, the traditional self-discipline management and public supervision mechanism of the credit reporting industry are difficult to effectively control [13]. This governance vacuum makes the

relevant platforms lack effective external constraints in terms of information collection and algorithm design and increases the risk of data abuse.

While the application of alternative data brings breakthroughs in technology, it also triggers profound ethical challenges. These problems are caused by the characteristics of the data itself. In the future, people need to further seek a balance between technology and ethics. On one hand, they should clarify the boundary of data rights and interests by improving laws and regulations. On the other hand, they need to inhibit algorithm discrimination through algorithm transparency and various governance mechanisms, to truly bring out the positive value of alternative data.

5 Conclusion

This paper finds that the application of alternative data has covered the whole process of credit. It improves default prediction accuracy in pre-loan risk identification, fills information gaps of borrowers to promote the development of Inclusive Finance and provides enterprise risk early warning in post-loan behavior monitoring. At the same time, personal data and business process data show a complementary relationship in credit scoring. Personal data is good at capturing individual behavior preferences, especially for borrowers who lack traditional credit reporting records, while business process data reflects the actual operation of enterprises and performs outstandingly in enterprise credit evaluation. The combination of the two types of data can build a more comprehensive and accurate credit evaluation system. However, the wide application of alternative data also triggers ethical issues related to privacy protection, algorithm discrimination and data governance. How to find a balance between technology and ethics becomes an urgent problem to be solved.

This paper reveals the differential role of different types of alternative data in credit evaluation and provides a theoretical reference for financial institutions to optimize the risk control model and regulators to improve the governance framework. However, most of the previous literature focuses on the prediction ability of single data, and the research on multiple data fusion applications is not sufficient. In the future, the fusion mechanism and synergy of personal data and business process data can be explored.

References

1. Wei, Y., Yildirim, P., Van den Bulte, C., Dellarocas, C.: Credit scoring with social network data. *Marketing Science* 35(2), 234-258 (2016)
2. Guo, W., Zhou, Z., Qian, C.: Application of mobile app list information in credit risk evaluation: an empirical study based on internet lending platforms. *Chinese Journal of Management Science* 30(12), 96-107 (2022)
3. Hu, J., Li, Q., Zeng, Y., Tian, X., Liu, S.: Digital footprint and default prediction: evidence from consumer credit. *Journal of Management Sciences in China* 28(2), 115-139 (2025)

4. Hjelkrem, L.O., De Lange, P.E., Nasset, E.: The value of open banking data for application credit scoring: case study of a Norwegian bank. *Journal of Risk and Financial Management* 15(12), 597 (2022)
 5. Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., Baesens, B.: The value of big data for credit scoring: enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing* 74, 26-39 (2019)
 6. Razavi, R., Elbahnasawy, N.G.: Unlocking credit access: using non-CDR mobile data to enhance credit scoring for financial inclusion. *Finance Research Letters* 73, 106682 (2025)
 7. Muñoz-Cancino, R., Bravo, C., Ríos, S.A., Graña, M.: On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance. *Expert Systems with Applications* 218, 119599 (2023)
 8. Muñoz-Cancino, R., Bravo, C., Ríos, S.A., Graña, M.: On the combination of graph data for assessing thin-file borrowers' creditworthiness. *Expert Systems with Applications* 213, 118809 (2023)
 9. Hu, C., Li, H.: Research on the correlation between supply chain stability based on transaction data and credit default risk of small and micro enterprises. *Zhongzhou Academic Journal* (5), 36-47 (2025)
 10. Wang, Y., Zhang, Y.: Credit risk assessment for small and micro-sized enterprises using kernel feature selection-based multiple criteria linear optimization classifier: evidence from China. *Complexity* 2020(1), 2394948 (2020)
 11. Rishchchi Fayyaz, M., Rasouli, M.R., Amiri, B.: A data-driven and network-aware approach for credit risk prediction in supply chain finance. *Industrial Management & Data Systems* 121(4), 785-808 (2021)
 12. Shang, B.: Application and enlightenment of alternative data in credit reporting in the United States. *Credit Reference* 39(10), 59-65 (2021)
 13. Yang, F.: Credit scoring as a new type of credit reporting: regulatory dilemmas and basic stance. *Modern Economic Research* (8), 124-132 (2021)
- Should this paper be considered for publication in the conference proceedings?

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

