Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/52/20230720

The Application of Alternative Data in Personal Consumption Credit

Siwen Chen[1,a,†], Renhui Wu[2,b,*,†] and Zhiyu Yin[3,c,†]

[1]No.4 Middle School of Tianjin, Tianjin, 300211, China
[2]China Economics and Management Academy, Central University of Finance and Economics, Beijing, 100081, China
[3]Guanghua Cambridge International School, Shanghai, 201319, China
a. 15022656165@139.com, b. 2020312290@email.cufe.edu.cn, c. yinzhiyu_ray@sina.com
**corresponding author*
[†]All the authors contributed equally.

Abstract: Credit evaluation has been an important part in personal consumption industry and developed since 1960s. Alternative data provides a new view of data mining and gets to be one of the research hotspots in this field. This paper will summarize the application situation and development status of alternative data in personal consumption credit, according to the specific usage in the researches. Alternative data has been more and more widely used in personal credit industry. It has the advantage of improving predictivity and decreasing discrimination. However, meanwhile, because of the incompleteness of relative law and the existence of Lucas Critique, the legality of the employment of alternative data and the reliability of the results are to be proven. Many researches based on machine learning or other AI algorithm couldn't provide enough power of explanatory. This paper will give financial candidates a distinct view of the pros and cons in the research about applications of alternative data in personal consumption credit and leads to an increasing accuracy in credit assessment.

Keywords: alternative data, personal credit, consumption loan

1. Introduction

Since the beginning of 21st century, people have been getting increasing requirement on consumption. In order to satisfy the desire of consumers, personal consumption loan has risen. This means that banks or other financial institutions take the form credit, mortgage, pledge guarantee or other promises to provide loans for personal consumptions so that applicants are provided credit in commodity currencies. One of the most important questions during the process is how to qualify the applicants before lending the money. There've been papers laying great emphasis of the analysis in personal credit evaluation in order to decrease default risk. Qualifying credit appliance helps cooperates in providing services as well as the consumers. Moreover, research has shown that credit scoring is one of the most successful areas applying quantitative analysis. It helps organizations to determine which applicant is creditworthy and should be provided loan, based on mathematic model [1].

Traditionally, the principal subject of personal consumption credit is the banks. Banks consider loan applicants' creditworthiness with a series of parameters. Individuals were classified as 'good' or

DOI: 10.54254/2754-1169/52/20230720

'bad' and provided loan after synthetical appraisal in a face-to-face pattern. With the progress of relative technology, research are trying different methods to predict the default situation of the consumptions. Based on similar database, scholars have been devoting themselves in building up new models and achieving a better ability of explanation, though the conclusions are similar [2].

Different from traditional banking loan, the provision of consumption loan has been expanded to internet cooperates, especially those based on online electronic platform with the technology of P2P (peer-to-peer). These cooperates are usually called FinTech companies. Trading Economics show that the size of consumption credit has reached 556.547 billion yuan in China in 2023, June, 498977 billion yen in Japan in 2023, March, and 1723.381 euro in France, 734.145 billion Canadian dollar in Canada in 2023, May. With the development of the industry, previous credit risk evaluating method can't satisfy the increasing consuming demand of people anymore. One of the focuses is data mining, meaning that data science has been of great importance. Big data has been seen as the research hotspot. With the arrival of information era, both cooperates, and investors are having increasing demand in all kinds of data.

"Science" created special column "Dealing with Data" in 2011 February, concentrating on the discussions about big data in science research, which implies the importance of big data in studies. In 2015, big data analysis and deep learning are implied to be the two points among data science [3]. The role of big data has never been raised such attention on. Both the public and private organizations are collecting all kinds of information in specific area. These data probably contain the sensitive ones, such as state secrets, those relative to international safety and personal privacy, and the untraditional ones, such as fraud examination, marketing, mechanical information and so on, which also arouses the attention of the public.

Those alternative data, which means the data different from traditional one, such as personal history consumption, weather or news, has the value for financial attenders. Nowadays many cooperates still use traditional data to estimate consumers' credit risk, however, traditional data has limitation in reality and university for the conclusions, which will lead to disadvantages in estimation. Alternative data has four characteristics—big volume, sizable variety, high velocity and great value so that it helps cooperates to process a more complete and more realistic evaluation in credit risk. But on the other hand, because of the huge volume and complication, how to apply alternative data in personal consumption credit is the one this paper is to answer.

The application of alternative data in finance is a wide field, among which is the credit evaluation. In recent years, research around alternative data in this domain has been a hot issue. Many researchers have tried different kinds of alternative data and examined whether there exists some that gives help to improve predictivity [4]. This paper will review the literature in the application of alternative data in personal consumption credit evaluation, summarize the conclusions of the papers and predicts the future development and restriction in this field. It is hoped that this paper will provide a summary in this area and a clear knowledge for financial attenders and improve the accuracy of personal consumption credit assessment.

2. Literature Review

2.1. A Brief Summary

Before the occurrence of the concept—alternative data, personal credit industry is used to qualify loan applicants with traditional financial data. FICO score, introduced by the company named FICO, serves as the most common indicator in personal credit risk. It mainly depends on three core factors: the debt level of consumer, credit history and income situation. Different financial institutions have their own fashions and scoring standard, combining with FICO score, to make decision in issuing a credit. Because of the restriction of data set, financial institutions and scoring organizations have been

DOI: 10.54254/2754-1169/52/20230720

trying to optimize traditional methods. With the development of time and the progress of technology, more and more data has been mined out and the academia has been making effort on the improvement of traditional models.

Lending club, a lending platform based on untraditional data, provides its unique credit score, with a correlation rate of 35%, however, also does well in prediction [5]. In 2016, Equifax created FICO score XD, cooperating with FICO and LexisNexis Risk. In 2019, Experian, with FICO and Finicity, built up UltraFICO score. Many evaluating models depending on machine learning algorithm also use numerous alternative data, besides traditional ones. Even though this type of studies may not find the key factors, they did make improvements in predictivity [6].

Generally speaking, the usage of alternative data has been a fashion in this industry. Alternative data does help in providing services for unscored applicants and improving the accuracy in credit rating [7].

2.2. Digital Footprint

A number of scholars have dealt with digital footprint, named after the data left behind when users have been online, as alternative data, to model personal consuming default. By introducing 'whether a borrower has at least one contact who has more than ten calls with borrower' and 'whether a borrower has at least one online shopping delivery address or not' as new digital footprint variable, the forecast of overdue payment can be improved. The implication is that if a borrower does have a constant relation or delivery address, the delay will affect the accessibility of his/her social capital or physical dwelling due to the default call [8]. 'The device used and time when a borrower applying a consumption loan', as alternative data, also has the predictability of default rate. After taking this digital footprint into consideration, the default rate of users significantly decreased. It also does well as a supplementary of traditional credit score [9]. The usage of mobile phone, to be specific, phone-call-related data, such as time stamp, spot and duration time of a call, as untraditional data, behaves better than traditional ones in model [10].

2.3. Discrimination

Race or gender should never be an explanatory variable in measuring consumption credit risk, however, in fact, the discrimination in credit industry is not a new issue. In the early 21st century, an independent female lender has a gap in the probability of rejection with the one that has a co-applicant. A female applicant of a minority race will be rejected than any other type of people [11]. Besides, lending services facing African or Latin applicants have increased, but the service quality decreased at the same time [12].

Some papers concludes that after taking advantage of alternative data in personal credit industry, discrimination will be decreased. Based on machine learning algorithms, credit business supported by untraditional data has a one third lower discrimination level than traditional face-to-face business. This kind of algorithm lenders have nearly no discrimination in rejecting loan applications. On the other hand, concerns about novel discrimination were raised by some research. If we see working experience as alternative data, since it's not a typical traditional variable, a part-time female will have a lower credit score, but for such an individual among a vulnerable group, this will absolutely increase the inequality [13].

3. Predictions and Comments

Research about alternative data do have some insufficiencies. The root reason is that alternative data is usually independent of traditional financial data. The question is even though statical results from alternative data may indicate some potential explanatory power, even though there's no direct

DOI: 10.54254/2754-1169/52/20230720

relationship. Thus, the deep reason can seldom be revealed sensibly, especially for those mainly focusing on machine learning or dealing with a large number of variables. These kinds of research usually lack of the explaining power.

The traditional evaluating method taken by banks has lots of limitations. For those with low credit score, traditional model will result in nearly the same default rate, which suggests that applied on low-credit-score users, traditional model shows weak ability of distinction. Meanwhile, traditional method relies on credit history, but about 11% adult America don't have enough credit history. Moreover, another 8.3% have credit history but don't have credit score. This means that 45 million American adults can't enter a traditional personal consumption loan [14]. Comparatively speaking, credit scoring with alternative data admits those consumers lack of traditional financial data to finish credit rating. Therefore, alternative data can support more general conclusions.

If we consider the credit industry as an economic question and comprehend it in an equilibrium, we can't avoid worrying about Lucas Critique. We may take the evaluating method as the policy and use game theory to explain. Alternative data includes data that doesn't directly associate with credit, such as the brand or usage time of devices [4]. This instantly raises the concern that applicants may pretend to be high-scored in credit by altering the information. The long-term reliability isn't promised.

Although there has been research proved that there exists an evaluation method that will not leak out privacy, this issue has long been a concern in this field. The credit scoring system in USA originated from the 1960s, but not until the publication of "California Consumer Privacy Act" (CCPA) was the collection of private information canonicalized. By the end of 2005, most of the countries in the world still haven't published law relative to data protection in this field [15]. In recent years, the increasing usage of alternative data indicates that the credit scoring authorities need richer data to achieve more accurate prediction, which requires advanced rules.

Alternative data is helpful in eliminating discrimination and expand the credit industry, and therefore improve the development. Comparative to traditional face-to-face one, the assessment fashion based on data is independent of gender and race. Only the data behaves well in predicting credit risk are considered in the process of assessment. However, considering that some alternative data has a positive correlation rate with advantaged or disadvantaged population, new discrimination may be raised out. One the other hand, the results of discrimination-relative studies vary from time to time. The reason is that the industry has always been trying to avoid inequality, in all aspects, because of the pressure from public voice or the political power, or the demand of more borrowers. Thus, similar to the discussed one, researchers should construct more convincing explanations.

4. Conclusion

This paper studies the application of alternative data in personal consumption credit. In recent years, alternative data has been paid much attention on. It is revealed that alternative data has done a lot of help in consumption credit risk prediction and used in the industry. Alternative data is able to decrease discrimination and expand target users. On the other hand, alternative data means novel data, which means that if more data to be mined and taken advantage of, larger extent of privacy will be invaded, but relative laws haven't been fully accomplished. Moreover, because of the existence of Lucas critique, when data that is not directly relative to credit is included in the process of credit scoring, consumers may change their behavior according to the policy. If that is the case, alternative data will no longer has the explanatory power.

Although many research take a positive opinion on the development of alternative data's application and, indeed, alternative data is getting widely-used by many organizations, as a fresh field, it requires the completion of law and more examination. Therefore, it is advised that agencies would issue laws and regulations restricting the financial institutions' collection of alternative data. It is also

DOI: 10.54254/2754-1169/52/20230720

hoped that more research rose to ensure the efficiency of alternative data. On the other hand, many studies depending on machine learning or big data are usually lack of the interpretation of key variables and doesn't possess enough explanatory power. That's what is to be perfected.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] *Noh, H.J., Roh, T.H. and Han, I., (2005) Prognostic personal credit risk model considering censored information. Expert Systems with Applications, 28(4), 753–762.*

[2] *Leo, M., Sharma, S. and Maddulety, K., (2019) Machine learning in banking risk management: A literature review. Risks, 7(1), 29.*

[3] *Najafabadi, M.M., Villanustre, F. and Khoshgoftaar, T.M., (2015) Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1), 1.*

[4] *Lili, D., Han, J. and Shi, J., (2022) Digital footprints as collateral for debt collection, Working Paper.*

[5] *Fitzpatrick, T. and Mues, C., (2021) How can lenders prosper? Comparing machine learning approaches to identify profitable peer-to-peer loan investments. European Journal of Operational Research. 294 (2), 711–722.*

[6] *Sreesouthry, S., Ayubkhan, A., Rizwan, M.M., Lokesh, D. and Raj, K.P., (2021) Loan prediction using logistic regression in machine learning. Annals of The Romanian Society for Cell Biology, 25(4), 2790 – 2794.*

[7] *Bradford, T., (2023) "Give Me Some Credit!": Using alternative data to expand credit access. Payments System Research Briefing, 1-6.*

[8] *Dai, L., Han, J., Shi, J., Zhang, B., (2020) Digital footprints as collateral for debt collection. Working Paper.*

[9] *Berg, T., Burg, V., Gombović, A. and Puri, M., (2020) On the rise of FinTechs: Credit scoring using digital footprints. Review of Financial Studies 33(7), 2845-2897.*

[10] *Björkegren, D. and Grissen, D., (2019) Behavior revealed in mobile phone usage predicts loan repayment. Working Paper.*

[11] *Pager, D. and Shepherd, H., (2018) The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. Annual Review of Sociology. 34(1), 181-209.*

[12] *Bartlett, R., Morse, A., Stanton, R. and Wallace, N., (2019). Consumer-lending discrimination in the FinTech era. National Bureau of Economic Research. Volume 143, Issue 1,Pages 30-56.*

[13] *Pasquale, F. and Citron, D.K., (2014) Promoting innovation while preventing discrimination: Policy goals for the scored society. 89 Washington Law Review, 1413, 1413-14.*

[14] *Brevoort, K.P., Philipp, G. and Michelle, K., (2016) Credit Invisibles and the Unscored. Cityscape, 18(2), 9–33.*

[15] *Jentzsch, N., (2006) The economics and regulation of financial privacy: An international comparison of credit reporting systems. Physica Verlag, Heidelberg, 19.*