The Value of Alternative Data in Credit Risk Prediction: Evidence from a Large Field Experiment

Tian Lu
Carnegie Mellon University, lutiansteven@gmail.com

Yingjie Zhang
University of Texas at Dallas, yingjie.zhang@utdallas.edu

Beibei Li
Carnegie Mellon University, beibeili@andrew.cmu.edu

The Value of Alternative Data in Credit Risk Prediction: Evidence from a Large Field Experiment

Completed Research Paper

Tian Lu
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA
tianlu@andrew.cmu.edu

Yingjie Zhang
The University of Texas at Dallas
800 W. Campbell Road, Dallas, Texas
Yingjie.Zhang@UTDallas.edu

Beibei Li
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA
beibeili@andrew.cmu.edu

Abstract

Recently, the high penetration of mobile devices and internet access offers a new source of fine-grained user behavior data (aka "alternative data") to improve the financial credit risk assessment. This paper conducts a comprehensive evaluation of the value of alternative data on microloan platforms with a large field experiment. Our machine-learning-based empirical analyses demonstrate that alternative data can significantly improve the prediction accuracy of borrowers' default behavior and increase platform profits. Cellphone usage and mobility trace information perform the best among the multiple sources of alternative data. Moreover, we find that our proposed framework helps financial institutions extend their service to more lower-income and less-educated loan applicants from less-developed geographical areas – those historically disadvantaged population who have been largely neglected in the past. Our study demonstrates the tremendous potential of leveraging alternative data to alleviate such inequality in the financial service markets, while in the meantime achieving higher platform revenues.

Keywords: Credit risk, alternative data, bias, machine learning, prediction, welfare

Introduction

Assessing credit risks has become a top priority in financial activities. An accurate prediction of credit risks is rewarded by sizable financial returns. Traditional financial firms like banks assess their users' credit risks by keeping tracks of their historical financial records. The introduction of FinTech has enabled more platforms and more users to get involved, but it also brings more challenges to the emerging financial services like microloans. One main reason is the lack of collaterals (i.e., something pledged as security for repayment of a loan, to be forfeited in the event of a default). This limitation lets financial service providers and lenders face a higher credit risk (Serrano-Cinca et al. 2015). The problem is growing more severe in developing countries because of the lack of individual's official credit records. The above challenges in credit

risk assessment could become more non-trivial with the current trends in FinTech, where many financial services have recently turned their attention to the vast untapped pool of unbanked users. Without an effective approach to evaluate those thin-file users, the unbanked are hampered from getting financial supports.

To achieve a satisfactory default risk prediction, practitioners and researchers have made great strides in seeking proper features, including loan characteristics (Serrano-Cinca et al. 2015), borrowers' "hard" characteristics such as credit history and demographic information (Iyer et al. 2015), and their "soft" characteristics such as social capital (Lin et al. 2013), linguistic features (Gao and Lin 2013), and personal appearance (Ravina 2007). We term these commonly-adopted information as "conventional data." In recent years, the high penetration of mobile devices and internet access offers a new and unparalleled source of fine-grained user behavior data (such as loan applicants' social media activities, mobile phone usage activities, app usage and browsing logs, location trajectories, or shopping behaviors), which we term as "alternative data." The conventional data typically cover data from a credit bureau, a credit application, or a lender's own files on existing data. The alternative data, instead, often come from public social media sites or private applications and devices and may not directly relate to a consumer's credit behavior. Nevertheless, such new, rich source of data have been shown to be able to reflect people's personalities, psychological, and socioeconomic status. Meanwhile, people's personalities, psychological, and socioeconomic status have likewise been found to highly correlated to their obligation fulfillment including repayment behaviors (e.g., Guzelian et al. 2015; Morse 2015). Thus these data could show significant potential to complement the conventional data in enhancing the accuracy of existing credit risk assessment (Carroll and Rehmani 2017). Moreover, recent studies have found that credit risk prediction suffers from unintended bias due to potential correlation between the input (observed) features and the sensitive attribute (such as race, gender, or income) (e.g., Dobbie et al. 2018, Fu et al. 2018). Such correlation to some extent is due to a lack of control in the unobservable. Leveraging alternative new source of behavioral data can help better control for the individual features that were previously omitted from the model, hence reducing bias in financial risk prediction.

Furthermore, prior work (e.g., Serrano-Cinca et al. 2015) on financial credit risk prediction mostly used training data that are heavily biased towards the successfully approved loan applicants – those applicants whose credit risks are perceived to be low enough to have successfully received a loan approval ("approved samples" hereafter). Whereas, for other applications who are initially perceived to be high-risk, their applications will be immediately rejected, hence no further loan payment data about these applicants will be recorded or included in model training later ("excluded samples" hereafter). Obviously, learning credit risk models using approved samples alone can be rather problematic. Compared with the true population of the loan applicants, the approved samples tend to have lower probabilities to default and may demonstrate significantly different characteristics (e.g., higher income, better educated). The patterns or relationships learned from such biased samples may have limited generalizability, and hence may lead to poor predictive performance for new applicants. Moreover, if the initial approved samples are biased (intendedly or unintendedly) towards certain sensitive attributes, such error could be further amplified by training with the biased sample.

Motivated by the current challenges of financial service markets, especially the microloan platforms, the present paper aims to answer the following questions:

1. Can new alternative data help improve the predictive performance in microloan credit risk assessment? Moreover, which type of information is the most valuable?
2. Can new alternative data help alleviate concerns from training sample bias (i.e., using approved samples only for model training)?
3. How can we leverage alternative data to achieve a more accurate risk assessment, and ultimately, better financial performance for microloan platforms?

Note that it remains costly for financial service providers to acquire, store, and process information (Loufield et al. 2018). To obtain individual's information from multiple sources, financial service providers have to establish close collaborations at a time with third party data providers, such as social media providers, telecommunication and mobile network operators, and other specialized data vendors. Moreover, the increasing size and data complexity of the alternative (and mostly, semi-structured or unstructured) information often require sophisticated techniques and multiple players to turn it into something of value. Last but not the least, the financial service provides might face potential information privacy concern and

security issues. Therefore, the ability to evaluate the credit risk of borrowers with minimal-accessible information is key to the burgeoning financial market. This is one unique contribution of our study. Our goal is to examine, and compare, the value from various types of alternative data for credit risk assessment in the microloan market. We aim to provide loan platforms important managerial insights on what information is most valuable, and hence, should be efficiently combined with conventional data to maximize profits and minimize potential prediction bias. To the best of our knowledge, no existing study has modeled such a holistic process of an individual loan borrower's behavior with these cross-channel, rich sources of information, let alone the comparisons among different dimensions of these alternative data.

In this study, we cooperated with a major microloan company in China to conduct a large randomized field experiment from December 2nd, 2017 to December 22nd, 2017. During the experimental period, the platform did not enable any filtering strategy, but instead, approved loan applications from all applicants (as opposed to the usual situation where the platform approves approximately only 40–45% applicants based on the predicted risks of the applicants). We then collected a fine-grained dataset with detailed records of individual loans over time since the approval. It is worth noting that by approving all loan applications and tracking borrowers' payment behaviors over time, we are now able to observe the counterfactual cases – those applicants who would otherwise be rejected under a normal situation. This unique setting enables us to form an unbiased sample for model training by including behavioral patterns from the entire loan applicant population, as well as to evaluate our model under various counterfactual scenarios that would otherwise be unobserved.

Also note that when calculating the profits of a microloan platform, we should consider not only the losses from default but also the revenues from a delinquent fine payment. Therefore, unlike previous studies with default indicators merely (e.g., Duarte et al. 2012), we define a multiclass categorical credit risk indicator that captures borrowers' repayment behavior of being `delinquent` (at least one installment was not paid in time), `delinquent but not in default` (at least one installment was not paid in time but the loan was fully paid finally), and `default` (loan was not fully paid finally). We also consider the `repayment rate` and `profit per loan` (or `loan profit`) as alternative numerical credit indicators. Another uniqueness in our dataset is that it describes an individual loan/borrower from multiple sources. Inspired from the existing literature, we construct and extract more than one hundred features, covering four main categories: commonly-adopted conventional data (e.g., borrower demographic and socioeconomic characteristics, credit history, and loan attributes), online shopping, cellphone usage and mobility traces, and social media usage. We apply the above features to train different state-of-the-art machine learning models and identify the value of different sources of information on credit risk assessment, under the contexts of delinquent and default cases. For comparison, we conducted similar analyses using a filtered sample (i.e., with approved applicants only) collected from the same platform. The comparison between approved sample and the full applicant sample enables us to identify the potential financial impact of training sample bias.

Our empirical analysis yields several interesting findings. *First*, our prediction results show that among the four sets of features we constructed, cellphone usage and mobility trajectory features present the highest predictive power, followed by shopping characteristics. For social media users, social media presence and sentiment are also valuable in predicting users' repayment behavior. At a more granular level, interestingly, we observe that the consumption on gaming-related products (e.g., game app usage, amounts spent on game cards) ranks the top among all alternative-data-related features.

Second, the platform welfare analysis indicates that, loan permission strategy based on credit risk prediction with cellphone usage and mobility traces information also yield 15% more economic gains to the microloan platform than simple on the conventional features. The platform can achieve a further 7% economic gains when we make loan approval decision based on credit risk prediction with all the feature sets. We also find that under certain loan approval rates, loan permission based on the predicted delinquent-but-not-default probabilities or numerical repayment rates and loan profits, may yield higher economic gains than the current industry practice, which is primarily based on the predicted default probabilities. This finding stems from and confirms that on the premise of accurate risk prediction with alternative data, lending to the borrowers with a certain level of delinquency risks despite relatively high default risk can also yield positive economic gains.

Third, we demonstrate bias indeed exists if using only approved samples or using only conventional data for model training, which can lead to significant loss in prediction accuracy as well as economic gains of

microloan platforms. Interestingly, we find that these existing approaches tend to favor higher-income and more-educated applicants from areas with more developed economy. By leveraging alternative data, we find that microloan platforms are more likely to include lower-income and less-educated loan applicants from less-developed geographical areas – those historically disadvantaged population who have been largely neglected in the past. Our study hence demonstrates the tremendous potential of leveraging alternative data to alleviate such inequality in the financial service markets, while in the meantime achieving higher platform revenues.

The contributions of our study are multifold. *First*, this study is the first to investigate the predictive power and financial value of multi-dimensional alternative data, including cellphone and mobile app usage, mobility trajectories, shopping behavioral, and social media information, on borrowers' credit risk assessment and microloan platforms' revenues.

Second, while previous studies simply focused on default probability, this study contributes to the literature with more sophisticated credit risk indicators. We identify separately delinquent and delinquent-but-not-default behavior. The extra information allows us to examine the trade-off between profits from delinquency and losses from default.

Third, our unique field experimental setting enables us to examine the "what-if" counterfactual scenarios under different risk assessment strategies. By comparing the final rankings of loan applicants based on predicted risk scores (hence the recommended approved loans) generated by different models, data or training sets, we are able to interpret not only "what" strategies but also "why" these strategies perform better and lead to higher economic returns to platforms. Such interpretability is critical and enables us to understand where potential prediction bias and economic loss may come from, and how we can address them.

Fourth, we offer a practice-oriented approach for microloan platforms to easily adopt a cost-effective solution based on what is easier to implement in practice. For example, training sample bias has been a major challenge from both prior research and industry practice due to practical data limitation. We demonstrate that incorporating alternative data can largely offset potential economic loss caused by the training sample bias and can lead to a significant improvement in platform revenues even when platforms have no access to the unbiased full sample of loan applicants during model training.

Context, Experiment, and Data

Field Experiment Design, Setup and Data

We conducted an experiment by cooperating with a Chinese microloan company from December 2nd, 2017 to December 22nd, 2017. During the experimental period, the platform randomly selected 40% from all loan applicants and approved all applications without any filtering strategies.

We obtained the full experimental sample containing 5,214 loans from 5,214 unique borrowers. The dataset includes three parts: (1) Conventional Data: including (i) loan attributes (i.e., loan amount, loan term, and interest rate), (ii) the corresponding borrower's demographic and socioeconomic characteristics (i.e., age, gender, education level, income level, marriage status, number of children, job, and contact information of at least one family members or close relationship), (iii) the borrower's self-reported purpose of the loan, and (iv) the borrower's loan history on the focal and other microloan platforms. (2) Alternative Data: for each loan applicant, the platform collects alternative sources of information covering the borrowers' personal behavior during the six months before loan applications upon their authorization. The platform considers the following three alternative data: (i) online shopping records (i.e., order time, product name, price, quantity, product type, and receiver information) from taobao.com and jd.com, which are the two largest online shopping platforms in China; (ii) cellphone-related records (i.e., call history, cellphone usage, detailed mobile app usage, GPS mobility trajectories); and (iii) social media usage (i.e., whether the borrower has accounts, (if yes) all posted messages with timestamps, and social media presence including the number of fans, followings, received comments, and received "likes") at weibo.com. (3) Repayment Information (collected during the 13 months after the loan issuance): including repayment behavior (i.e., due date and repayment rate) of each loan at the installment level (monthly). On the focal platform, borrowers must repay installments every month until the loan is paid off. If a loan was not paid off three months after the loan due date, the loan default was confirmed.

Theoretically, a loan becomes *delinquent* when the borrower makes a payment late, while a loan goes into *default* if the borrower misses several installment payments over a period and fails to keep up with ongoing loan obligations. A typical strategy most financial service providers are adopting is to impose a (relatively high rate of) fine on borrowers for a delinquent payment. That is to say, although credit risk arises financial losses, from the platform profit perspective, a delinquent borrower with a certain level of credit risk could possibly be valuable if she repays the installment and fine. Hence, we deem that the `goal` of a comprehensive prediction on credit risk should identify accurately not only borrowers with low default probabilities (to avoid high credit risk) but also those who are likely to delinquent but not in default (to increase potential revenues). Therefore, we considered three loan-level indicators to measure borrowers' credit risks and repayment behavior: (1) `Delinquent/Default`: it is a multiclass categorical factor (1 = not delinquent, 2 = delinquent but not in default, and 3 = default). Note that a non-delinquent loan means there was no delinquent installment repayment across the entire repayment period. (2) `Repayment Rate:` it measures the proportion of repaid monthly installments. Unlike the categorical factor which captures an overall repayment performance of a loan, this numerical indicator delivers finer-grained details of repayment behavior as suggested by Drozd and Serrano-Padial (2017). (3) `Loan Profit`: we calculate the platform's profit from one loan by taking into account both the revenue and cost. The revenue from each loan includes gains of interests and possible penalty (fines) for late payment (delinquency). The cost includes the loss of principal capitals and potential opportunity cost from a default or delinquent loan.

In our sample, 784 (15.04%) loans were not delinquent, 1,329 (25.45%) had been delinquent but not in default, and 3,101 (59.47%) defaulted. The average repayment rate was 0.606, and the average profit per loan was -729 RMB (approximately 109 USD). 639 borrowers had never paid their loans. 2,375 loans yielded positive profits less than 1,000 RMB (approximately 150 USD, majorly are revenue from the interests), and totally 1,615 loans resulted in financial losses between 2,000 RMB and 4,000 RMB. Moreover, the average delinquency duration of paid installments is approximately 29 days. In sum, the full applicant pool on the focal microloan platform performed high risk.

Constructed and Extracted Features

Based on our unique dataset, we construct and extract 117 features covering four main categories.

Commonly-adopted Conventional Characteristics (`Fc`): We construct most of the commonly-adopted characteristics in prior studies. Concretely, for borrowers' demographic and socioeconomic characteristics, we code age, gender, education level, marriage status, number of children, house ownership, type of occupation, monthly income, the disposable personal income (DPI) of borrowers' living cities in 2017, and whether having insurance. For loan attributes, we have loan amount, loan period, interest rate, and income-to-debt ratio. We also code three features indicating whether the installment payment due date was during holidays, weekends, or the beginning/end of a month respectively. For loan histories, we code whether borrowers had microloan experience with the focal platform and other platforms, whether they defaulted in prior microloans, contacting frequency with microloan platforms, as well as whether they have credit cards and regular payment behavior. Besides, we extract borrowers' self-reported loan purpose via text mining techniques and code it as a binary feature, indicating whether the loan was used for (high) consumption or for needs of dealing with emergencies (e.g., healthcare, accidents, and business turnover) (1 = consumption, 0 = otherwise).

Online Shopping Characteristics (`Fs`): We start from several general characteristics, including the total amounts transferred out and in on Alipay (i.e., the main payment tool on Taobao). We then decompose borrower's shopping behavior by considering different types of products. For each type, we aggregate the total number of transactions, average amount and quantity of purchased products, and diversity of purchased product categories. Specifically, we consider four types, including durable goods, game products (suggesting indulgent intention (Kim et al. 2008)), special products such as medicines, caffeine, and tobaccos (Amonini and Donovan 2005), and products for others.

Cellphone Usage and Mobility Traces (`Fp`): Several studies have successfully built the relationship between cellphone usage and mobility data and people's social-economic status (e.g., Blumenstock et al. 2015), friendship and social ties (e.g., Cho et al. 2011), and psychological and personality traits (e.g., Chittaranjan et al. 2013). This set of features covers three aspects. The first one includes cellphone call and message related features from call log, including average monthly cellphone expenses, frequency and duration of

incoming and outgoing calls, number of sent and received text messages, and frequency of getting contacts with the close relationship registered at loan application stage. Second, inspired by Ma et al. (2018), we extract app usage features, including average weekly frequency and duration of different kinds of app usage (e.g., financial and payment app, news app, game app, entertainment app, and social media app), as well as data traffic usage. The third subset of features captures individual's offline trajectories (Tan et al. 2016)) extracted from the fine-grained GPS data. Specifically, we calculate the number of borrowers' visited cities, and the average weekly frequency of appearance in different kinds of locations including official buildings/areas, commercial, entertainment/recreational, and public service places.

Social Media Information (Fm): In our sample, there are 1,618 (31.03%) borrowers who had registered on the microblog. We consider two types of social-media-related features. Similar to Ge et al. (2017), we first extract presence features including number of fans, followings, reciprocities, received comments, and "likes". Moreover, we extract sentiment valence (from -1 of extremely negative to 1 of extremely positive) and sentiment variance of each textual message the borrower post on the microblog (Wei et al. 2015).

Predictive Power of Alternative Data

Machine Learning Framework for Credit Risk Prediction

We implement multiclass classification (one-versus-all) and numerical regression algorithms for the pre-defined categorical and numerical outcome variables respectively. Specifically, we implemented diverse widely-accepted machine learning models, including logistic and linear regression (L&R), support vector machine (SVM), k-nearest neighbors (k-NN), multi-layer perceptron (MLP), two ensemble methods: random forest (RF), and XGBoost. The XGBoost model, proposed by Chen and Guestrin (2016), is a boosting tree ensemble model that consists of multiple regression trees. The mechanisms of nonlinear optimization, sparse regularization, and iterative boosting endow XGBoost with advantageous learning abilities such as low computation complexity, high generalizability, and desirable prediction accuracy (Chen and Guestrin 2016). These models have been widely employed in credit risk prediction (e.g., Abdelmoula 2015; Serrano-Cinca and Gutiérrez-Nieto 2016). Followed by Cui et al. (2018), we use a 10-fold cross-validation approach to tune the hyper-parameters of our models. We use grid search to choose the value of the hyper-parameter that gives the best performance.

To delve into the predictive value of different feature categories, we first include each individual feature category (Fc, Fs, Fp and Fm) respectively, of which conventional features (Fc) are the benchmark. Then, we combine features from different categories. Because only approximately one in three borrowers in our sample have microblog records, we evaluate the prediction performance with two combinations: combining all features except microblog-related features ($Fc \cup Fs \cup Fp$) for the whole sample, and combining all four categories ($Fc \cup Fs \cup Fp \cup Fm$) for the micro-blogger subsample. Notably, to address the potential self-selection issue regarding the microblog usage, we followed the two-stage framework proposed by Heckman (1977) to first predict whether to use microblog or not and then include it in the second stage of credit risk predictions.

All features are normalized to ensure the comparability of results. Due to the imbalanced distribution of each class in the categorical credit risk indicator, we implement the over sampling strategy for those minority class to balance the trade-off among all classes. To avoid overfitting issues, we first implement feature selections via the L1-norm-based regularized sparse model before training any machine learning models. We randomly partition our sample into two parts, two in threes (3,476 loans) serving as training sample, and the rest 1,738 loans as validation sample. We then applied 10-fold cross-validation to train various models based on the training sample and evaluate the prediction performance using the validation sample.

We applied several commonly adopted metrics to evaluate prediction performances. Specifically, we considered precision, recall, and F1 score for the multiclass categorical risk indicator (delinquent/default), and mean absolute error (MAE), root mean squared error, and R^2 for numerical risk indicators (repayment rate and loan profit). R^2 also indicates the explanatory power of information on credit risk assessment.

Prediction Performance

Tables 1(a), 1(b) and 1(c) report the prediction performance on the proposed categorical credit risk indicator, numerical `repayment rate`, and `loan profit`, respectively. First, we observe that among all machine learning models we have considered, XGBoost shows the best performance. This finding is consistent across metric and different combinations of features sets. Second, by comparing the prediction performances (e.g., R^2) on different credit risk indicators, we find that machine learning models have explicitly better performance on `repayment rate` than `loan profit`, which implies that many unobservable interferences may exist when directly predicting loan profits.

Regarding the predictive power of diverse feature sets, our results yield several interesting findings. First, Table 1 indicates that the predictive powers of the three alternative feature sets (i.e., `Fs`, `Fp`, and `Fm`) are significantly stronger compared to the benchmark conventional features. Among all individual feature sets, cellphone usage and mobility traces features (`Fp`) have the highest predictive power. Online shopping features (`Fs`) improve the prediction performance significantly even though their influence is weaker than the features from telecommunication carriers. For social media usage borrowers, social media features likewise have strong predictive power on credit risk prediction. Furthermore, although we found strong predictive power of the new feature sets from alternative data on credit risk assessment, we do not observe any significant improvement when we combine different feature sets. It suggests that the microloan platform may only need access to feature sets from one alternative information source such as telecommunication carriers, which can lead to satisfactory credit risk assessment with the minimum cost. The paired t-tests on the performance of different alternative feature sets against the conventional feature set showed the differences were overall significant.

Table 1. Prediction Performances on Credit Risk Indicators

(a) On Categorical Risk Indicator

Model	**L&R**			**SVM**			**k-NN**		
Feature set	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Fc	0.357	0.362	0.359	0.360	0.363	0.361	0.338	0.319	0.328
Fs	0.401	0.407	0.404	0.397	0.402	0.399	0.390	0.371	0.380
Fp	0.536	0.583	0.559	0.548	0.587	0.567	0.555	0.550	0.552
Fm	0.525	0.569	0.546	0.508	0.569	0.537	0.541	0.490	0.514
Fc∪Fs∪Fp	0.538	0.586	0.561	0.556	0.590	0.572	0.559	0.554	0.556
Fc∪Fs∪Fp∪Fm	0.540	0.588	0.563	0.557	0.592	0.574	0.560	0.555	0.557
Model	**MLP**			**RF**			**XGBoost**		
Feature set	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Fc	0.351	0.350	0.350	0.424	0.361	0.390	**0.425**	0.370	**0.396**
Fs	0.408	0.420	0.414	0.449	0.418	0.433	**0.479**	0.438	**0.458**
Fp	0.725	0.607	0.661	0.627	0.613	0.620	**0.737**	0.638	**0.684**
Fm	**0.663**	0.583	**0.620**	0.564	0.560	0.562	0.549	**0.585**	0.566
Fc∪Fs∪Fp	0.728	0.616	0.667	0.625	0.619	0.622	**0.738**	0.641	**0.686**
Fc∪Fs∪Fp∪Fm	0.729	0.617	0.668	0.627	0.622	0.624	**0.740**	0.643	**0.688**

(b) On Repayment Rate

Model	**L&R**			**SVM**			**k-NN**		
Feature set	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
Fc	0.393	0.193	0.068	0.386	0.212	0.062	0.424	0.241	0.049
Fs	0.384	0.188	0.103	0.358	0.194	0.077	0.386	0.215	0.088
Fp	0.237	0.093	0.604	0.227	0.089	0.629	0.184	0.099	0.590
Fm	0.267	0.104	0.537	0.250	0.097	0.588	0.270	0.133	0.402

Fc∪Fs∪Fp	0.232	0.092	0.605	0.224	0.086	0.630	0.180	0.098	0.593
Fc∪Fs∪Fp∪Fm	0.230	0.090	0.609	0.222	0.083	0.635	0.178	0.095	0.596
Model	**MLP**			**RF**			**XGBoost**		
Feature set	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
Fc	0.446	0.224	0.068	0.381	0.194	0.075	**0.368**	**0.186**	**0.120**
Fs	0.353	0.174	0.110	**0.345**	0.173	0.193	0.348	**0.169**	**0.214**
Fp	0.155	0.062	0.775	0.127	0.059	0.780	**0.124**	**0.053**	**0.788**
Fm	0.250	0.113	0.513	**0.160**	**0.064**	**0.754**	0.179	0.066	0.739
Fc∪Fs∪Fp	0.152	0.062	0.776	0.124	0.059	0.784	**0.121**	**0.052**	**0.791**
Fc∪Fs∪Fp∪Fm	0.150	0.060	0.777	0.123	0.057	0.785	**0.118**	**0.052**	**0.793**
(c) On Loan Profit									
Model	**L&R**			**SVM**			**k-NN**		
Feature set	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
Fc	2,194.97	2,423.17	0.017	2,148.37	2,559.94	0.019	2,286.84	2,755.75	0.013
Fs	2,155.81	2,297.30	0.055	2,063.18	2,499.06	0.068	2,153.09	2,581.96	0.054
Fp	1,733.12	1,857.79	0.250	1,596.94	1,780.95	0.289	1,756.33	2,107.54	0.223
Fm	1,879.70	2,031.62	0.208	1,884.31	2,082.55	0.251	1,916.44	2,201.37	0.192
Fc∪Fs∪Fp	1,681.02	1,846.84	0.255	1,575.57	1,778.883	0.293	1,746.15	2,105.57	0.226
Fc∪Fs∪Fp∪Fm	1,664.76	1,830.83	0.256	1,563.99	1,762.844	0.294	1,733.10	2,092.48	0.227
Model	**MLP**			**RF**			**XGBoost**		
Feature set	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
Fc	2,126.61	2,512.90	0.021	**2,072.87**	2,442.13	0.030	2,108.35	**2,389.13**	**0.048**
Fs	2,082.53	2,303.29	0.104	2,026.81	2,242.34	0.124	**2,019.25**	**2,239.39**	**0.146**
Fp	1,494.86	1,700.78	0.310	1,380.14	1,673.15	0.376	**1,250.34**	**1,626.07**	**0.384**
Fm	1,660.23	1,914.64	0.285	1,583.00	1,840.79	0.311	**1,341.63**	**1,799.88**	**0.342**
Fc∪Fs∪Fp	1,473.96	1,684.80	0.312	1,366.38	1,653.15	0.377	**1,214.18**	**1,611.11**	**0.389**
Fc∪Fs∪Fp∪Fm	1,417.55	1,672.61	0.314	1,351.10	1,631.06	0.379	**1,201.10**	**1,580.12**	**0.390**

Note. Values in bold indicate the best model among evaluation metrics for each feature set.

We next explore, at a more granular level, which features are the most relevant in the financial credit risk prediction based on XGBoost model. We implement a permutation feature attribution method (Fisher et al. 2018). The feature importance analysis shows that living city DPI ranks the highest in predicting credit risk behavior. Consistent with previous findings, we show the importance regarding features in the cellphone usage and mobility trace category, as well as in the shopping behavior category. Interestingly, game relevant behaviors, i.e., frequency of using game apps and the average amount of purchasing game card, play significant roles (i.e., only second to living city DPI) in predicting borrowers' credit risk. Following them, borrowers' mobility trajectories including the appearance frequencies in official buildings/areas, commercial places, and recreational places are also important. In general, conventional features (except living city DPI and monthly income level) are less important than the new alternative sources of information. When considering social media features for social media usage sample, we learn that several social media usage characteristics are likewise quite important in predicting financial credit risk, including the number of fans a borrower maintains, sentiment valence of the originally generated messages, and number of "likes" received in her microblog.

Platform Welfare Analysis with Alternative Data

To help microloan platforms control the credit risks as well as increase revenues, we next propose and

evaluate the optimal filtering strategies adopted on all applicants. In practice, the microloan platform makes approval decisions about whether to offer an applicant the loan product. The most commonly implemented strategy is to evaluate a borrower's default probability. This strategy emphasizes the necessity of excluding applicants with high credit risks. We consider this default-based strategy as our benchmark. In addition, as we discussed before, given that delinquent fine is another important source of revenue, we propose an alternative business strategy to filter applicants based on their probabilities of delinquent-but-not-default behavior. The above two strategies can be easily implemented using our categorical credit risk indicator. Moreover, numerical indicator may deliver more credit risk information, especially at the installment level, we thus also apply `repayment rate` as a criterion for platform profit analysis. Finally, inspired by Papouskova and Hajek (2019), we also directly use the predicted values of `loan profit` as a criterion in making approval decisions.

For each of the above four filtering strategies, we first predict the corresponding probabilities/values using the XGBoost model with different sets of features. We then rank all applicants based on our predicted values (from the best to the worst) and assume the ranking is the only criterion in the approval decision-making process. To evaluate the performance, we then calculate the actual platform profits (in thousand RMB) and relative profits (i.e., the ratio of loan profit to loan amount) by choosing top 5%, 10%, 15%, … 100% best loans with different thresholds. Then we calculate platform profits of the total granted loans under different approval rates.

Table 2 reports the platform profits from the approved loans (in validation sample) based on the predicted values of different credit risk indicators using various feature sets. We find that the microloan platform can achieve the highest profits at 45% of loan approval rate across different credit risk indicators and feature sets. When the loan approval rate is higher than 65%, the microloan business will become unprofitable. More importantly, we find that loan permission based on cellphone usage and mobility traces information brings the highest economic gains to the microloan platform, and approximately 15% more economic gains than simply on conventional information. Unlike the prediction analysis, we observe a significant increase in revenues when we incorporate more features in designing the filtering strategies. When we apply all feature sets to predict credit risk and make loan approval decision, it yields 22% more economic gains to the microloan platform than simply using conventional features. This finding implies that although the combination of diverse alternative features might not be helpful in improving the overall prediction accuracy, it is potentially more effective in identifying valuable borrowers, who are more likely to repay the loan on time or even earlier. It could be hard to achieve using conventional features merely, while the alternative features (capturing individuals' shopping behavior and mobility traces) are capable of anticipating borrowers' repayment decisions. From this profit welfare analysis, we demonstrate the financial value of our proposed feature set in credit risk assessment.

Table 2. Platform Profit of Alternative Data Based on (Microblog Usage Sample Case)

(a) Based on default prediction

Feature	Approved loans (%)													
	5%	10%	15%	20%	25%	30%	35%	40%	**45%**	50%	55%	60%	65%	70%
Fc	11.0	24.5	38.9	49.9	64.0	69.8	82.8	86.6	**92.8**	98.2	90.2	56.8	13.1	-37.7
Fs	15.1	32.4	49.0	64.6	76.9	84.2	95.7	100.8	**105.0**	107.7	95.1	60.5	13.3	-40.1
Fp	16.5	36.3	53.7	69.6	82.9	91.0	102.2	108.2	**113.1**	112.7	98.3	61.8	12.8	-45.8
Fm	16.4	36.1	53.1	68.8	82.0	89.3	100.1	106.7	**109.7**	112.7	98.0	62.3	13.0	-43.3
Fc∪Fs∪Fp	19.1	39.1	57.8	75.2	87.8	97.2	106.5	115.1	**116.0**	114.5	97.5	61.3	11.8	-47.4
Fc∪Fs∪Fp∪Fm	20.0	40.5	59.3	76.2	88.7	99.4	108.9	116.4	**118.4**	114.7	97.3	60.3	11.6	-48.7

(b) Based on delinquent-but-not-default prediction

Feature	Approved loans (%)													
	5%	10%	15%	20%	25%	30%	35%	40%	**45%**	50%	55%	60%	65%	70%
Fc	12.6	28.1	43.4	53.5	66.9	72.5	83.6	85.0	**90.3**	95.4	87.9	57.4	16.3	-33.4
Fs	17.1	36.6	56.0	68.8	80.9	86.3	96.2	99.1	**102.4**	104.6	93.0	62.0	15.8	-35.7
Fp	19.0	41.9	60.5	74.0	86.0	93.7	102.5	106.0	**110.0**	108.6	95.7	61.7	15.4	-41.0
Fm	18.5	41.1	60.0	73.2	84.7	92.0	100.4	103.9	**106.5**	108.1	95.7	63.1	16.1	-39.8

Fc∪Fs ∪Fp	22.9	44.5	64.9	80.3	91.9	100.7	106.6	112.7	**112.5**	110.1	95.4	61.8	14.8	-42.5
Fc∪Fs ∪Fp∪ Fm	23.4	46.3	66.7	81.5	92.5	102.6	109.6	113.7	**115.4**	110.9	94.9	60.8	13.7	-44.4
(c) Based on repayment rate prediction														
Feat-ure	**Approved loans (%)**													
	5%	10%	15%	20%	25%	30%	35%	40%	**45%**	50%	55%	60%	65%	70%
Fc	10.9	24.4	38.4	50.3	64.0	69.1	83.4	87.5	**93.2**	98.7	88.0	55.1	9.9	-39.9
Fs	15.2	32.3	47.7	64.7	76.4	83.2	96.2	102.1	**106.1**	107.1	93.0	58.0	10.1	-42.0
Fp	16.2	36.0	53.7	68.7	82.3	90.5	103.3	109.5	**113.8**	112.1	95.6	58.4	10.0	-47.7
Fm	15.9	35.3	53.2	67.8	81.0	88.8	100.7	107.4	**110.9**	111.9	95.3	59.7	10.5	-46.4
Fc∪Fs ∪Fp	19.2	38.9	57.8	75.1	87.3	96.8	107.6	115.7	**116.1**	114.3	94.5	58.2	9.0	-50.2
Fc∪Fs ∪Fp∪ Fm	19.9	40.0	58.5	75.5	87.9	98.8	109.1	116.8	**119.0**	114.3	94.6	57.4	8.5	-51.7
(d) Based on loan profit prediction														
Feat-ure	**Approved loans (%)**													
	5%	10%	15%	20%	25%	30%	35%	40%	**45%**	50%	55%	60%	65%	70%
Fc	14.3	29.4	43.5	52.6	66.0	70.4	82.0	84.5	**91.2**	96.7	89.8	58.6	15.7	-36.2
Fs	19.6	38.8	55.0	68.1	79.2	84.9	94.7	98.4	**103.1**	106.1	94.6	62.5	15.9	-38.5
Fp	21.5	43.9	60.3	73.2	85.4	91.2	99.9	106.0	**111.0**	110.9	97.8	65.1	15.1	-42.4
Fm	21.3	43.5	59.5	72.5	84.5	90.2	98.5	104.4	**107.7**	111.0	97.5	64.3	15.6	-41.5
Fc∪Fs ∪Fp	24.9	46.8	64.7	79.3	90.5	98.1	105.5	112.3	**113.9**	112.7	97.0	63.2	14.1	-45.5
Fc∪Fs ∪Fp∪ Fm	26.0	48.5	66.5	80.3	91.4	100.3	107.9	113.6	**116.3**	113.0	96.8	62.3	13.9	-46.8
Note. Values in bold indicate the highest profits for the various loan approval rates.														

We further illustrate the comparison among the four proposed filtering strategies in Figure 1 using the full sample and the combined feature set Fc∪Fs∪Fp. Figure 1 presents the platform profit in RMB. From Figure 1, we find that overall the four credit risk criteria show similar trends. Specifically, when loan the approval rate is lower than 35%, loan permission strategies using the loan profit indicator and delinquent-but-not-default probabilities yield higher platform profits than default probabilities and repayment rates. Whereas when more than 35% of loans are granted, loan permission using default probabilities or repayment rates leads to higher profits. One potential explanation is that the loan profit indicator and delinquent-but-not-default loans are most economically valuable, especially when we apply them to guide loan permission decisions.[1] The loan permission using repayment rates realizes the highest (and maximum) profit at the loan approval rate of 45%. It is reasonable that compared to default probabilities, repayment rate captures the installment-level repayment characteristics, and is more capable of locating delinquent-but-not-default cases, which are probably profitable though with a certain level of credit risk.

In practice, the executed loan approval rate majorly depends on the platform's budgets (i.e., available funds to lend out). Our results thus indicate that microloan platforms could realize optimal platform profits with their budgets (i.e., loan approval rates) by using personalized credit risk indicators. For example, if the platform budgets allow them to approve the maximum top 20% best loans, then granting loans using the predicted delinquent-but-not-default probabilities can achieve the highest economic gains. On the other hand, if the budgets allow granting more than 45% of loan applications, approving the top 45% best loans using the predicted repayment rates can achieve the highest economic gains.

[1] Filtering loans directly by loan profits should be theoretically the ideal (most accurate) approach, however, for loan approval rate between 15% and 35%, we found delinquent-but-not default probabilities lead to higher platform profits than does the direct loan profit indicator, which shows lower accuracy in the prediction analysis.

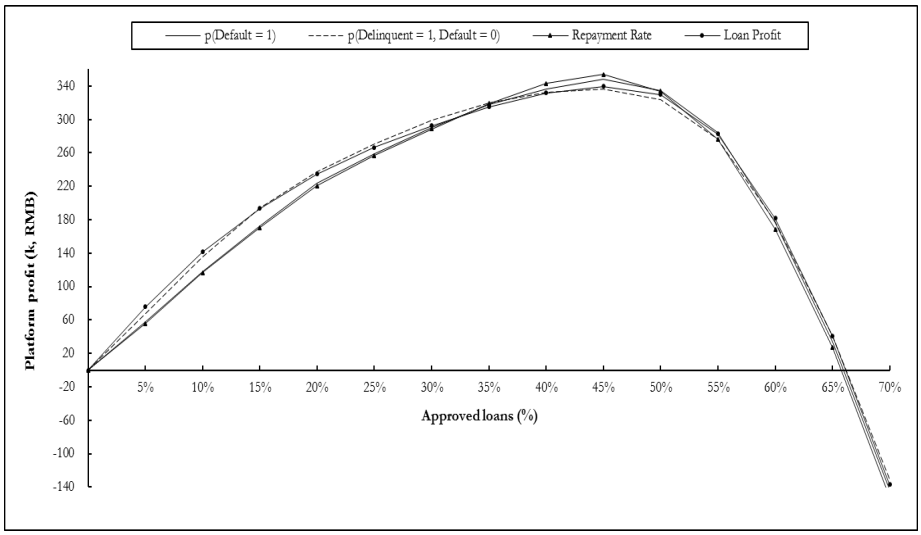

Figure 1. Platform Profit Analysis (Full Sample Case, Based on Feature Set: $\mathbf{F_c} \cup \mathbf{F_s} \cup \mathbf{F_p}$)

Value of Alternative Data in Coping with Sample Bias

Approved Sample

Recall that prior studies and current industry practice on credit risk prediction tend to use the approved samples, which are easily accessible and cleaner compared to the full applicant samples. It becomes necessary to identify any potential prediction biases from the approval samples. This identification would also help us justify the value of the full-sample prediction with a more comprehensive feature set, and figure out how alternative data can help alleviate concerns from training sample bias.

We collect a secondary data set of the approved sample loans from the same microloan platform. The current filtering strategy the focal platform is adopting is simply based on the personal experience of platform staffs, who make decisions manually with borrowers' conventional information such as age, education level, income level, and the cities they are living in. The loan approval rate on the focal platform is approximately 40%. The data set contains 5,378 randomly-selected borrowers with 5,450 approved loans issued from September 1st, 2017 to September 30th, 2017. The dataset contains the same set of information as the experimental data which we described in Section 3. 1,813 (i.e., 33.71%) borrowers in the approved sample are microblog users.

The loan attributes in the approved sample are akin to those of the full sample. However, as expected, compared to the full sample, the demographic and socioeconomic features of the filtered sample have overall lower variances. The filtered sample has a larger proportion of male borrowers, and the mean values of their socioeconomic characteristics such as education level, income level, and living city DPI, are obviously higher. They also performed better in microloan history. The approval samples show better repayment performance. Among the approved loans, 1,701 (i.e., 31.21%) had no delinquent installments, 1,785 (i.e., 32.75%) belonged to the delinquent-but-not-default class, and the other 1,964 (i.e., 36.04%) loans defaulted. The average repayment rate of these loans is 0.792 (standard deviation is 0.348), and the average loan profit of these loans is 75 RMB (approximately 11 USD, standard deviation is 1.105). These loans have explicitly overall lower risks than our full experimental samples.

Cross-Sample Prediction Accuracy and Platform Welfare Analysis

We apply the exact same training strategies on the approval samples as what we did on the full experimental sample, except that we implement the training coefficients to predict the credit risk of the full sample validation set as used in Section 4.[2] Although XGBoost still outperforms the other machine learning models,

[2] We also did the training and prediction on both the approved sample using 10-fold cross-validation as did prior

we find from Table 3 that compared to the full-sample prediction, the cross-sample prediction shows significantly poorer prediction performance. The performance gap is quite large (approximately 55% – 70%) when we applied conventional features merely. The performance gap becomes smaller when we apply other features extracted from alternative data. The prediction biases are only between 20% and 30% when cellphone usage and mobility traces features are applied. That is, alternative data can help alleviate the training sample bias. It might be due to the more accurate and less biased assessment of borrower's credit risk with the abundant alternative information despite for the approved sample. This finding strengthens our argument regarding the effectiveness of alternative information in credit risk assessment.

Table 3. Prediction Performance Comparison (Based on XGBoost)

Feature	Delinquent/Default (F1 score)				Repayment rate (R^2)				Loan profit (R^2)			
	Cross sample	Full sample	Bias (%)	P-values on bias	Cross sample	Full sample	Bias (%)	P-values on bias	Cross sample	Full sample	Bias (%)	P-values on bias
Fc	0.17	0.40	56.31	0.002***	0.04	0.12	69.98	< 0.001***	0.02	0.05	58.33	0.008***
Fs	0.37	0.46	20.09	0.075*	0.17	0.21	19.63	0.080*	0.10	0.15	32.19	0.036**
Fp	0.55	0.68	20.32	0.062*	0.63	0.79	20.43	0.070*	0.27	0.38	30.47	0.045**
Fm	0.33	0.57	41.52	0.016**	0.48	0.74	34.91	0.030**	0.20	0.34	42.98	0.013**
Fc∪Fs∪Fp	0.54	0.69	21.43	0.063*	0.63	0.79	20.73	0.066*	0.27	0.39	30.59	0.046**
Fc∪Fs∪Fp∪Fm	0.54	0.69	21.51	0.061*	0.63	0.79	20.55	0.067*	0.27	0.39	29.74	0.050**

Note. Bias = (Full sample – Cross sample)/Full sample. We report the P-values of pair-wise T test on the performance between cross-sample and full-sample prediction. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We further compare the platform profits among different credit risk indicators in pure full-sample prediction and cross sample prediction (Figure 2). Because of the downward bias of cross-sample prediction, it yields lower platform profits than does pure full-sample prediction under all kinds of feature sets. Specifically, under the optimal loan approval rate (i.e., 45%, 243 loans in validation sample), the platform would have an opportunity loss of 22.79 thousand RMB (approximately 3.42 thousand USD, i.e., 24.55%) if they implemented cross sample prediction with conventional features, whereas the loss was 12.16 thousand RMB (approximately 1.82 thousand USD, i.e., 10.27%) if they used all features to conduct credit risk prediction. More interestingly, we find even with sample bias, the economic gains of applying multiple sources of data (106.26 thousand RMB) is much larger than applying conventional data merely without sample bias (92.83 thousand RMB). These findings indicate that alternative data can help shrink the economic loss caused by sample bias, and their economic values exceed the usage of conventional data merely even without sample bias.

studies. The model performances show similarities with those with full sample data.

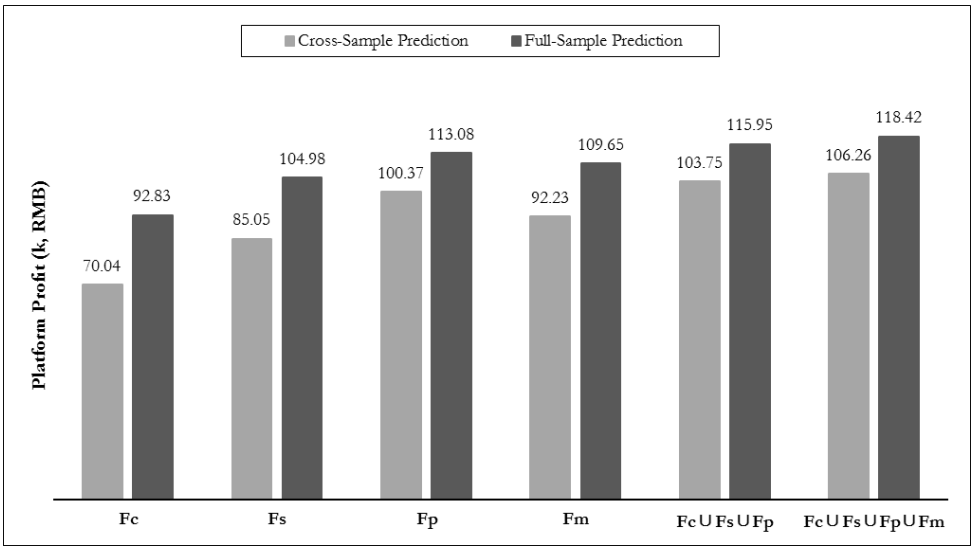

Figure 2. Platform Profit Comparison (Based on Microblog Usage Subsample, XGBoost, Default-based Prediction, Loan Approval Rate 45%)

Characteristic Comparisons of Approved Loans (Social Welfare Analysis)

We have demonstrated the significance of alternative data in improving credit risk prediction accuracy, platform profit welfare, and alleviating the influence of sample bias. We wonder *why* alternative data perform these advantages. To this end, we make a further examination on the approved loans/borrowers selected by predicted credit scores with conventional feature set versus all feature sets (containing alternative data) under cross-sample prediction and pure full-sample prediction.

We compare four groups of the approved loans/borrowers by different loan filtering strategies, that is, the top 45% best loans filtered by default-based full-sample prediction or cross-sample prediction with all features (i.e., including alternative data) versus conventional features merely, respectively. Table 4 reports the mean values of several sample conventional features of the unique borrowers (i.e., without overlaps across four groups) in the four groups. There are 126 (i.e., 51.85%) loans that are selected by all four filtering strategies. We learn from Table 4 that the approved loans based on cross-sample prediction with conventional features (group 1) have the least ratio of overlap to full-sample prediction with all features (group 4), which is only 58.44%; whereas the approved loans based on cross-sample prediction with all features (group 3) have much higher (87.24%) overlap to full-sample prediction with all features. The number/ratio of overlapped loans to group 4 for group 3 is larger than that for full-sample prediction with conventional features (group 2). These results show the primary reasons of different economic gains of microloan platforms cross the four filtering strategies as shown in the previous subsection. The larger overlap of approved borrowers with full-sample prediction using all features (i.e., applying abundant information and without sample bias), the higher financial returns it can achieve.

Theoretically, as aforementioned, the focal microloan platform's current filtering strategy is majorly on conventional features, which leads to the approved samples are those with overall better performance of conventional features. Consequently, training on these approved samples would tend to distort the true influence of conventional features especially when we use the training coefficients to predict the credit risk of the full sample. By comparing the conventional features of the unique approved loans/borrowers cross the four filtering strategies (Table 4), we find that prediction with training sample bias and with conventional information only indeed overemphasize the "goodness" of some conventional features. Table 4 suggests that actually a few borrowers who have relatively "worse" performance on certain conventional features (e.g., lower living city DPI, monthly income level, educational level, and house ownership, and higher loan-to-income ratio) can yield higher economic gains. Their loan applicants should have been approved but got rejection due to the biased filtering strategy. However, the loan filtering strategy using all features can alleviate the negative effect caused by sample bias to a large extent, for group 3 not only has the largest number of overlapped approved borrowers with group 4 among groups 1 – 3, the unique

borrowers in group 3 display the most similar conventional features to those in group 4.

Table 4. Comparisons of Top 45% Best Approved Loans (Based on Microblog Usage Subsample, XGBoost, Default-based Prediction)

		Loan Filtering Strategy			
		Group 1	Group 2	Group 3	Group 4
		Cross-sample prediction with conventional features	Full-sample prediction with conventional features	Cross-sample prediction with all features	Full-sample prediction with all features
# (ratio) of overlap loans to Group 4		142 (58.44%)	170 (69.96%)	212 (87.24%)	243 (100%)
Means of sample conventional features of the unique borrowers	Living city DPI	5,4016.24	52,039.85	46,778.60	42,540.59
	Monthly income level	5.54	5.28	4.57	4.35
	Loan-to-income ratio	1.05	1.18	1.33	1.40
	Education level	4.26	4.10	3.95	3.73
	House ownership	0.57	0.50	0.41	0.34

Note. Sample features are those showing significantly different mean values across groups.

In a nutshell, a further examination on the approved sample across different loan filtering strategies suggest that sample biases would lead to abandonment of some low-risk and profitable users who have been thought "bad" because they have certain characteristics that are incorrectly trained as high-risk predictors. Full-sample prediction with alternative data could minimize the sample bias and yield higher economic profits to the microloan platforms. At the same time, it also possibly helps reduce the unfairness caused by individual experience or machine learning algorithms, because an accurate loan filtering strategy can help financial institutions extend their service to more lower-income and less-educated loan applicants from less-developed geographical areas – those historically disadvantaged population who have been largely neglected in the past.

Conclusions

By conducting a large quasi-field experiment and supported by unique datasets containing multiple alternative sources of borrower information in a major microloan platform, we construct and extract more than one hundred features to compare the value of alternative data on financial credit risk based on the full loan applicant sample without prior filtering. We define a multiclass categorical and two numerical credit indicators as financial credit risk indicators, under the idea that a certain level of delinquency risks may also yield positive economic gains. The prediction results with state-of-the-art machine learning models suggest that cellphone usage and mobility trajectory features and shopping characteristics have significantly better predictive performance than conventional loan attributes and borrowers' demographic and socioeconomic features. Cellphone usage and mobility traces features have the highest predictive power. For social media users, social media presence and sentiment are also quite valuable in predicting users' repayment behaviors. The platform welfare analysis with experimental data indicates that alternative data yield higher economic gains than conventional data to microloan platforms. It would be beneficial for the platform to adopt alternative data if the economic gains from a better credit risk assessment with alternative data outweigh the cost of data collection. Moreover, we find that for certain loan approval rates, loan permission based on the predicted delinquent-but-not-default probabilities, and the predicted numerical repayment rates and loan profits, may yield higher economic gains than based on the conventionally adopted predicted default probabilities. This finding confirms that on premise of accuracy risk prediction with alternative data, lending to the borrowers with a certain level of delinquency risks despite relatively high default risk can also yield positive economic gains.

On a broader note, our study demonstrates the tremendous potential of leveraging alternative data to increase social welfare in the financial market. Previous approaches in financial credit risk assessment, using only approved samples or primarily conventional data for model training, tend to favor higher-income and more-educated applicants from areas with more developed economy. Importantly, we show the potential of alternative data to alleviate such inequality in the financial service markets, while in the

meantime achieving higher platform revenues.

Acknowledgement

T. Lu thanks the National Natural Science Foundation of China (Grants 71872050, 91546104 and 71531006).

References

Abdelmoula, A. K. 2015. "Bank credit risk analysis with k-nearest-neighbor classifier: Case of Tunisian banks," *Accounting and Management Information Systems* (14:1), pp. 79.

Amonini, C., and Donovan, R. J. 2005. "The Relationship between Youth's Moral and Legal Perceptions of Alcohol, Tobacco and Marijuana and Use of These Substances," *Health Education Research* (21:2), pp. 276-286.

Blumenstock, J., Cadamuro, G., and On, R. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata," *Science* (350:6264), pp. 1073-1076.

Carroll, P., and Rehmani, S. 2017. "Alternative Data and the Unbanked," https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2017/may/Alternative_Data_And_The_%20Unbanked.pdf.

Chen, T., and Guestrin, C. 2016, August. "Xgboost: A Scalable Tree Boosting System," In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 785-794.

Chittaranjan, G., Blom, J., and Gatica-Perez, D. 2013. "Mining Large-scale Smartphone Data for Personality Studies," *Personal and Ubiquitous Computing* (17:3), pp. 433-450.

Cho, E., Myers, S. A., and Leskovec, J. 2011, August. "Friendship and Mobility: User Movement in Location-based Social Networks," In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1082-1090.

Cui, R., Gallino, S., Moreno, A., and Zhang, D. J. 2018. "The Operational Value of Social Media Information," *Production and Operations Management* (27:10), pp. 1749-1769.

Dobbie, W., Liberman, A., Paravisini, D., and Pathania, V. 2018. "Measuring Bias in Consumer Lending," National Bureau of Economic Research, No. w24953.

Duarte, J., Siegel, S., and Young, L. 2012. "Trust and Credit: The Role of Appearance in Peer-to-Peer Lending," *The Review of Financial Studies* (25:8), pp. 2455-2484.

Fisher, A., Rudin, C., and Dominici, F. 2018. "Model Class Reliance: Variable Importance Measures for Any Machine Learning Model Class, from the 'Rashomon' Perspective," arXiv preprint arXiv:1801.01489.

Fu, R., Huang, Y., and Singh, P. V. 2018. "Crowd, Lending, Machine, and Bias," Available at: https://ssrn.com/abstract=3206027.

Gao, Q., and Lin, M. 2013. "Linguistic Features and Peer-to-Peer Loan Quality: A Machine Learning Approach," Available at SSRN.

Ge, R., Feng, J., Gu, B., and Zhang, P. 2017. "Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending," *Journal of Management Information Systems* (34:2), pp. 401-424.

Guzelian, C. P., Stein, M. A., and Akiskal, H. S. 2015. "Credit Scores, Lending, and Psychosocial Disability," *Boston University Law Review* (95), pp. 1807-1868.

Heckman, J. J. 1977. "Sample Selection Bias as a Specification Error (with an Application to the Estimation of Labor Supply Functions)".

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. 2015. "Screening Peers Softly: Inferring the Quality of Small Borrowers," *Management Science* (62:6), pp. 1554-1577.

Kim, E. J., Namkoong, K., Ku, T., and Kim, S. J. 2008. "The Relationship between Online Game Addiction and Aggression, Self-Control and Narcissistic Personality Traits," *European Psychiatry* (23:3), pp. 212-218.

Lin, M., Prabhala, N. R., and Viswanathan, S. 2013. "Judging Borrowers By the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending," *Management Science* (59:1), pp. 17-35.

Loufield, E., Ferenzy, D., and Johnson, T. 2018. "Accelerating Financial Inclusion with New Data," Center for Financial Inclusion. http://www. centerforfinancialinclusion. org/storage/AcceleratingFINewData_Final_2018, 6.

Ma, L., Zhao, X., Zhou, Z., and Liu, Y. 2018. "A New Aspect on P2P Online Lending Default Prediction using Meta-level Phone Usage Data in China," *Decision Support Systems* (111), pp. 60-71.
Morse, A. 2015. "Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending," *Annual Review of Financial Economics* (7), pp. 463-482.
Papouskova, M., and Hajek, P. 2019. "Two-stage Consumer Credit Risk Modelling using Heterogeneous Ensemble Learning," *Decision Support Systems* (118), pp. 33-45.
Ravina, E. 2007. "Beauty, Personal Characteristics, and Trust in Credit Markets."
Serrano-Cinca, C., Gutiérrez-Nieto, B., and López-Palacios, L. 2015. "Determinants of Default in P2P Lending," *PloS One* (10:10), e0139427.
Serrano-Cinca, C., and Gutiérrez-Nieto, B. 2016. "The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending," *Decision Support Systems* (89), pp. 113-122.
Tan, T., Bhattacharya, P., and Phan, T. 2016. "Credit-worthiness Prediction in Microfinance using Mobile Data: A Spatio-network Approach," In *2016 International Conference on Information Systems*.
Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2015). Credit scoring with social network data. Marketing Science 35(2), 234-258.