Invisible Primes: Fintech Lending with Alternative Data*

Marco Di Maggio[†], Dimuthu Ratnadiwakara [‡], Don Carmichael [§]

December 11, 2022

Abstract

We exploit anonymized administrative data provided by a major fintech platform to investigate whether using alternative data to assess borrowers' creditworthiness results in broader credit access. Comparing actual outcomes of the fintech platform's model to counterfactual outcomes based on a "traditional model" used for regulatory reporting purposes, we find that the latter would result in a 70% higher probability of being rejected and higher interest rates for those with low credit scores. The borrowers most positively affected are the "invisible primes"–borrowers with low credit scores and short credit histories, but also a low propensity to default. About two thirds of the effects are due to the inclusion of additional data, while the remainder is due to a more sophisticated underwriting model. We show that funding loans to these borrowers leads to better economic outcomes for the borrowers and to positive returns for the platform.

Keywords: Financial Inclusion, Fintech Lending, Alternative Data, Machine Learning, Underwriting models

JEL Classification: D14, H52, H81, J24, I23

---

*We thank seminar and conference participants at the University of Chicago Booth, Yale SOM, MIT (Sloan), USC (Marshall), Colorado Leeds, University of Houston, Florida University Business School, Baruch College, Louisiana State University, UCLA (Anderson), Bank of England, 11th FDIC Consumer Research Symposium, 2022 GSU-RFS Fintech Conference, FIRS 2022, NBER Summer Institute 2022, NFA 2022, GSU-Atlanta Fed Annual Financial Stability Conference, and 4th Future of Financial Information Conference for helpful comments. We thank Upstart Network, Inc. for providing access to its data. The authors and Upstart structured the agreement in a way that maintains the authors' independence.

†Harvard Business School and NBER: mdimaggio@hbs.edu.

‡E. J. Ourso College of Business, Louisiana State University: dratnadiwakara2@lsu.edu.

§Upstart Network, Inc.: don@upstart.com.

1. Introduction

Credit markets have experienced significant disruption consequent to the rise of online intermediaries and financial technology (fintech) companies. A key feature of fintech companies is the substitution of algorithms and alternative data for in-person interaction between lender and borrower. Fintech lending has grown rapidly, and they have acquired significant market share across lending products (Berg et al., 2021). Online lender Quicken Loans, for example, is the largest mortgage originator in the United States, and fintech lenders account for about 40% of the unsecured personal loan market (TransUnion, 2019). Despite their growing prominence, a clear understanding of how these new intermediaries affect credit availability and household financial health is lacking. Traditionally, consumers with high credit scores have reaped the benefits of having multiple low-rate credit options. However, underscored-individuals who are equally creditworthy face limited and expensive options (Brevoort et al., 2016).

The advent of fintech lenders has the potential to change this circumstance. Alternative underwriting models used by fintech lenders might be able to identify invisible primes–creditworthy individuals currently overlooked by standard measures of creditworthiness, such as credit scores. Former director of the Consumer Financial Protection Bureau, Richard Cordray, observed: "Adding ... alternative data into the mix ... holds out the promise of opening up credit for millions of additional consumers."[1] Relying primarily on credit scores excludes a large fraction of Americans from credit markets altogether. According to Fair Isaac Corporation, a leading provider of credit scores, 28 million Americans have files with insufficient data to generate credit scores and 25 million Americans have no credit file at all.[2] The Consumer Financial Protection Bureau (CFPB) has responded by encouraging lenders to develop innovative means of increasing fair, and equitable, nondiscriminatory access to credit, particularly for *credit invisibles*: those limited by their credit history or lack thereof. In addition, the emergence of alternative underwriting models has raised a number of related policy-relevant questions such as impact of expanding credit access on household financial health and whether fintech lenders are subsidizing market share growth at the expense of funding unprofitable loans to invisible primes.

[1] https://www.consumerfinance.gov/about-us/newsroom/prepared-remarks-cfpb-director-richard-cordray-alternative-data-field-hearing/

[2] https://www.fico.com/blogs/leveraging-alternative-data-extend-credit-more-borrowers

Empirical evidence on these issues is scarce. The ideal setting would entail observing fintech lenders' information set, underwriting model, and the ability to differentiate between funded and non-funded loan applicants. Such a setting would enable researchers to investigate the main drivers of the new credit models and whether those lead to broader credit access. Access to the necessary data has been elusive as their underwriting models are an important part of fintech lenders' competitive edge and a key intellectual property asset. Even given access to such a setting, it would be difficult to answer the counterfactual: whether borrowers funded by fintech lenders would have been rejected by traditional financial institutions. Absent counterfactual information, assessing whether fintechs' lending decisions are *expanding* access to credit remains a challenging task. Moreover, researchers interested in the impact of credit on household financial health would need to follow applicants over time, which would require not only cross-sectional data at the time of origination, but longitudinal information about the same set of applicants.

This paper makes substantial progress on these research questions using a unique dataset from a major fintech platform, Upstart Network, Inc. (henceforth the "Platform"), which provided access to its anonymized administrative data. Operating in the unsecured personal loan space, the fastest growing category in consumer lending, the Platform originated more than $3 billion in personal loans from April 2019 to March 2020.[3,4] It excels as a research setting because it advertises the use of alternative data, including education and job history, as one of the main pillars of its underwriting process. The dataset provided by the Platform is unique in a number of respects. It covers rejected as well as funded loan applications, provides a panel for both funded *and* rejected individuals, and facilitates assessment of the impact of the Platform's underwriting model on credit access by comparing actual outcomes of the Platform's model to counterfactual outcomes based on a "traditional model" that does not use the Platform's alternative features. Rather than make assumptions about banks' underwriting practices, we employ the counterfactual model developed in coordination with the Consumer Financial Protection Bureau (CFPB) which has been used for regulatory reporting purposes. Access to this counterfactual model's outcomes reassures us of its representativeness with respect to traditional lenders' credit decisions. We also provide an additional benchmark model provided by a large bank with about $200 billion of assets and show that the model developed by the

[3]https://www.experian.com/blogs/ask-experian/research/personal-loan-study/
[4]https://www.sec.gov/Archives/edgar/data/1647639/000119312520285895/d867925ds1.htm

Platform in coordination with the CFPB is significantly more complex and nuanced than what even large banks adopt.

We begin our analysis by exploring how standard metrics of creditworthiness, such as credit score, capture borrowers' probability of defaulting. We show that credit score, in general, is a good predictor of default: there is a monotone negative relationship between credit score and default for credit cards and mortgages, products not offered by the Platform. Examining personal loans funded by the Platform, we find no relationship between the probability of defaulting and credit scores below 700. This suggests that the Platform borrowers on the left side of the credit score distribution look the same when examined using such a standard metric: the likelihood of being delinquent or of a loan being charged off is predicted to be constant. However, the Platform's underwriting model is able to identify significant differences in creditworthiness even among borrowers with similar credit scores. We find a monotone positive relationship between the default behavior and the Platform's measure of creditworthiness (the Platform default probability henceforth) after controlling for the credit score and other standard measures of borrowers' risk. We also show the area under the curve, a common measure of model predictability, significantly larger for the Platform default probability compared to models using traditional measures of creditworthiness.

This comparison suggests that, for some individuals, credit score may not paint an accurate picture of *future* creditworthiness, and the Platform's more sophisticated underwriting model is able to better capture their future creditworthiness. This might be because the characteristics used to compute the credit score are not reliably informative or predictive of future behavior of some individuals.[5] For instance, recent college graduates starting their first job, or recent immigrants, despite being potentially creditworthy, might exhibit low credit scores primarily due to their shorter credit histories.

A natural next step is to investigate the main drivers of the Platform's credit model. We iteratively identify the main variables driving the the Platform default probability using the machine learning algorithm "recursive feature elimination with random forest". We find that the main variables that account for improved predictability include non-traditional information, such as education and employment history, that supplement traditional credit report variables. Consistent with Berg et al. (2020), we find information extracted from borrowers digital footprint also impact the Platform de-

[5] We also show that these results are not confined to the set of funded applicants but also hold for the larger set of rejected loans.

fault probability. This non-traditional information has large economic effects on the Platform default probability after controlling for borrower information in the credit report.

Although the new credit model is assessed to be more accurate for the Platform-funded borrowers, a key question for policy makers and academics is whether it improves credit inclusion -are there borrowers granted credit by the Platform that would have been denied by a traditional lender? A natural hypothesis is that a better credit model could be used to compete with traditional lenders by attracting the best borrowers who could be offered lower rates. Counter to this hypothesis, we find that about 25% of borrowers funded by the Platform over our sample period with credit scores of less than 660 would have been rejected by the traditional model. Furthermore, this fraction declines as credit scores increase, which suggests that the mismatch between the traditional and the Platform models is magnified among low-credit score borrowers. Borrowers with credit scores lower than 640 who are granted loans by the Platform have a 70% probability of being rejected by traditional lenders. Only the subset of borrowers with credit scores higher than 740 experience a similar approval rate at the Platform and at a traditional lender. Comparing the outcomes with the simpler model provided by a large bank makes these findings even starker as borrowers with credit scores less than 660 would have been rejected with 100% probability.[6]

Differences between the traditional and alternative models might not be confined to approval decisions. Interest rates might also differ and it is not clear in which direction. On the one hand, the Platform might be able to more accurately price risk and so reduce the rates charged to some borrowers, but on the other hand, the less intense competition for that same set of borrowers gives the Platform the ability to charge a higher interest rate (Ben-David et al., 2022). Higher rates charged by the Platform, construed as a risk premium, could provide a plausible explanation for the expansion of credit. We address this question by comparing the rates predicted by the traditional model and the Platform. Our finding that low-credit score borrowers funded by the Platform would have incurred a significantly higher interest rate under the traditional model indicates that differences between models affect both the extensive and intensive margins.

We also characterize how the composition of the funded borrowers pool changes as a result of the different underwriting model. We find that thin credit file individuals (those with short credit

[6] 660 credit score is typically considered as the threshold that determines prime status for unsecured consumer loans and all major banks have a minimum credit score threshold of 660 or higher (Ben-David et al., 2022)

histories) and those with advanced degrees and salaried jobs benefit more due to the Platform's underwriting model. We also find results to be stronger in regions with a higher fraction of minorities and foreign-born individuals. Overall, our findings highlight how alternative underwriting models could identify "invisible primes"—borrowers who, evaluated on the basis of standard metrics, would either be denied credit or be granted credit on unfavorable terms. At the same time, we also characterize the borrowers that might potential lose access to credit if more alternative data are adopted by standard credit models, i.e., those that would only be accepted by the traditional model. These borrowers tend to be older, have a lengthier credit history, but are less likely to be college educated. Thus, the same information that is able to expand credit access to the invisible primes might come at the cost of excluding borrowers that are currently served by traditional lenders.

One question at this point is to what extent the expansion of credit is due to the access to alternative data or the development of a more sophisticated credit model. This question also has policy implications since traditional financial institutions could be encouraged to look at other indicators and borrowers' information other than the ones contained in the credit score if data is the main obstacle towards a more inclusive lending system. To address this question, we construct an augmented traditional model that also includes alternative variables. In this way, the differences in predicted outcomes between the Platform's model and this one will identify the contribution of adopting a more sophisticated algorithm.[7] We find that adding the alternative variables does induce the traditional model to fund a significantly higher number of loans among low credit score applicants. A comparison of the AUC for the traditional model with the augmented version's AUC and the Platform's one shows that about two thirds of the improvement results from the use of alternative data, while the remaining one third is due to the more sophisticated model.

Next, we explore the effects of expanded credit access on borrower outcomes. To determine whether novel underwriting models are able to generate gains from trade between borrowers and lenders, we need to assess the extent to which borrowers benefit from improved access to credit. We do so in our setting by quantifying the extent to which applicants' ability to meet future obligations (as measured by subsequent credit card defaults), credit scores, and likelihood of purchasing a first home improves after being funded by the Platform. This analysis includes observations of subsequent

[7] As explained in more details in Section 4, the traditional model is a logistic regression while the Platform's model adopts a random forest algorithm.

outcomes for the set of applicants denied credit.

We first employ an entropy balance algorithm to match the funded and denied borrowers on an extensive set of observables. We find funded, low-credit score applicants to be 2.8% less likely than disqualified applicants to default on credit card payments. This is economically large compared to an average delinquency rate of 14%. We further observe low-credit score borrowers compared to similar disqualified applicants to experience an increase in credit scores within 12 months of loan origination and a higher propensity to make a first-time home purchase. To account for unobservable borrower characteristics that might drive these results, we exploit two key features of the Platform's business operation for identification purposes: applicants with debt-to-income ratios greater than 50% and applicants who filed for bankruptcy in the last 36 months are automatically denied funding. We use these cutoffs independently to implement regression discontinuity frameworks in which outcomes for borrowers with debt-to-income ratios (month since bankruptcy) lower than 50% (36 months) are compared to those for borrowers with debt-to-income ratios (months since bankruptcy) above 50% (36 months) at the time of application. Comparing borrowers with similar characteristics near the cutoffs, we confirm that borrowers able to obtain financing from the Platform experience an improvement in financial health.

We also provide evidence that those borrowers who are rejected by the Platform's model, but are able to secure a loan from another lender, tend to experience worse financial outcomes than funded borrowers, e.g., they are more likely to default on ther debts. This holds for both low and high credit score individuals, which suggests that alternative data might also be instrumental in identifying *hidden risk* among applicants considered safe by standard metrics.

Given the expansion of credit and lower rates charged to higher-risk borrowers, one might wonder whether the Platform is subsidizing market share growth at the expense of funding unprofitable loans. We complement our previous results by investigating whether its loans negatively affect the Platform's profitability or its borrowers. Addressing the first question by computing the internal rate of return (IRR), the key metric tracked by the Platform, for borrowers with different credit scores, we do not find underperformance of loans granted to borrowers with low credit scores. In other words, misalignment between the traditional and alternative models does not result in loans that generate lower profits. We further find that the IRR is higher, holding credit score constant, for borrowers with advanced degrees and salaried employees than for other borrowers with similar credit characteristics. Finally, the IRR of the later cohorts has been significantly higher, which is consistent with the model being further trained and refined. Our results are different from the finding in Ben-David et al. (2022) that the *average* fintech lender generates a significantly higher return by charging a higher interest rate from borrowers with credit scores less than 660 as there is less competition in this market segment.

To support the external validity of our findings, we complement the previous analysis with mortgage data from Quicken Loans and the top four traditional banks. Using this sample, we replicate some of our main findings. For example, we find the credit score to be a good predictor of performance for the loans funded by traditional institutions, but not helpful for differentiating Quicken Loans' borrowers. Consistent with our earlier analysis, we also find that, relative to traditional institutions, Quicken Loans generates significantly higher returns from low-credit score borrowers. These results suggest that even in the case of mortgages, fintech lenders might be able to better price borrowers deemed less creditworthy on the basis of standard metrics.

Our results have important policy implications related to the advent of fintech and the debate on fair access to credit. We show that algorithmic underwriting based on alternative data can result in expanded opportunities for individuals currently underserved by traditional institutions. Our findings also suggest a need for reform in how credit scores are currently computed and utilized. Examples of highly creditworthy individuals with low credit scores or none at all, typically younger individuals or recent immigrants with thin credit profiles, are easy to find. Such individuals are often denied access to credit or burdened with unfavorable terms. It seems likely that incorporating new data in evaluation criteria might identify these and many other credit-invisible people as a shadow group of prime borrowers.

Our findings contribute to an emerging literature on fintech lending.[8] One strand of this literature examines the usefulness of alternative data in predicting borrowers' default behavior. For example, Berg et al. (2020) exploit data from a German e-commerce company to show digital footprint variables (e.g., computer type, distribution channel) to be important predictors of default and usefully complement credit bureau information. Using data from a large Indian fintech firm, Agarwal et al.

[8]See Morse (2015) for an early review of this strand of the literature.

(2019) show that mobile and social footprint data can substitute for traditional credit bureau scores. Ghosh et al. (2021) show that information produced by cashless payments is helpful in screening loan applications. However, Di Maggio and Yao (2020) show that the *average* fintech lender relies mostly on the information in the credit report in their underwriting models. We show that there exists a significant fraction of individuals with low credit scores who are substantially more creditworthy than their credit scores suggest, and alternative data can be successfully employed to improve credit access to this segment. The reliance on conventional credit bureau information might risk under-serving this important segment of the population.[9] Our work is also related to the recent work by Blattner and Nelson (2021) who document that traditional credit scores are statistically noisier indicators of default risk for borrowers with thin credit files, and estimate a structural model of lending to quantify the gains from addressing this disparity for the US mortgage market. We exploit the granularity of our data to quantify the effects of alternative models and data on financial inclusion.

This study also contributes to the literature on benefits of expanding credit access. Economic theory and existing empirical literature do not offer an unambiguous prediction on benefits of expanding credit access to invisible primes. Melzer (2011) finds that increased access to payday loans inhibited low-income households' ability to pay important bills due to large debt service burden. Karlan and Zinman (2010) show that high interest consumer loans produced net financial benefits to borrowers. We provide evidence that expanding credit access to invisible primes improves their financial conditions.

More generally, our paper is related to recent studies focused on whether fintech lenders and traditional banks are substitutes or complements (see, for example, Buchak et al. (2018), Fuster et al. (2019), and Tang (2019a)). Our evidence suggests that alternative data allows fintech companies to complement traditional bank lending leading to more inclusive and expanded credit access.[10]

A few recent papers investigating the role machine learning algorithms play in the underwriting process have emphasized the potential for discrimination. Fuster et al. (2020), for example, studied the distributional consequences of the adoption of machine learning techniques in the mortgage market. Their finding that a white, non-Hispanic group experienced lower estimated default propensities with

[9]Our findings are also consistent with the theoretical predictions of Hau et al. (2019).

[10]Related papers in this literature include Danisewicz and Elard (2018); De Roure et al. (2019); Balyuk (2019), Iyer et al. (2016), Mariotto (2016), Wolfe and Yoo (2018), Vallee and Zeng (2019), Hertzberg et al. (2018), and Balyuk and Davydenko (2019).

machine learning than with less sophisticated technology, did not generalize to other ethnic groups. Their paper overcomes the data limitations discussed earlier by developing a model that traces how changes in predicted default propensities map to real outcomes. The issue of potential discrimination due to the use of new underwriting models in the mortgage market has also been explored by Bartlett et al. (2019).[11] Our results suggest that invisible primes in regions with a higher fraction of minorities and foreign born individuals are the most likely beneficiaries of the expansion of credit.

The remainder of the paper is organized as follows. In Section 2, we describe the data and its main features. We compare traditional and fintech credit models in terms of predictive power in Section 3, and in terms of resulting credit access in Section 4. In Section 5, we demonstrate that the broader credit access afforded by fintech credit models does not come at the expense of the Platform's profitability. In Section 6, we investigate whether outcomes improve over time for borrowers who obtain credit. Concerns about external validity are discussed in Section 7, Section 8 addresses additional robustness, and Section 9 concludes.

2. Data

Lack of data has been a key challenge in trying to ascertain whether the emergence of new financial institutions employing alternative data and novel credit models has affected credit access. Fintech lenders' use of alternative information in their underwriting models has been noted by Buchak et al. (2018), Fuster et al. (2019), and Jagtiani and Lemieux (2019), among others. Absent access to the data fed into these underwriting models and the internal measures of creditworthiness used, it is difficult to assess the role alternative information plays in fintechs' lending decisions.

We use a de-identified administrative dataset from a major fintech platform operating in the personal loan space that includes information about approved and rejected applicants, the credit model used to assess borrower creditworthiness, and subsequent credit reports for both approved and rejected applicants. To the best of our knowledge, no similar data has previously been made available to academics.

Founded in 2012, the Platform provides unsecured personal loans to borrowers throughout the

[11] Dobbie et al. (2018) show that substituting loan officers with machine learning could address issues related to bias.

United States, and it is one of the few public online lending platforms. The Platform's underwriting model differs from those used by traditional lenders in that its artificial intelligence driven pricing algorithm incorporates alternative data from unconventional sources. To forestall running afoul of existing regulations, the Platform applied for and obtained a No-Action Letter ("NAL") from the CFPB in 2017 (see Ficklin and Watkins (2019)). As part of the NAL review process, the CFPB analyzed the Platform's automated unsecured underwriting model and its features, and compared its outcomes (i.e., approval rates and interest rates) to those generated by a counterfactual model that did not use alternative data. The CFPB found that its review did not warrant any supervisory or enforcement action against the Platform at the time.[12] In late 2020, the Platform was approved for an additional three years under the NAL program, and the terms of the agreement required the Platform to notify the CFPB of significant changes to its "artificial intelligence" model prior to their implementation.[13] [14]

Our dataset begins in 2014 and ends in the first quarter of 2021. It includes information about borrower characteristics at the time of origination and monthly loan performance for more than 900,000 loans originated by the Platform. Our analytic sample is restricted to 770,523 loans for which the traditional credit score is available. In recent years, the Platform experienced significant growth in terms of the number of loans and total amount loaned. Figure 1, Panel A shows growth in the number of loans originated, which increased from approximately 75,000 in 2017 to nearly 300,000 in 2020; Panel B shows growth in the volume of loans originated, which increased from less than $ 100 million in 2014 to almost $ 3.5 billion in 2020. We also have information about the platform's customer

[12]Note that the NAL is not an endorsement of the Platform model, in fact, the CFPB specifies: "The No-Action Letter references the application of the Equal Credit Opportunity Act (ECOA) and its implementing regulation, Regulation B, to the NAL Recipient's use of alternative data and machine learning for its underwriting and pricing model. This No-Action Letter is specific to the facts and circumstances of the NAL Recipient and does not serve as an endorsement of the use of any particular variables or modeling techniques in credit underwriting and pricing. In addition, neither the No-Action Letter nor this blog post serve as an endorsement of the NAL Recipient or the products or services it offers" (see Ficklin and Watkins (2019)).

[13]"On April 13, 2022, the Platform notified the CFPB that it intended to add a significant number of new variables to its underwriting and pricing model. The CFPB needed sufficient time to review and rigorously evaluate the implications of the changes to the Platform's model. In response, the Platform requested termination of the "no-action letter," effectively ending the company's special regulatory status, and allowing it to be able to make changes to its model without need for CFPB review and approval." see https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-order-to-terminate-Upstart-no-action-letter/

[14]The no action letter published by the CFPB can be found here https://files.consumerfinance.gov/f/documents/201709_cfpb_upstart-no-action-letter.pdf.

acquisition channels. The Platform reaches borrowers in multiple ways. Although a borrower can complete an application directly on the website, and the Platform also mails marketing offers, most funded loans, as can be seen in the Panel C of Figure 1, are generated through lending aggregators like Credit Karma—platforms that enable borrowers to search for the loan product that best fits their needs. Credit Karma sources multiple offers from lenders, significantly reducing search costs for borrowers who complete an application. As shown in Figure 1 Panel D, most loans are secured for credit card refinancing. The Platform's market reach is plotted in Figure 2. Darker, shaded counties on the US map capture a higher number of loans per capita; although the Platform operates in all states, loan issuance is higher in particular counties, notably those in Washington, California, Nevada, and Colorado. A similar pattern is observed for other fintech lenders (Di Maggio and Yao, 2020). Low issuances in Iowa and West Virginia are due to the Platform's main bank partner, Cross River Bank, not being active in those states. Our specifications control for zipcode by year fixed effects, which take into account differences across regions as well as time differences correlated with origination activity.

We complement the previous description with a discussion of the main borrower characteristics presented in Table 1 Panel A. On average, the Platform's loans are about \$11,700. The standard deviation of \$10,000 indicates significant heterogeneity among borrowers, some individuals borrowing significantly larger amounts. The average contract is characterized by an APR of 22% with a four-year maturity. Borrowers tend to have an average credit score of 653 at origination. That even the top quartile exhibits a score barely above 680 shows the Platform's focus to be on individuals other than those traditionally regarded as most creditworthy. Most borrowers are between 28 and 46 years old, with a median age of 35 years. We also observe verified income information about the borrowers. Mean annual income is \$67,000 (with a standard deviation of \$173,000), and debt to income ratio about 18%. Approximately 44% of borrowers have a college degree and average tenure at the same job is five years. As an indicator of access to credit markets, the average borrower holds 18 accounts on file.

A key advantage of our data is that it includes all application information for both funded and unfunded applicants. Table 2 compares the key characteristics of both funded borrowers and those immediately disqualified because they do not meet the Platform's eligibility criteria. On average, credit score is 65 points higher and annual income $12,700 higher for funded than for non-funded applicants. Total liabilities and credit balance are also higher for funded ($120,000) than for non-funded ($68,000) applicants. Funded borrowers are also more likely to be college educated, less likely to be hourly employees, and more likely to use a computer and use the loan for debt consolidation.

The data also afford access to applicants' credit report data both at origination and for subsequent months. The Platform pulls credit reports regularly for borrowers who are current and monthly for borrowers who have missed payments. The Platform is authorized to pull credit reports for disqualified applicants up to 12 months from the time of the application, and typically does so several times during this period with a final pull around the 12-month mark. Credit reports in our sample are all from the same credit bureau (TransUnion).

To provide evidence of external validity, we supplement our analysis using mortgage performance data from Moody's Analytics and Freddie Mac, and mortgage application data provided under the Home Mortgage Disclosure Act (HMDA). Moody's Analytics data provides loan-level data at origination and monthly performance data for mortgages underlying non-agency residential mortgage-backed securities. We restrict the sample to 30-year fixed rate mortgages and the sample period to post-2000. Summary statistics for these samples are reported in Table A1.

3. Predicting Defaults

We begin our analysis by comparing traditional measures and procedures for assessing borrowers' creditworthiness to the Platform's model using alternative data.

3.1. Credit Score and Traditional Credit Models

We first explore how borrowers' creditworthiness is currently captured by credit score, which is based on outstanding debt, payment history, length of credit history, and types of credit currently utilized. A higher score implying less risk.

We begin by investigating the relationship between credit score and credit card defaults for the sample of borrowers disqualified by the Platform. Specifically, we select loan applicants not approved by the Platform who had no delinquencies at the time of application. This is useful as a benchmark for understanding what to expect in the case of personal loans. Figure 3 plots estimates of the effects of credit score and corresponding 95% confidence intervals of β_{cs} in the following regression:

$$Default_{i,s,t} = \sum_{cs} \beta_{cs} \times cs_i + \mu_{s,t} + \epsilon_{i,s,t}$$

where subscripts i,cs,s, and t represent the individual, 5 point credit score bin, state, and application year respectively. *Default* is a dummy variable that indicates whether an applicant defaulted on at least one credit card account within 12 months of the application. $\mu_{s,t}$ represents $state \times year$ fixed effects. Standard errors are clustered at the state level. The coefficient β_{cs} captures the propensity to default for borrowers whose credit score is in bin cs relative to the omitted credit score bin–620 to 624.

As expected, there is a clear and significant negative relationship between credit score and credit card defaults. The economic magnitude is also important, likelihood of default being about 5% less for borrowers with a 700 than for borrowers with a 620 credit score. We find similar effects in the estimation reported in Table 3, which controls for borrower characteristics and zip by year fixed effects. This pattern is not unique to credit cards. We observe a similar result for traditional mortgage lenders as reported in Figure A1 and in Table A2 in the appendix. Overall, these results suggest that the credit score is, in general, a good predictor of default across multiple loan products.

The use of credit score as measure of creditworthiness is not, however, without limitations. For example, a credit score may not paint an accurate picture of *future* creditworthiness. The characteristics used to compute credit score simply might not be informative for some individuals. Moreover, a higher credit score implies a higher creditworthiness in the past, but a lower credit score does not necessarily imply high future credit risk. Recent college graduates starting their first jobs or recent immigrants may exhibit low credit scores, primarily due to shorter credit histories. Credit score can also be compromised by a simple oversight; credit scores can drop as much as 100 points for a single 30-day late payment, and take more than two years to get back on track.[15]

Exclusive reliance on credit score could also exclude a large fraction of Americans from credit markets altogether. According to recent estimates by the company that owns the FICO algorithm, 28

[15]https://www.nerdwallet.com/article/finance/late-bill-payment-reported

million Americans have files with insufficient data to generate credit scores and 25 million Americans have no credit file at all. These unscorable populations likely include many potentially creditworthy individuals. Such "credit-invisibles" would be automatically disqualified by lenders that rely primarily on credit score. For example, traditional banks typically have a minimum credit score requirement of 660 to qualify for a personal loan.[16] In the mortgage market, Fannie Mae guidelines have specific credit score cut-offs.[17]

Recent advances in technology and big data enable lenders to use data not typically found in credit reports to help identify these otherwise "credit invisible" applicants. Such information include education, employment history, monthly cash flow, type of the device used (Berg et al., 2020), call logs (Agarwal et al., 2019), and time of day the credit application was completed (Berg et al., 2020). For instance, online lender Social Finance (SoFi) has announced that its credit decisions do not take credit score into account at all.[18]

We begin our main analysis by checking the predictive power of the credit score for the Platform's loan sample. Specifically, we estimate equation (1) using the Platform's loan performance data with two measures of default—charge-offs and delinquency—as outcome variables. Charge-off captures instances in which the outstanding balance of a loan is written off as a loss; delinquency is defined as missed payments of 90 days or more. Three measures each of charge-off and delinquency were defined using the 12-month loan history, 24-month loan history, and complete available loan history. Figure 4 plots the estimated β_{cs} with corresponding 95% confidence intervals. Panel A uses the three measures of delinquency as the dependent variables, and Panel B uses the three measures of charge-off as the dependent variables. In contrast to what was reported for credit cards, Figure 4 plots flat curves between 600 and 660, indicating that the probability of default is the same for a Platform borrower with a credit score of 620 as for a Platform borrower with credit score of 660. The relationship is even weaker for charge-offs, borrowers with credit scores of 700 being only slightly less likely to be charged off than borrowers with a credit score of 620. The negative relationship between credit score and delinquency is restored when credit scores improve beyond 660.

We report this specification in a regression framework in Table 4, in which to capture any time-

[16] https://www.creditkarma.com/personal-loans/i/personal-loan-credit-score

[17] See, for instance, the information provided at https://singlefamily.fanniemae.com/media/20786/display

[18] https://www.americanbanker.com/news/will-fintechs-kill-the-fico-score

varying local heterogeneity, we include a number of borrower and loan characteristics as well as zip code by year fixed effects. Panel A uses the default measures calculated using the 12 month history since loan origination, and Panel B reports the coefficient of 'Credit score/100' using 24 month and complete loan histories. We confirm our previous observation in Figure 4–the credit score to not be as predictive of defaults for the Platform as for traditional lenders, especially for low-credit score individuals. In the next section, we unpack this result and explore other factors that drive the lending decisions.

3.2. New Credit Model

We now examine the Platform's credit model. We provide evidence that the new credit model outperforms credit score in predicting defaults and explore how this improved predictability is achieved by supplementing traditional variables with alternative data.

We begin by running a regression similar to the one in equation (1), but with the Platform's measure of borrowers' creditworthiness, which is expressed as a probability of default, as the main independent variable. We refer to this as the Platform default probability henceforth. This regression uses the Platform default probability bins instead of the credit score bins in equation (1). The β_{pd} estimates and their corresponding 95% confidence intervals are plotted in Figure 5. Panels A and B use the entire sample and separate dummy variables to indicate charge-off and delinquency as dependent variables. For both measures, in contrast to the case for credit score we find a monotonically increasing relationship, which Panels C and D show to be robust across credit score categories. This suggests that the Platform is able to achieve more granular pricing even among low-credit score borrowers.

We next investigate the main drivers of the Platform's credit model. To determine whether the Platform default probability is really driven by alternative data, we first consider the extent to which the new measure is captured by other variables that can be observed in the credit report. This hypothesis is tested in Table 5 by regressing the two measures of default on the Platform default probability. We include a number of credit report variables including credit score, borrower income, number of accounts, number of recent inquiries, and outstanding debt balances as well as zip code by year fixed effects to absorb any time-varying regional heterogeneity. Given the different pattern identified for low credit score individuals, we distinguish between borrowers with credit scores below

660 (Columns (1) and (3)) and those with scores above 660 (Columns (2) and (4)).

The dependent variable in columns (1) and (2) is a dummy variable indicating whether a loan has at any time been 90 days or more delinquent and the dependent variable in columns (3) and (4) a dummy variable indicating whether the loan was charged off within 12 months of loan origination. Panel B uses longer time horizons to calculate the measures of default. We suppress other coefficients in Panel B to conserve space. The results show that the Platform default probability is a highly significant predictor of default for both low- and high-credit score borrowers, even after controlling for a number of traditional measures of credit quality. Thus, these findings suggest that its information content is not subsumed by these other traditional variables. We find a one standard deviation increase in the Platform default probability to be associated with a 7% increase in being delinquent and 4.3% increase in the probability of a loan being charged off.

To further corroborate the interpretation that it captures information not contained in the credit report, we plot the distribution of the Platform default probability for different subsamples of credit score. The Panel A of Figure 6 shows that even in the presence of some correlation between credit score and probability of default, as indicated by the flatter right tail for high-credit score borrowers, there is substantial variation in the Platform default probability within a given credit score bin. In other words, the credit score would deem to be similarly creditworthy borrowers who exhibit quite different levels of risk based on the Platform default probability. This is key to providing a superior ability to price risk correctly. The Panel B of Figure 6, which presents the actual default rates for each Platform default probability-credit score bin, clearly shows that the Platform's model is able to differentiate borrowers with different risk levels even when they have very similar credit scores. The Platform's model is able to identify high-risk borrowers even among applicants with high credit scores.

We further show that the new credit model performs better at predicting borrower creditworthiness (using 12-month charged-off as the outcome variable) by investigating the area under the curve (AUC). Ranging from 50% to 100%, the AUC is routinely used in machine learning applications to quantify the predictive power of a metric (Berg et al., 2020). A 50% AUC implies that a metric has no predictive power (i.e., it is random); a 100% AUC implies perfect predictive ability. The samples in Panel A and Panel B of Table 6 include borrowers with credit scores less than 660 and greater than

660 respectively. Each row in the table presents the AUC using credit score and the Platform default probability, with differences indicated in the last column. Column (1) in Panels A and B shows credit score to have little predictive power for all borrowers, and the Platform default probability to have reasonable predictive power of more than 60 across all sub-categories.

We now consider the nature of the information used to supplement the standard content of the credit report. We use Recursive Feature Elimination with Random Forests (RFE-RF) to select the most relevant variables in the Platform default probability. The RFE-RF procedure performs feature selection by iteratively training a random forest model, then ranking the different features and removing the lowest ranking ones–the ones that do not improve the predictive power of the model. This procedure accommodates non-linear effects of, and includes interactions among, the different features, which might be missed in a simpler setting.

We perform the RFE-RF procedure for 3-year loans originated by the Platform using 38 variables, both traditional and non-traditional. Figure 7 Panel A shows that the model improvement becomes marginal beyond 15 variables. The top 15 variables include level of education, type of job, and loan purpose in addition to other variables obtained from the credit report. The level of education is shown to be the most important non-traditional variable that affects the Platform default probability. In Panel B, in addition to the 38 variables, we include pair-wise interactions among the top five variables in Panel A as inputs. The level of education remains one of the top predictors of the Platform default probability.

Table 7 shows the individual contribution of the selected variables by regressing the Platform default probability on both traditional and non-traditional variables. The base model in column (1) includes only variables captured from the credit report; columns (2) through (5) include non-traditional variables. Due to intellectual property concerns, we include these variables as fixed effects to avoid reporting the coefficient estimates. Dummy variables representing level of education are included in column (2), indicating type of employment in column (3), and representing categories of loan purpose and device/technology used in columns (4) and (5), respectively. Column (6) includes all non-traditional dummy variables. We also include zip code by year fixed effects and loan term by year fixed effects.

The results suggest that non-traditional variables are important predictors in the Platform default probability, even after controlling for a host of variables extracted from credit reports. Consistent with the RFE-RF analysis, education has the greatest impact on the Platform default probability, as indicated by the significant increase of the R^2 in column (2). Based on unreported coefficient estimates of the fixed effects and conditional on other controls, the Platform default probability can change by as much as 4.2% depending on level of education. This effect is highly economically significant given the mean Platform default probability of 22%. Device type or technology can move the Platform default probability by about 4.7%, employment type by about 2.8%. These economic impacts are reported in the last row of Table 7.

Our results show the Platform default probability to constitute a significant improvement over credit score, and that improvement to not be solely explained by information available in the credit report. The use of alternative data is in fact critical.

4. Alternative Models and Credit Access

Having described the main features of the Platform's credit model, we now consider the impact on credit availability of basing credit decisions on non-traditional models. Although a hotly debated topic, whether the use of alternative data improves credit access is yet to be informed by empirical evidence. Key questions include whether unconventional underwriting models help potentially creditworthy borrowers invisible to traditional measures to obtain credit, possibly at lower interest rates, and whether better access to credit improves these borrowers' financial outcomes.

In addition to challenges related to accessing proprietary internal data related to fintechs' credit models, the absence of a benchmark makes addressing these questions problematic. The main uncertainty revolves around distinguishing outcomes between credit models that employ alternative data and those that rely on traditional measures.

We tackle this challenge in a unique way. We assess the impact of the Platform's underwriting model on credit access by comparing its outcomes with counterfactual outcomes generated by a traditional model that does not employ the Platform's alternative underwriting features. Rather than building a proxy for the traditional model, we obtained outcomes from a counterfactual model developed in coordination with the CFPB for regulatory reporting purposes (see Ficklin and Watkins

(2019)). This traditional model is our main benchmark. Utilizing the regulatory agency's assessment frees us from basing our analysis on beliefs about traditional lenders' underwriting practices or relying on outcomes provided by a single lender. We observe not only the credit decisions (approved or disqualified) of the traditional model, but also the interest rate based on the traditional model for loan applications. Both the traditional and the Platform model provide a one-to-one mapping between the predicted probability of default and interest rates. If the interest rate predicted by the model is above the usury limit, about 36 percent, then the model denies the application.

The traditional model is a logistic regression model based on more than 1,400 variables available in the credit report (and their combination and slight variations). In addition, the model also includes the loan amount and annual income of the borrower as predictors of default. For comparison, we have plotted the predictive power of two additional models: basic model and traditional + alternative variables model. The basic model is a logistic regression that uses the credit score, number of inquiries, number of accounts, income, debt-to-income ratio, and loan amount as predictors. The traditional + alternative variables model supplements the predicted probability of default from the traditional model with alternative variables: education, employment type, employment industry, loan purpose, device, and technology. The 'test sample' is a 30% random sample used to evaluate the out-of-sample predictive power of these two additional models.

Figure 8 shows that the predictive power of the traditional model can be improved significantly with the addition of alternative variables as predictors. Figure 8 shows that the predictive power of the Platform's model as measured by the Area Under the Curve (AUC) is about 10% greater than that of the traditional model. This analysis is also helpful in decomposing the improvement in default predictability between data and model. In principle, the Platform's model differs from the traditional model along two dimensions: first, it includes additional variables; second, it is a random forest so more advanced than the logistic regression used by the traditional model. Understanding what drives the improvement can be informative for policy makers as well, because they could encourage traditional lenders to rely less on the credit score and the other traditional variables and include additional features in their models. By comparing the AUC for the three models, we find that about two thirds of the improvement is due to the inclusion of alternative data (i.e., going from 60% to 64%), while the remaining one third results from changes in the model (i.e., from 64% to 66%).

Having shown that the Platform's model is more accurate, we can investigate whether this improvement also results in significantly different lending decisions. Figure 9 plots the outcomes of applications based on the two models: the traditional model and the Platform model. Compared to the traditional model, the Platform's model is more likely to approve applicants with lower credit scores. More than 20% of the applicants with lower credit scores who were approved by the Platform's model would have been denied credit if the lender relied on the traditional model. There is an important fraction of individuals for which the two models would predict similar lending decision: borrowers would either be accepted or rejected by both models. Finally, there is also a fraction of higher-credit score borrowers who would be accepted only by traditional lenders. More formally, Panel A of Figure 10 plots the coefficients and corresponding 95% confidence intervals for the regression of the rejection rate based on the traditional model on credit score bin dummy variables with zip code fixed effects. For comparison, we have also included the same estimates using the underwriting model of a larger bank (light blue line) and the traditional model with alternative variables (dark blue line).[19] This figure shows that more than 75% of borrowers with credit scores less than 640 funded by the Platform over the sample period would have been rejected by the traditional model, relative to the borrowers in the highest credit score bin. This fraction declines as the credit score increases, that is, the mismatch between the traditional and the Platform model is magnified among low-credit score borrowers. The differences are larger with the large bank's model, and smaller with the traditional model with alternative data. This finding highlights the potentially uneven consequences of using alternative data: those who benefit are those who are most in need. Examining the subset of funded loans in Panel B using the same regression specification with interest rate as the dependent variable reveals a difference in APR as well; low-credit score borrowers funded by the Platform would have incurred significantly higher interest rates under the traditional model. Hence, the differences between models affect both the extensive and intensive margins. Finally, in Figure 12, we provide further evidence that non-traditional factors, including education and loan purpose (debt consolidation), are among the top five factors that explain the difference between the traditional vs the Platform model using a Recursive Feature Elimination strategy using Random Forests (RFE-RF).

[19]This model is provided by a US bank with $200 billion of assets and is significantly simpler than the traditional model developed by the CFPB. It is a coarse matrix based on credit score and debt-to-income ratio.

4.1. Winners and Losers from the use of alternative data

Next, we explore additional heterogeneity between the traditional and the Platform underwriting models. This allows us to characterize how the composition of the borrowers pool might change in response to changes in the credit underwriting model, such as the adoption of alternative data. Specifically, we can investigate who the borrowers most likely to gain and to lose from the adoption of new data and tools are.

Figure 11 plots the predicted alternative borrower characteristics by outcome for each credit score bin. This approach enables us to investigate whether the use of alternative data improves credit inclusion on the extensive margin by funding loans that would otherwise be rejected, and whether this non-conventional information helps to lower the cost of credit for individuals who presently face expensive credit options. The figure shows that applicants exclusively approved by the Platform's model have a higher level of education, are more likely to be consolidating debt, and are less likely to be an hourly worker. This figure confirms that non-traditional factors such as education, employment type, loan purpose, and the type of the device used are key drivers of the outcome differences.

Table 8 reports the results of regressions that are aimed at understanding the types of borrowers who benefit the most from the Platform's use of alternative data. Panel A of Table 8 regresses the outcome (columns (1) and (2)) and interest rate differences (columns (3) and (4)) on borrower characteristics and zip code fixed effects. The results suggest that higher educated, salaried, and borrowers with thin credit files are more likely to benefit from the use of alternative data. These effects are magnified for borrowers with low credit scores. Panel B of Table 8 use county level measures that proxy for traditionally under-served fractions of the population: minorities, renters, and foreign born. The results suggest that these types of borrowers are more likely to benefit from the use of alternative data.

However, Figure 9 shows that there are borrowers who would have been accepted by the traditional model but are rejected by the Platform. Intuitively, by adding more information to the underwriting models, we could expect some borrowers to be reclassified as riskier than previously expected. Table 9 compares the key alternative characteristics of losers (applicants who would have only been approved by the traditional underwriting models) to the winners (applicants who were only approved by the Platform's model). The regressions control for variables extracted from the credit reports, the five-point credit score bin fixed effects, and year fixed effects. The results suggests that all else equal, losers are less likely to be college educated, less likely to use a computer, less likely to be consolidating debt, older, and have longer credit histories. These findings suggest that, as alternative data might be key in identifying borrowers who are creditworthy but mis-classified by traditional metrics, the same data leads to some borrowers being deemed riskier and to their application being denied.

Table 10 is aimed at better understanding the ex-post financial performance of the *losers* of the Platform's underwriting model. These are the applicants who would have been approved by traditional lenders, but denied by the Platform. This analysis is helpful in better understanding whether the new model and data are helpful in mitigating credit risk for those who tend to be approved based on standard creditworthiness metrics. In principle, the alternative data might not be as helpful in understanding the risks of these higher-score borrowers, and so it might lead lenders to deny borrowers who should have gotten credit. We restrict our analysis to the denied applicants who were able to obtain a personal loan from another lender, and compare subsequent financial performance of these applicants to the subsequent financial performance of the Platform funded applicants. Restricting attention to rejected applicants, who were able to obtain credit from other lenders, allows us to analyze differences in subsequent financial performance without attributing them to access to new credit since both groups received personal loans.

For this analysis, we exploit one of the unique features of our data which contains the subsequent credit report data available for both rejected and approved applicants. We use rejected applicants credit reports in the three months following the rejection to identify applicants who were able to obtain a personal loan from another lender. The outcome variables we are interested in are: 1) whether or not individuals defaulted on credit card debt around 12 months after the application, 2) credit score changes, and 3) whether applicants who did not own a home at the time of application subsequently obtained a mortgage. The results presented in Table 10 suggest that the Platform funded borrowers have performed significantly better financially 12 months since application compared to rejected applicants who were able to secure a personal loan from another lender. The fact that we find similar results for applicants with both high and low credit scores suggests that the alternative data is not only able to identify invisible primes, but it is also able to identify *hidden risk* among applicants with high credit scores. Taken together, these results show that exploiting alternative data is instrumental in expanding credit access for underserved borrowers, but that some borrowers might find themselves with fewer options if more information is incorporated into the credit models.

5. The Effects of Credit Access

Does better access to credit improve borrower outcomes? The finding in the previous section that alternative data models expand access to credit is a desirable outcome only to the extent that less expensive credit benefits newly-enfranchised borrowers, which is ultimately an empirical question. Easier access to credit can ease financial distress by enabling individuals to better smooth income or manage consumption shocks, or potentially be mismanaged or adversely affect borrower choices through debt overhang.

We address this question by quantifying the extent to which being funded by the Platform improves applicants' credit scores and ability to meet future obligations (measured in terms of subsequent credit card defaults) and make a first home purchase. One challenge in assessing the effect of access to credit is that the set of rejected applicants is not typically observed, either at the point of rejection or subsequently. An additional challenge is that comparisons of funded and rejected borrowers are likely to be biased due to heterogeneity. The Platform data allows us to observe both rejected and approved applicants' credit reports for a period up to at least 12 moths. We address the endogeneity concerns in three ways.

First, we exploit the granularity of the data to control for the main borrower characteristics likely to drive individuals' behavior. We ensure that funded and disqualified applicants are comparable by entropy balancing applicants' credit score, age, length of credit history, total liabilities, monthly debt payment, and number of accounts. The sample is restricted to applicants with credit reports available approximately 12 months after the date of application and no delinquencies at the time of the application. We compare outcomes in the year after the application. We are not allowed to go beyond 12 months since the Platform is not allowed to pull credit reports of disqualified applicants 12 months after the initial application.

Table 11 presents the results of this analysis. Columns (1) through (3) examine low-credit score (less than 660) and columns (4) through (6) higher-credit score (greater than 660) borrowers. The dependent variable in columns (1) and (4) is a dummy variable that indicates whether an applicant has been delinquent on at least one credit card within 12 months of the applications. This captures the notion that personal loans that ease borrowers' financial constraints reduce the probability of defaulting on other accounts. The dependent variable in columns (2) and (5) is the change in credit score relative to the time of application, which constitutes another measure of whether borrowers were able to improve their creditworthiness over time. The dependent variable in columns (3) and (6) indicates, for applicants who did not have a mortgage at the time of the application, whether a new mortgage was obtained within 12 months.

The results in column (1) suggest that relative to disqualified applicants low-credit score funded applicants are 1.2% less likely to default on credit cards in the 12 months following loan origination. This is economically significant compared to 17% of applicants who become delinquent on at least one credit card within 12 months of the application date. Low-credit score borrowers also see a slight increase in their credit scores within 12 months of the loan. Moreover, low-credit score borrowers are 0.7% more likely than similar disqualified applicants to make a first-time home purchase. This effect is economically significant compared to the 3.5% of applicants who obtain a first time mortgage within 12 months of making a loan application. The effects are muted for higher credit score borrowers. These results show access to credit to benefit low-credit score borrowers who do not have many outside options.

One concern with the previous results is that, although we control in a non-parametric way for the variables available to lenders at the time of origination, borrowers are likely to differ on unobservable characteristics that may be correlated with both the probability of being funded and later outcomes. In other words, we know that funding decisions are not random. We address this concern by exploiting two key features of the Platform's business operation: that applicants with debt-to-income ratio greater than 50% and/or applicants who filed for bankruptcy within 3 years of application are automatically denied.[20]

Panel A of Figure 15, which plots the percent of the funded applications against the debt-to-income bins, shows the discontinuity in the probability of funding at the debt-to-income ratio of

[20] The maximum allowed debt-to-income ratio in Colorado, Connecticut, Maryland, New York, and Vermont is 45% https://upstarthelp.upstart.com/questions/108501-what-are-the-minimum-credit-requirements-to-receive-a-loan

50%. While the probability of funding declines gradually as the debt-to-income ratio reaches the 50% threshold, it drops sharply to 0% at the cutoff and above. An applicant with a debt-to-income ratio just below the 50% cutoff has about a 20% probability of being funded, and an applicant who has debt-to-income ratio just above the cutoff has a 0% probability of being funded. Panel B of Figure 15, shows the discontinuity in probability of approval at 36 months since the bankruptcy. The probability of approval jumps for about 5% just below 36 months to 13% just after.

Both the discontinuities are not sharp discontinuities—but there is a discontinuity in the probability of being funded at the cutoff. Further, the impact of the debt-to-income ratio on the probability of funding is different below and above the cutoff as can be seen by the different slopes in Panel A of Figure 15. Panel B suggests that the relationship between probability of being funded and months since bankruptcy is quadratic. These institutional feature suggests a fuzzy regression discontinuity design given by the following specification for test that exploits the debt-to-income cutoff. To capture the differences in slopes, $I(\text{Debt-to-income} > 50\%)$ is interacted with the Debt-to-income ratio.

$$Funded_i = \beta_0 + \beta_1 \times I(\text{Debt-to-income} > 50\%) \times \text{Debt-to-income} + \boldsymbol{\beta_2 X} + \mu_{zt} + \eta_i \tag{1}$$

$$Y_i = \gamma_0 + \gamma_1 \times \widehat{Funded}_i + \gamma_2 \times \text{Debt-to-income} + \boldsymbol{\Gamma_3 X} + \mu_{zt} + \mu_i \tag{2}$$

The fuzzy regression discontinuity design for the test that exploits the discontinuity in the months since bankruptcy would be as follows.

$$Funded_i = \beta_0 + \beta_1 \times I(\text{Months since bankruptcy} < 36) + \sum_{i=1}^{2} \beta_{i+1} \times \text{Months since bankruptcy}^i + \boldsymbol{\beta_4 X} + \mu_{zt} + \eta_i \tag{3}$$

$$Y_i = \gamma_0 + \gamma_1 \times \widehat{Funded}_i + + \sum_{i=1}^{2} \gamma_{i+1} \times \text{Months since bankruptcy}^i + \boldsymbol{\Gamma_4 X} + \mu_{zt} + \mu_i \tag{4}$$

Equations 1 and 3 are the first stage regressions which predicts the probability of being funded based on the cutoff. $Funded_i$ is a dummy variable that takes the value one if the application was funded. The dummy variables $I(\text{Debt-to-income} > 50\%)$ and $I(\text{Months since bankruptcy} < 36)$ indicate whether the debt-to-income ratio is greater than the cutoff and months since bankruptcy is less than the cutoff, respectively. μ_{zt} represents zip code $\times$ year fixed effects. Equations 2 and 4 use the predicted $\widehat{Funded}_i$ to estimate the causal effect of credit on outcome variables, Y_i. The outcome variables, Y_i, are the same as in the previous specifications–default on credit cards, change in the credit score, and new mortgage. X contains the same set of control variables as other regressions.

Panels A and B in Figure 16 show that there is no discontinuity at both the cutoffs when we consider other applicants' characteristics, which provides support for our identifying assumption.

Table 12 reports the estimation results of equation 2. The results in columns (1), (2), and (3) are consistent with the results of the entropy balancing. The economic effects are much larger than the estimated effects from entropy balancing (Table 11). Specifically, we find that applicants who get funded are significantly less likely to default on their credit cards, by about 20%, their credit score increases by about 9%, and 13.4% more likely to obtain a mortgage, compared to applicants who were not funded. Table 13 presents similar results when using the discontinuity in months since bankruptcy for identification. Overall, these findings strongly suggest that having access to credit, thanks to the more predictive underwriting models of the Platform, results to an improvement in the borrowers' financial health.

6. Profitability

Having shown that subprime borrowers benefit from the Platform's novel pricing strategy, we consider whether it is profitable for the lenders to identify and lend to invisible primes. In principle, fintech lenders might have an incentive to acquire market share by lending to riskier and potentially less profitable borrowers.

We begin by computing the average internal rate of return for each origination year-credit score bin for loans originated by the Platform (as in Jansen et al. (2019)). The sample being restricted to loans originated before March 2020, loans have at least a one-year performance history to enable computation of the IRR. Three-year loans are chosen as the main sample, as they have a longer history compared to the loan term.[21] Panel A in Figure 13 plots the computed IRR against the credit score bins using the entire available loan history. Each line represents the loans originated in

[21] We obtain similar results for the five-year loan sample.

a particular year. For loans still active as of March 2021, we use the outstanding amount as the final cash flow. The figure shows the IRR to be weakly negatively correlated with credit score across years. In Panel B and C, cash flows are restricted to one and two years since loan origination, respectively. We use the outstanding amount at the end of one-year since origination as the terminal cash flow. This enables us to compare the IRR over the same time horizon for each vintage. The figure also shows the return from low- relative to high-credit score borrowers to improve over time.

Table 14 presents similar results in regression form. We regress the IRR($\times$100) of each origination year-credit score bin-loan term portfolio on a dummy variable indicating borrowers with credit scores less than 660 interacted with dummy variables indicating loan origination year. Column (1) uses the entire available loan history, columns (2) and (3) the 12-month and 24-month history since loan origination, for the IRR calculation. The results suggest that, on average, low-credit score borrowers generate slightly higher returns. The results further show returns from low-credit score borrowers to be about 1 percentage point higher in 2019 than in 2018. Low-credit score loans originated in 2020 generated returns approximately 2 percentage points higher than in 2018. This evidence suggests that identifying creditworthy individuals among those deemed too risky by traditional models may provide additional returns to the platform. Figure 14 compares the IRR of loans approved only by the Platform model to IRR of loans approved by both the traditional model (see section 4 for more details) and the Platform model.

Table 15 examines how IRR varies with borrower characteristics. The sample is restricted to loans with at least a 12-month history. The IRR was calculated for each loan (as opposed to each credit score bin in the previous analysis). For loans current at the end of the sample period, the outstanding principal was used as the terminal cash flow. Columns (1) through (4) include the most important alternative variables in the Platform's underwriting model as per section 3.2 as well as the traditional determinants of borrower creditworthiness. The results suggest that applicants with higher levels of education who use funds for debt consolidation, are salaried employees, and use a computer to complete the application generate significantly higher returns even after accounting for lower interest rates.

7. External Validity

A natural question is whether these results are specific to the Platform and thus not applicable to other institutions. Although the data provided by the Platform is unique, we provide suggestive evidence that similar patterns might hold more generally.

Specifically, we report the default behavior of privately securitized mortgages as well as mortgages in the Freddie single-family loan dataset. We compare the performance of mortgages originated by the three largest banks (Bank of America, Chase, and Wells Fargo) to the performance of mortgages originated by Quicken Loans, which, being the dominant fintech mortgage lender and processing loan applications entirely online, is likely to be better positioned to leverage alternative data sources. We find that for traditional lenders the FICO score is a good predictor of default (light blue lines in A1) for both subprime (Panel A) and prime (Panel B) borrowers. Similarly to the Platform, the performance of mortgages originated by Quicken is generally flat (dark blue lines in A1), particularly for low-FICO score borrowers, which suggests that FICO score is not a good predictor of default for fintech lenders like Quicken. Figure A2 shows the FICO distribution for originated loans to differ between Quicken and the banks, the former tending towards the left side of the distribution.

In Table A3, using mortgage application data provided under the Home Mortgage Disclosure Act (HMDA) for years 2018 and 2019, we show that, compared to other lenders, fintech mortgage lenders are more likely to lend to borrowers whose creditworthiness is better assessed using alternative information.[22].

Using the Freddie Mac dataset, we compare profitability between mortgages originated by Quicken Loans and those originated by the three largest banks. Because the Freddie Mac sample identifies Quicken Loans starting in 2012, we restrict the sample to 30-year mortgages originated after 2011.[23] In Table A4, we regress the IRR (100) of each mortgage on *Quicken*, a dummy variable that indicates whether the mortgage was originated by Quicken Loans. We control for loan amount, debt-to-income ratio, loan-to-value ratio, FICO score, loan purpose (new purchase vs. refinance), and zip code by year fixed effects. Column (1) uses the complete sample; columns (2) through (5) estimate the same regression for subsamples based on FICO score. The results suggest that mortgages originated by

[22] We follow Buchak et al. (2018) in identifying fintech lenders

[23] The originator's name is populated if the lender originated at least 1% of the loans in a given quarter.

Quicken Loans, compared to those originated by other lenders, generate about 12bp higher return, which is 3.3% of the mean IRR of 3.6%.

Overall, these results are consistent with our main findings not being confined to the Platform sample; they suggest that use of alternative data has enabled fintech lenders to provide credit to less creditworthy borrowers and in doing so to achieve higher profitability.

8. Robustness

8.1. Selection

One concern with our findings is that a specific type of borrowers are self-selecting into the Platform and that is the reason for better performance of the Platform loans, not the alternative data or the better underwriting model. On the one hand, borrowers who apply to a fintech lender might be more sophisticated than those applying to traditional banks and can then be less likely to default. On the other, the borrowers that apply to new lenders might have been forced out of the traditional financial system and can then be riskier. In this section, we exploit the data about the way in which borrowers apply to address these potential selection issue.

To address this concern, we estimate our main results separately for applicants coming through loan aggregators, online advertisements on Google or Facebook, and directly from the Platform's website. Intuitively, loan aggregators collect information from borrowers and then present the best available offers to them. Then, for those borrowers, the decision to apply to the Platform is not driven by potentially higher sophistication but just by the automatic comparison provided by a loan aggregator like Credit Karma. We find that results are similar across these three types of applicants. First, we evaluate the predictability of credit score for the borrowers with credit scores less than 660. The predictability of credit score is low across the three types of applicants as indicates by AUCs of 51.02%, 51.14%, and 52.55% for applicants coming thorough loan aggregator websites, online ads, and the Platform's website, respectively. In contrast the AUC of the Platform's probability of default is 59.57%, 60.56%, and 60.38% for applicants coming thorough loan aggregator websites, online ads, and the Platform's website, respectively. We find similar consistent results for borrowers with credit scores greater than 660.

In Figures 17, 18, and 19, we replicate the results that compare model outcomes of the traditional model and the Platform's model separately for applicants coming thorough loan aggregator websites, online ads, and the Platform's website. We do not find any noticeable difference in any of the measures across these three types of applicants. Similarly, the Figure 20 compares the predictability of different measures of creditworthiness. Since we confirm the main results of the paper, irrespective of the channel used by the loan applicant, we believe the selection into the applicant pool is not a significant concern.

8.2. Alternative Data about Unfunded Borrowers

The main results of this paper focus on the borrowers who received funding from the Platform, and the Platform has a minimum credit score criteria of 600. A natural question is whether alternative variables are informative for individuals whose credit score is lower than 600 and for unscored individuals. In other words, we can test whether the role of alternative data holds also for the non-funded loans, which would suggest that the improvements we have documented, in terms of financial inclusion and default prediction, hold for a wider population than just the one funded by the Platform. We attempt to answer this question by focusing on disqualified the Platform applicants with credit scores between 500 and 600 at the time of application. We restrict our sample to applicants who did not have any delinquency at the time of application and whose credit reports are available approximately 12 months since application date.

In Table 16 columns (1) through (3), we regress whether there is any delinquency approximately 12 months since application on selected alternative variables. We include standard control variables and Zip code $\times$ year fixed effects. In addition, we also include fixed effects representing five-point credit score bins. The dependent variable in columns (4) through (6) takes the value one if there are credit cards that are 90+ days delinquent approximately after 12 months since application. We find that level of education, type of employment, and type of device used to fill out the application are strong predictors of default even in this sample.

8.3. Robustness to economic cycles

One potential defense of the traditional models is that their simplicity might potentially make them more resilient during bad economic times. In other words, it is possible that the Platform is funding ‘high-beta’ applicants whose performance is significantly better during good economic times, but perform worse during bad economic times. This concern goes against intuition since the population of funded borrowers is heavily skewed towards more educated borrowers and salaried, rather than hourly, employees, who are usually more recession-proof than others. However, to directly address this concern, we compare the performance of the Platform’s loans to the performance of personal loans originated by other lenders during the economic fallout caused by the COVID-19 pandemic in early 2020.

For this exercise we rely on a DV01 report that provided insights into consumer loan performance during COVID-19 pandemic.[24] In Figure 21 we compare the average impairment rate of consumer loans originated by *online lenders* to the impairment rate of the Platform loans. An impairment is defined to any loan that is delinquent or has an active modification status, and impairment rate is calculated as number of impaired loans as a percentage of total outstanding loans. Panel A compares the performance of all the Platform loans to the performance of all loans originated by online lenders. Panel B compares the performance of the Platform loans with credit score less than 660 to the bottom 7% of online lenders’ loan portfolio in terms of the credit score. Both figures suggest that impairment rate increase in the Platform loans is less than that of the average increase in the impairment rate. This gap is significantly higher for borrowers with low credit scores.

9. Conclusion

Fintech lenders’ increasing prominence in the market for unsecured loans is elevating the importance of understanding their methods and the implications of their participation. Pursuit of this understanding has been hampered by the lack of detailed administrative data about fintech lenders’ operations.

We exploit unique data from a large fintech platform to shed new light on this sector. We show that implementing alternative data leads to substantially more predictive power with respect to the

[24]Dv01 Insights. (2021). Covid-19 performance report (Vol. 29, Issue November).

likelihood of default than credit score, the standard metric traditionally used to judge borrower creditworthiness. We further show that the superior ability to predict default rates translates into broader access to credit, particularly for borrowers with low credit scores. These effects are detected at the extensive margin, that is, whether an individual is able to access credit, and in the pricing of loans, that is, the interest rates at which lenders are willing to fund loans. The beneficiaries of fintech underwriting models are low-credit score borrowers who would otherwise likely be denied credit under traditional underwriting models or subjected to high interest rates. That granting credit to these individuals translates into higher returns for the Platform suggests that there might be a private incentive to adopt the new credit models. Low-credit score individuals able to access credit also exhibit a much lower probability of defaulting on other liabilities, such as credit cards, and are often able to subsequently improve their credit scores, affording greater access to credit from traditional lending sources.

If the benefits of modern underwriting algorithms that use alternative data and artificial intelligence are potentially large, so are the risks. Regulators and consumer advocates have been concerned, for example, with the potential for biased treatment, which would violate fair lending regulations. Furthermore, using information about education, utility bills, or bank transactions could inadvertently reduce credit access for *some* households. Although our setup does not allow us to address arguments around potential concerns about privacy and statistical discrimination, our results demonstrate that alternative data models deliver quantifiable benefits to both low-score borrowers and lenders, while also leading to the identification of hidden risk among high-score borrowers, who would have been funded according to traditional models.

References

Agarwal, S., Alok, S., Ghosh, P., and Gupta, S. (2019). Financial inclusion and alternate credit scoring: role of big data and machine learning in fintech. *Indian School of Business*.

Balyuk, T. (2019). Financial innovation and borrowers: Evidence from peer-to-peer lending. *Rotman School of Management Working Paper*, (2802220).

Balyuk, T. and Davydenko, S. A. (2019). Reintermediation in fintech: Evidence from online lending.

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2019). Consumer-lending discrimination in the fintech era. Technical report, National Bureau of Economic Research.

Ben-David, I., Johnson, M. J., Lee, J., and Yao, V. (2022). Fintech lending with lowtech pricing.

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

Berg, T., Fuster, A., and Puri, M. (2021). Fintech lending. *Annual Review of Financial Economics*, 14.

Bilan, A., Degryse, H., O'Flynn, K., and Ongena, S. (2019). Fintech and the future of banking. In *Banking and Financial Markets*, pages 179–199. Springer.

Blattner, L. and Nelson, S. (2021). How costly is noise? data and disparities in consumer credit. *arXiv preprint arXiv:2105.07554*.

Brevoort, K. P., Grimm, P., and Kambara, M. (2016). Credit invisibles and the unscored. *Cityscape*, 18(2):9–34.

Buchak, G., Matvos, G., Piskorski, T., and Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483.

CFPB (2017). Cfpb explores impact of alternative data on credit access for consumers who are credit invisible. https://www.consumerfinance.gov/about-us/newsroom/cfpb-explores-impact-alternative-data-credit-access-consumers-who-are-credit-invisible. Accessed: 2020-07-19.

Danisewicz, P. and Elard, I. (2018). The real effects of financial technology: Marketplace lending and personal bankruptcy. *Available at SSRN 3208908*.

De Roure, C., Pelizzon, L., and Thakor, A. V. (2019). P2p lenders versus banks: Cream skimming or bottom fishing?

Di Maggio, M. and Yao, V. (2020). Fintech borrowers: Lax-screening or cream-skimming? *The Review of Financial Studies (forthcoming)*.

Dobbie, W., Liberman, A., Paravisini, D., and Pathania, V. (2018). Measuring bias in consumer lending. Technical report, National Bureau of Economic Research.

Ficklin, P. and Watkins, P. (2019). An update on credit access and the bureau's first no-action letter. https://www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter/. Accessed: 2020-07-19.

FICO (2015). Can alternative data expand credit access?

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2020). Predictably unequal? the effects of machine learning on credit markets.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The role of technology in mortgage lending. *The Review of Financial Studies*, 32(5):1854–1899.

Ghosh, P., Vallee, B., and Zeng, Y. (2021). Fintech lending and cashless payments. In *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

Hau, H., Huang, Y., Shan, H., and Sheng, Z. (2019). How fintech enters china's credit market. In *AEA Papers and Proceedings*, volume 109, pages 60–64.

Hertzberg, A., Liberman, A., and Paravisini, D. (2018). Screening on loan terms: evidence from maturity choice in consumer credit. *The Review of Financial Studies*, 31(9):3532–3567.

Homonoff, T., O'Brien, R., and Sussman, A. B. (2019). Does knowing your fico score change financial behavior? evidence from a field experiment with student loan borrowers. *Review of Economics and Statistics*, pages 1–45.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577.

Jagtiani, J. and Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the lending club consumer platform. *Financial Management*, 48(4):1009–1029.

Jansen, M., Nguyen, H., and Shams, A. (2019). Human vs. machine: Underwriting decisions in finance.

Karlan, D. and Zinman, J. (2010). Expanding credit access: Using randomized supply decisions to estimate the impacts. *The Review of Financial Studies*, 23(1):433–464.

Liao, L., Wang, Z., Yan, H., and Zhou, C. (2020). When fintech meets privacy: The consequence of personal information misuse in debt collection.

Mariotto, C. (2016). Competition for lending in the internet era: The case of peer-to-peer lending marketplaces in the usa. *Available at SSRN 2800998*.

Melzer, B. T. (2011). The real costs of credit access: Evidence from the payday lending market. *The Quarterly Journal of Economics*, 126(1):517–555.

Morse, A. (2015). Peer-to-peer crowdfunding: Information and the potential for disruption in consumer lending. *Annual Review of Financial Economics*, 7:463–482.

Tang, H. (2019a). Peer-to-peer lenders versus banks: substitutes or complements? *The Review of Financial Studies*, 32(5):1900–1938.

Tang, H. (2019b). The value of privacy: Evidence from online borrowers.

TransUnion (2019). Fintechs continue to drive personal loan growth.

Vallee, B. and Zeng, Y. (2019). Marketplace lending: a new banking paradigm? *The Review of Financial Studies*, 32(5):1939–1982.

Wolfe, B. and Yoo, W. (2018). Crowding out banks: Credit substitution by peer-to-peer lending. *Available at SSRN 3000593*.

Figure 1. Platform's Loan Growth and Source

Panel A of this figure plots the number of loans originated by the platform in each year, Panel B the total amount lent by the platform in each year, Panel C the referring domain associated with each application, and Panel D the loan purpose distribution. Panels C and D use the subsample of loans originated in or after year 2019.

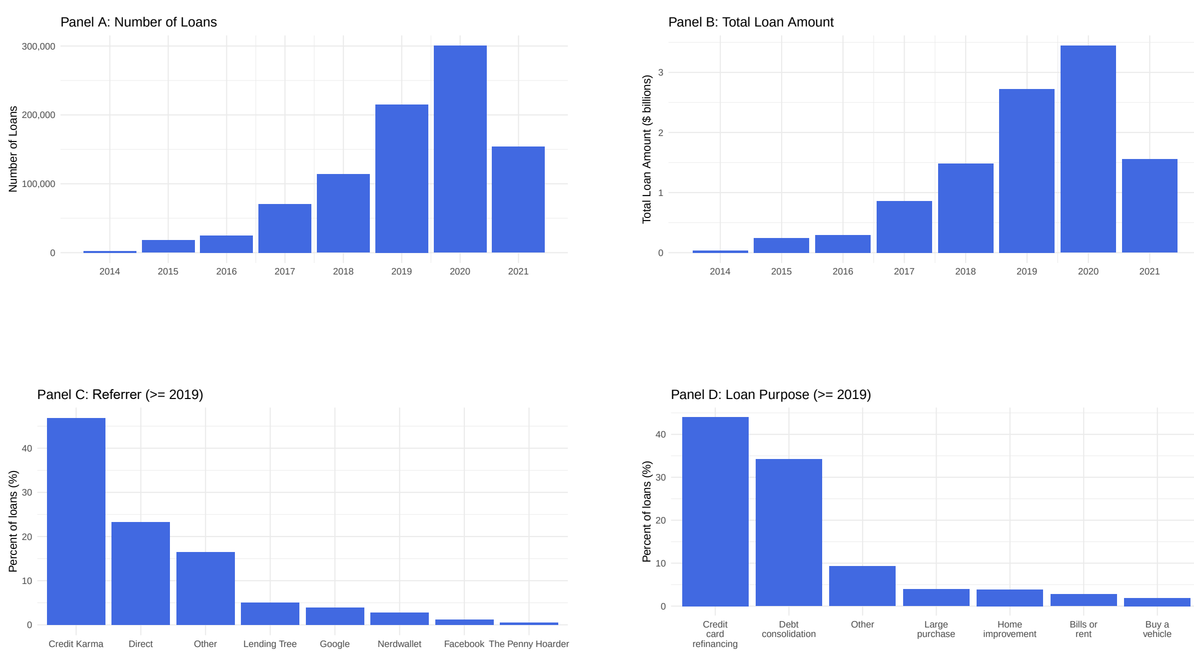


Figure 2. Geographic Coverage

This map shows the Platform's geographic coverage. The figure plots the total number of loans originated per 100 people in each county.

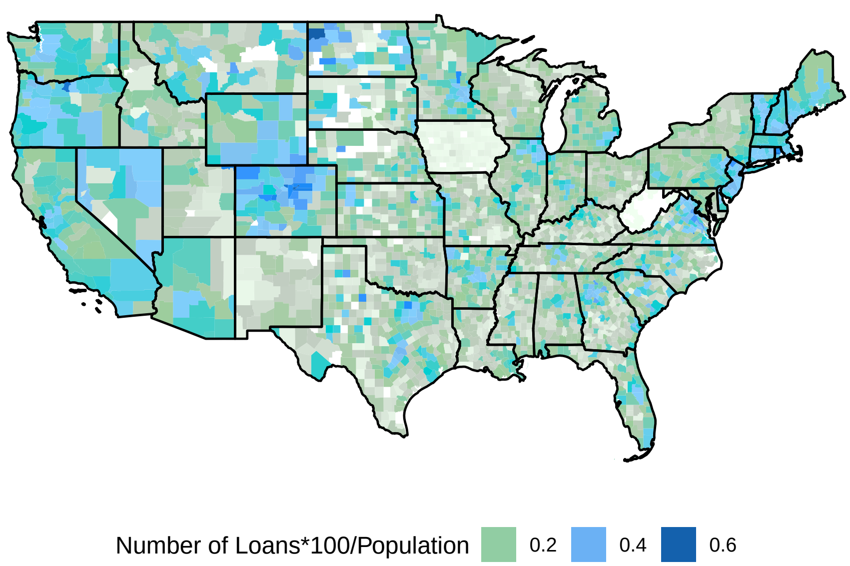


Figure 3. Predictability of Credit Score: Credit Cards

This figure plots the estimates of β_{cs} and corresponding 95% confidence intervals in the following estimation using the sample of applicants rejected by the Platform. The sample excludes applicants who had delinquent accounts at the time of application. Subscripts i,cs,s, and t represent the applicant, credit score bin, state, and application year, respectively. *Default* is a dummy variable that takes the value of one if applicant i defaulted on at least one credit card within 12 months of the application. $\mu_{s,t}$ represents $state \times year$ fixed effects. Standard errors are clustered at state level.

$$Default_{i,s,t} = \sum_{cs} \beta_{cs} \times cs_i + \mu_{s,t} + \epsilon_{i,s,t}$$

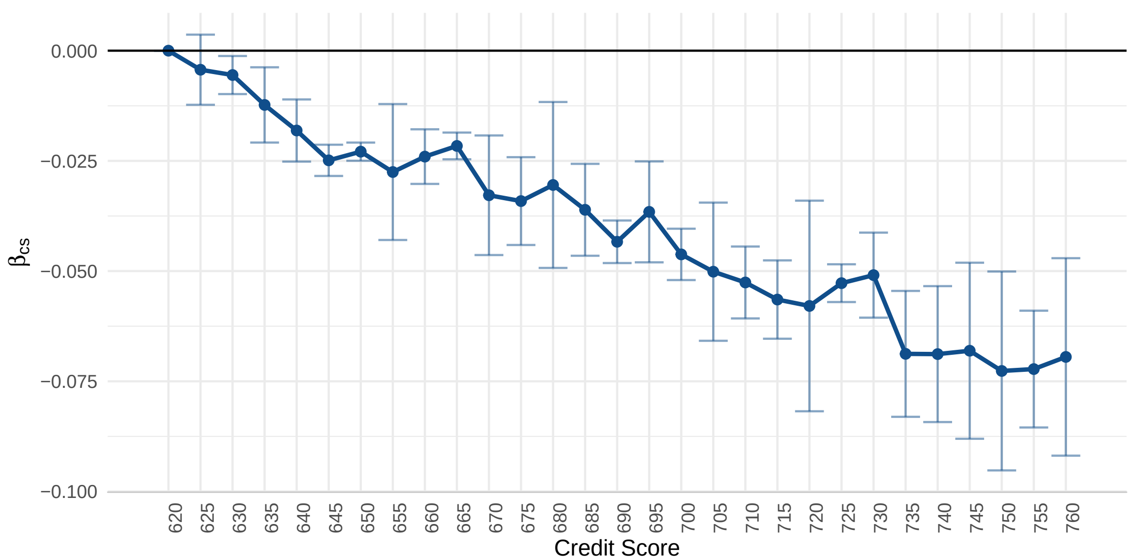
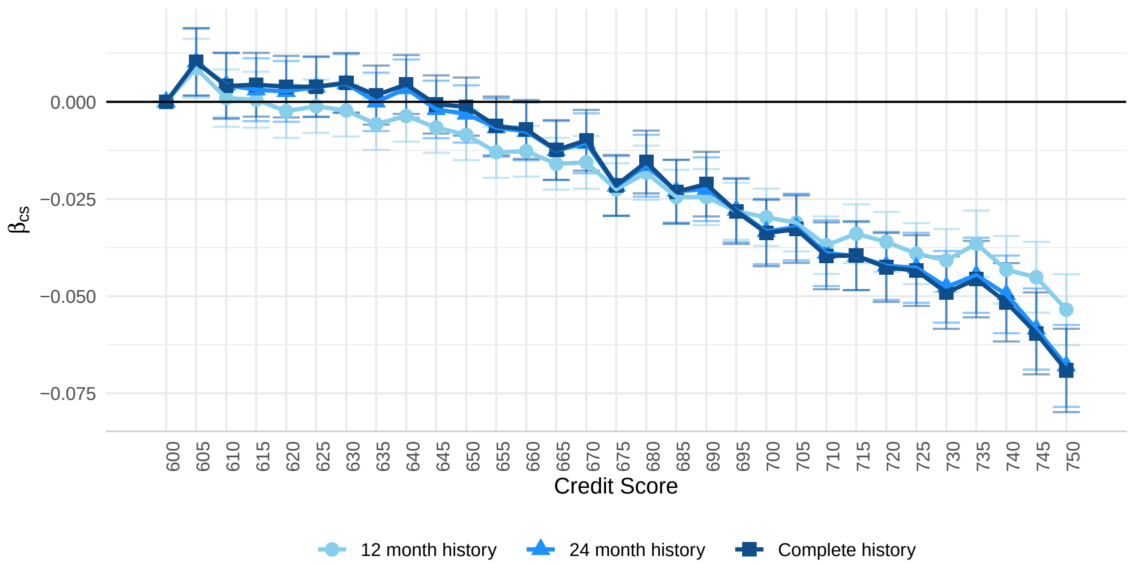


Figure 4. Predictability of Credit Score: the Platform Loans

This figure plots the estimates of β_{cs} and corresponding 95% confidence interval in the following estimation using the Platform's loan portfolio. Subscripts i,cs,s, and t represent the borrower, credit score bin, state, and loan application year respectively. Y is a dummy variable that takes the value of one if loan i was 90 days or more delinquent (Panel A) at any point after origination, or charged-off (Panel B). $\mu_{s,t}$ represents $state \times year$ fixed effects. Standard errors are clustered at state level.

$$Y_{i,s,t} = \sum_{cs} \beta_{cs} \times cs_i + \mu_{s,t} + \epsilon_{i,s,t}$$

Panel A: Dependent variable: Delinquent



Panel B: Dependent variable: Charged-off

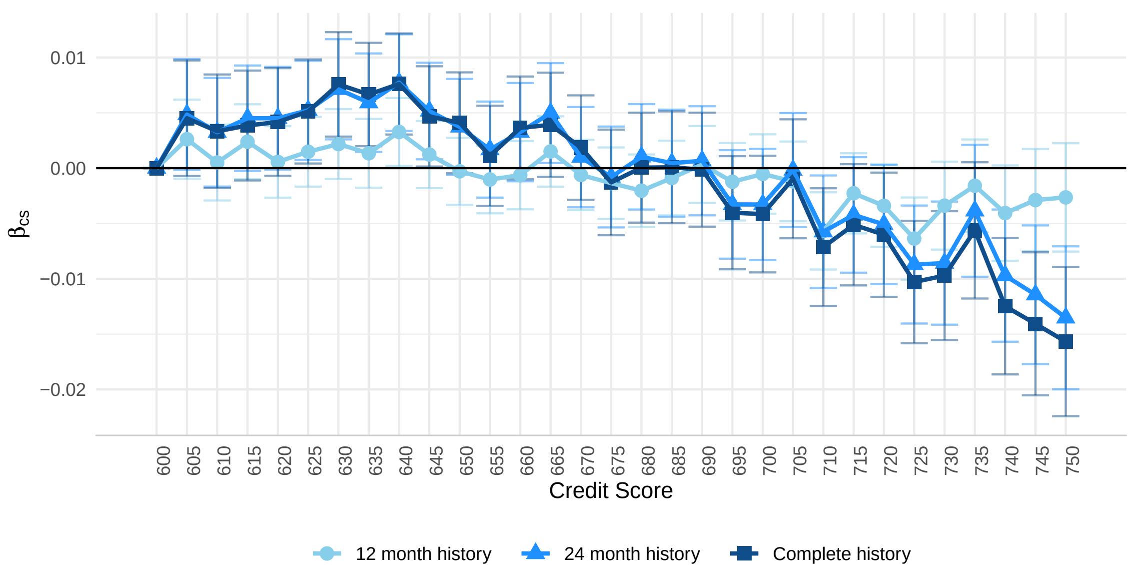


Figure 5. Predictability of the Platform Default Probability

This figure plots the estimates of β_{pd} and corresponding 95% confidence interval of the following estimation. Subscripts i,pd,s, and t represent the borrower, the Platform default probability bin, state, and loan application year, respectively. Y is a dummy variable that takes the value of one if applicant i defaulted or charged-off depending on the specification. $\mu_{s,t}$ represents $state \times year$ fixed effects. Panel A uses the entire sample of loans; Panel B estimates the regression separately for each credit score category using charge-off as the dependent variable. Standard errors are clustered at the state level.

$$Y_{i,s,t} = \sum_{pd} \beta_{pd} \times pd_i + \mu_{s,t} + \epsilon_{i,s,t}$$

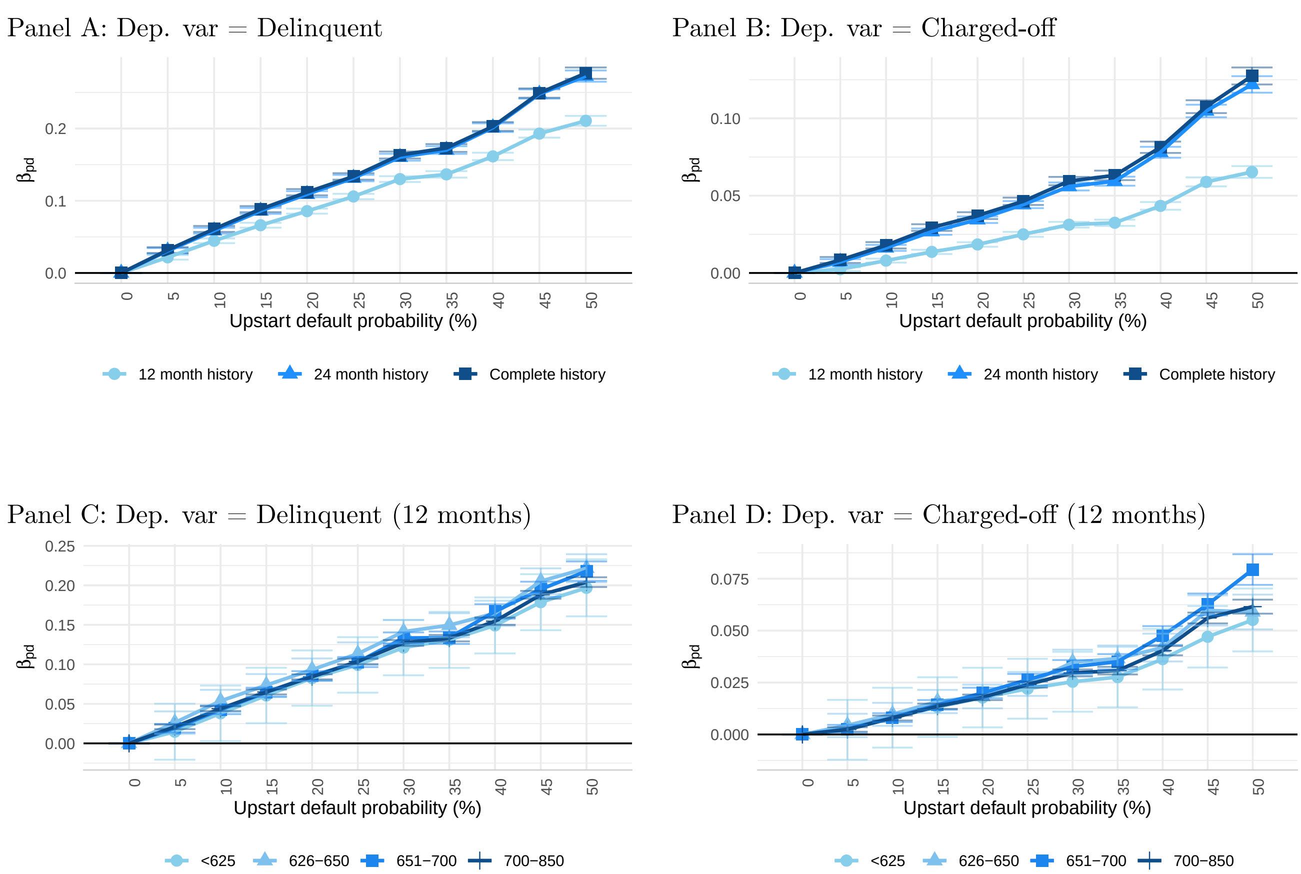
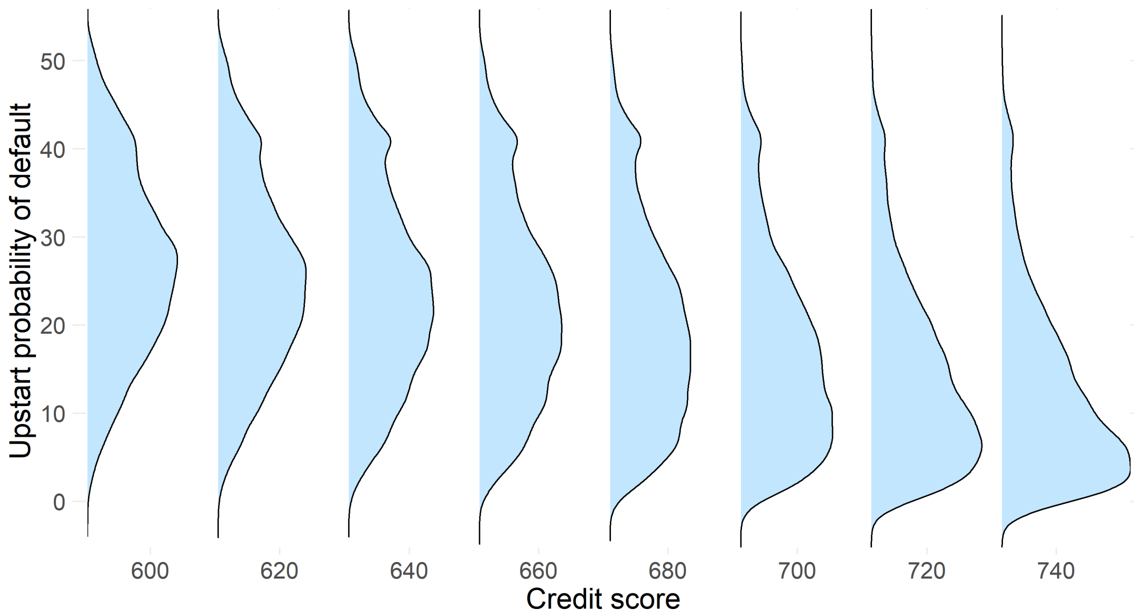


Figure 6. Credit Score and the Platform Default Probability

Panel A of this figure plots the distribution of the the Platform default probability for loans originated by the Platform separately for each credit score category. Panel B plots the actual default rates for each credit score - Platform default probability bin.

Panel A: Distribution of the the Platform Default Probability



Panel B: Default Rate

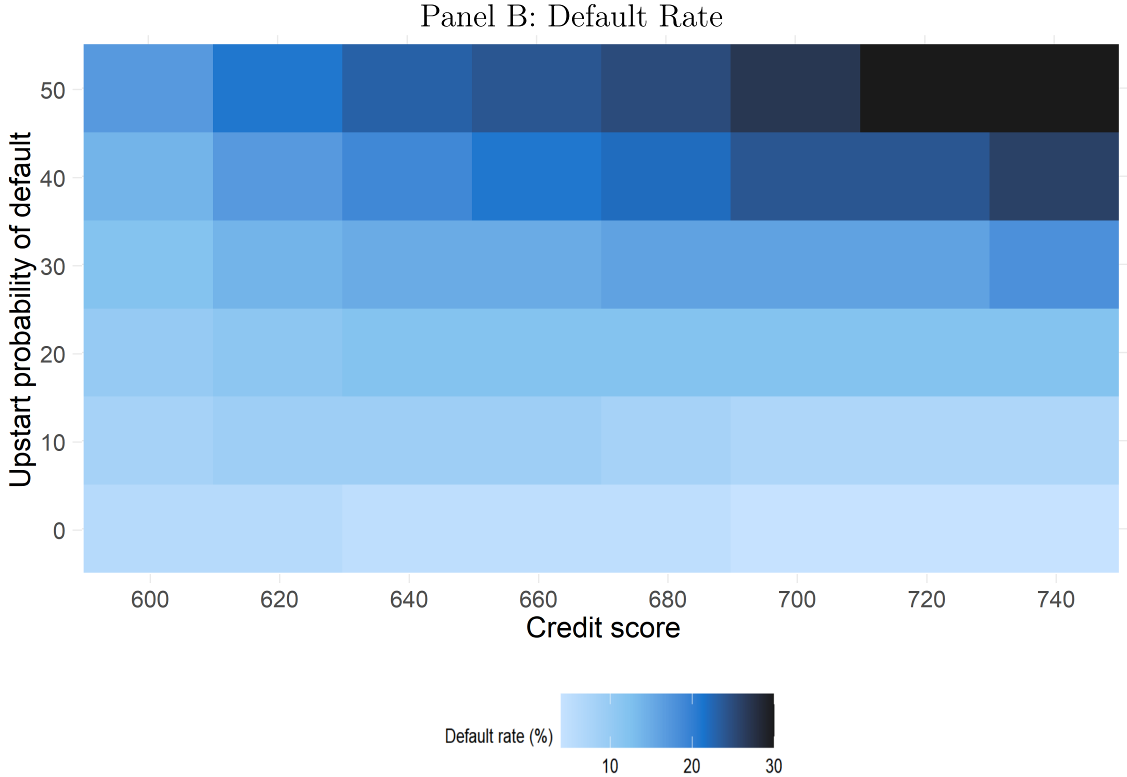


Figure 7. Recursive feature elimination: Root-mean-square deviation

This figure plots the incremental improvement of the model performance when new predictive variables are added to the model that predicts the Platform default probability. Predictors are added in the order of importance, and include both traditional and non-traditional variables. The model in Panel A only use linear variables as inputs and the model in Panel B includes pair-wise interactions among the top five variables in Panel A.

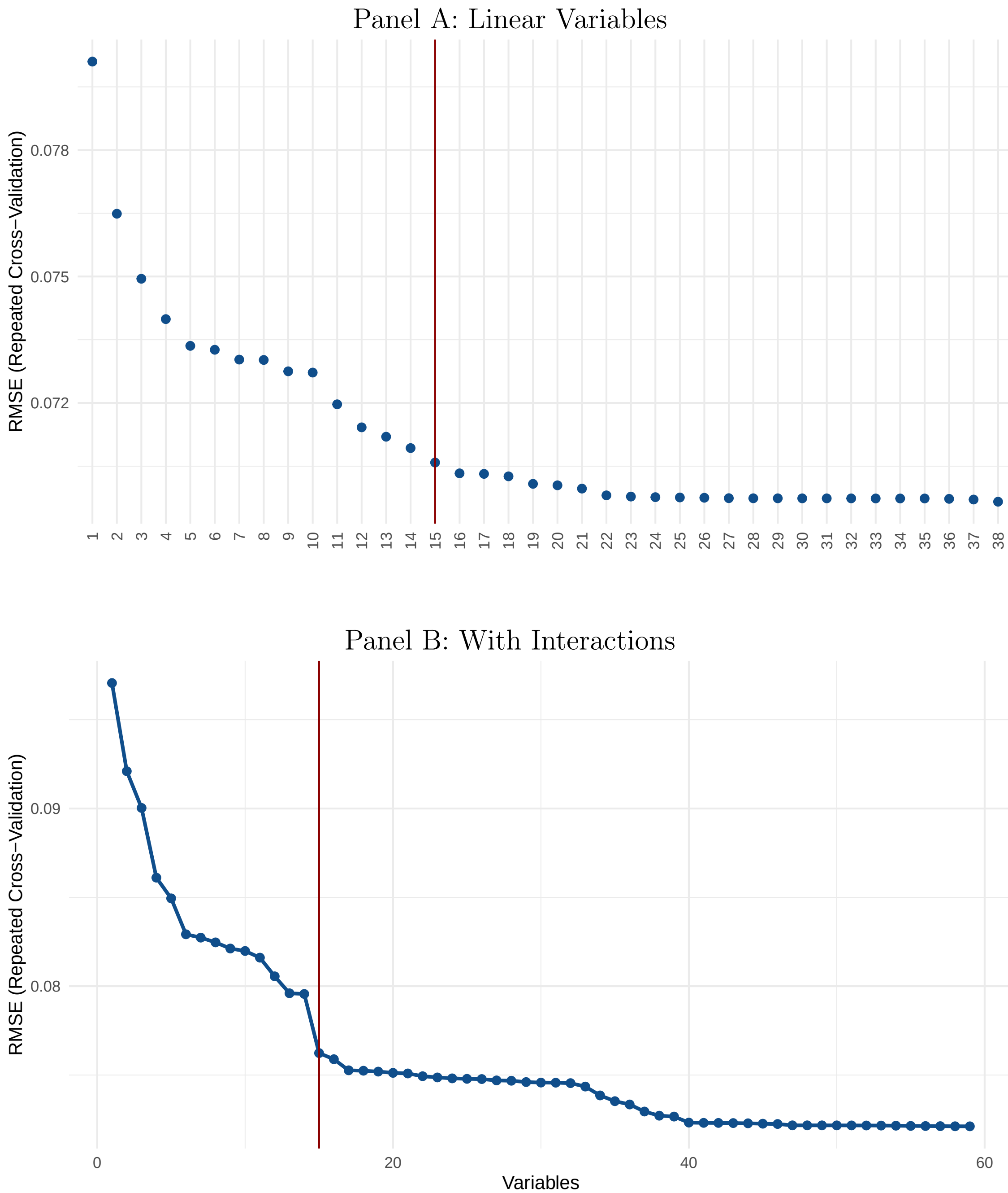


Figure 8. Predictive power of different models

This figure plots the area under the curve (AUC) for the funded Platform sample using credit worthiness measures estimated using different models.

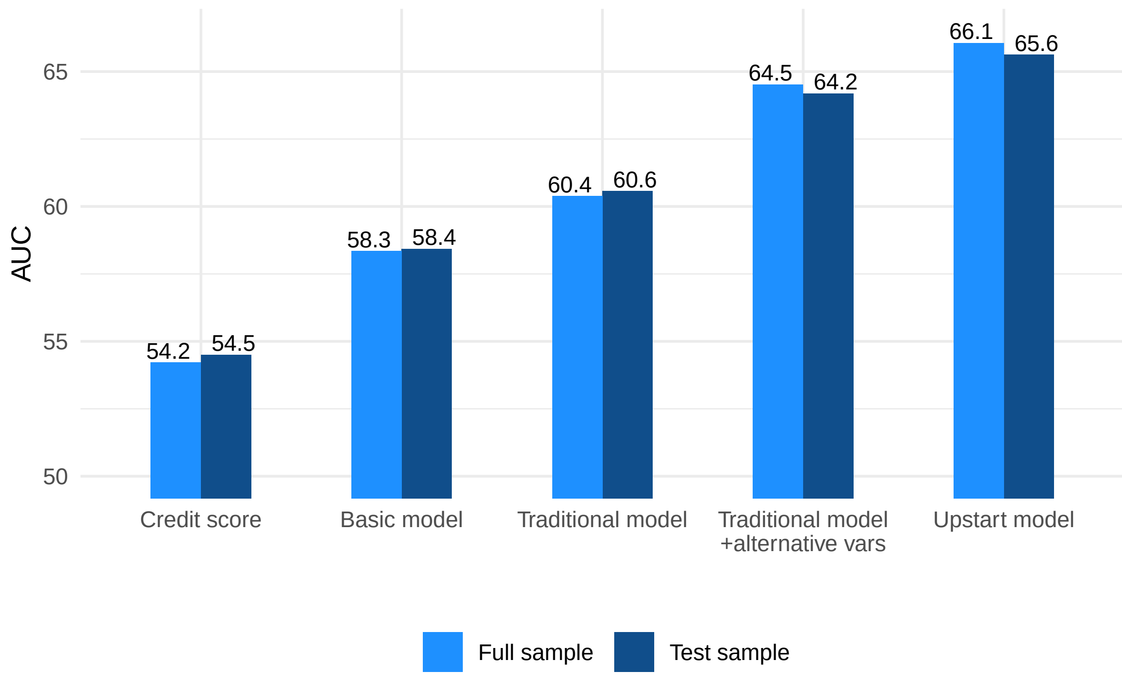


Figure 9. Traditional Model vs. Platform's Underwriting Model

This figure plots the outcome distribution under each model. The figure plots the share of each outcome against the credit score bin.

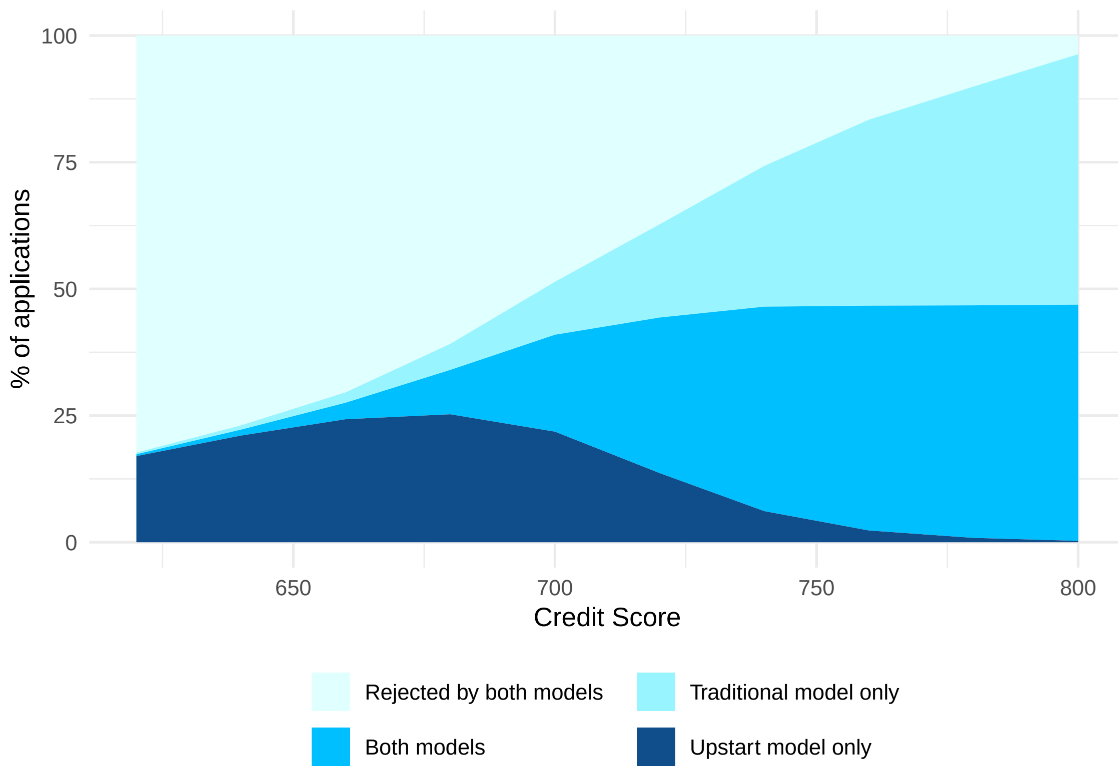
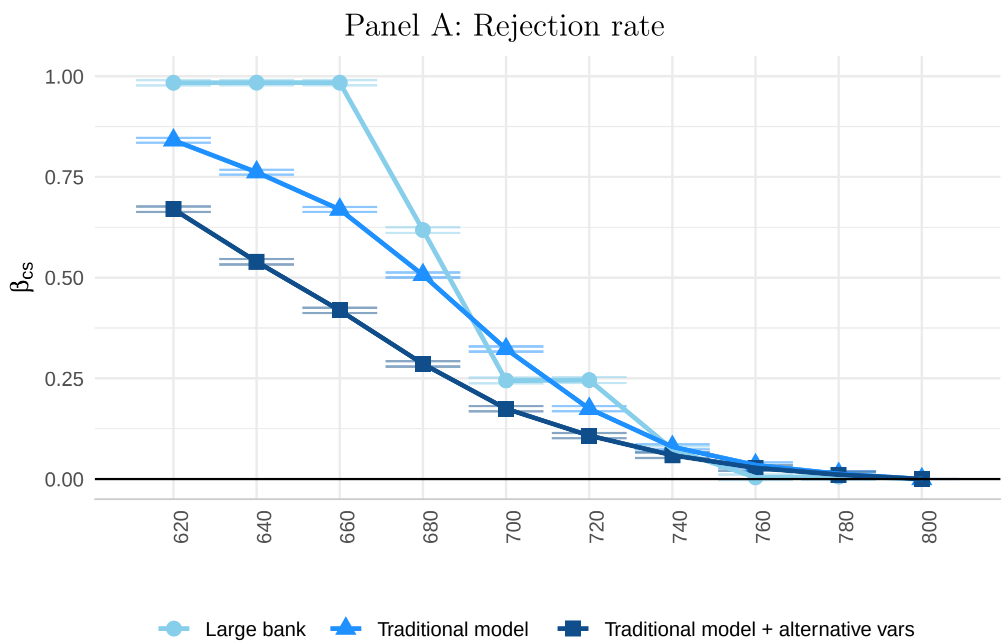


Figure 10. Model Outcome Comparison

Panel A of this figure plots the coefficients and corresponding 95% confidence intervals of the regression that regresses the rejection rate based on the traditional model on credit score bin dummy variables with zip code fixed effects. Panel B plots the estimation results for regressions that regress the interest rate under each model on the same variables. The sample consists of loans funded by the Platform for which the traditional model outcomes are available. Standard errors are clustered at the zip code level.



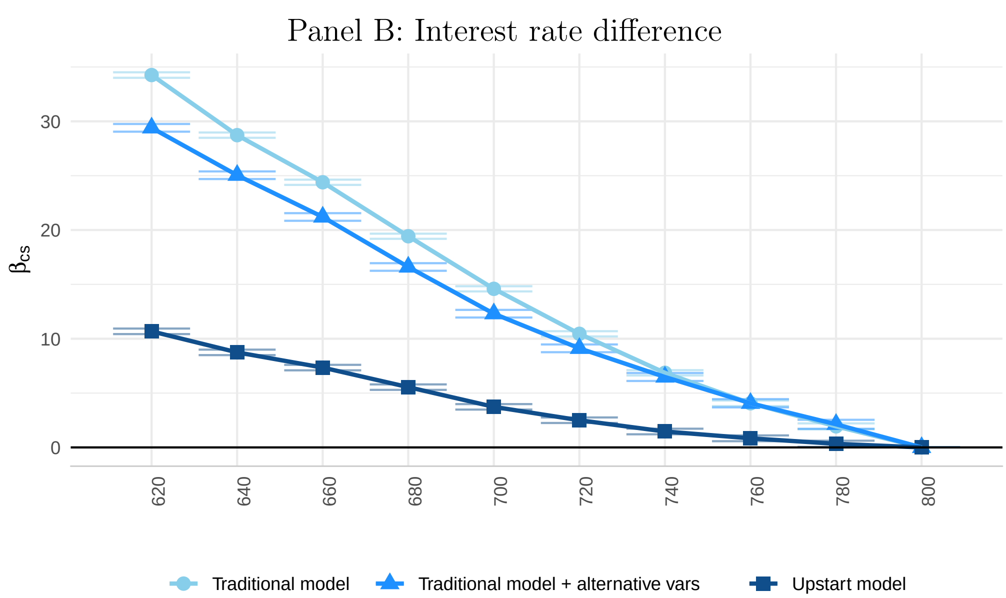


Figure 11. Traditional Model vs. the Platform's Underwriting Model: Outcome Heterogeneity

This figure plots how alternative borrower characteristics differ based on the model outcomes. Each panel plots the predicted characteristic on the y-axis based on a regression that regresses the characteristic on the outcome interacted with the credit score bin dummy.

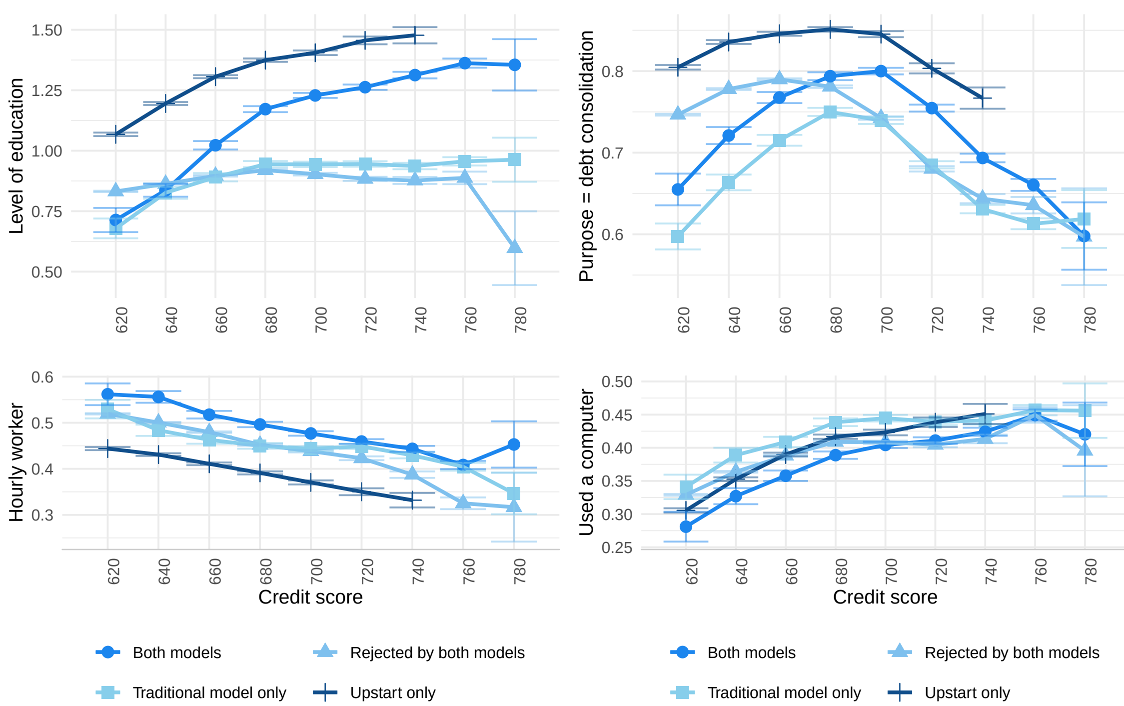


Figure 12. Traditional Model vs. the Platform's Underwriting Model: Contributing Factors

This figure plots the incremental improvement of the model performance when new predictive variables are added to the model that predicts whether a loan application is rejected by the traditional model, but approved by the Platform's model.

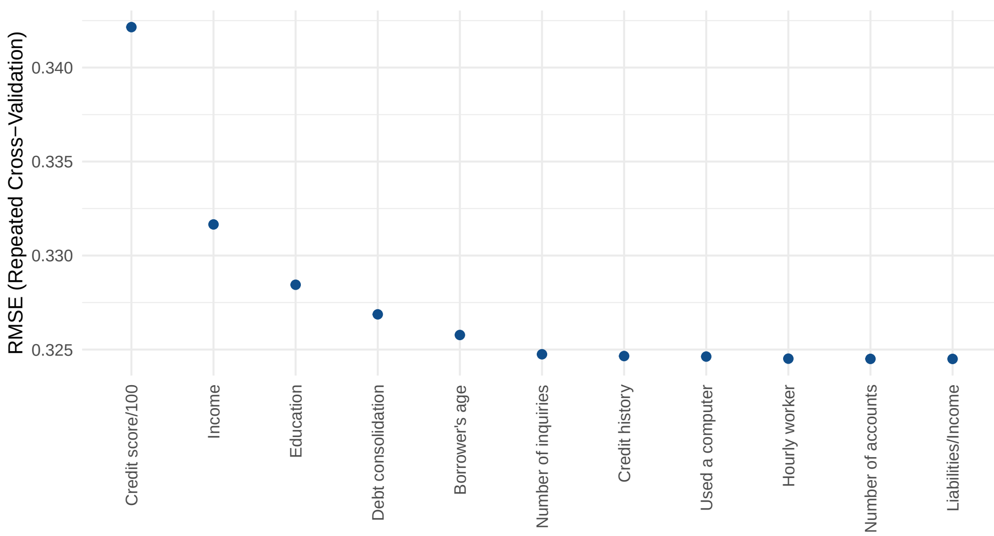


Figure 13. Internal Rate of Return: 3-Year Loans

This figure plots the average internal rate of return for each origination year-Credit score bin for 3-year loans originated by the Platform. Panel A users the full available history to calculate the IRR; Panel B and C limit the history to one and two years after loan origination, respectively. If the loan is current at the end of the period, the outstanding capital amount is used as the terminal cash flow.

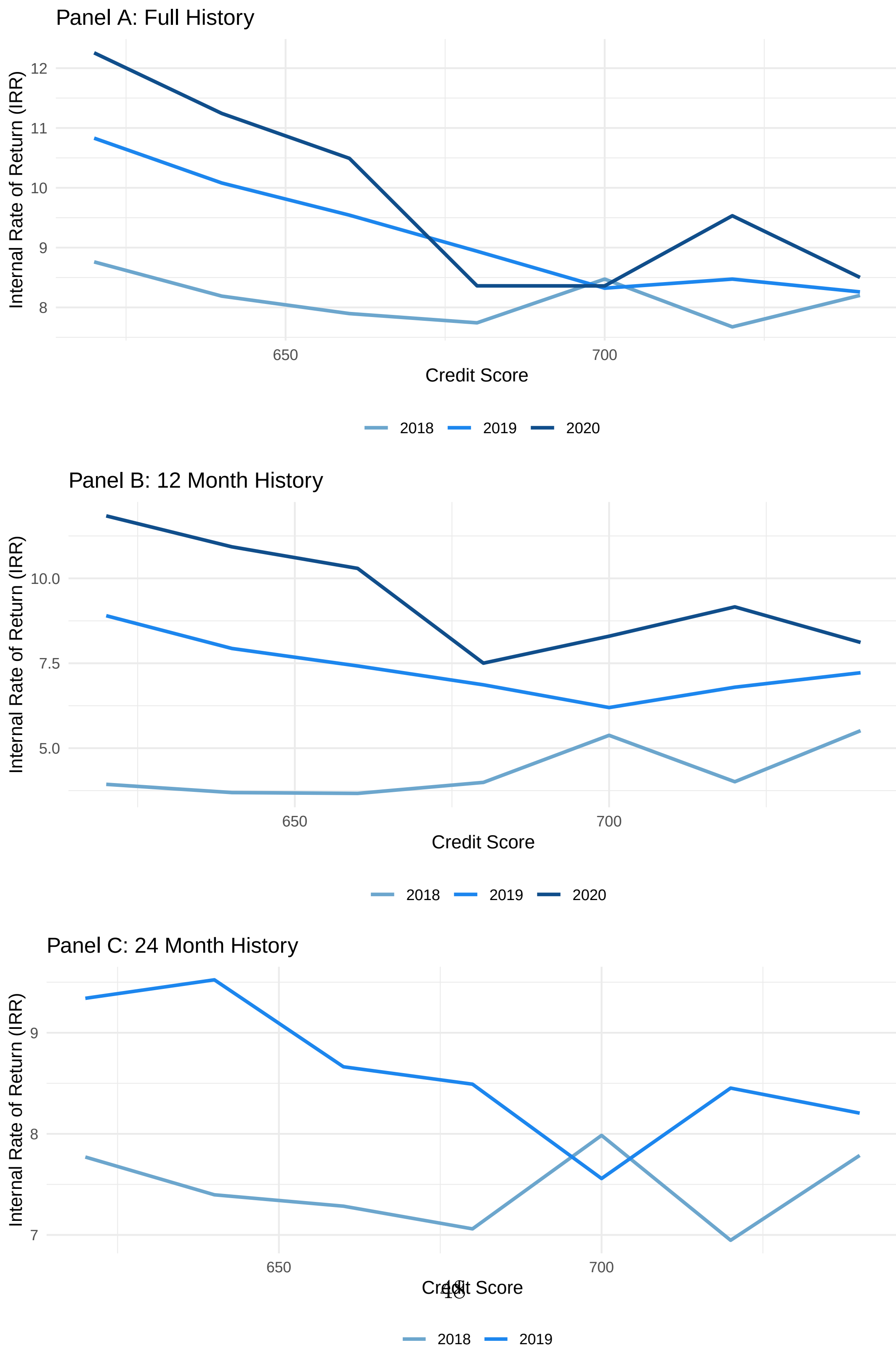


Figure 14. Internal Rate of Return: By Model Outcome

This figure plots the average internal rate of return for each Credit score bin for 3-year loans originated by the Platform separately by the outcome of the underwriting model. The loan history is restricted to 12 months since origination, and if the loan is current at the end of the period, the outstanding capital amount is used as the terminal cash flow.

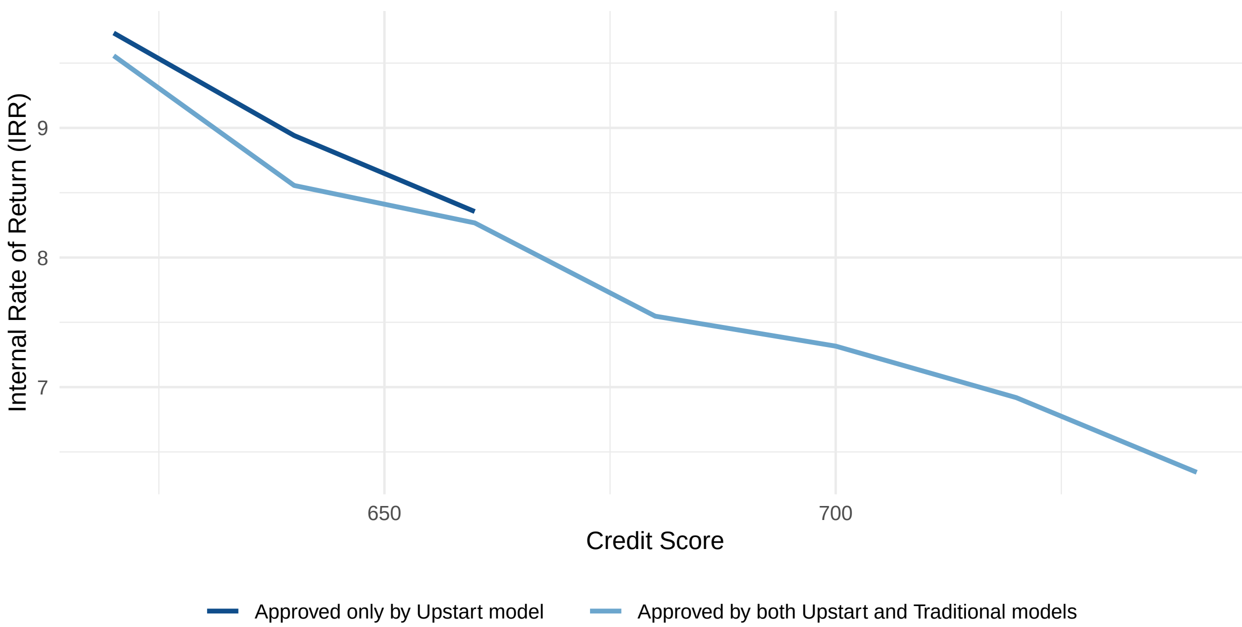
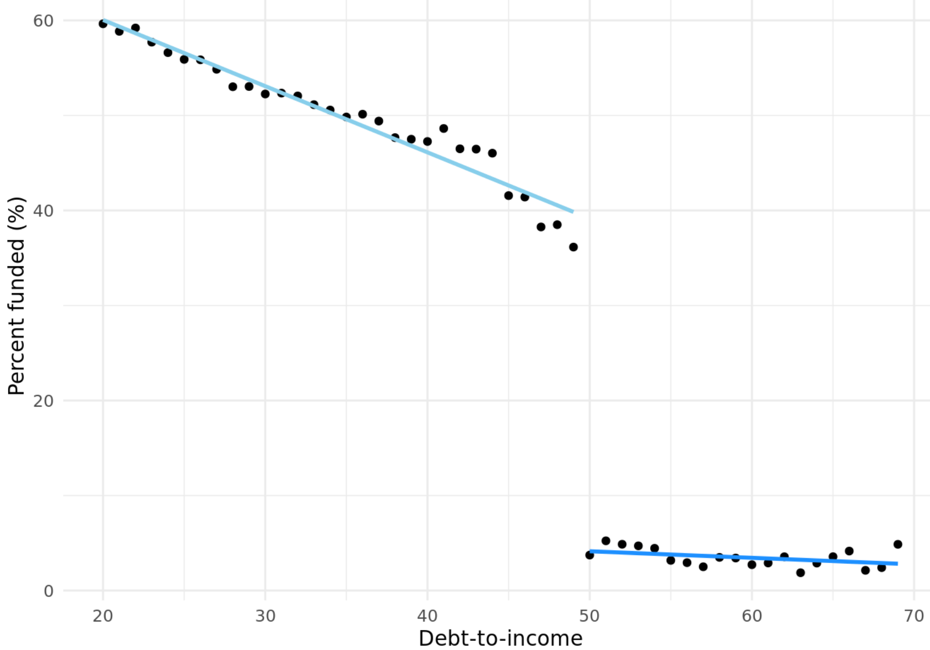


Figure 15. Discontinuity of Probability of Approval

This figure shows the discontinuity of the probability of approval for applicants based on the debt-to-income ratio (Panel A) and months since bankruptcy (Panel B) at the time of the application.

Panel A: Discontinuity at 50% debt-to-income ratio



Panel B: Discontinuity at 36 months since bankruptcy

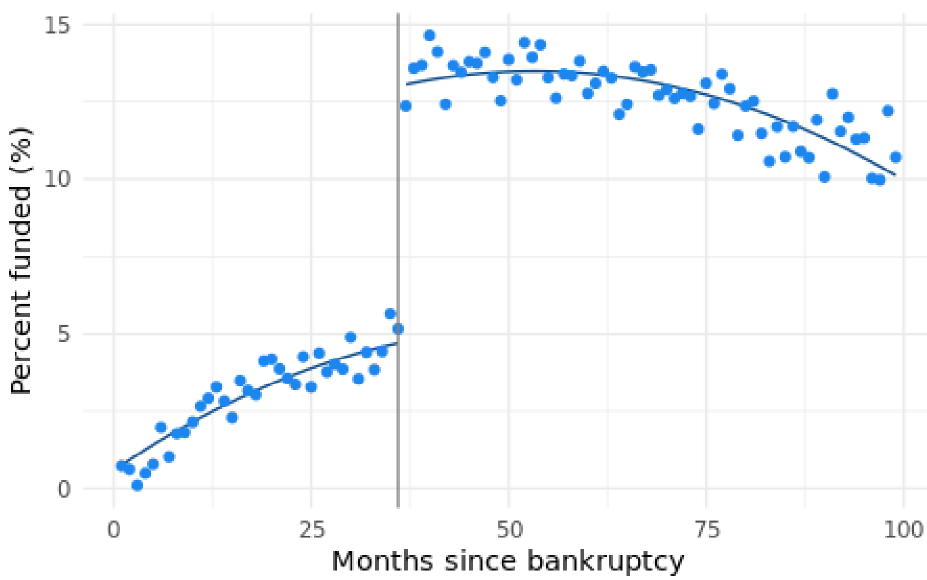
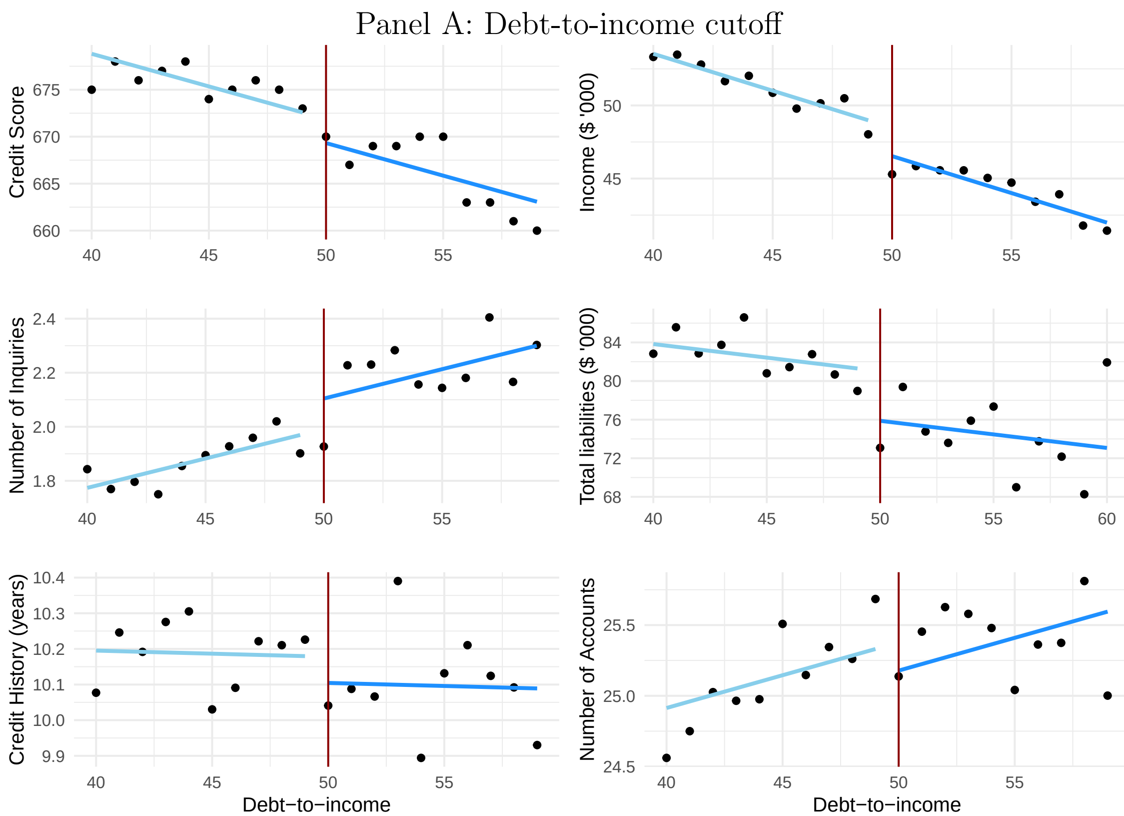


Figure 16. Discontinuity of other Variables

This figure shows the discontinuity of other variables that are commonly used to measure creditworthiness at the 50% debt-to-income cutoff (Panel A) and 36 months since bankruptcy (Panel B).



Panel B: Debt-to-income cutoff

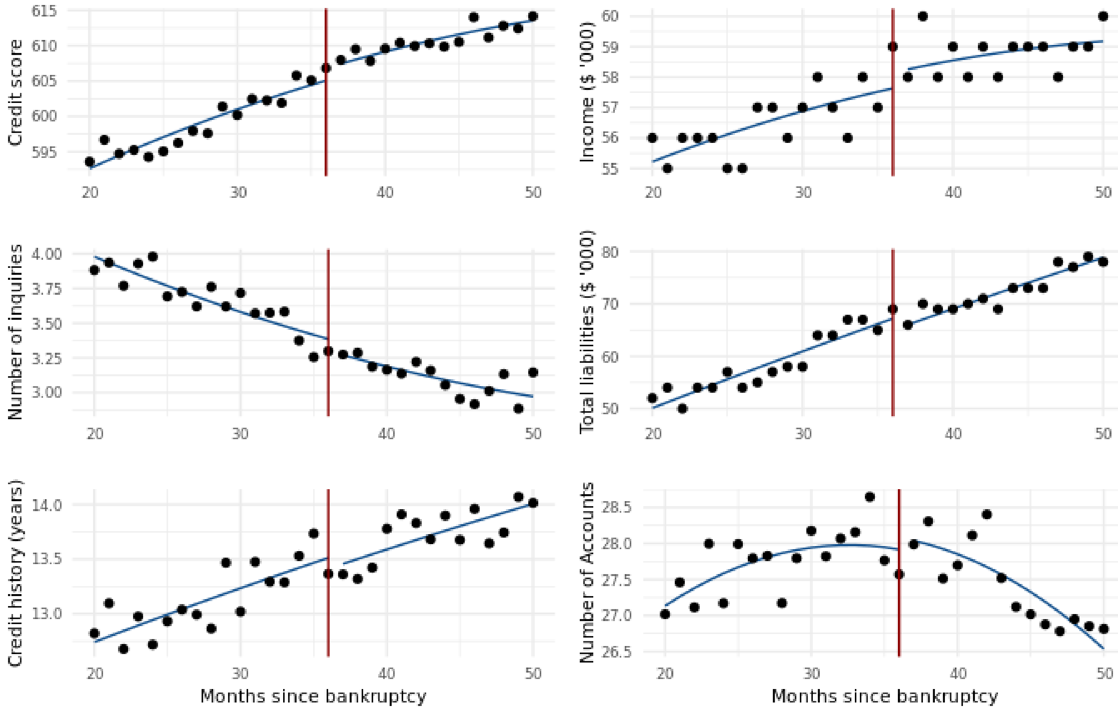
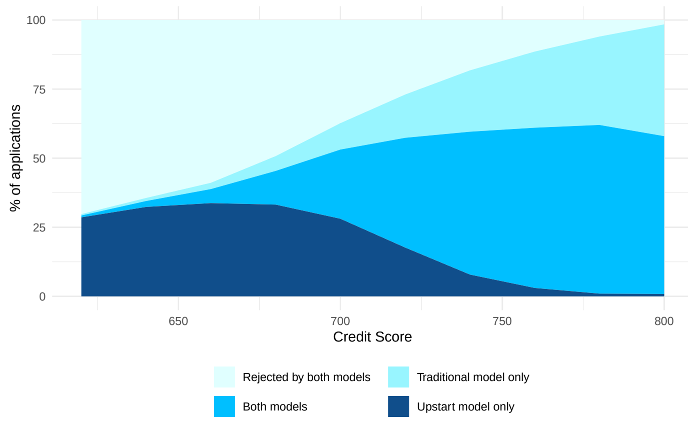


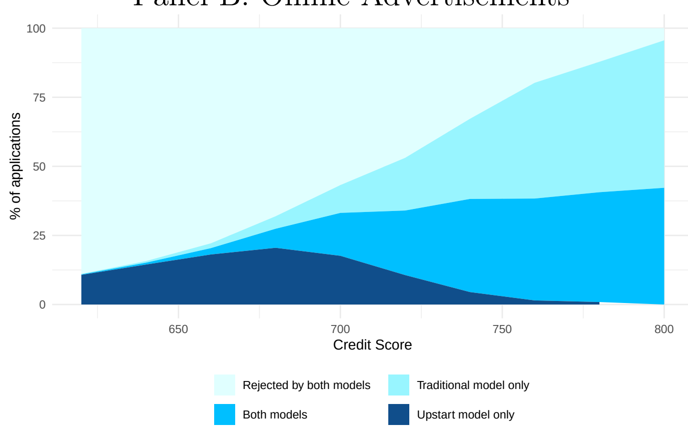
Figure 17. Traditional Model vs. the Platform's Underwriting Model (Robustness)

This figure plots the outcome distribution under each model separately for applicants who came through loan aggregators, online advertisements, and the Platform's website in Panels A, B, and C, respectively. The figures plot the share of each outcome against the credit score bin.

Panel A: Loan Aggregators



Panel B: Online Advertisements



Panel C: Direct Applicants

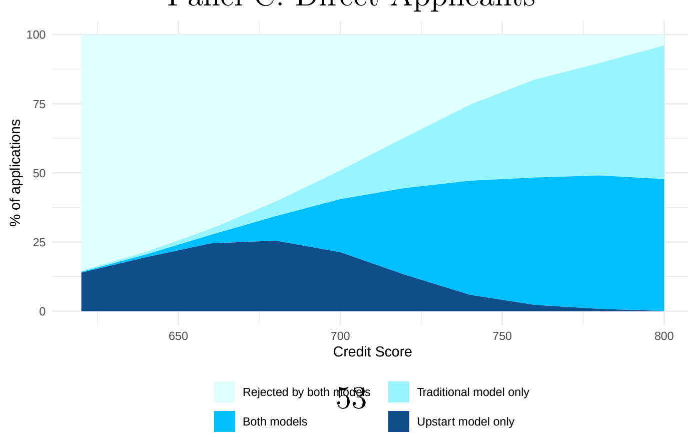


Figure 18. Robustness: Model Outcome Comparison (Funding Decision)

This figure plots the coefficient and corresponding 95% confidence intervals of the regressions that regress the rejection rate based on the traditional model on credit score bin dummy variables with zip code fixed effects. Panel A, B, and C plot use applicants who came thorough loan aggregators, online advertisements, and the Platform website, respectively.

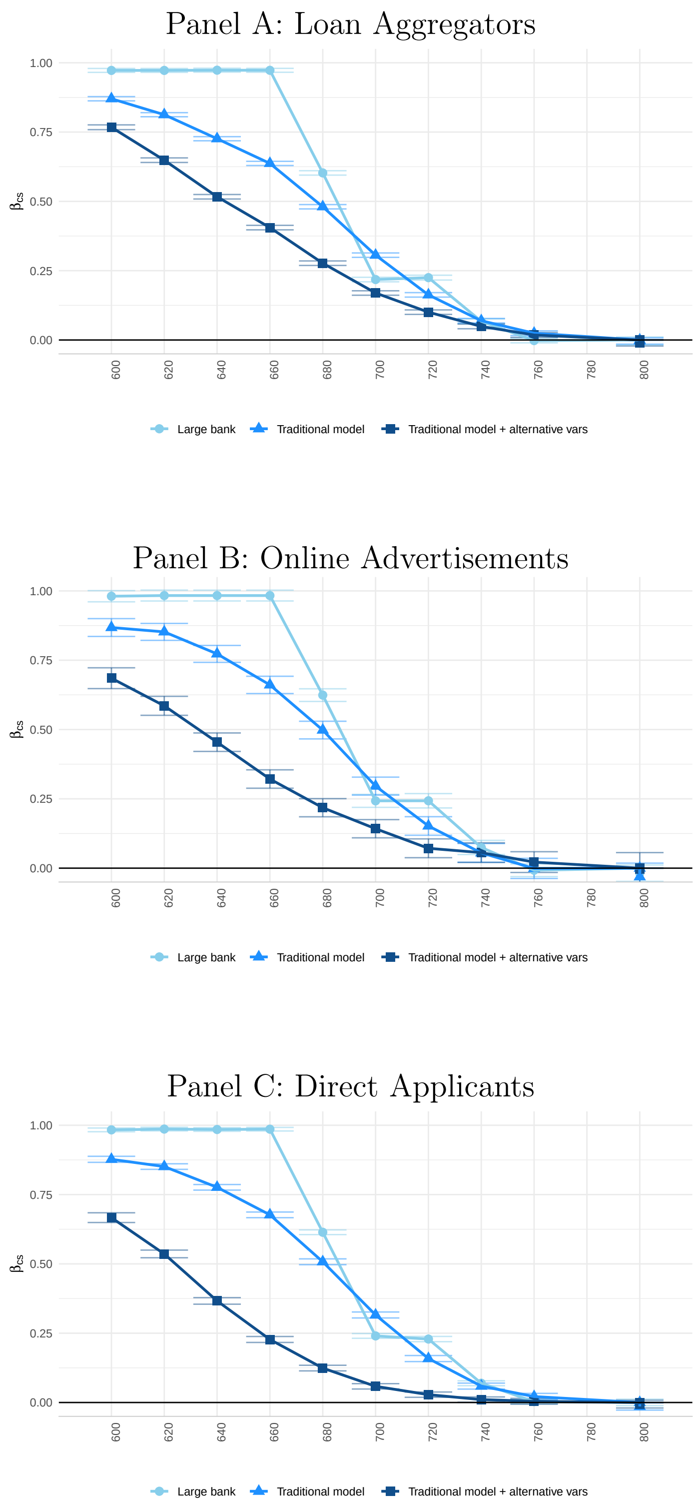
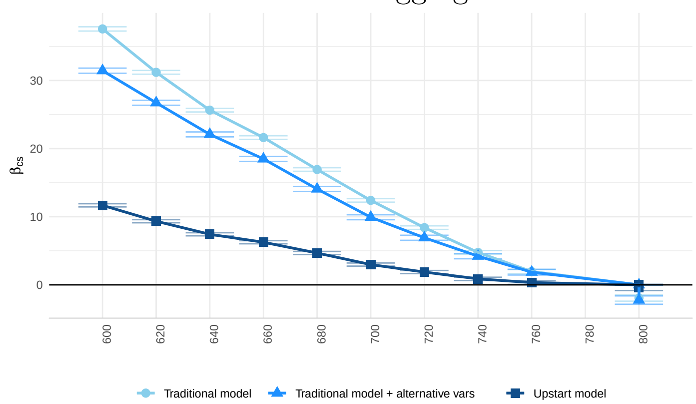


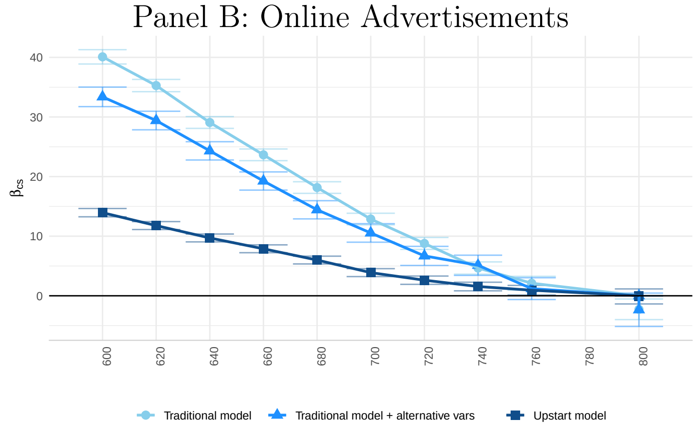
Figure 19. Robustness: Model Outcome Comparison (Interest Rate)

This figure plots the estimation results of regressions that regress the interest rate under each model on credit score bin dummy variables with zip code fixed effects. Panel A, B, and C plot use applicants who came thorough loan aggregators, online advertisements, and the Platform website, respectively.

Panel A: Loan Aggregators



Panel B: Online Advertisements



Panel C: Direct Applicants

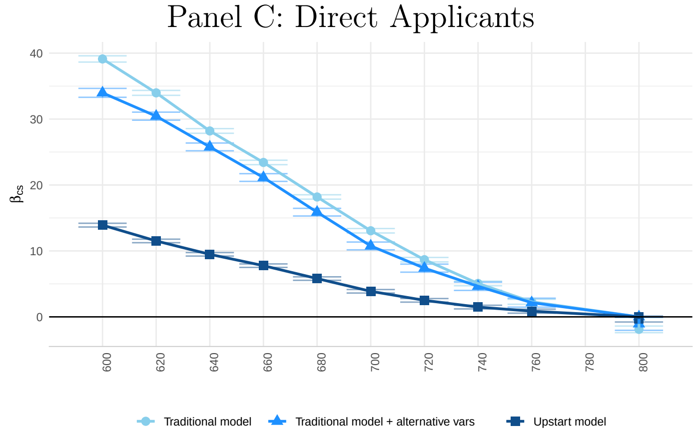
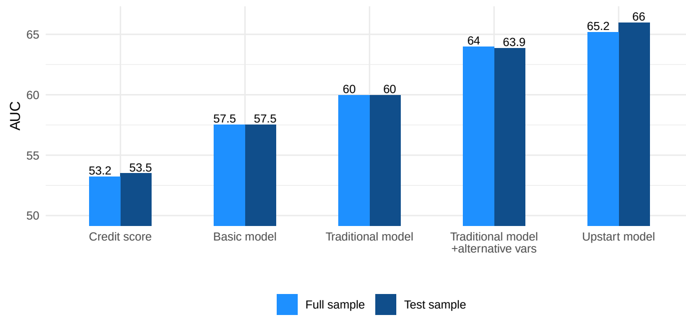


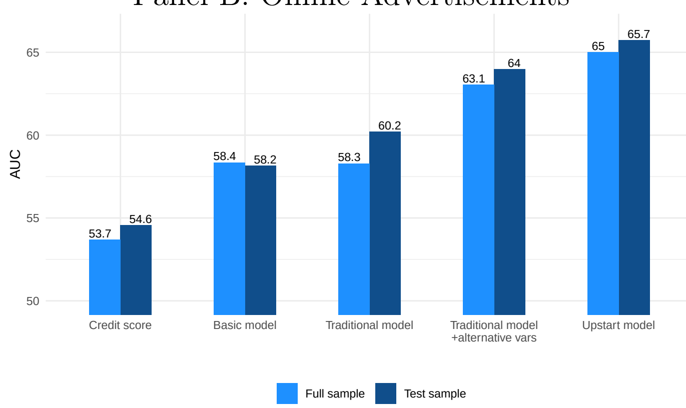
Figure 20. Robustness: Predictive power of different models

This figure plots the area under the curve (AUC) for the funded Platform sample using credit worthiness measures estimated using different models. Panel A, B, and C plot use applicants who came thorough loan aggregators, online advertisements, and the Platform website, respectively.

Panel A: Loan Aggregators



Panel B: Online Advertisements



Panel C: Direct Applicants

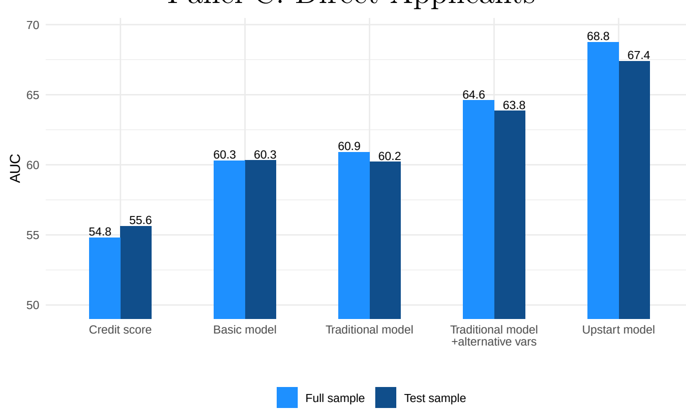
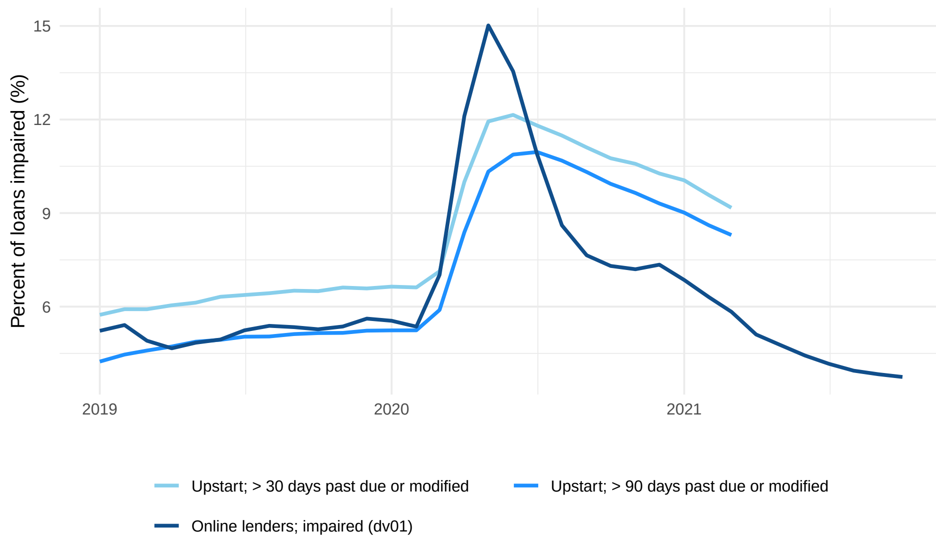


Figure 21. Robustness: Performance during COVID-19

This figure compares the performance of the Platform loans to the performance of loans originated by all online lenders during the initial stages of COVID-19 pandemic. A loan is defined as an impaired loan if it is delinquent or modified. Panel A compares all the Platform loans to all online loans. Panel B compares the Platform loans with credit score less than 660 to borrowers in the bottom 7% of credit score distribution.

Panel A: Full Sample



Panel B: The Platform - Credit score $<$ 660; dv01 - Bottom 7%

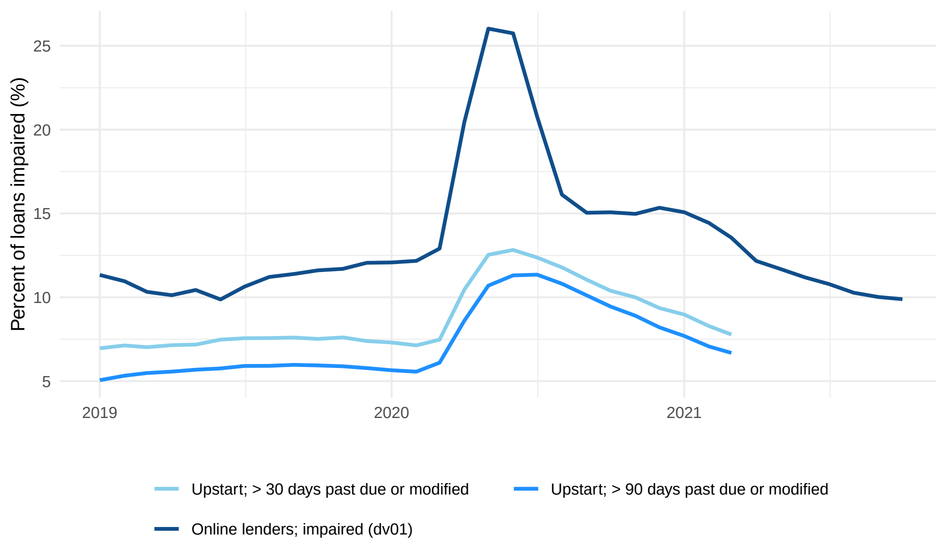


Table 1: **Descriptive Statistics: The Platform Sample**

This table presents the summary statistics of selected loan and borrower characteristics of the loans funded by the Platform (Panel A) and disqualified borrower characteristics (Panel B). Our main measure of credit score is the VantageScore. Ranging from 300 (poor) to 850 (excellent), the VantageScore is provided by VantageScore Solutions LLC, which is jointly owned by TransUnion, Experian, and Equifax, the three major consumer credit reporting companies.

Statistic	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Borrower characteristics						
Credit Score	770,523	653.996	46.904	623	651	683
Age of the borrower	770,511	37.674	12.054	28	35	46
Annual income	770,523	66,958	173,828	39,000	55,000	80,000
Debt-to-income	770,516	18.237	17.820	9.460	16.400	24.980
Number of accounts	770,523	18.624	13.011	9	16	25
Credit history in years	770,523	11.014	6.910	6	10	15
Total credit balance	770,523	120,394	160,851	19,615	51,518	170,554.5
Credit card utilization	770,523	66.940	29.455	47	75	92
Has a mortgage	770,523	0.293	0.455	0	0	1
Inquiries (last 6 months)	770,523	1.037	1.716	0	0	1
College degree	770,523	0.445	0.497	0	0	1
Hourly worker	748,993	0.451	0.498	0	0	1
Years at job	748,996	5.367	7.771	1	3	7
Purpose = consolidation	770,523	0.788	0.409	1	1	1
Used device type = computer	770,523	0.324	0.468	0	0	1
Used a Mac	249,620	0.283	0.450	0	0	1
Used an iPhone	437,949	0.644	0.479	0	1	1
Loan characteristics						
Loan amount	770,523	11,712	10,419	5,000	8,500	15,000
Interest rate	770,523	22.012	6.940	16.320	21.860	27.880
Contract years	770,523	4.250	1.003	3	5	5
Platform default probability	770,523	0.256	0.127	0.156	0.247	0.344
Delinquent (12 months)	770,523	0.083	0.275	0	0	0
Charged-off (12 months)	770,523	0.018	0.133	0	0	0
Charged-off (life time)	770,523	0.041	0.197	0	0	0

Table 2: **Funded vs. Disqualified Applications**

This table compares the key characteristics of funded and disqualified borrowers.

var	Disqualified	Funded	P-Value
Number of Obs	2,374,912	770,523	0
Credit Score	589.058	653.996	0
Age of the borrower	37.106	37.674	0
Annual income	54,258	66,958	0
Debt-to-income	19.861	18.237	0
Number of accounts	17.050	18.624	0
Credit history in years	9.163	11.014	0
Total credit balance	68,263	120,394	0
Credit card utilization	58.366	66.940	0
Has a mortgage	0.151	0.293	0
Inquiries (last 6 months)	2.589	1.037	0
College degree	0.247	0.445	0
Hourly worker	0.562	0.451	0
Years at job	4.362	5.367	0
Purpose = consolidation	0.629	0.788	0
Used device type = computer	0.263	0.324	0
Used a Mac	0.228	0.283	0
Used an iPhone	0.585	0.644	0

Table 3: **Predictability of Credit Score in General**

This table reports the results of the regressions that examine the relationship between credit score and propensity to default. The table uses the sample of applicants rejected by the Platform, and excludes applicants with any delinquent accounts at the time of application. The dependent variable indicates whether applicant i defaulted on at least one credit card within 12 months of the application at time t. Column (1) uses the sub-sample of borrowers with credit scores less than 660, column (2) the sub-sample of borrowers with credit scores greater than or equal to 660. Standard errors are clustered at zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Credit score $<$ 660	Credit score $>=$ 660
	(1)	(2)
Credit score/100	-0.085^{**}	-0.044^{*}
	(0.009)	(0.013)
log(Annual income)	-0.007	0.002
	(0.004)	(0.003)
Debt-to-income	-0.00001^{**}	-0.00000^{*}
	(0.00000)	(0.00000)
Age of the borrower	-0.003^{**}	-0.004^{**}
	(0.0004)	(0.0004)
Age of the borrower^2	0.00003^{**}	0.00003^{***}
	(0.00000)	(0.00000)
log(Number of accounts)	-0.005	-0.007^{**}
	(0.005)	(0.001)
log(Number of inquiries)	0.015^{**}	0.003
	(0.002)	(0.002)
log(Total balance)	-0.002	0.001
	(0.002)	(0.0004)
log(Credit history)	0.006	0.003^{*}
	(0.002)	(0.001)
Zip code $\times$ Year	Y	Y
N	59,538	32,152
Adjusted R^2	0.004	0.004

Table 4: **Predictability of Credit Score: the Platform**

This table reports the results of the regressions that examine the relationship between credit score and propensity to default using the loans originated by the Platform. The dependent variable in columns (1) and (2) is whether loan i was 90 days or more delinquent within 12 months of loan origination. The dependent variable in columns (3) and (4) indicates whether the loan was charged-off by the lender within 12 month of loan origination. Panel B uses default measures using longer time horizons as dependent variables. Other coefficients are suppressed in Panel B to conserve space. Standard errors are clustered at zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: 12 month history

	Dep. Var = Delinquent		Dep. Var = Charged-off	
	Credit score < 660	Credit score >= 660	Credit score < 660	Credit score >= 660
	(1)	(2)	(3)	(4)
Credit score/100	-0.006^{**}	-0.025^{***}	0.003^{**}	-0.003^{***}
	(0.003)	(0.002)	(0.001)	(0.001)
log(Annual income)	-0.003^{**}	-0.021^{***}	-0.002^{***}	-0.006^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Debt-to-income	0.00002	0.0001	0.00001	0.0001^{***}
	(0.00001)	(0.00005)	(0.00001)	(0.00003)
Age of the borrower	-0.003^{***}	-0.001^{***}	-0.001^{***}	-0.0003^{*}
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Age of the borrower2	0.00003^{***}	0.00002^{***}	0.00001^{***}	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
log(Number of accounts)	-0.015^{***}	-0.010^{***}	-0.003^{***}	-0.002^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
log(Number of inquiries)	-0.004^{***}	0.005^{***}	-0.0003	0.003^{***}
	(0.001)	(0.001)	(0.0004)	(0.001)
log(Total liabilities)	0.005^{***}	-0.002^{***}	0.004^{***}	-0.0005
	(0.001)	(0.001)	(0.0004)	(0.0003)
log(Credit history)	-0.002	-0.008^{***}	-0.002^{***}	-0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
log(No of recently opened accounts)	0.007^{***}	0.012^{***}	0.003^{***}	0.004^{***}
	(0.001)	(0.001)	(0.0004)	(0.001)
log(Pct. of revolving liabilities)	0.001^{*}	-0.006^{***}	0.005^{***}	0.0003
	(0.001)	(0.001)	(0.0004)	(0.0003)
log(Pct. of mortgage liabilities)	0.001^{*}	-0.0004	0.0002	-0.0001
	(0.0004)	(0.0004)	(0.0002)	(0.0002)
log(Credit card utilization)	0.005^{***}	0.001^{***}	-0.0001	-0.0003
	(0.001)	(0.0005)	(0.0003)	(0.0002)
log(Pct. trades ever delinquent)	0.002^{***}	0.003^{***}	0.00004	0.0003^{***}
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
log(Loan amount)	0.006^{***}	0.014^{***}	-0.001^{***}	0.003^{***}
	(0.001)	(0.001)	(0.0004)	(0.0005)
Zip code×Year	*Y*	*Y*	*Y*	*Y*
Loan term	*Y*	*Y*	*Y*	*Y*
N	391,682	283,620	391,682	283,620
Adjusted R^2	0.013	0.018	0.011	0.012

Panel B: Credit score/100 coefficient using 24 month and complete history

24 month history	0.004	-0.044^{***}	0.007^{***}	-0.014^{***}
	(0.003)	(0.003)	(0.002)	(0.002)
Complete history	0.005	-0.046^{***}	0.007^{***}	-0.015^{***}
	(0.003)	(0.003)	(0.002)	(0.002)

Table 5: **Predictability of the Platform Default Probability**

This table reports the results of the regressions that examine the relationship between the Platform's estimate of probability of default and propensity to default using the loans originated by the Platform. The dependent variable in columns (1) and (2) is whether loan i was 90 days or more delinquent within 12 months of loan origination. The dependent variable in columns (3) and (4) indicates whether the loan was charged-off by the platform. Panel B uses longer time horizons to calculate dependent variables. Standard errors are clustered at zip code level and reported in parentheses below coefficient estimates. We use *,**, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: 12 month history

	Dep. Var = Delinquent		Dep. Var = Charged-off	
	Credit score < 660	Credit score >= 660	Credit score < 660	Credit score >= 660
	(1)	(2)	(3)	(4)
Platform default probability	0.369***	0.398***	0.114***	0.146***
	(0.005)	(0.007)	(0.003)	(0.004)
Credit score/100	0.014***	0.004**	0.010***	0.009***
	(0.003)	(0.002)	(0.001)	(0.001)
log(Annual income)	0.013***	0.002	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Debt-to-income	−0.0001*	−0.0003***	−0.00002	−0.00004*
	(0.00004)	(0.0001)	(0.00001)	(0.00002)
Age of the borrower	−0.003***	−0.003***	−0.001***	−0.001***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Age of the borrower2	0.00003***	0.00003***	0.00001***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
log(Number of accounts)	−0.008***	−0.002	−0.002***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
log(Number of inquiries)	0.0002	−0.002*	0.001**	0.00004
	(0.001)	(0.001)	(0.0004)	(0.001)
log(Total liabilities)	0.003***	0.002***	0.003***	0.001***
	(0.001)	(0.001)	(0.0004)	(0.0003)
log(Credit history)	0.006***	0.001	0.001	−0.0004
	(0.001)	(0.001)	(0.001)	(0.001)
log(No of recently opened accounts)	−0.003***	−0.004***	−0.001	−0.002***
	(0.001)	(0.001)	(0.0004)	(0.001)
log(Pct. of revolving liabilities)	−0.001*	−0.001**	0.004***	0.002***
	(0.001)	(0.001)	(0.0004)	(0.0003)
log(Pct. of mortgage liabilities)	0.001***	0.001***	0.0004***	0.001***
	(0.0003)	(0.0004)	(0.0002)	(0.0002)
log(Credit card utilization)	−0.00003	−0.001	−0.002***	−0.001***
	(0.001)	(0.0005)	(0.0003)	(0.0002)
log(Pct. trades ever delinquent)	0.001***	0.001***	−0.0002***	−0.0002***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
log(Loan amount)	0.003***	0.006***	−0.002***	0.0005
	(0.001)	(0.001)	(0.0004)	(0.0005)
Zip code×Year	*Y*	*Y*	*Y*	*Y*
Loan term	*Y*	*Y*	*Y*	*Y*
N	391,682	283,620	391,682	283,620
Adjusted R^2	0.051	0.047	0.017	0.022

Panel B: Platform default probability coefficient using 24 month and complete history

24 month history	0.439***	0.500***	0.188***	0.243***
	(0.006)	(0.008)	(0.004)	(0.005)
Complete history	0.437***	0.503***	0.190***	0.248***
	(0.006)	(0.008)	(0.004)	(0.005)

Table 6: **Predictive Power of Credit Score and the Platform Default Probability**

This table compares predictability of credit score and the Platform's estimate of the probability of default using the Area Under the Curve (AUC). Panel A uses borrowers who had a credit score less than 660 at origination, Panel B borrowers who had a credit score greater than or equal to 660 at origination.

	AUC (Credit Score)	AUC (the Platform default probability)	AUC Diff
Panel A: Credit Score <660			
Full Sample	50.97	65.13	14.16
Age <30	50.15	65.64	15.49
Age >= 30	51.17	65.00	13.83
No college education	53.02	63.34	10.32
College educated	51.51	67.11	15.60
Credit history <3	53.31	61.42	8.11
Credit history 3-7	50.62	64.93	14.31
Credit history >7	50.89	65.51	14.62
Income <55k	52.42	62.84	10.42
Income >= 55k	50.48	66.84	16.36
Panel B: Credit Score >= 660			
Full Sample	51.23	72.00	20.77
Age <30	51.75	72.29	20.54
Age >= 30	50.97	71.83	20.86
College educated	51.93	74.37	22.44
No college education	50.91	69.58	18.67
Credit history <3	49.75	68.12	18.37
Credit history 3-7	52.92	70.63	17.71
Credit history >7	51.60	72.96	21.36
Income <55k	51.25	70.15	18.90
Income >= 55k	51.95	73.71	21.76

Table 7: **Key Drivers of the Platform Default Probability**

This table reports the results of the regressions that examine the determinants of the Platform default probability. The dependent variable is the Platform default probability. Standard errors are clustered at state level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Credit score/100	-0.112^{***}	-0.111^{***}	-0.113^{***}	-0.115^{***}	-0.111^{***}	-0.115^{***}
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
log(Annual income)	-0.045^{***}	-0.042^{***}	-0.038^{***}	-0.047^{***}	-0.042^{***}	-0.038^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Debt-to-income	0.001^{***}	0.0005^{***}	0.0005^{***}	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Age of the borrower	0.004^{***}	0.003^{***}	0.005^{***}	0.004^{***}	0.004^{***}	0.004^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Age of the borrower^2	-0.00003^{***}	-0.00003^{***}	-0.00004^{***}	-0.00003^{***}	-0.00003^{***}	-0.00003^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
log(Number of accounts)	0.004^{***}	0.009^{***}	0.005^{***}	0.005^{***}	0.003^{***}	0.010^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(Number of inquiries)	0.035^{***}	0.033^{***}	0.035^{***}	0.033^{***}	0.034^{***}	0.032^{***}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
log(Total balance)	-0.008^{***}	-0.006^{***}	-0.008^{***}	-0.007^{***}	-0.008^{***}	-0.006^{***}
	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0004)
log(Credit history)	-0.027^{***}	-0.025^{***}	-0.028^{***}	-0.027^{***}	-0.026^{***}	-0.026^{***}
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0003)
log(Loan amount)	0.00004	0.002^{***}	0.0002	0.003^{***}	0.0002	0.004^{***}
	(0.001)	(0.0005)	(0.001)	(0.0005)	(0.0005)	(0.0005)
Zip code×Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
Loan Term × Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
Educational attainment	*N*	*Y*	*N*	*N*	*N*	*Y*
Employment type	*N*	*N*	*Y*	*N*	*N*	*Y*
Loan purpose	*N*	*N*	*N*	*Y*	*N*	*Y*
Device/Technology	*N*	*N*	*N*	*N*	*Y*	*Y*
N	770,299	770,299	748,796	770,299	687,370	667,777
R^2	0.431	0.451	0.435	0.439	0.439	0.463
Maximum economic impact		0.042	0.028	0.028	0.047	0.056

Table 8: **Traditional Model vs. the Platform default probability: Who benefits?** This table reports the results of the regression that aimed at understanding who benefits most by the platform's use of alternative data. Panel A studies the differences in borrower characteristics and Panel B studies the differences in regional level differences. Columns (1) and (2) in Panel A and columns (1) through (3) in Panel B regress the dummy variable 'rejected by traditional model' on borrower/regional characteristics. Columns (3) and (4) in Panel A and columns (5) through (7) in Panel B regress the percentage difference between the APR suggested by the traditional model and the actual APR on borrower/regional characteristics. Standard errors are clustered at zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Borrower Differences

	Dep. var = Rejected by Traditional Model		Dep. var = Traditional APR - Platform APR	
	(1)	(2)	(3)	(4)
Credit score/100	-0.660^{***}		-15.229^{***}	
	(0.001)		(0.033)	
Credit score < 660		0.332^{***}		9.067^{***}
		(0.003)		(0.068)
Income < 55k	-0.108^{***}	-0.128^{***}	-4.996^{***}	-4.859^{***}
	(0.001)	(0.002)	(0.028)	(0.033)
Advanced degree	0.067^{***}	0.046^{***}	4.153^{***}	3.615^{***}
	(0.002)	(0.002)	(0.041)	(0.051)
College degree	0.036^{***}	0.016^{***}	2.762^{***}	2.521^{***}
	(0.001)	(0.002)	(0.029)	(0.035)
Salaried employee	0.027^{***}	0.032^{***}	1.648^{***}	1.506^{***}
	(0.001)	(0.002)	(0.027)	(0.033)
Thin credit file	0.032^{***}	0.005^{***}	-0.578^{***}	-0.774^{***}
	(0.001)	(0.002)	(0.027)	(0.033)
Credit score < 660 × Income < 55k		0.048^{***}		-0.615^{***}
		(0.003)		(0.064)
Credit score < 660 × Advanced degree		0.024^{***}		1.085^{***}
		(0.003)		(0.095)
Credit score < 660 × College degree		0.024^{***}		0.100
		(0.003)		(0.067)
Credit score < 660 × Salaried employee		-0.013^{***}		0.476^{***}
		(0.002)		(0.064)
Credit score < 660 × Thin credit file		0.047^{***}		-0.285^{***}
		(0.003)		(0.064)
Zip code	*Y*	*Y*	*Y*	*Y*
N	717,524	717,524	717,524	717,524
Adjusted R^2	0.267	0.149	0.335	0.243

Panel B: Regional Differences

	Dep.var = Rejected by trad. model			Dep. var = Trad. APR - Platform APR		
	(1)	(2)	(3)	(4)	(5)	(6)
Minority fraction in middle third	0.014***			0.569***		
	(0.003)			(0.065)		
Minority fraction in top third	0.023***			0.823***		
	(0.004)			(0.125)		
Fraction of renters in middle third		0.007**			0.228***	
		(0.003)			(0.076)	
Fraction of renters in top third		0.014***			0.610***	
		(0.003)			(0.104)	
Foreign born fraction in middle third			0.012***			0.671***
			(0.003)			(0.088)
Foreign born fraction in top third			0.022***			0.960***
			(0.004)			(0.109)
Credit score	−0.006***	−0.006***	−0.006***	−0.141***	−0.141***	−0.141***
	(0.00004)	(0.00004)	(0.00004)	(0.002)	(0.002)	(0.002)
log(Annual income)	0.137***	0.138***	0.137***	7.248***	7.271***	7.243***
	(0.003)	(0.003)	(0.003)	(0.060)	(0.059)	(0.061)
Debt-to-income	0.002***	0.002***	0.002***	0.029**	0.029**	0.029**
	(0.001)	(0.001)	(0.001)	(0.011)	(0.011)	(0.011)
State	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	736,896	736,896	736,902	736,896	736,896	736,902
Adjusted R^2	0.283	0.283	0.283	0.368	0.368	0.368

Table 9: **Characteristics of Winners and Losers**

This table is aimed at understanding the differences of losers of the Platform's underwriting model compared to winners. Each column regresses key alternative variables on the dummy variable *Loser* which takes the value of one if the application is approved only by the traditional model. The sample consists of applicants approved only by the Platform's model and the applicants approved only by the traditional model. The dependent variables are given in the first row. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	College	Hourly worker	Used a computer	Debt consolidating	Age	log(Credit History)
	(1)	(2)	(3)	(4)	(5)	(6)
Loser	−0.154***	−0.004	−0.012***	−0.137***	5.761***	0.096***
	(0.004)	(0.004)	(0.002)	(0.003)	(0.035)	(0.005)
Credit score/100	−0.022	0.013	0.001	0.041*	2.703**	0.344***
	(0.024)	(0.024)	(0.021)	(0.022)	(0.895)	(0.023)
log(Annual income)	0.095***	−0.227***	0.015***	−0.022***	3.937***	0.120***
	(0.002)	(0.008)	(0.001)	(0.004)	(0.133)	(0.006)
Age of the borrower	−0.015***	0.0002	−0.009***	0.003***		
	(0.001)	(0.001)	(0.0004)	(0.0004)		
Age of the borrower2	0.0001***	−0.00005***	0.0001***	−0.00003***		
	(0.00001)	(0.00001)	(0.00000)	(0.00000)		
log(Number of accounts)	0.110***	−0.036***	0.008***	0.104***	4.382***	0.527***
	(0.003)	(0.002)	(0.001)	(0.004)	(0.102)	(0.009)
log(Number of inquiries)	−0.013***	−0.006***	−0.010***	−0.040***	−0.245***	−0.031***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.042)	(0.002)
log(Total liabilities)	0.042***	−0.016***	0.008***	0.008***	−0.383***	0.044***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.041)	(0.005)
Credit history	0.005***	−0.001***	0.001***	−0.001		
	(0.0003)	(0.0002)	(0.0001)	(0.0004)		
log(No of recently opened accounts)	−0.038***	0.023***	−0.012***	−0.016***	−0.395***	−0.155***
	(0.003)	(0.002)	(0.001)	(0.002)	(0.064)	(0.008)
log(Pct. of revolving liabilities)	0.008**	−0.009***	0.008***	0.072***	0.212**	0.006
	(0.003)	(0.002)	(0.001)	(0.002)	(0.092)	(0.008)
Pct. of mortgage liabilities	−0.007	0.002*	−0.001	−0.002	0.318**	0.006***
	(0.004)	(0.001)	(0.001)	(0.001)	(0.133)	(0.002)
Credit card utilization	−0.0003***	0.0002***	−0.0002***	−0.0001	0.009	0.002***
	(0.0001)	(0.0001)	(0.00005)	(0.0002)	(0.005)	(0.0003)
log(Pct. trades ever delinquent)	−0.024***	0.005***	−0.004***	−0.011***	1.404***	0.144***
	(0.003)	(0.001)	(0.0004)	(0.001)	(0.027)	(0.004)
Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
5 point credit score	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	396,832	396,832	396,832	396,832	396,832	396,862
Adjusted R^2	0.199	0.190	0.019	0.095	0.243	0.599

Table 10: **Winners vs. Losers**

This table is aimed at understanding the losers of the Platform's underwriting model. Each regression compares the subsequent financial health of the Platform funded applicants to the Platform rejected applicants who were able to secure personal loans from other lenders soon after the Platform's rejection. The sample used in columns (1) through (3) consists of borrowers with credit scores less than 660, the sample used in columns (4) through (6) of borrowers with credit scores greater than 660. The dependent variables are given in the second row. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Credit score $<$ 660			Credit score $>=$ 660		
	Credit card delinq	Credit score change	Mortgage	Credit card delinq	Credit score change	Mortgage
	(1)	(2)	(3)	(4)	(5)	(6)
Platform funded	-0.086^{***}	0.042^{***}	0.002	-0.096^{***}	0.046^{***}	0.003^{*}
	(0.005)	(0.001)	(0.003)	(0.002)	(0.003)	(0.002)
Credit score/100	-0.237^{**}	-0.087^{***}	-0.005	-0.105^{**}	-0.061^{***}	0.017
	(0.070)	(0.013)	(0.026)	(0.037)	(0.008)	(0.032)
log(Annual income)	0.006	0.009^{***}	0.036^{***}	-0.015^{***}	0.011^{***}	0.043^{***}
	(0.003)	(0.0004)	(0.002)	(0.002)	(0.0005)	(0.002)
Age of the borrower	-0.001	-0.001^{***}	0.003^{***}	-0.001^{*}	-0.0001	0.002^{***}
	(0.001)	(0.0001)	(0.0001)	(0.0004)	(0.0002)	(0.0003)
Age of the borrower^2	0.00001	0.00001^{**}	-0.00003^{***}	0.00001^{*}	0.00000	-0.00002^{***}
	(0.00001)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
log(Number of accounts)	0.042^{***}	0.004^{***}	0.015^{***}	0.016^{***}	0.006^{***}	0.021^{***}
	(0.008)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
log(Number of inquiries)	0.022^{***}	-0.010^{***}	0.028^{***}	0.021^{***}	-0.016^{***}	0.037^{***}
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
log(Total liabilities)	-0.015^{***}	0.001^{**}	-0.003	-0.009^{***}	0.002^{***}	-0.004^{***}
	(0.001)	(0.0003)	(0.002)	(0.001)	(0.0002)	(0.0003)
Credit history	0.001^{***}	0.001^{***}	0.0005	0.001^{**}	0.001^{***}	-0.0002
	(0.0003)	(0.0001)	(0.0003)	(0.0002)	(0.0001)	(0.0002)
log(No of recently opened accounts)	0.028^{***}	-0.015^{***}	-0.016^{***}	0.030^{***}	-0.018^{***}	-0.012^{***}
	(0.002)	(0.001)	(0.002)	(0.004)	(0.001)	(0.001)
log(Pct. of revolving liabilities)	-0.004	0.0002	-0.004	-0.003	0.0004	-0.008^{***}
	(0.005)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Pct. of mortgage liabilities	0.015^{**}	0.003^{***}		0.0004	0.001	
	(0.006)	(0.0003)	(0.000)	(0.001)	(0.001)	(0.000)
Credit card utilization	-0.0001	-0.0002^{***}	-0.0001^{*}	0.0003^{***}	-0.0002^{***}	0.0001^{**}
	(0.0001)	(0.00002)	(0.0001)	(0.00004)	(0.00001)	(0.00004)
log(Pct. trades ever delinquent)	0.064^{***}	-0.002^{***}	0.002^{***}	0.048^{***}	-0.004^{***}	0.004^{***}
	(0.002)	(0.0003)	(0.0004)	(0.003)	(0.0002)	(0.0004)
Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
5 point credit score	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	82,569	83,456	65,132	174,274	174,796	121,618
Adjusted R^2	0.087	0.098	0.022	0.075	0.089	0.023

Table 11: **The Effects of Credit Access: Entropy Balancing**

This table compares the financial outcomes of funded and disqualified applicants. Disqualified applicants are matched with funded applicants using entropy balancing. The table reports the estimates of regressions aimed at understanding the effect of credit access on financial outcomes. The sample used in columns (1) through (3) consists of borrowers with credit scores less than 660, the sample used in columns (4) through (6) of borrowers with credit scores greater than or equal to 660. The dependent variables are given in the second row. Standard errors are clustered at state level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Credit score < 660			Credit score >= 660		
	Credit card delinq	Credit score change	Mortgage	Credit card delinq	Credit score change	Mortgage
	(1)	(2)	(3)	(4)	(5)	(6)
Funded	−0.012***	0.010***	0.007***	−0.006***	−0.006***	−0.001
	(0.001)	(0.0004)	(0.001)	(0.002)	(0.0005)	(0.001)
Credit score/100	−0.002***	−0.001***	0.0002***	0.0001***	−0.001***	0.0002***
	(0.00002)	(0.00001)	(0.00001)	(0.00002)	(0.00001)	(0.00002)
log(Annual income)	−0.018***	0.005***	0.024***	0.019***	0.004***	0.033***
	(0.001)	(0.0003)	(0.001)	(0.002)	(0.0004)	(0.001)
Debt-to-income	−0.00001**	−0.00000	0.00000	0.00000	−0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Age of the borrower	−0.005***	−0.001***	0.001***	−0.003***	0.001***	0.002***
	(0.0003)	(0.0001)	(0.0002)	(0.0004)	(0.0001)	(0.0003)
Age of the borrower2	0.0001***	0.00001***	−0.00001***	0.00002***	−0.00001***	−0.00002***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
log(Number of accounts)	0.018***	−0.0003	0.010***	0.031***	0.002***	0.001
	(0.002)	(0.0005)	(0.001)	(0.002)	(0.0005)	(0.001)
log(Number of inquiries)	0.025***	−0.022***	0.003***	0.0004	−0.018***	0.023***
	(0.001)	(0.0002)	(0.001)	(0.001)	(0.0003)	(0.001)
log(Total liabilities)	0.014***	−0.007***	−0.012***	−0.001	−0.004***	−0.016***
	(0.001)	(0.0002)	(0.0004)	(0.001)	(0.0002)	(0.001)
log(Credit history)	−0.066***	0.019***	0.006***	−0.015***	0.020***	0.013***
	(0.001)	(0.0004)	(0.001)	(0.002)	(0.0004)	(0.001)
log(No of recently opened accounts)	−0.056***	−0.007***	0.011***	−0.001	−0.021***	0.037***
	(0.001)	(0.0003)	(0.001)	(0.001)	(0.0003)	(0.001)
log(Pct. of revolving liabilities)	0.036***	−0.018***	−0.031***	0.027***	−0.021***	−0.074***
	(0.001)	(0.0002)	(0.0004)	(0.001)	(0.0003)	(0.001)
log(Pct. of mortgage liabilities)	0.005***	0.0001		0.007***	−0.002***	
	(0.0004)	(0.0001)		(0.001)	(0.0001)	
log(Credit card utilization)	0.00004	−0.022***	0.003***	−0.004***	−0.015***	0.024***
	(0.0005)	(0.0002)	(0.0002)	(0.001)	(0.0001)	(0.001)
log(Pct. trades ever delinquent)	0.055***	−0.011***	−0.002***	0.080***	−0.008***	−0.001***
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Zip code × Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	458,213	458,212	351,782	258,543	258,542	184,157
Adjusted R^2	0.323	0.313	0.105	0.422	0.352	0.182

Table 12: **The Effects of Credit Access: Debt-to-income ratio Discontinuity**

This table reports the estimation results of the second stage of the fuzzy regression discontinuity specification 2. The sample is restricted to applicants with debt-to-income ratios between 40% and 60%. The sample used in columns (1) through (3) includes borrowers with credit scores less than 660, the sample used in columns (4) through (6) borrowers with credit scores greater than 660. The dependent variables are given in the second row. Standard errors are clustered at the zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Credit score < 660			Credit score >= 660		
	Credit card delinq	Credit score change	Mortgage	Credit card delinq	Credit score change	Mortgage
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{Funded}$	-0.198^{*}	0.087^{***}	0.134^{*}	-0.074	0.012	0.010
	(0.102)	(0.026)	(0.072)	(0.074)	(0.015)	(0.087)
Debt-to-income	-0.003^{*}	0.001^{***}	0.002^{*}	-0.004	0.001^{*}	0.003
	(0.002)	(0.0004)	(0.001)	(0.003)	(0.001)	(0.003)
Credit score	-1.949	-6.627^{***}	1.039	3.851^{***}	-3.843^{***}	2.323^{**}
	(1.574)	(0.382)	(1.155)	(1.135)	(0.209)	(1.113)
Credit score2	9.455^{***}	0.323	-0.822	-0.283	-0.672^{***}	0.854
	(0.812)	(0.206)	(0.527)	(0.699)	(0.137)	(0.628)
log(Annual income)	-0.030^{***}	-0.003	0.031^{***}	-0.035^{**}	0.008^{**}	0.025
	(0.010)	(0.002)	(0.006)	(0.017)	(0.003)	(0.017)
Age of the borrower	-0.004^{**}	-0.00003	0.002^{*}	-0.001	-0.001	0.002
	(0.002)	(0.0005)	(0.001)	(0.003)	(0.001)	(0.003)
Age of the borrower^2	0.00004^{**}	0.00000	-0.00002	-0.00001	0.00000	-0.00003
	(0.00002)	(0.00000)	(0.00001)	(0.00003)	(0.00001)	(0.00003)
log(Number of accounts)	0.048^{***}	-0.006^{***}	0.022^{***}	0.023	-0.004	0.021
	(0.009)	(0.002)	(0.006)	(0.018)	(0.004)	(0.015)
log(Number of inquiries)	0.008	-0.010^{***}	0.007^{**}	-0.005	-0.010^{***}	0.027^{***}
	(0.006)	(0.002)	(0.003)	(0.011)	(0.002)	(0.010)
Total liabilities	7.150^{***}	-1.333^{***}	1.326	3.817^{***}	-0.719^{***}	-4.806
	(0.973)	(0.239)	(2.720)	(1.046)	(0.216)	(3.207)
Credit history	-5.817^{***}	2.187^{***}	-0.203	-2.042^{*}	1.577^{***}	2.205^{*}
	(0.936)	(0.231)	(0.604)	(1.124)	(0.227)	(1.173)
log(No of recently opened accounts)	-0.067^{***}	-0.010^{***}	0.004	-0.018	-0.027^{***}	0.020^{*}
	(0.008)	(0.002)	(0.005)	(0.015)	(0.003)	(0.011)
log(Pct. of revolving liabilities)	0.055^{***}	-0.018^{***}	-0.053^{***}	0.044^{***}	-0.022^{***}	-0.082^{***}
	(0.005)	(0.001)	(0.004)	(0.009)	(0.002)	(0.010)
log(Pct. of mortgage liabilities)	0.020^{***}	-0.004^{***}		0.008	-0.004^{**}	
	(0.005)	(0.001)		(0.010)	(0.002)	
Credit card utilization	9.374^{***}	-8.078^{***}	1.465^{***}	0.016	-5.398^{***}	3.480^{***}
	(0.726)	(0.182)	(0.435)	(0.876)	(0.182)	(0.833)
log(Pct. trades ever delinquent)	0.055^{***}	-0.010^{***}	-0.003^{***}	0.068^{***}	-0.009^{***}	-0.002
	(0.001)	(0.0003)	(0.001)	(0.002)	(0.0004)	(0.002)
Zip code × Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	29,692	29,692	21,183	13,171	13,171	7,890
Adjusted R^2	0.320	0.304	0.039	0.317	0.498	0.231

Table 13: **The Effects of Credit Access: Months Since Bankruptcy Discontinuity**
This table reports the estimation results of the second stage of the fuzzy regression discontinuity specification 4. The sample is restricted to applicants filed for bankruptcy in the last 10 years since application date. The sample used in columns (1) through (3) includes borrowers with credit scores less than 660, the sample used in columns (4) through (6) borrowers with credit scores greater than 660. The dependent variables are given in the second row. Standard errors are clustered at the zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Credit score < 660			Credit score >= 660		
	Credit card delinq	Credit score change	Mortgage	Credit card delinq	Credit score change	Mortgage
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{Funded}$	-0.264^{***}	0.017	0.086^{*}	-0.001	-0.112^{**}	0.108
	(0.075)	(0.022)	(0.045)	(0.017)	(0.044)	(0.098)
Months since bankruptcy-36	0.602^{***}	0.784^{***}	2.731^{***}	0.065	0.416^{***}	0.630
	(0.110)	(0.146)	(0.443)	(0.050)	(0.130)	(0.449)
(Months since bankruptcy-36)^2	-0.484^{***}	0.288	-3.399^{***}	-0.011	-0.749^{***}	-0.085
	(0.102)	(0.195)	(0.452)	(0.078)	(0.205)	(0.463)
Credit score/100	-0.036^{***}	-0.051^{***}	0.045^{***}	-0.009^{***}	-0.039^{***}	-0.017
	(0.002)	(0.004)	(0.005)	(0.003)	(0.007)	(0.015)
Debt-to-income	-0.001^{**}	0.003^{***}	0.026^{***}	0.001	-0.008^{***}	0.030^{***}
	(0.0004)	(0.001)	(0.003)	(0.001)	(0.003)	(0.005)
log(Annual income)	0.0002^{***}	-0.001^{***}	-0.0004^{***}	0.0003^{***}	-0.001^{***}	-0.001^{**}
	(0.00004)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0004)
Age of the borrower	-0.0004^{**}	0.0001	0.001	-0.001^{***}	0.002^{***}	-0.003
	(0.0002)	(0.0003)	(0.001)	(0.0002)	(0.0004)	(0.002)
Age of the borrower^2	0.00000^{**}	0.00000	-0.00001	0.00001^{***}	-0.00002^{***}	0.00001
	(0.00000)	(0.00000)	(0.00001)	(0.00000)	(0.00000)	(0.00002)
log(Number of accounts)	-0.001^{***}	0.010^{***}	0.010^{**}	-0.001	0.005^{**}	0.014^{**}
	(0.001)	(0.002)	(0.004)	(0.001)	(0.002)	(0.007)
log(Number of inquiries)	0.002^{***}	-0.013^{***}	0.023^{***}	0.002	-0.019^{***}	0.062^{***}
	(0.001)	(0.001)	(0.003)	(0.001)	(0.003)	(0.009)
Total liabilities	0.002^{***}	0.002^{***}	-0.001	-0.0004	0.006^{***}	0.002
	(0.0003)	(0.001)	(0.002)	(0.0004)	(0.001)	(0.004)
Credit history	-0.001	0.010^{***}	-0.001	-0.002	0.012^{***}	0.009
	(0.001)	(0.001)	(0.004)	(0.001)	(0.003)	(0.007)
log(No of recently opened accounts)	0.004^{***}	-0.016^{***}	-0.010^{**}	0.003^{**}	-0.020^{***}	-0.015
	(0.001)	(0.002)	(0.004)	(0.001)	(0.004)	(0.011)
log(Pct. of revolving liabilities)	0.002^{***}	0.001^{**}	-0.003	0.001^{*}	0.003^{**}	-0.013^{***}
	(0.0003)	(0.001)	(0.002)	(0.0004)	(0.001)	(0.004)
log(Pct. of mortgage liabilities)	-0.0004^{**}	0.002^{***}		-0.0002	0.003^{***}	
	(0.0002)	(0.0004)		(0.0002)	(0.0004)	
Credit card utilization	0.001^{***}	-0.003^{***}	-0.004^{**}	0.001^{*}	-0.001	-0.001
	(0.0003)	(0.001)	(0.001)	(0.0003)	(0.001)	(0.003)
log(Pct. trades ever delinquent)	-0.0002	0.001^{***}	0.003^{***}	0.0002^{*}	0.001^{*}	0.0001
	(0.0002)	(0.0003)	(0.001)	(0.0001)	(0.0003)	(0.001)
Zip code * Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	41,663	41,611	35,514	14,653	14,643	9,712
Adjusted R^2	0.185	0.213	0.029	0.101	0.086	0.024

Table 14: **Internal Rate of Return by Credit Score**

This table presents the results of regressions that examine whether the platform generates higher returns from low-credit borrowers. The IRR ($\times$ 100) is regressed on the interaction of credit scores less than 660 dummy and loan origination year. Column (1) uses the entire available history of loans at least 12 months old, column (2) and (3) limit the time period to 12 and 24 months since loan origination, respectively. Standard errors are clustered at year level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Full History (1)	12 Month History (2)	24 Month History (2)
Credit score $<$ 660	0.432***	0.353	0.095***
	(0.000)	(0.374)	(0.000)
Year = 2019	1.168***	3.670***	1.881***
	(0.000)	(0.125)	(0.000)
Year = 2020	1.312***	5.780***	
	(0.000)	(0.125)	
Credit score $<$ 660 $\times$ Year = 2019	1.374***	1.148***	1.098***
	(0.000)	(0.374)	(0.000)
Credit score $<$ 660 $\times$ Year = 2020	2.125***	2.212***	
	(0.000)	(0.374)	
Loan Term	Y	Y	Y
N	42	38	28
Adjusted R^2	0.800	0.865	0.687

Table 15: **Internal Rate of Return Heterogeneity**

This table presents the results of regressions that examine how the internal rate of return varies with borrower characteristics. The dependent variable is the internal rate or return (× 100) for each loan. Standard errors are clustered at zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
College or higher	4.301***			
	(0.464)			
Debt consolidation		2.665***		
		(0.636)		
Salaried employee			3.404***	
			(0.513)	
Used a computer				2.184***
				(0.432)
Credit score	0.053***	0.057***	0.043***	0.052***
	(0.006)	(0.006)	(0.006)	(0.006)
log(Income)	3.580***	4.191***	2.480***	3.940***
	(0.573)	(0.576)	(0.631)	(0.571)
Debt-to-income	−0.092***	−0.105***	−0.118***	−0.098***
	(0.026)	(0.026)	(0.027)	(0.026)
Age of the borrower	0.259*	0.177	0.082	0.200
	(0.148)	(0.148)	(0.176)	(0.148)
Age of the borrower^2	−0.006***	−0.005***	−0.004*	−0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
log(No of accounts)	−2.686***	−2.310***	−2.447***	−2.189***
	(0.522)	(0.520)	(0.538)	(0.519)
log(No of inquiries)	−6.414***	−6.426***	−6.314***	−6.424***
	(0.461)	(0.461)	(0.476)	(0.461)
log(Total liabilities)	1.106***	1.232***	1.012***	1.236***
	(0.202)	(0.201)	(0.202)	(0.201)
log(Credit history)	4.858***	4.952***	5.620***	4.955***
	(0.549)	(0.549)	(0.565)	(0.549)
log(Loan amount)	−2.869***	−3.087***	−2.646***	−2.828***
	(0.327)	(0.344)	(0.336)	(0.328)
Contract years	−4.566***	−4.481***	−4.573***	−4.553***
	(0.222)	(0.222)	(0.227)	(0.221)
Zip code*Year	*Y*	*Y*	*Y*	*Y*
N	323,854	323,854	295,357	323,854
Adjusted R^2	0.006	0.006	0.004	0.006

Table 16: **Alternative Data: 500 to 600 Credit Score Sample**

This table reports the estimation results of the regressions that shows alternative variables are informative in predicting default behavior of individuals with credit score less than 600. The sample is restricted to disqualified applicants with credit score between 500 and 600 and no delinquencies at the time of application. The dependent variable in columns (1) through (3) is a dummy variable that indicates if there are any delinquencies approximately after 12 months since the application. In columns (4) through (6), the dependent variable indicates if there is a credit card with 90 days or more past due. Standard errors are clustered at zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

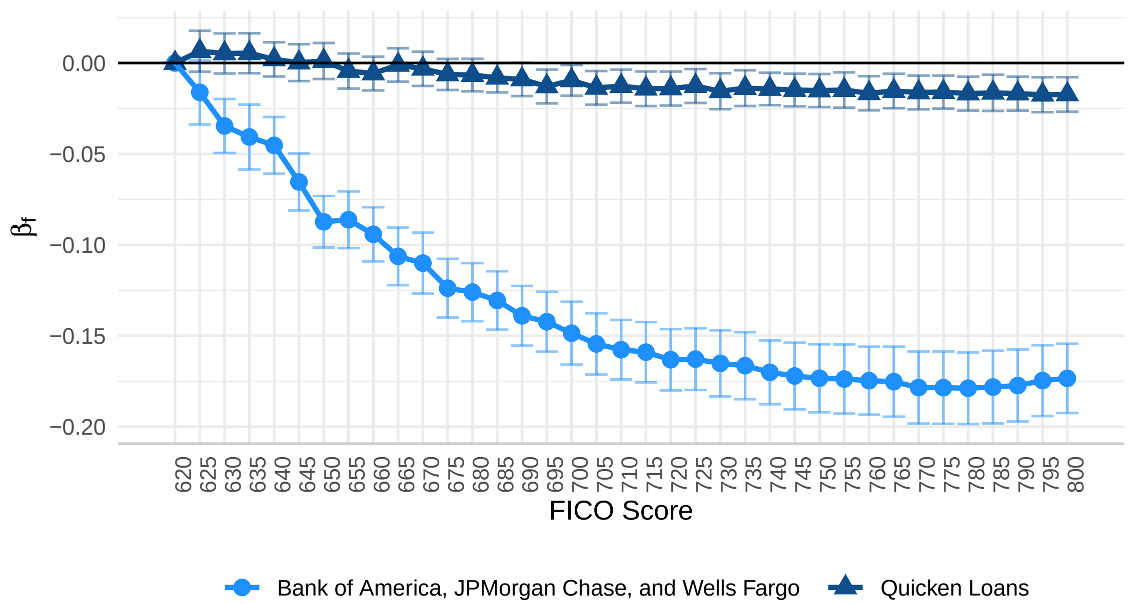
	Any delinquency			Credit card 90+ past due		
	(1)	(2)	(3)	(4)	(5)	(6)
College Degree	-0.021^{***}			-0.043^{***}		
	(0.004)			(0.005)		
Advanced Degree	-0.030^{***}			-0.075^{***}		
	(0.009)			(0.009)		
Salaried worker		-0.007			-0.023^{***}	
		(0.004)			(0.005)	
Computer			-0.014^{***}			-0.010^{**}
			(0.004)			(0.005)
log(Annual income)	0.006^{**}	0.005	0.005^{*}	-0.065^{***}	-0.062^{***}	-0.067^{***}
	(0.003)	(0.003)	(0.003)	(0.008)	(0.008)	(0.008)
Age of the borrower	0.001	0.002^{*}	0.001	0.001	0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age of the borrower^2	-0.00002^{*}	-0.00003^{**}	-0.00002^{*}	0.00001	0.00001	0.00001
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00001)
Debt-to-income	-0.00000	-0.00000	-0.00000	-0.006^{***}	-0.006^{***}	-0.006^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.001)	(0.001)	(0.001)
log(Number of accounts)	0.023^{***}	0.019^{***}	0.021^{***}	0.060^{***}	0.055^{***}	0.053^{***}
	(0.003)	(0.004)	(0.003)	(0.007)	(0.008)	(0.007)
log(Number of inquiries)	0.051^{***}	0.052^{***}	0.051^{***}	0.047^{***}	0.047^{***}	0.048^{***}
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
log(Total liabilities)	0.022^{***}	0.022^{***}	0.021^{***}	0.044^{***}	0.042^{***}	0.042^{***}
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
log(Credit history)	-0.009^{***}	-0.010^{***}	-0.010^{***}	-0.047^{***}	-0.052^{***}	-0.049^{***}
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Zip code×Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
Five point credit score	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	84,920	79,328	84,920	78,113	73,199	78,113
Adjusted R^2	0.026	0.025	0.025	0.110	0.108	0.108

Figure A1. Predictability of FICO: Quicken Loans vs. Banks

Panels A and B of this figure plot the estimates of β_f and corresponding 95% confidence interval in the following estimation using a sample of subprime mortgage borrowers and prime mortgage borrowers, respectively. Subscripts i, f, s, and t represent the borrower, FICO bin, state, and loan application year, respectively. *Default* is a dummy variable that takes the value one if the borrower i was 90 days or more delinquent at any time after origination. $\mu_{s,t}$ represents $state \times year$ fixed effects. The green lines denote mortgages originated by Quicken Loans, other lines mortgages originated by large banks. Standard errors are clustered at state level.

$$Default_{i,s,t} = \sum_f \beta_f \times f_i + \mu_{s,t} + \epsilon_{i,s,t}$$

Panel A: Non-agency mortgages



Panel B: Mortgages sold to Freddie Mac

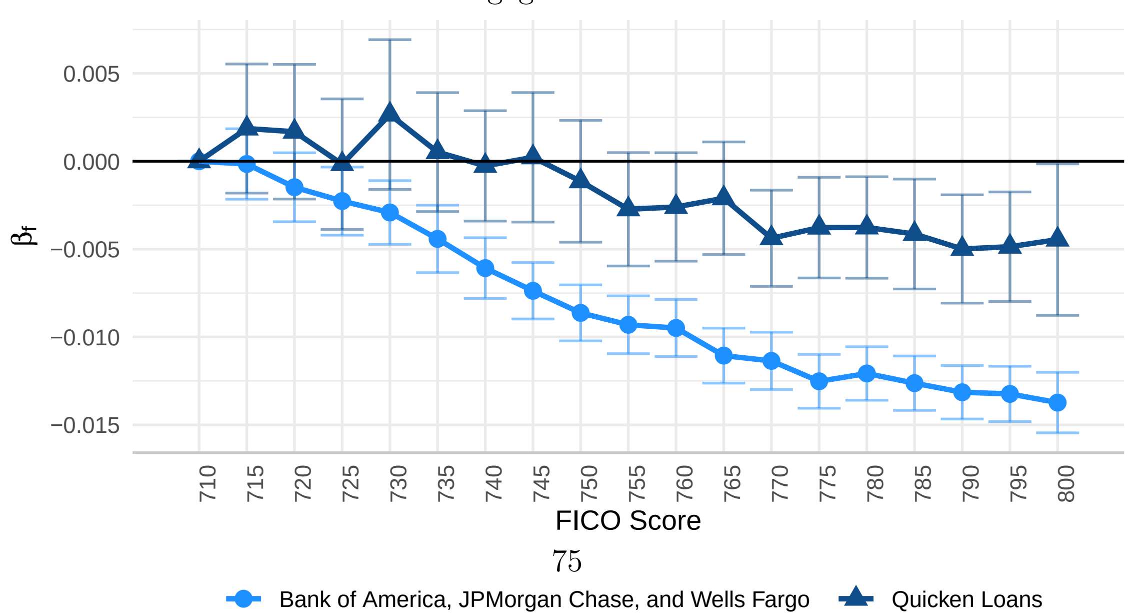


Figure A2. FICO Score Distribution: Quicken Loans vs. Large Banks

This figure compares the FICO score distribution of mortgages originated by Quicken Loans and the banks using the Freddie Mac sample.

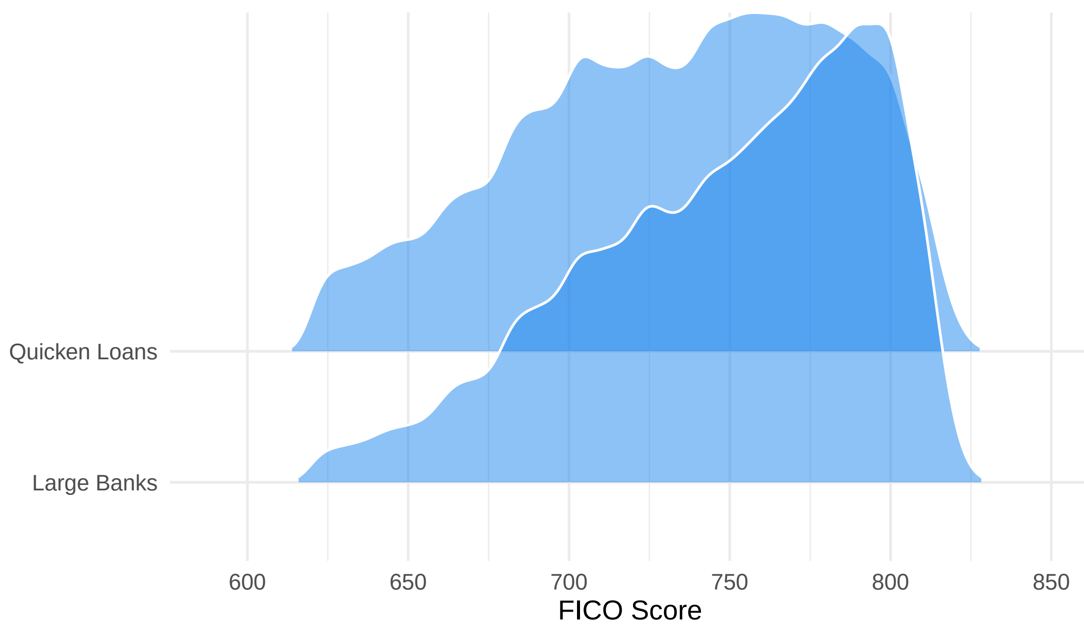


Table A1: **Descriptive Statistics - Mortgage Samples**

This table reports the descriptive statistics of mortgage samples used in this paper. Panel A contains descriptive statistics of non-agency mortgages, Panel B descriptive statistics of mortgages sold to Freddie Mac (agency mortgages). Panel C contains descriptive statistics of mortgage applications data.

Panel A: Moody's Sample

Statistic	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Delinquent	10,353,772	0.08	0.26	0.00	0.00	0.00
FICO Score	10,353,772	735	49	699	743	777
Loan amount	10,353,772	245,393	162,161	132,000	212,000	329,000
New purchase mortgage	10,353,772	0.56	0.50	0.00	1.00	1.00
Loan-to-value	10,353,772	77.52	16.80	73.00	80.00	89.00
Interest rate	10,352,224	5.10	1.84	3.88	4.38	6.00
Year	10,353,772	2012	5	2006	2013	2015

Panel B: Freddie Mac Sample

Statistic	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Delinquent	18,196,672	0.04	0.20	0.00	0.00	0.00
FICO Score	18,196,672	741	39	710	746	774
Loan amount	18,196,672	199,992	109,647	118,000	176,000	260,000
Loan-to-value	18,196,672	75	17	68	80	85
Debt-to-income	18,196,672	46.69	109.10	27.00	35.00	43.00
Interest rate	18,196,672	5.56	1.23	4.50	5.63	6.50
New purchase mortgage	18,196,672	0.47	0.50	0.00	0.00	1.00
Year	18,196,672	2008	6	2003	2007	2013

Panel C: HMDA Sample

Statistic	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Approved	17,593,112	0.62	0.48	0.00	1.00	1.00
Loan amount	17,593,112	227,169	131,387	135,000	205,000	305,000
Annual income	17,593,112	108,940	3,368,695	53,000	80,000	120,000
Non-white	17,591,130	0.29	0.45	0.00	0.00	1.00
Applicant's age	17,272,470	47	15	30	50	60
Debt-to-income	12,494,604	0.38	0.12	0.33	0.38	0.45
New purchase	17,593,112	0.54	0.50	0.00	1.00	1.00
More than 20% college educ.	17,201,167	0.49	0.50	0.00	0.00	1.00
Joint application	17,593,112	0.42	0.49	0.00	0.00	1.00
Interest rate	11,971,136	4.82	114.37	3.88	4.38	4.88
Conventional mortgage	17,593,112	0.81	0.39	1.00	1.00	1.00

Table A2: **Predictability of FICO in General**

This table reports the results of the regressions that examine the relationship between the FICO score and propensity to default. Panel A uses a sample of 30-year fixed rate privately securitized mortgages. The dependent variable in Panel A is $Default_{i,s,t}$, which indicates whether loan i in state s originated in year t was 90 days or more delinquent within five years of origination. Panel B uses the sample of applicants rejected by the Platform. The dependent variable in Panel B indicates whether applicant i defaulted on a credit card within 12 months of the application at time t. Column (1) uses the sub-sample of borrowers with FICO scores less than 660, column (2) the sub-sample of borrowers with credit scores greater than or equal to 660. Standard errors are clustered at zip code level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Mortgages

	Moody's Sample		Freddie Mac Sample
	(1)	(2)	(3)
	620 <= FICO < 660	660 <= FICO < 800	660 <= FICO < 800
FICO Score/100	-0.100***	-0.054**	-0.044***
	(0.007)	(0.023)	(0.001)
Loan-to-value	0.028***	0.053***	0.001***
	(0.005)	(0.008)	(0.00003)
log(Loan amount)	0.020***	0.004	-0.011***
	(0.003)	(0.002)	(0.0002)
New purchase	0.044***	-0.0003	-0.012***
	(0.006)	(0.001)	(0.0003)
Zip code × Year	✓	✓	✓
Observations	959,287	9,165,010	18,195,428
Adjusted R^2	0.196	0.260	0.087

Table A3: **Mortgage Approvals**

This table reports the results of the regressions that compare mortgage approval decisions and interest rates between fintech and other lenders. In columns (1) through (4), the dependent variable indicates whether the mortgage application was approved. The sample is restricted to mortgage applications for 2018 and 2019. The dependent variable in columns (5) though (8) is the interest rate ($\times$ 100). The sample consists only of originated mortgages. Standard errors are clustered at census tract level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Var: Approved				Dependent Var: Interest Rate			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fintech×(Age<45)	0.025***				0.059			
	(0.001)				(0.107)			
Fintech×(College frac. > 0.2)		0.021***				-0.113		
		(0.001)				(0.101)		
Fintech×Joint application			0.015***				0.048	
			(0.001)				(0.096)	
Fintech × (Debt-to-income>0.4)				0.023***				-0.222*
				(0.001)				(0.120)
log(Income)	0.058***	0.057***	0.057***	0.057***	0.253**	0.252**	0.252**	0.128
	(0.000)	(0.000)	(0.000)	(0.000)	(0.106)	(0.107)	(0.107)	(0.087)
log(Loan amount)	-0.011***	-0.011***	-0.011***	0.012***	-0.073	-0.077	-0.078	-0.026
	(0.000)	(0.000)	(0.000)	(0.000)	(0.143)	(0.138)	(0.138)	(0.130)
Age of the applicant		-0.0005***	-0.0005***	-0.0004***		-0.002	-0.002	-0.002
		(0.000)	(0.000)	(0.000)		(0.001)	(0.001)	(0.002)
Joint application	0.024***	0.024***	0.022***	0.007***	0.024***	0.024***	0.022***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Conventional mortgage	0.055***	0.055***	0.055***	0.050***	0.177***	0.182***	0.182***	0.106**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.031)	(0.035)	(0.034)	(0.045)
Refinance mortgage	-0.102***	-0.101***	-0.101***	-0.102***	-0.358***	-0.344***	-0.343***	-0.352***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.103)	(0.090)	(0.089)	(0.088)
Cash-out refinance mortgage	-0.084***	-0.084***	-0.084***	-0.094***	0.229	0.249	0.25	0.246
	(0.000)	(0.000)	(0.000)	(0.000)	(0.189)	(0.209)	(0.210)	(0.211)
Race: Asian/Other	-0.040***	-0.040***	-0.040***	-0.027***	-0.174**	-0.181**	-0.181**	-0.177*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.082)	(0.090)	(0.090)	(0.092)
Race: Black	-0.059***	-0.059***	-0.059***	-0.058***	0.035	0.039	0.038	0.059
	(0.001)	(0.001)	(0.001)	(0.001)	(0.100)	(0.098)	(0.099)	(0.101)
Race: Hispanic	-0.032***	-0.032***	-0.032***	-0.028***	-0.043	-0.048	-0.049	-0.04
	(0.000)	(0.000)	(0.000)	(0.000)	(0.039)	(0.041)	(0.041)	(0.043)
Race: Not provided	-0.085***	-0.085***	-0.085***	-0.041***	-0.128	-0.131	-0.13	-0.127
	(0.000)	(0.000)	(0.000)	(0.000)	(0.085)	(0.088)	(0.088)	(0.089)
Age>45	0.010***				-0.027			
	(0.000)				(0.058)			
Debt-to-income>0.40				-0.091***				0.338***
				(0.000)				(0.110)
Census tract × year	✓	✓	✓	✓	✓	✓	✓	✓
Lender	✓	✓	✓	✓	✓	✓	✓	✓
Observations	16,906,221	16,878,144	16,906,221	12,228,427	9,907,916	9,891,657	9,907,916	9,907,916
Adjusted R^2	0.190	0.190	0.190	0.176	0.003	0.003	0.003	0.003

Table A4: **Internal Rate of Return - Mortgages**

This table presents the results of regressions that examined whether the Platform generates higher returns from low-FICO borrowers. Column (1) regresses the Platform's IRR (*100) on a dummy variable that takes the value of one when the borrowers FICO score is less than 660. Column (2) regresses the IRR (*100) on the interaction of the less than 660 dummy and loan origination year. Standard errors are clustered at year level and reported in parentheses below coefficient estimates. We use *, **, and *** to denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	All loans	FICO $<$ 725	725 $<$ FICO $<$ 750	750 $<$ FICO $<$ 775	FICO $>$775
	(1)	(2)	(3)	(4)	(5)
Quicken	0.125***	0.074***	0.154***	0.127***	0.140***
	(0.008)	(0.008)	(0.011)	(0.014)	(0.019)
FICO Score	-0.001***	-0.004***	-0.003***	-0.003***	-0.00001
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Loan-to-value	0.008***	0.009***	0.006***	0.006***	0.007***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Debt-to-income	0.008***	0.005***	0.005***	0.008***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(loan amount)	-0.076***	-0.061***	-0.048***	-0.051***	-0.063***
	(0.011)	(0.014)	(0.014)	(0.018)	(0.019)
New purchase	-0.221***	-0.235***	-0.170***	-0.162***	-0.228***
	(0.009)	(0.011)	(0.014)	(0.016)	(0.017)
Zipcode×Year	✓	✓	✓	✓	✓
Observations	1,010,780	287,458	152,139	201,029	363,072
Adjusted R^2	0.014	0.029	0.015	0.014	0.005