HARVARD | BUSINESS | SCHOOL

Invisible Primes: Fintech Lending with Alternative Data

Marco Di Maggio
Dimuthu Ratnadiwakara
Don Carmichael

Working Paper 22-024

Invisible Primes: Fintech Lending with Alternative Data

Marco Di Maggio
Harvard Business School

Dimuthu Ratnadiwakara
Louisiana State University

Don Carmichael
Upstart

Working Paper 22-024

Funding for this research was provided in part by Harvard Business School. We thank Upstart Network, Inc for providing access to its data

Invisible Primes: Fintech Lending with Alternative Data*

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

October 6, 2021

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 60% higher probability of being rejected and higher interest rates for those approved. 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. We show that funding loans to these borrowers leads to better economic outcomes for the borrowers and higher returns for the fintech platform.

Keywords: Fintech Lending, Alternative Data, Machine Learning, Algorithm Bias

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

*We thank Upstart Network, Inc for providing access to its data.

[†]Harvard Business School and NBER: mdimaggio@hbs.edu.
[‡]E. J. Ourso College of Business, Louisiana State University: dratnadiwakara2@lsu.edu.
[§]Upstart: 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. Their online presence has enabled these new companies to cut costs and acquire significant market share across lending products. Online lender Quicken Loans, for example, is the largest mortgage originator in the United States, and fintech lenders account for a quarter of the personal credit market. 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, while equally creditworthy, but overlooked if underscored, individuals faced limited and expensive, if any, options. The advent of fintech lenders has the potential to change this circumstance.

The emergence of this new class of intermediaries has raised a number of policy-related questions. A key question revolves around the impact on credit availability of credit models that employ alternative data and algorithmic underwriting. Alternative data sources and a more automated underwriting approach could reduce loan origination costs which might translate into lower rates for borrowers. Alternative underwriting models might also be able to identify individuals currently overlooked by standard measures of creditworthiness, such as credit score. Observed former director of the Consumer Financial Protection Bureau Richard Cordray: "Adding this kind of alternative data into the mix thus holds out the promise of opening up credit for millions of additional consumers."[1] Relying primarily on credit score has the potential to exclude a large fraction of Americans from the 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, equitable, nondiscriminatory access to credit, particularly for credit *invisibles* and those limited by their credit history or lack thereof.

If the benefits of these innovations are potentially large, so are the risks. Regulators and consumer

[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

advocates have been concerned, for example, with the potential for biased treatment, which would violate fair lending regulations. Using information about education, utility bills, or bank transactions could inadvertently reduce credit access for some households, and the extent to which new data fed into a model is correlated with information that could result in discriminatory practices is largely unknown. Such concerns are exacerbated by the fact that, like credit scores, the new underwriting models are proprietary.

Empirical evidence on these issues is scarce. The ideal setting would entail observing fintech lenders' lending decisions 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 regulators' concerns are corroborated by evidence. 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, namely, 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 an elusive 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 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 "Upstart" or the "Platform"), which provided access to its anonymized administrative data. Operating in the personal loan space, the fastest growing category in consumer lending, Upstart 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 Upstart is unique in a number of respects. It covers rejected as well as funded loans, provides a panel for both funded *and* rejected individuals, and facilitates assessment of the impact of Upstart's underwriting model on credit access by comparing actual outcomes of Upstart's model to counterfactual outcomes based on a "traditional model" that does not use Upstart's

[3] https://www.experian.com/blogs/ask-experian/research/personal-loan-study/

[4] https://www.sec.gov/Archives/edgar/data/1647639/000119312520285895/d867925ds1.htmtoc867925$_4$

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 (see Ficklin and Watkins (2019)). Indeed, Upstart applied for and received the first CFPB No-Action Letter ("NAL") in 2017. As part of the NAL review process, the CFPB analyzed Upstart'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 Upstart. In late 2020, Upstart was approved for an additional three years under the NAL program. Access to this counterfactual model's outcomes reassures us of its representativeness with respect to traditional lenders' credit decisions.

We begin our analysis by exploring how standard metrics of creditworthiness, such as credit score, capture a borrower's probability of defaulting. The objective is to investigate whether and how alternative data can usefully complement existing borrower information for underwriting purposes. We begin by investigating the relationship between credit score and default for credit cards, a product not offered by Upstart. As expected, a monotone negative relationship is found, that is, borrowers with higher scores exhibit a significantly lower probability of defaulting. Examining personal loans funded by Upstart, we find no relationship between the probability of defaulting and credit scores below 700. In other words, 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.

This comparison suggests that, for some individuals, credit score may not paint an accurate picture of *future* creditworthiness. This might be because the characteristics used to compute credit score are not reliably informative or predictive of future behavior. 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 history.

A natural next step is to investigate the credit model used by the fintech platform. We first show that the new credit model outperforms the credit score in predicting delinquencies and charge-offs, that is, it identifies significant differences among all borrowers, even those with low credit scores. We demonstrate this in multiple ways. For example, we plot the distribution of the Upstart default probability by fine credit score bins and show the distributions to have fat tails, that is, the new score identifies important differences in creditworthiness even among borrowers with similar credit score. We also show the area under the curve, a common measure of model predictability, to differ between a model using standard metrics and the one used by Upstart.

We next investigate the main drivers of Upstart's credit model. One hypothesis is that the model might just capture, in addition to credit score, a combination of other borrower information in the credit report correlated with defaults, such as number of accounts, outstanding credit balance, and history of defaults. We continue to find unexplained variation, however, even after accounting for all the information in credit reports.

Motivated by this evidence, we employ a machine learning algorithm termed "recursive feature elimination with random forest" to iteratively identify the main variables driving the Upstart default probability. 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.

Although the new credit model is assessed to be more accurate, a key question for policy makers and academics is whether it improves credit inclusion, that is, are there borrowers granted credit by Upstart who 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 more than 30% of borrowers with credit scores of less than 680, funded by Upstart over our sample period would have been rejected by the traditional model. We further find that this fraction declines as credit score increases, that is, the mismatch between the traditional and Upstart model is magnified among low-credit score borrowers. Borrowers with credit scores lower than 640 who are granted loans by Upstart have a 60% probability of being rejected by traditional lenders. Only the subset of borrowers with credit scores higher than 740 experience a similar approval rate at Upstart and at a traditional lender. This suggests that alternative data can be helpful in assessing the creditworthiness of borrowers whose credit scores deem them riskier.

Differences between the traditional and alternative model might not be confined to approval decisions. Interest rates might also differ. Higher rates charged by Upstart, construed as a risk premium, could provide a plausible explanation for the expansion of credit. We address this question by comparing for a subset of funded loans the respective rates predicted by the traditional model and Upstart. Our finding that low-credit score borrowers funded by Upstart 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 find that other characteristics that predict a significant differential between the performance of the traditional and Upstart models are magnified for thin credit file individuals (those with short credit histories) and those with advanced degrees and salaried jobs. 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.

Given the expansion of credit and lower rates charged to higher-risk borrowers, one might wonder whether Upstart is subsidizing market share growth at the expense of funding unprofitable loans. We complement our previous results by investigating whether its loans negatively affect Upstart's profitability or its borrowers. Addressing the first question by computing the internal rate of return (IRR), the key metric tracked by Upstart, 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.

Our analysis also explores 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 Upstart. This analysis includes observations of subsequent 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 a key feature of Upstart's business operation for identification purposes, specifically, that applicants with debt-to-income ratios greater than 50% are automatically denied funding. This cutoff provides a natural setting for a regression discontinuity strategy in which outcomes for borrowers with debt-to-income ratios lower than 50% are compared to those for borrowers with debt-to-income ratios above 50% at the time of application. Comparing borrowers with similar characteristics near the cutoff, we confirm that borrowers able to obtain financing from Upstart experience an improvement in financial health.

The fact that the data provided by Upstart is unique and difficult to replicate in other settings raises the concern that our results may be specific to a particular fintech platform and not generalizable. 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 generate 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. Although we acknowledge concerns around potential consequences of using alternative information in funding decisions, for example, with respect to privacy and bias, our findings also indicate 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 indentify these and many other credit-invisible people as a shadow group of prime borrowers.

Our findings contribute to an emerging literature on fintech lending.[5] 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 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).[6] Recently, Blattner and Nelson (2021) 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. Our paper exploits administrative and regulatory data which provide us with the unique opportunity to directly quantify the benefits derived from introducing models based on alternative data on credit availability. In addition, we identify in the invisible primes the most likely beneficiaries of this expansion of credit.

Other research examines specific information fintech lenders use to make underwriting decisions. 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. Counter to our findings, Di Maggio and Yao (2020) remark fintech lenders' reliance on information provided in credit reports to fully automate their lending decisions. Furthermore, individuals who borrow from fintech lenders exhibit a higher propensity to default. 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)). We contribute evidence to this strand of the literature that alternative data can be successfully employed to improve credit access and reliance on conventional credit bureau information might risk underserving an important part of the population.[7]

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

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

[7] Related papers in this literature include Danisewicz and Elard (2018); De Roure et al. (2019); Balyuk (2019),

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 Upstart'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 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 sets of applicants. To the best of our knowledge, no similar data has previously been made available to academics.

Founded in 2012, Upstart, one of a few public online lending platforms, provides personal loans to borrowers throughout the United States, and it is one of the few public online lending platforms. Upstart's underwriting model differs from those used by traditional lenders in that its pricing algorithm incorporates alternative data from unconventional sources. To forestall running afoul of existing regulations, Upstart applied for and obtained a No-Action Letter from the CFPB in 2017. The CFPB analyzed Upstart'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 Upstart. In late 2020, Upstart was approved for an additional three years

Iyer et al. (2016), Mariotto (2016), Wolfe and Yoo (2018), Vallee and Zeng (2019), Hertzberg et al. (2018), and Balyuk and Davydenko (2019).

under the NAL program.[8]

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 900,000 loans originated by Upstart. Our analytic sample is restricted to 770,523 loans for which the traditional credit score is available. In recent years, Upstart 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 acquisition channels. Upstart reaches borrowers in multiple ways. Although a borrower can complete an application directly on the website, and Upstart 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. Upstart's market reach is plotted in Figure 2. Darker, shaded counties on the US map capture a higher number of loans per capita; although Upstart 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 Upstart's main bank partner, Cross River Bank, not being active in those states. Our specifications control for zipcode by year fixed effects, which takes 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. On average, Upstart'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 Upstart's focus to be on other than the individuals

[8] The no action letter published by the CFPB can be found here https://files.consumerfinance.gov/f/documents/201709$_c fpb_u pstart - no - action - letter.pdf$.

traditionally regarded as most creditworthy. Most borrowers are between 28 and 46 years old, with a median age of 37 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 on the same job tenure of 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 reports main statistics for both funded borrowers, and those immediately disqualified because they do not meet Upstart's eligibility criteria. On average, credit score is 70 points higher and annual income $14,000 higher for funded than for non-funded applicants. Total liabilities and credit balance are also higher for funded ($118,000) than for non-funded ($65,000) 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. Upstart pulls credit reports regularly for borrowers who are current and monthly for borrowers who have missed payments. Upstart 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 Upstart'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 Upstart. Specifically, we select loan applicants not approved by Upstart 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.[9]

Exclusive reliance on credit score could also exclude a large fraction of Americans from the credit markets altogether. According to Fair Isaac Corporation, the company that owns the FICO algorithm, 28 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-invisible" would be automatically disqualified by lenders that rely primarily on credit score. Fannie Mae guidelines, for example, have specific credit score cut-offs.[10]

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). Online mortgage lender Social Finance (SoFi) has announced that its credit decisions do not take credit score into account at all.[11]

We next check the predictive power of the credit score for Upstart's loan sample. Specifically, we estimate equation (1) using Upstart'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. Figure 4 plots the estimated β_{cs} with corresponding 95% confidence intervals. 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 an Upstart borrower with a credit score of 620 as for an Upstart 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.

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

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

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

We report this specification in a regression framework in Table 4, in which to capture any time-varying local heterogeneity, we include a number of borrower and loan characteristics as well as zip code by year fixed effects. We find the credit score coefficient to not be statistically significant for borrowers with lower credit scores.

Our results show credit score to not be as predictive of defaults for fintech 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 Upstart'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 Upstart's measure of borrowers' creditworthiness, which is expressed as a probability of default, as the main independent variable. We refer to this as Upstart default probability henceforth. This regression uses Upstart default probability bins instead of the credit score bins in equation (1). The β_{us} estimates and their corresponding 95% confidence intervals are plotted in Figure 5. Panel A uses 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 Panel B shows to be robust across credit score categories. This suggests that Upstart is able to achieve more granular pricing even among low-credit score borrowers.

We next investigate the main drivers of Upstart's credit model. To determine whether the Upstart default probability might simply capture, in addition to credit score, a combination of other borrower characteristics correlated with defaults, 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 Upstart 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. That the results show the Upstart default probability to be 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, suggests that its information content is not subsumed by these other variables. We find a one standard deviation increase in the Upstart 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 Upstart default probability for different subsamples of credit score. 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 low-credit score borrowers, there is substantial variation in the Upstart 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 Upstart default probability. This is key to providing a superior ability to price risk correctly.

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 Upstart 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, that is, 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 Upstart 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 Upstart 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 Upstart default probability.

Table 6 shows the individual contribution of the selected variables by regressing the Upstart 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. Column (7) includes only non-traditional variables. Standard errors are clustered at zip code level.

The results suggest that non-traditional variables are important predictors in the Upstart 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 Upstart default probability, as indicated by the significant increase of the adjusted R^2 in column (2). Based on unreported coefficient estimates of the fixed effects and conditional on other controls, the Upstart default probability can change by as much as 4.2% depending on level of education. This effect is highly economically significant given the mean Upstart default probability of 22%. Device type or technology can move the Upstart default probability by about 4.7%, employment type by about 2.8%. These economic impacts are reported in the last row of Table 6.

We further show the new credit model to perform better at predicting borrower creditworthiness (using 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 7 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 Upstart default probability, with differences indicated in the first column. Column (1) in Panels A and B shows credit score to have little predictive power for all borrowers, and the Upstart default probability to have reasonable predictive power of more than 60 across all sub-categories.

Our results show the Upstart 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 Upstart'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 Upstart's underwriting model on credit access by comparing its outcomes with counterfactual outcomes generated by a traditional model that does not employ Upstart's alternative underwriting features. Rather than build 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)). 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 credit decisions (approved or disqualified) and interest rate based on the traditional model for loan applications in the second half of 2019. 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. Figure 8 plots the outcome distribution under both the traditional model and Upstart's underwriting model for each 20-point credit score bin. The figure suggests that about 25% of the applications with low credit scores would have been denied if not for Upstart's underwriting model.

For this part of the analysis we focus on the 98,671 funded loans in the sample for which the traditional model outcomes are available. We examine the fraction of funded loans that would have been rejected by the traditional model and compare the differences in interest rate proposed by both models. Panel A of Figure 9 shows that approximately 60% of borrowers with credit scores less than 640 funded by Upstart over the sample period would have been rejected by the traditional model. We further find that this fraction declines as credit score increases, that is, the mismatch between the traditional and Upstart model is magnified among low-credit score borrowers. 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 reveals a difference as well in APR; low-credit score borrowers funded by Upstart would have incurred significantly higher interest rates under the traditional model. Hence, the differences between models affect both the extensive and intensive margins.

Figures 10 and 11 further explore the heterogeneity of differences between the traditional and Upstart underwriting models. Figure 10 plots the percent of funded loans that would have been rejected for different subsamples. Panels A, B, C, and D show that borrowers who earn more than $55,000, are more educated, employed in salaried positions, and have thin credit files are more likely to be rejected by the traditional model. A similar pattern is observed in Figure 11 with respect to differences in interest rates; borrowers who earn more than $55,000 and are more educated and employed in salaried positions would be assigned higher interest rates by the traditional model. Panel A in Table 8 confirms the same result in regression form. The dependent variable in column (1), *Rejected*, takes the value of one if the loan would have been rejected by the traditional model. The dependent variable in column (2) is the difference between the implied interest rate based on the traditional model and the interest rate as calculated. The main independent variables are borrower characteristics that might be helpful in identifying the beneficiaries of the alternative model. Panel B of Table 8 confirms the effects of alternative data, such as education, to be stronger for low-credit score individuals.

Overall, these results show exploiting alternative data to be instrumental in expanding credit access.

5. Profitability

Having shown that subprime borrowers benefit from Upstart's novel pricing strategy, we consider whether it is profitable for the lenders to provide credit based on alternative data.

We begin by computing the average internal rate of return for each origination year-credit score bin for loans originated by Upstart (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.[12] Panel A in Figure 12 plots the computed IRR against the credit score bins. Each line represents the loans originated in 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, cash flows are restricted to one year since loan origination, and 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 9 presents similar results in regression form. We regress the IRR($\times$100) for each origination year-credit score bin-loan term 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, column (2) the 12-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

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

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.

Table 10 shows 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 Upstart'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.

6. 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 Upstart 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. We address these concerns in two 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 Upstart is not allowed to pull credit reports of disqualified applicants 12 months after the initial application.

Table 11 presents the results of a regression aimed at understanding the effects of credit access on the borrowers' outcome variables. 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.

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 a key feature of Upstart's business operation: that applicants with debt-to-income ratio greater than 50% are automatically denied. The maximum allowed debt-to-income ratio in Colorado, Connecticut, Maryland, New York, and Vermont is 45% and these five states are excluded from the following analysis.[13] Panel A of Figure 13 shows the distribution of debt-to-income ratio of funded and disqualified applicants. Not all applicants with debt-to-income ratios less than 50% are funded; some of the borrowers in this category are disqualified due to other reasons.

Panel B of Figure 13, which plots the percent of the funded applications against the debt-to-income bins, shows a discontinuity in the probability of funding at the debt-to-income ratio of 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.

The discontinuity at the 50% cutoff is not a sharp discontinuity—but there is a discontinuity in the probability of being funded at the 50% debt-to-income 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 B of Figure 13. These institutional feature suggests a fuzzy regression discontinuity design given by the following specification.

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

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

Equation 1 is the first stage regression which predicts the probability of being funded based on the 50% debt-to-income cutoff. To capture the differences in slopes, $I(\text{Debt-to-income} > 50\%)$ is interacted with the Debt-to-income ratio. $Funded_i$ is a dummy variable that takes the value one if the application was funded. The dummy variable $I(\text{Debt-to-income} > 50\%)$ indicates whether

[13] https://upstarthelp.upstart.com/questions/108501-what-are-the-minimum-credit-requirements-to-receive-a-loan

the debt-to-income ratio is greater than the cutoff, and μ_{zt} represents zip code $\times$ year fixed effects. Equation 2 uses 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. The control variables include–in addition to the debt-to-income ratio–credit score, income, total liabilities, number of accounts, credit history, age of borrower, and monthly debt payment.

The first column of Table 12 reports the estimation results of equation 1. The sample is restricted to applications where the debt-to-income ratio is between 40 and 60%, and credit report data is available approximately 12 months since the application date. The results suggest that the probability of funding drops by about 19% (0.890 - 50 $\times$ 0.014) at the 50% debt-to-income cutoff. Columns (2) through (6) use other key control variables as dependent variables to show that there is no discontinuity at the cutoff when we consider other applicants' characteristics. Figure 14 shows the results reported in columns (2) through (6) graphically.

Table 13 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.

7. External Validity

A natural question is whether these results are specific to Upstart and thus not applicable to other institutions. Although the data provided by Upstart 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 (blue lines in A1) for both subprime (Panel A) and prime (Panel B) borrowers. Similarly to Upstart, the performance of mortgages originated by Quicken is generally flat (green 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.[14].

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.[15] 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 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 Upstart 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.

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

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

8. 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 major fintech lender to shed new light on this sector. We show that alternative data used by Upstart exhibits substantially more predictive power with respect to likelihood of default than credit score, the standard metric traditionally used to judge borrower creditworthiness. We further show that 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 Upstart 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.

Our results inform the debate on the use of alternative data by providing evidence of positive effects in the screening process for loan approval. Although they do not address arguments around potential concerns about privacy and statistical discrimination, our results do demonstrate that alternative data models deliver quantifiable benefits to both borrowers and lenders.

References

Agarwal, S., Alok, S., Ghosh, P., and Gupta, S. (2019). Fintech and credit scoring for the millennials: Evidence using mobile and social footprints. *Available at SSRN 3507827.*

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.

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.

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

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.

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.

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*.

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.

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.

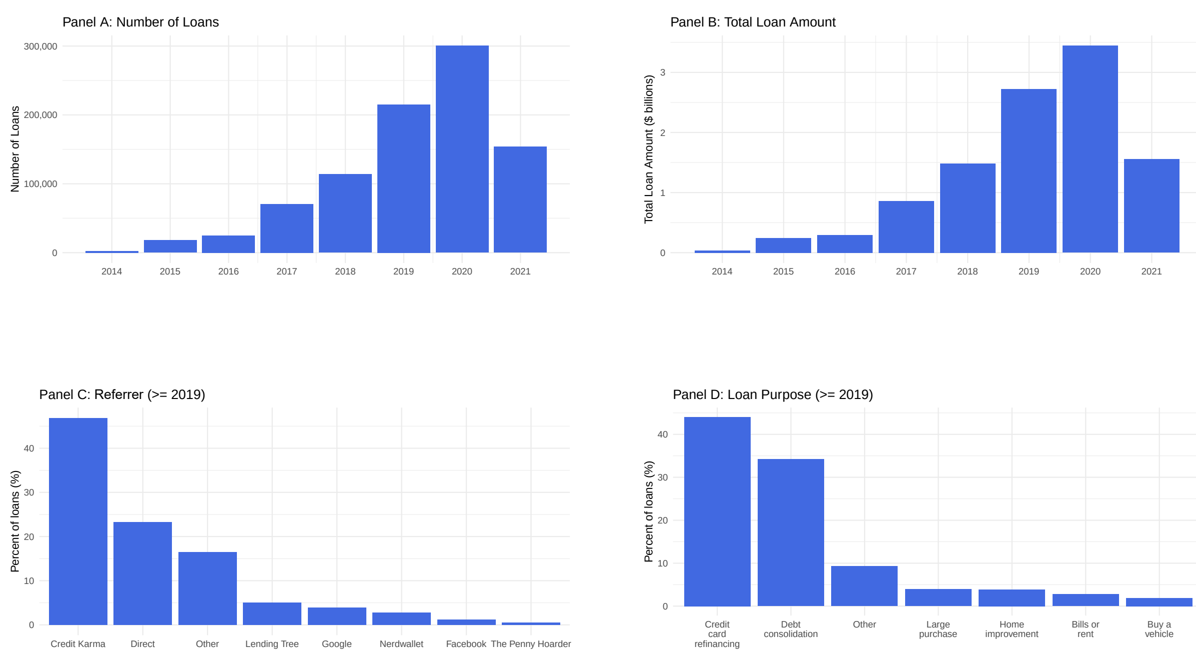


Figure 2. Geographic Coverage

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

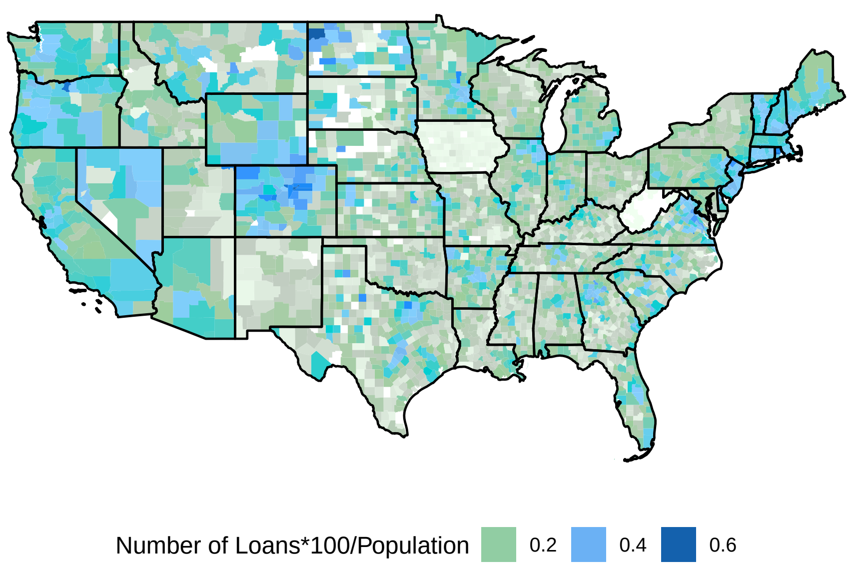


Figure 3. Predictability of Credit Score: Credit Cards

This figure plots the estimates of β_{cs} and corresponding 95% confidence interval in the following estimation using the sample of applicants rejected by Upstart. The sample excludes applicants who have 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.

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

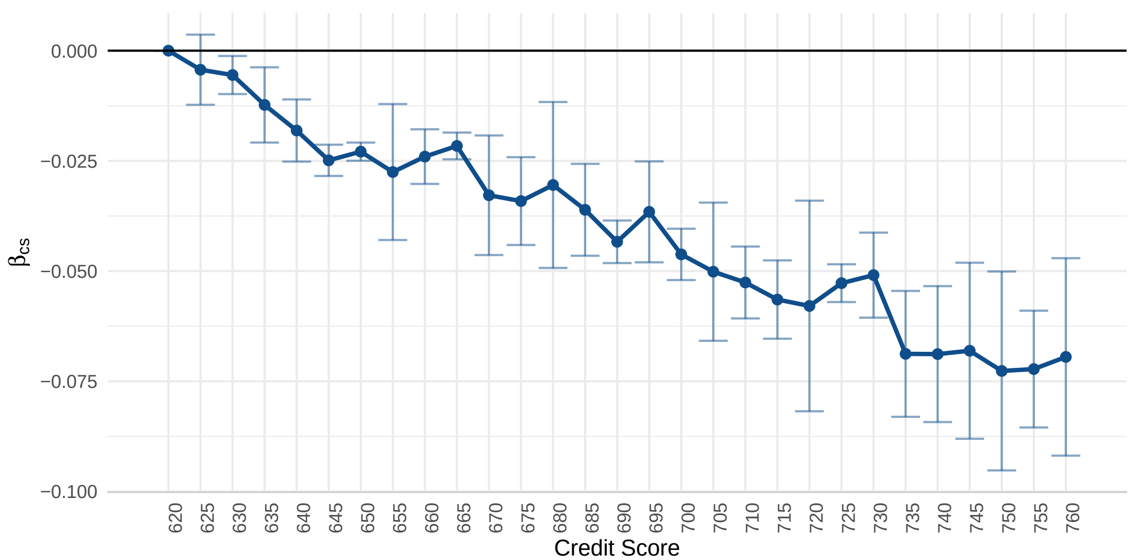


Figure 4. Predictability of Credit Score: Upstart Loans

This figure plots the estimates of β_{cs} and corresponding 95% confidence interval in the following estimation using Upstart'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 (dark line) at any point after origination, or charged-off (light line). $\mu_{s,t}$ represents $state \times year$ fixed effects. Standard errors are clustered at state level.

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

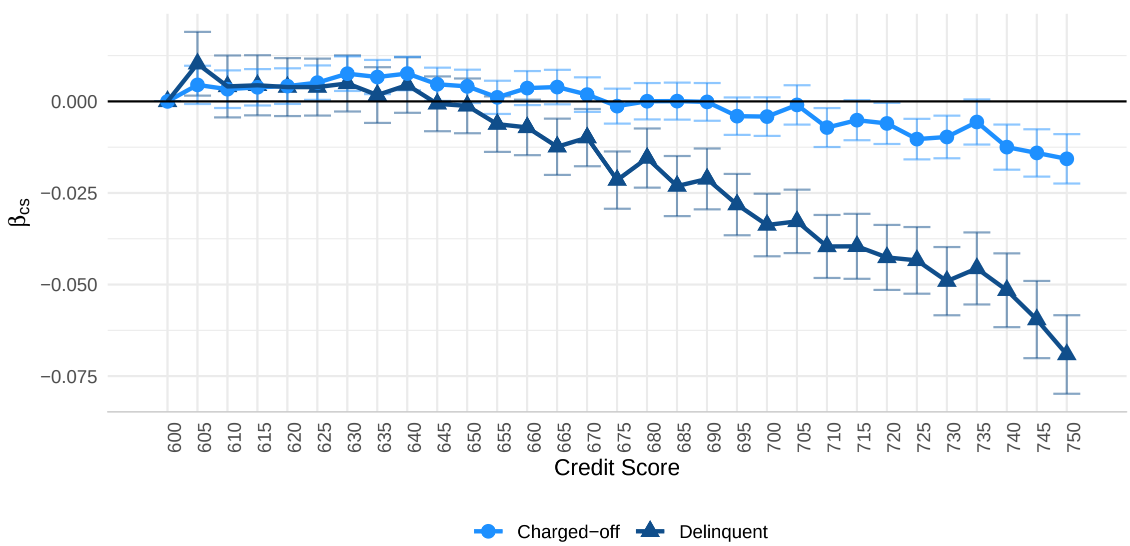
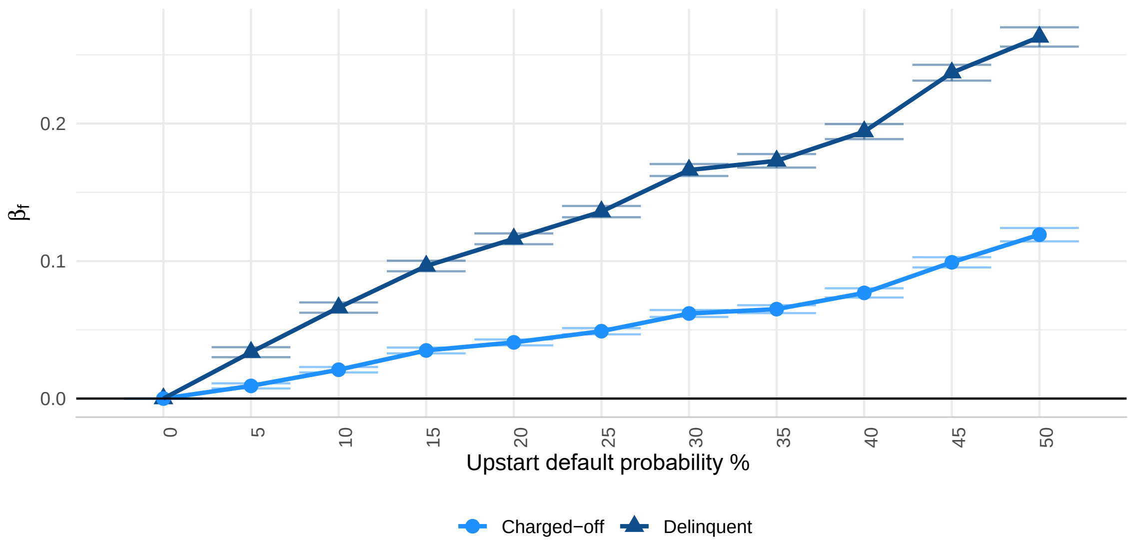


Figure 5. Predictability of Upstart Default Probability

This figure plots the estimates of β_{us} and corresponding 95% confidence interval of the following estimation. Subscripts i,us,s, and t represent the borrower, Upstart 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.

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

Panel A: All Credit Score Categories



Panel B: Charge-off by Credit Score Category

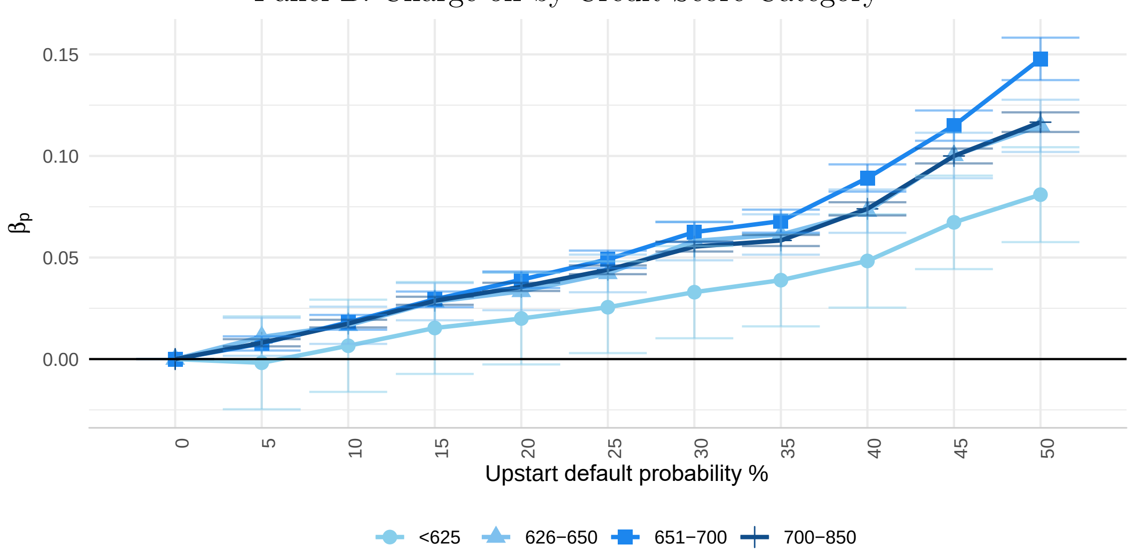


Figure 6. Credit Score and Upstart Default Probability

This figure plots the distribution of the Upstart default probability for loans originated by Upstart separately for each credit score category.

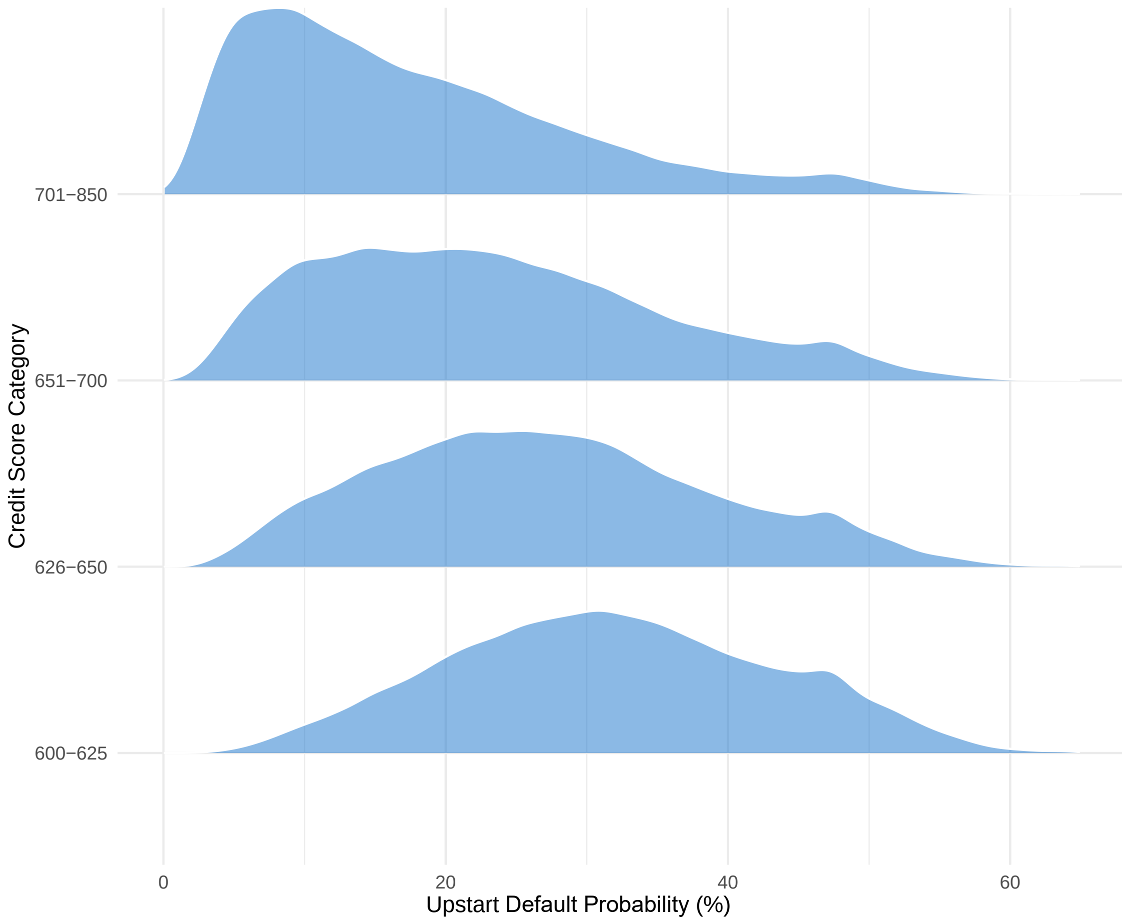
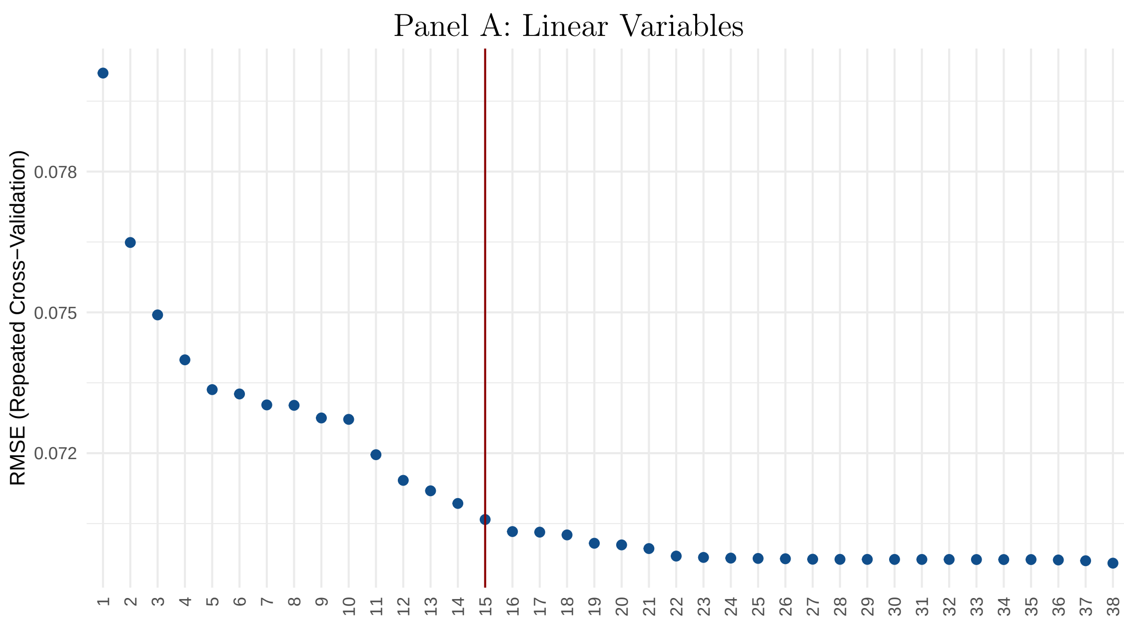


Figure 7. Recursive feature elimination: Root-mean-square deviation

This figure plots the incremental improvement of the model performance when new predictive variables are added to the model that predicts the Upstart 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.



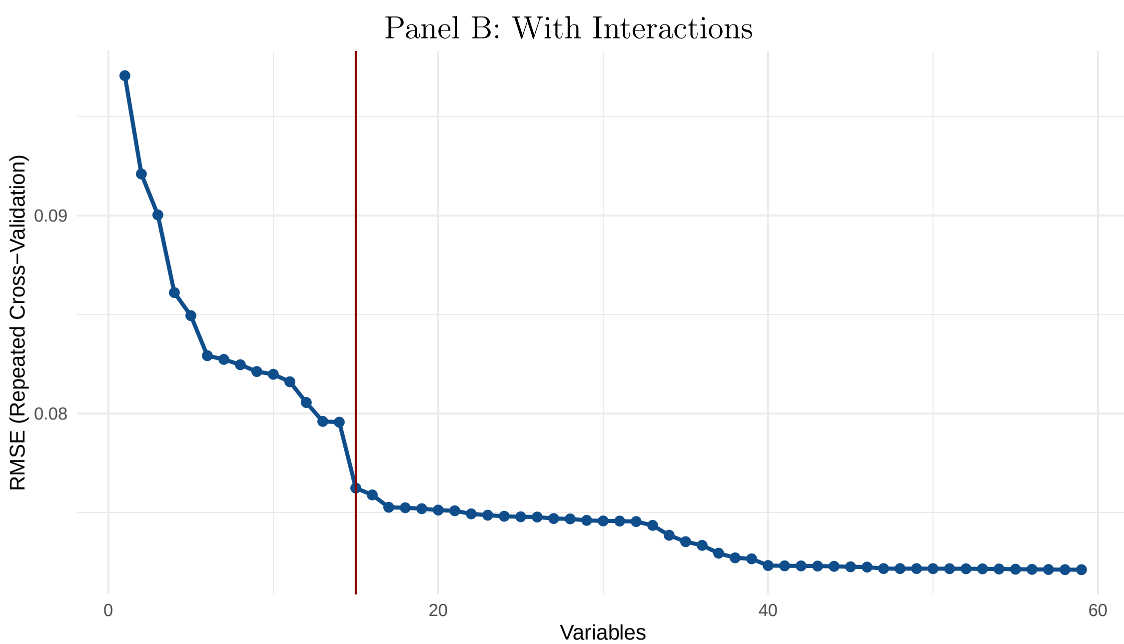


Figure 8. Traditional Model vs. Upstart's Underwriting Model

This figure plots the outcome distribution of loan applications for traditional model estimates area available. The figure plots the share of each outcome against the credit score bin.

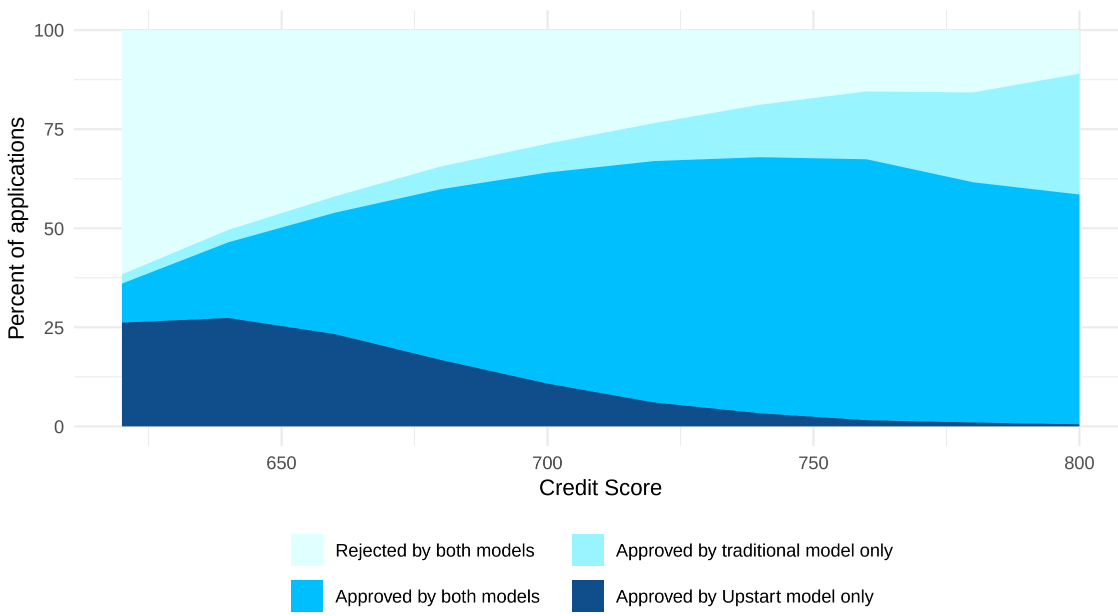
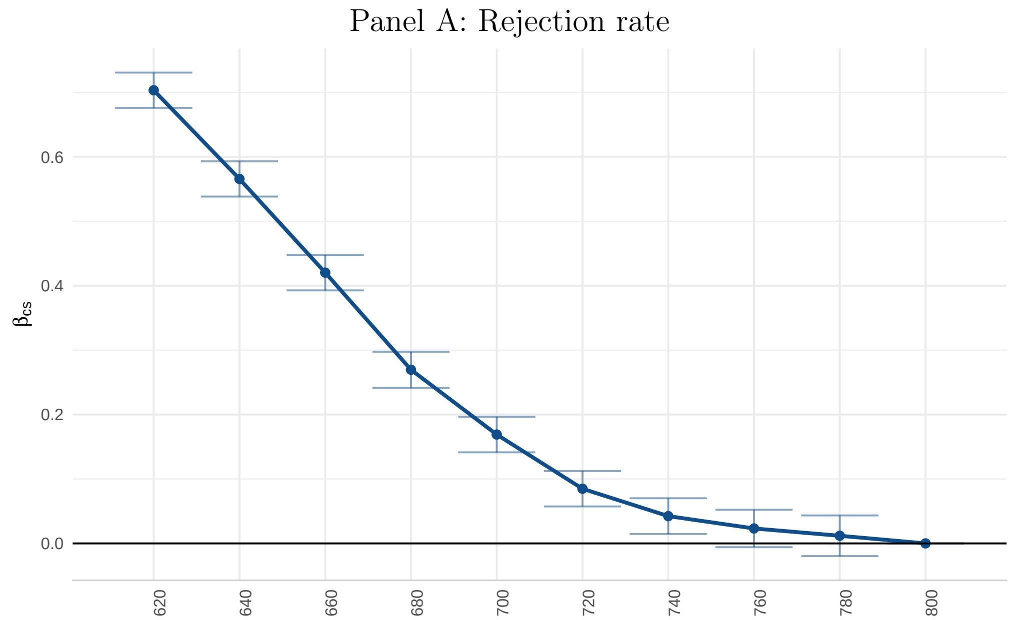


Figure 9. Traditional Model vs. Upstart's Underwriting Model: Outcome Comparison

Panels 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 same variables. The sample consists of loans funded by Upstart for which the traditional model outcomes are available. Standard errors are clustered at zip code level.



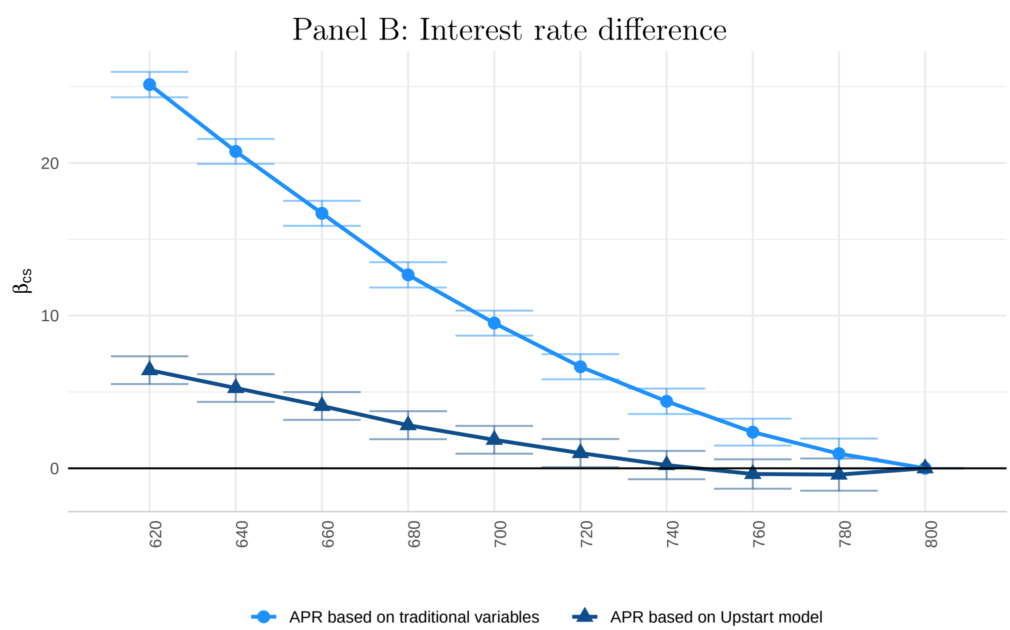


Figure 10. Percent Rejected: Heterogeneity

This figure plots the coefficients and corresponding 95% confidence interval of the regression that regresses the difference between the interest rate under the traditional model and Upstart model on credit score bin dummy variables with zip code fixed effects for different sub samples. Sample splits are given in the panel titles.

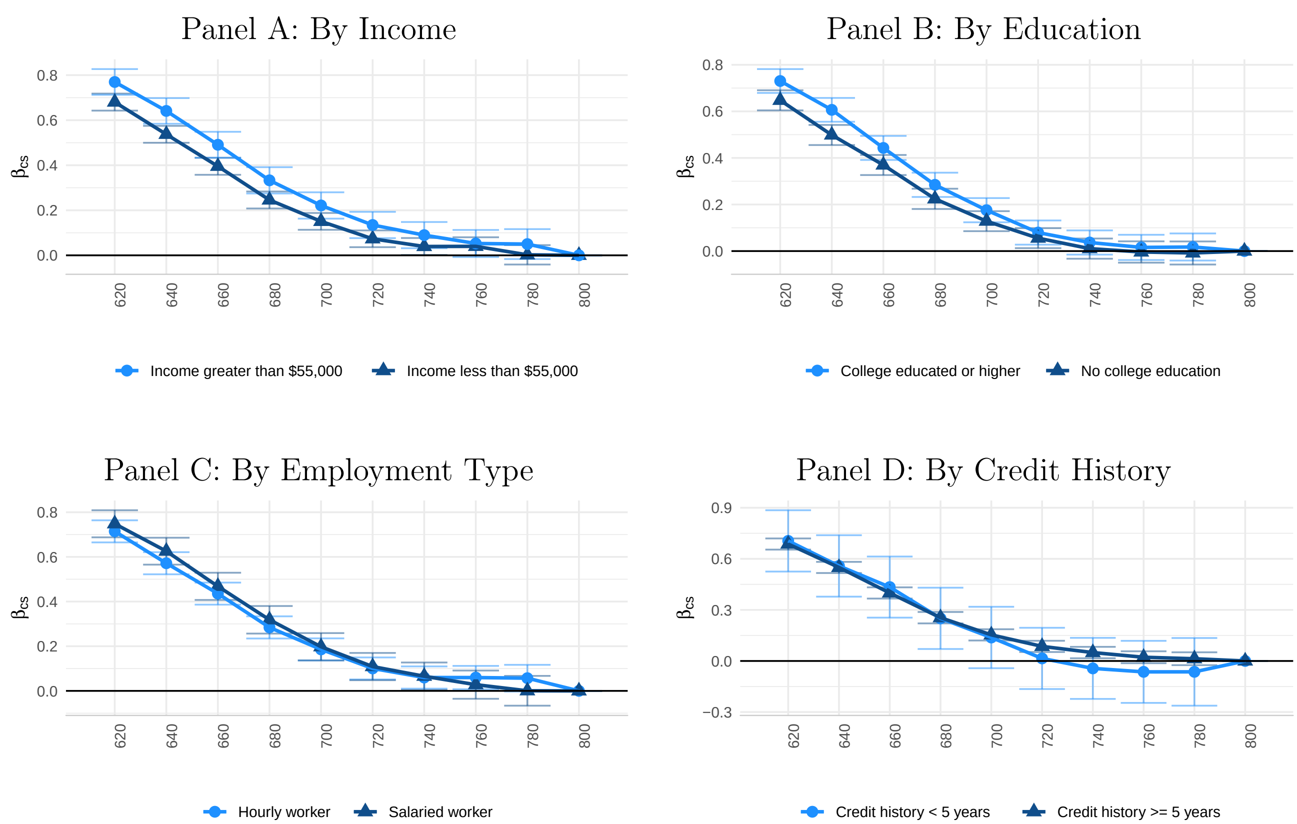


Figure 11. Rate Difference: Heterogeneity

This figure plots the coefficients and corresponding 95% confidence interval of the regression that regresses the rejection rate based on the traditional model on credit score bin dummy variables with zip code fixed effects for different sub samples. Sample splits are given in the panel titles.

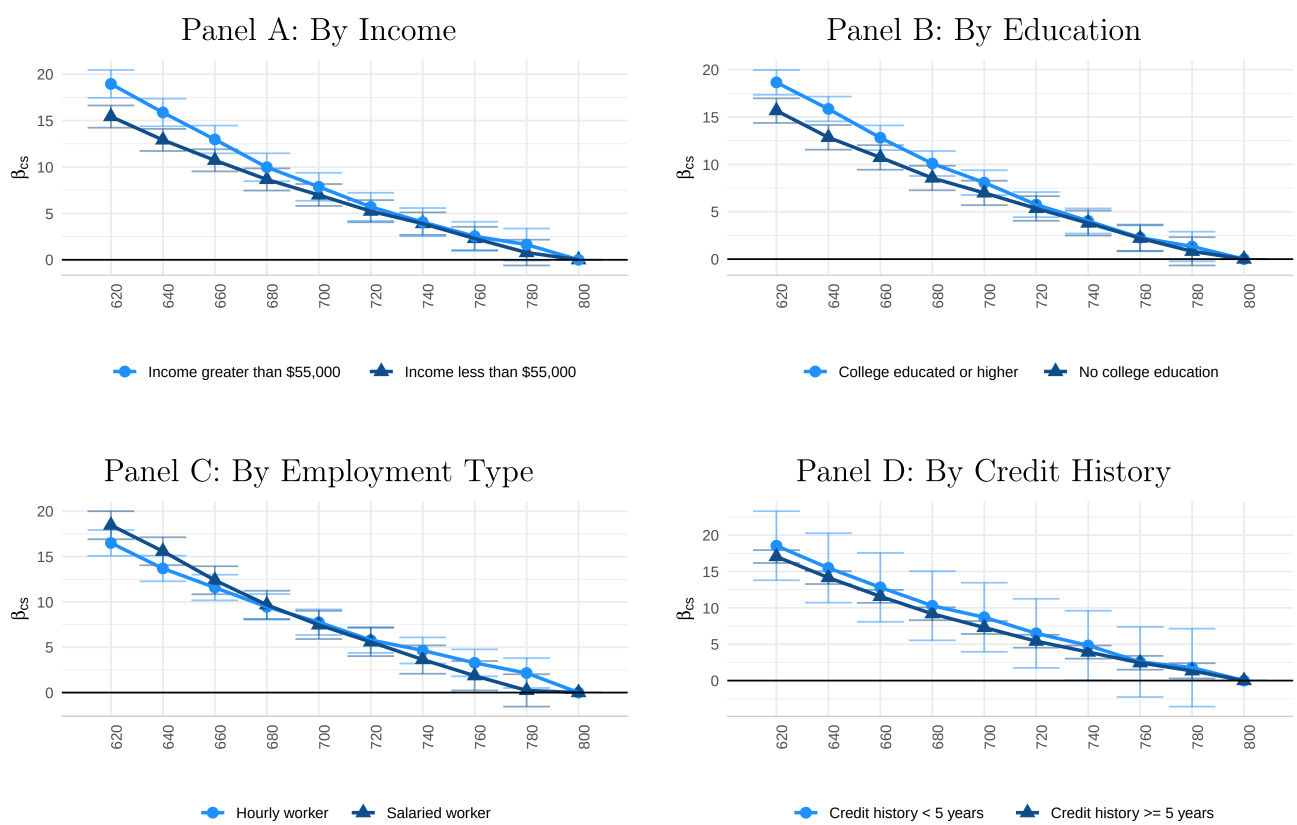
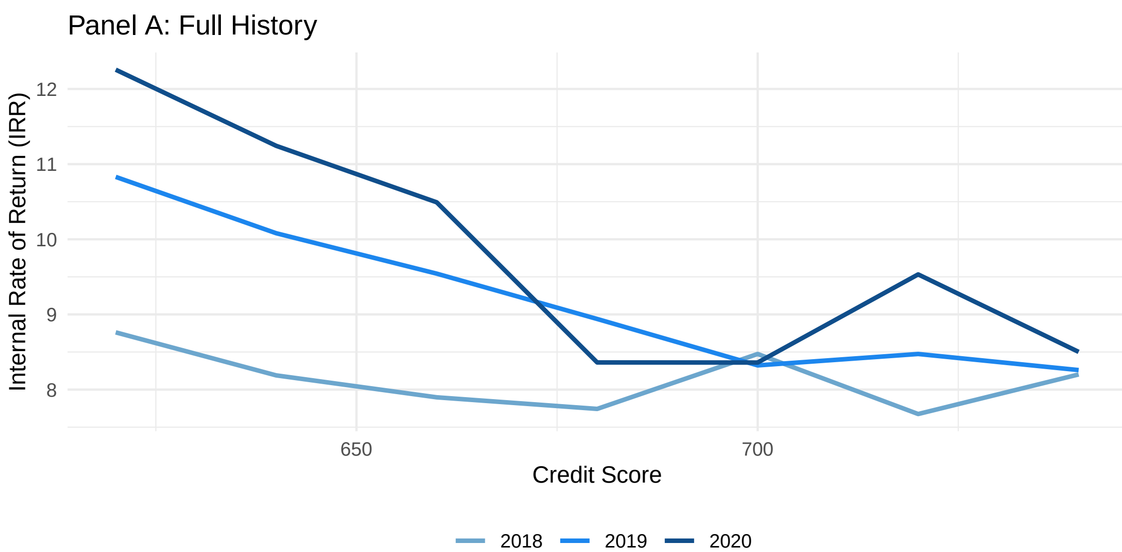


Figure 12. Internal Rate of Return: 3-Year Loans

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



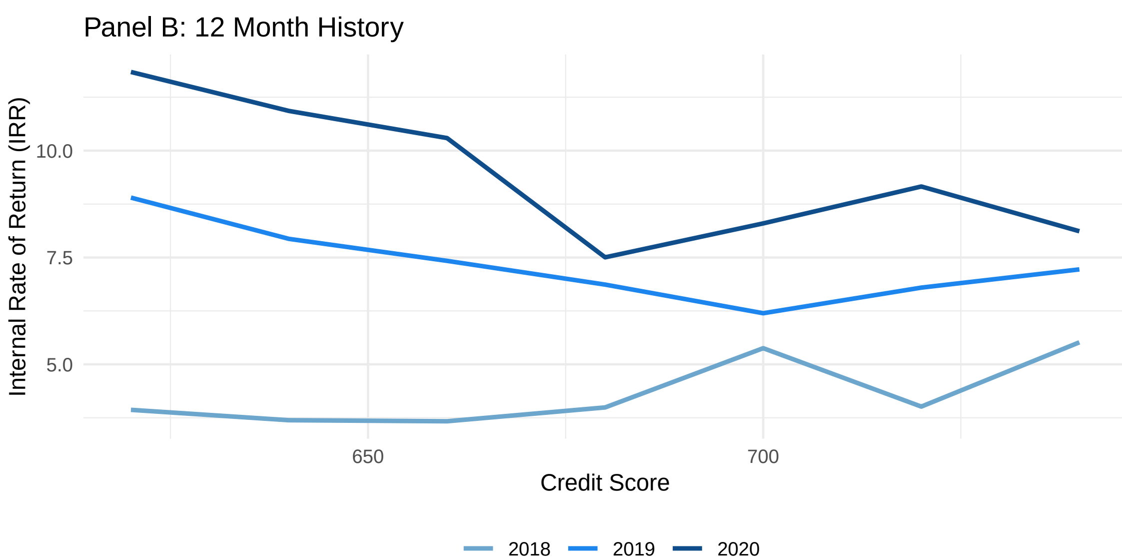


Figure 13. Discontinuity of Probability of Approval

This figure shows the discontinuity of the probability of approval for applicants based on the debt-to-income ratio at the time of the application. Panel A plots the density of debt-to-income ratio for funded applicants and applicants disqualified due to higher debt-to-income ratios. Panel B plots the fraction of applications funded (y-axis) for each debt-to-income bin (x-axis).

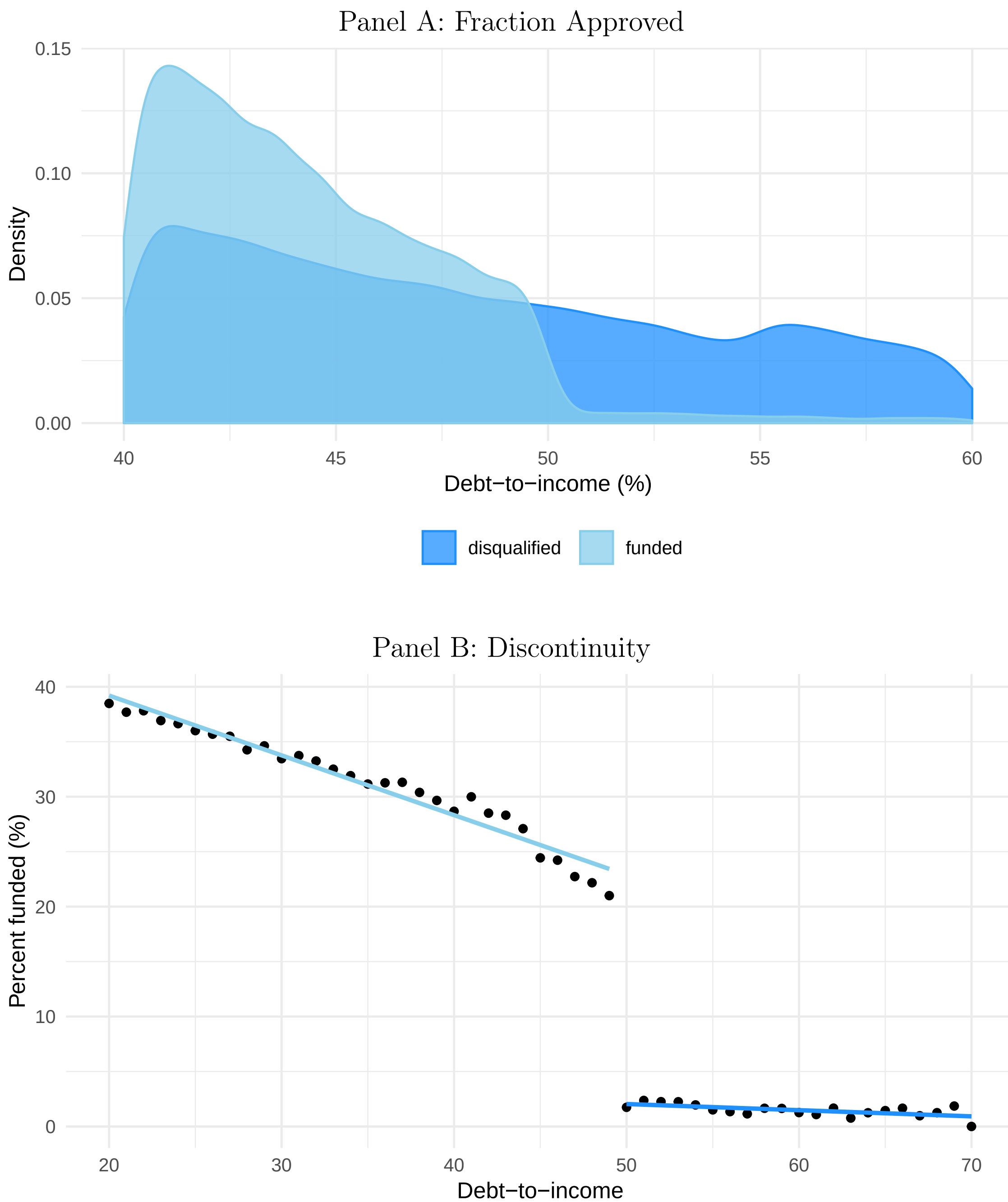


Figure 14. Discontinuity of other Variables

This figure shows the discontinuity of other variables that are commonly used to measure creditworthiness at the 50% debt-to-income cutoff.

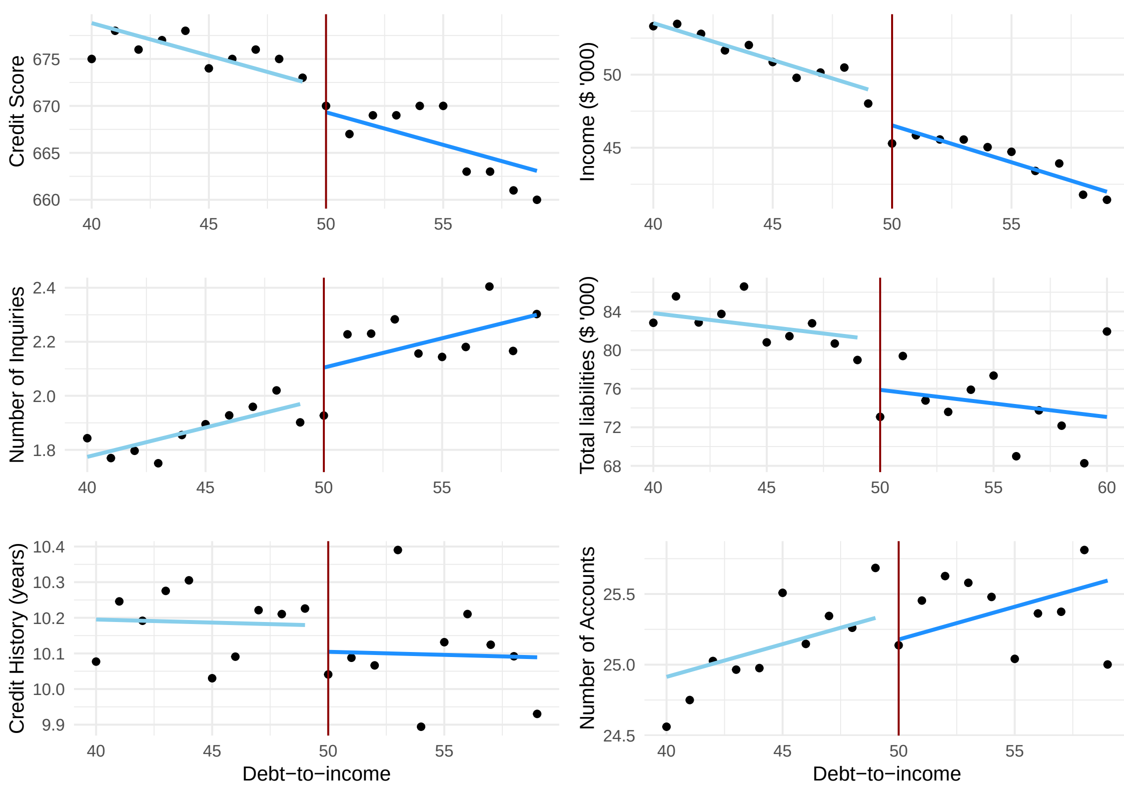


Table 1: **Descriptive Statistics: Upstart Sample**

This table presents the summary statistics of selected loan and borrower characteristics of the loans funded by Upstart. 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)
Loan amount	770,523	11,711.810	10,418.500	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
Credit Score	770,523	653.996	46.904	623	651	683
Age of the borrower	770,511	37.674	12.054	28.000	35.000	46.000
Annual income	770,523	66,958.410	173,828.100	39,000	55,000	80,000
Debt-to-income	770,516	18.237	17.820	9.460	16.400	24.980
College degree	770,523	0.445	0.497	0	0	1
Hourly worker	748,993	0.451	0.498	0.000	0.000	1.000
Years at job	748,996	5.367	7.769	1.000	3.000	7.000
Number of accounts	770,523	18.624	13.011	9	16	25
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	0.000	0.000	1.000
Used an iPhone	437,949	0.644	0.479	0.000	1.000	1.000
Credit history in years	770,523	11.014	6.910	6	10	15
Total credit balance	770,523	120,393.700	160,850.600	19,615	51,518	170,554.5
Number of inquiries	770,523	1.037	1.716	0	0	1

Table 2: **Funded vs. Disqualified Applications - Upstart Sample**

This table compares the characteristics of funded and disqualified borrowers.

	Disqualified	Funded	P-Value
Number of Obs	$3,017,377$	$783,171$	0
Credit Score	580.9	653.9	0
Age of the borrower	40.7	40.1	0
Annual income	$51,461$	$65,427$	0
Debt-to-income	19.043	18.106	0
College degree	0.234	0.449	0
Hourly worker	0.569	0.452	0
Years at job	4.410	5.222	0
Number of accounts	16.1	18.9	0
Purpose = consolidation	0.611	0.790	0
Used device type = computer	0.253	0.333	0
Used a Mac	0.223	0.283	0
Used an iPhone	0.578	0.643	0
Credit history in years	9.5	11.5	0
Total credit balance	$64,799$	$118,237$	0
Number of inquiries	2.510	1.074	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 Upstart, 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: Upstart**

This table reports the results of the regressions that examine the relationship between credit score and propensity to default using the loans originated by Upstart. The dependent variable in columns (1) and (2) is whether loan i was 90 days or more delinquent at any time. The dependent variable in columns (3) and (4) indicates whether the loan was charged-off by the lender. 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.

	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	0.001	−0.054***	0.014***	−0.018***
	(0.003)	(0.003)	(0.002)	(0.002)
log(Annual income)	−0.017***	−0.042***	−0.016***	−0.023***
	(0.002)	(0.002)	(0.001)	(0.001)
Debt-to-income	0.0001*	0.0003***	−0.0001*	0.0002***
	(0.0001)	(0.0001)	(0.00004)	(0.00005)
Age of the borrower	−0.002***	−0.001*	−0.0002	0.0003
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Age of the borrower2	0.00002***	0.00001***	0.00001**	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
log(Number of accounts)	−0.012***	−0.012***	−0.003***	−0.003***
	(0.002)	(0.002)	(0.001)	(0.001)
log(Number of inquiries)	−0.004***	0.009***	−0.003***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
log(Total liabilities)	0.004***	−0.003***	0.007***	−0.00003
	(0.001)	(0.001)	(0.001)	(0.0005)
log(Credit history)	−0.019***	−0.019***	−0.015***	−0.012***
	(0.002)	(0.002)	(0.001)	(0.001)
log(No of recently opened accounts)	0.009***	0.021***	0.007***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
log(Pct. of revolving liabilities)	0.001	−0.010***	0.011***	0.001**
	(0.001)	(0.001)	(0.001)	(0.001)
log(Pct. of mortgage liabilities)	0.0001	−0.003***	−0.0001	−0.002***
	(0.0005)	(0.0005)	(0.0003)	(0.0003)
log(Credit card utilization)	0.010***	0.003***	0.001	0.0001
	(0.001)	(0.001)	(0.001)	(0.0004)
log(Pct. trades ever delinquent)	0.003***	0.004***	−0.00003	0.001***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
log(Loan amount)	0.020***	0.029***	0.007***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)
Zip code × Year	*Y*	*Y*	*Y*	*Y*
Loan Term × Year	*Y*	*Y*	*Y*	*Y*
N	362,882	249,355	362,882	249,355
Adjusted R^2	0.013	0.018	0.011	0.012

Table 5: **Predictability of Upstart Default Probability**

This table reports the results of the regressions that examine the relationship between Upstart's estimate of probability of default and propensity to default using the loans originated by Upstart. The dependent variable in columns (1) and (2) is whether loan i was 90 days or more delinquent within three years of loan origination. The dependent variable in columns (3) and (4) indicates whether the loan was charged-off by the platform. The sample is restricted to loans that were originated prior to 2019. 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.

	Dep. Var = Delinquent		Dep. Var = Charged-off	
	Credit score < 660	Credit score >= 660	Credit score < 660	Credit score >= 660
	(1)	(2)	(3)	(4)
Upstart default probability	0.519***	0.557***	0.222***	0.274***
	(0.007)	(0.008)	(0.005)	(0.006)
Credit score/100	0.010***	−0.003	0.012***	0.008***
	(0.003)	(0.002)	(0.002)	(0.002)
log(Annual income)	0.007***	−0.007***	−0.003***	−0.003***
	(0.002)	(0.002)	(0.001)	(0.001)
Debt-to-income	−0.0003**	−0.001***	−0.0002**	−0.0001***
	(0.0001)	(0.0001)	(0.0001)	(0.00005)
Age of the borrower	−0.003***	−0.003***	−0.001**	−0.001***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Age of the borrower2	0.00003***	0.00003***	0.00001***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
log(Number of accounts)	−0.007***	−0.002	−0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
log(Number of inquiries)	0.0001	−0.003*	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
log(Total liabilities)	0.002	0.003***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.0005)
log(Credit history)	−0.001	−0.003*	−0.006***	−0.004***
	(0.002)	(0.002)	(0.001)	(0.001)
log(No of recently opened accounts)	−0.003**	−0.002*	0.001*	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
log(Pct. of revolving liabilities)	−0.005***	−0.003***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
log(Pct. of mortgage liabilities)	0.0004	−0.001***	−0.0001	−0.001***
	(0.0005)	(0.0005)	(0.0003)	(0.0003)
log(Credit card utilization)	0.001	0.00002	−0.004***	−0.002***
	(0.001)	(0.001)	(0.001)	(0.0004)
log(Pct. trades ever delinquent)	0.002***	0.002***	−0.0002*	−0.0004***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
log(Loan amount)	0.016***	0.019***	0.004***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Zip code×Year	*Y*	*Y*	*Y*	*Y*
Maturity	*Y*	*Y*	*Y*	*Y*
N	362,882	249,355	362,882	249,355
Adjusted R^2	0.080	0.079	0.067	0.059

Table 6: **Key Drivers of Upstart Default Probability**

This table reports the results of the regressions that examine the determinants of the Upstart default probability. The dependent variable is the Upstart 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
Adjusted 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 7: **Predictive Power of Credit Score and Upstart 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 (Upstart default probability)	AUC Diff
Panel A: Credit Score <660			
Full sample	51.63	63.18	11.550
Annual income <55k	51.61	61.03	9.420
Annual income >= 55k	52.81	64.32	11.510
Borrowers' age <30	51.33	61.66	10.330
Borrowers' age >= 30	52.53	60.57	8.040
Credit History (years) <10	51.68	61.12	9.440
Credit History (years) >= 10	52.62	60.8	8.180
No college education	55.15	67.83	7.520
College educated	50.00	65.09	15.090
Panel B: Credit Score >= 660			
Full sample	54.17	68.56	14.390
Annual income <55k	54.46	66.45	11.990
Annual income >= 55k	54.78	70.81	16.030
Borrowers' age <30	55.00	66.14	11.140
Borrowers' age >= 30	54.84	65.25	10.410
Credit History (years) <10	55.00	65.43	10.430
Credit History (years) >= 10	54.94	65.63	10.690
No college education	52.58	67.83	15.250
College educated	53.78	71.26	17.480

Table 8: **Traditional Model vs. Upstart default probability: Heterogeneity**
This table compares rejection rate and difference in interest rate between the traditional model and Upstart's underwriting model. In both Panels A and B, Column (1) regresses a dummy variable that takes the value of one if the loan would have been rejected using the traditional model. The dependent variable in column (2) is the difference between the interest rate the borrower would have received based on the traditional model and the actual interest rate. Regressors in Panel B include interactions of regressors in Panel A with a dummy variable that takes the value of one if the borrower's credit score is less than 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:

	Rejected by Traditional Model	Traditional APR - Upstart APR
	(1)	(2)
Credit score/100	-0.591^{***}	-0.127^{***}
	(0.003)	(0.001)
Income >= 55k	-0.068^{***}	-0.042^{***}
	(0.004)	(0.001)
Advanced degree	0.047^{***}	0.039^{***}
	(0.005)	(0.001)
College degree	0.020^{***}	0.025^{***}
	(0.003)	(0.001)
Salaried employee	0.028^{***}	0.018^{***}
	(0.004)	(0.001)
Credit history - 5 to 10 years	-0.073^{***}	-0.009^{***}
	(0.004)	(0.001)
Credit history - more than 10 years	-0.093^{***}	-0.013^{***}
	(0.004)	(0.001)
Zip code	Y	Y
N	96,969	96,969
Adjusted R^2	0.287	0.375

Panel B

	Rejected by Traditional Model	Traditional APR - Upstart APR
	(1)	(2)
Credit score < 660	0.453***	0.093***
	(0.010)	(0.002)
Income >= 55k	−0.080***	−0.037***
	(0.005)	(0.001)
Advanced degree	0.022***	0.029***
	(0.008)	(0.002)
College degree	0.004	0.023***
	(0.005)	(0.001)
Salaried employee	0.021***	0.012***
	(0.005)	(0.001)
Credit history - 5 to 10 years	−0.027***	0.005***
	(0.006)	(0.001)
Credit history - more than 10 years	−0.053***	−0.005***
	(0.006)	(0.001)
Credit score < 660 × Income >= 55k	−0.009	−0.017***
	(0.007)	(0.002)
Credit score < 660 × Advanced degree	0.042***	0.018***
	(0.010)	(0.002)
Credit score < 660 × College degree	0.037***	0.007***
	(0.007)	(0.002)
Credit score < 660 × Salaried employee	0.027***	0.016***
	(0.007)	(0.002)
Credit score < 660 × Credit history - 5 to 10 years	−0.053***	−0.015***
	(0.009)	(0.002)
Credit score < 660 × Credit history - more than 10 years	−0.051***	−0.010***
	(0.009)	(0.002)
Zip code	*Y*	*Y*
N	96,969	96,969
Adjusted R^2	0.222	0.302

Table 9: **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. Columns (1) and (2) regress the IRR ($\times$ 100) 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) limits the time period to 12 months since loan origination, for the IRR calculation. 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.

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

Table 10: **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 borrower2	−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 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: **Funding Discontinuity at 50% Debt-to-Income**

This table shows the discontinuity of probability of funding at 50% debt-to-income ratio cutoff. The dependent variable is given in the column header and the sample is restricted to applicants with debt-to-income ratio between 40% and 60%. Column (1) is the first stage regression results of the fuzzy regression discontinuity design given by equation (1). 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.

	Funded	Credit score	log(Income)	log(Number of inquiries)	log(Total liabilities)	Credit history	Number of accounts
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Debt-to-income > 50%	−0.867***	−3.338	0.120	−0.046	−0.249	−1.712	1.624
	(0.104)	(10.783)	(0.104)	(0.190)	(0.190)	(1.075)	(1.189)
Debt-to-income × > 50% Debt-to-income	0.014***	0.003	−0.003	0.001	0.005	0.028	−0.027
	(0.002)	(0.207)	(0.002)	(0.004)	(0.004)	(0.021)	(0.023)
Debt-to-income	−0.013***	−0.087	−0.009***	0.003	0.005***	0.022**	−0.003
	(0.001)	(0.107)	(0.001)	(0.002)	(0.002)	(0.010)	(0.011)
Credit score	22.256***		−2.029***	−6.855***	4.232***	140.718***	−63.246***
	(0.626)		(0.624)	(1.139)	(1.139)	(6.195)	(6.891)
log(Annual income)	0.016**	2.558***		0.104***	0.788***	1.179***	−0.537***
	(0.007)	(0.747)		(0.013)	(0.012)	(0.076)	(0.081)
Age of the borrower	−0.007***	−1.255***	0.031***	0.004*	−0.004*	0.459***	0.012
	(0.001)	(0.145)	(0.001)	(0.002)	(0.002)	(0.019)	(0.012)
Age of the borrower^2	0.00003**	0.013***	−0.0003***	−0.00003	0.00001	−0.002***	−0.0001
	(0.00001)	(0.002)	(0.00001)	(0.00003)	(0.00003)	(0.0002)	(0.0001)
log(Number of accounts)	0.040***	−6.796***	0.238***	0.191***	0.652***	3.262***	26.432***
	(0.007)	(0.764)	(0.007)	(0.013)	(0.012)	(0.067)	(0.142)
log(Number of inquiries)	−0.029***	−4.953***	0.031***		−0.065***	−0.680***	0.392***
	(0.004)	(0.408)	(0.004)		(0.007)	(0.037)	(0.048)
Total liabilities	7.099***	1,571.420***	45.917***	−9.669***		207.267***	165.750***
	(0.791)	(86.835)	(0.721)	(1.439)		(8.322)	(9.668)
Credit history	8.930***	1,963.317***	9.305***	−22.435***	28.703***		
	(0.773)	(80.555)	(0.768)	(1.399)	(1.398)		
log(No of recently opened accounts)	−0.092***	−4.272***	0.042***	0.150***	−0.381***	−1.242***	−2.576***
	(0.004)	(0.460)	(0.004)	(0.008)	(0.008)	(0.043)	(0.048)
log(Pct. of revolving liabilities)	−0.006*	5.887***	0.057***	−0.109***	−0.485***	0.542***	0.211***
	(0.003)	(0.443)	(0.003)	(0.006)	(0.005)	(0.035)	(0.041)
log(Pct. of mortgage liabilities)	−0.001	3.198***	0.050***	−0.001	−0.173***	0.246***	−0.126***
	(0.004)	(0.435)	(0.004)	(0.007)	(0.007)	(0.038)	(0.041)
Credit card utilization	−3.605***	−3,710.047***	0.288	15.268***	29.000***	−2.255	−14.455**
	(0.636)	(67.009)	(0.634)	(1.153)	(1.139)	(6.137)	(7.341)
log(Pct. trades ever delinquent)	−0.005***	−4.615***	−0.006***	0.012***	−0.026***	0.125***	−0.042***
	(0.001)	(0.081)	(0.001)	(0.001)	(0.001)	(0.008)	(0.008)
Zip code×Year	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*
N	42,660	42,660	42,660	42,660	42,660	42,660	42,660
Adjusted R^2	0.442	0.446	0.531	0.152	0.664	0.511	0.861

Table 13: **The Effects of Credit Access: Regression Discontinuity Design**

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.

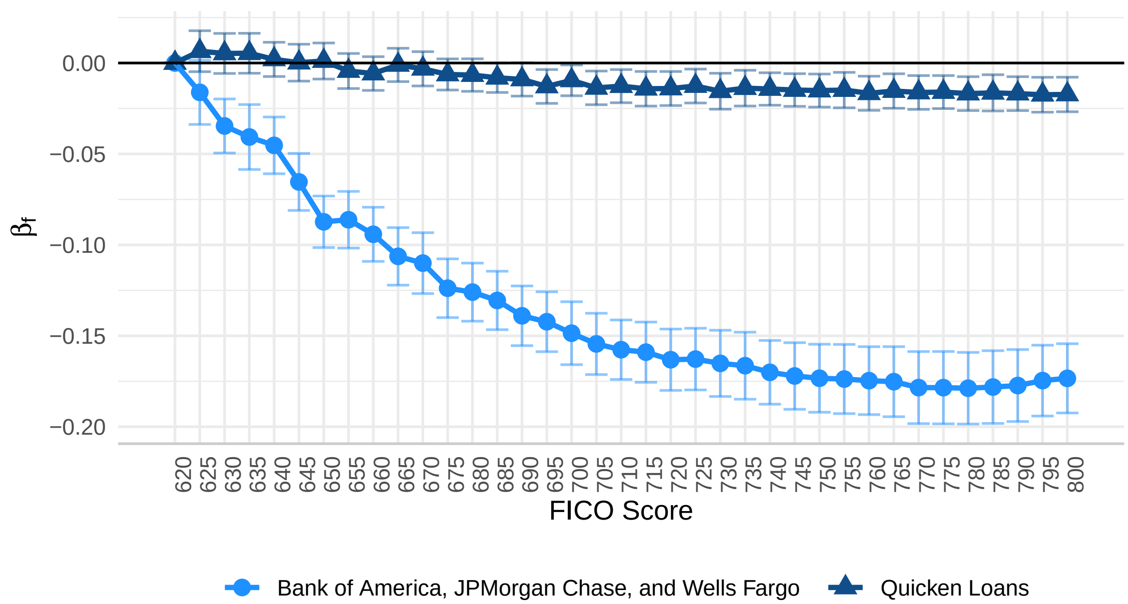
	Credit score < 660			Credit score >= 660		
	Credit card delinq	Credit score change	Mortgage	Credit card delinq	Credit score change	Mortgage
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{Funded}$	-0.198^{*}	0.087^{***}	0.134^{*}	-0.074	0.012	0.010
	(0.102)	(0.026)	(0.072)	(0.074)	(0.015)	(0.087)
Debt-to-income	-0.003^{*}	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

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

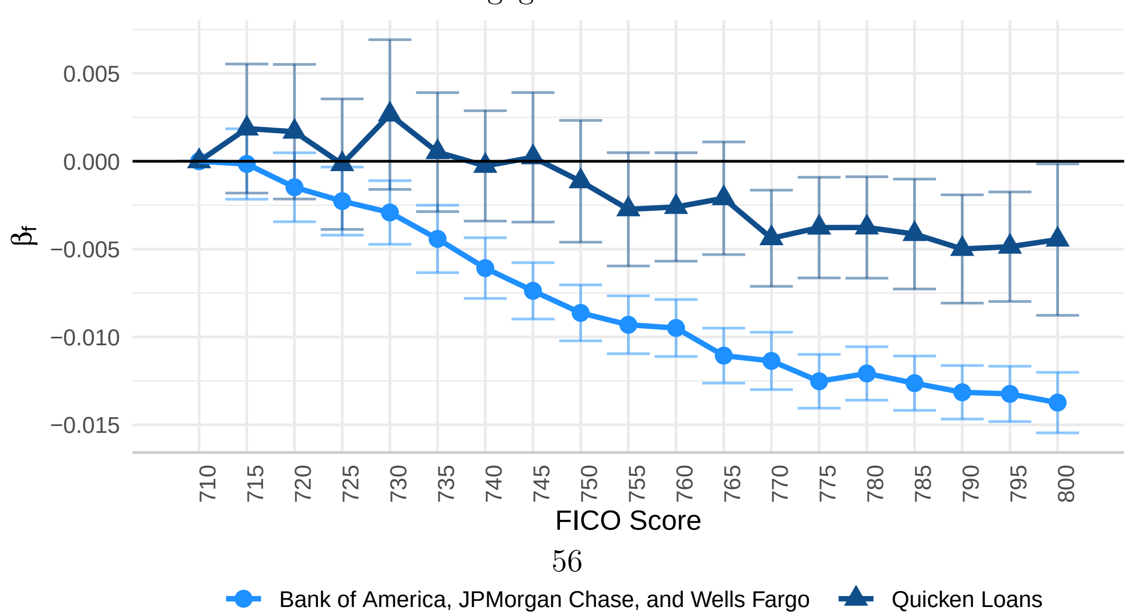


Figure A2. FICO Score Distribution: Quicken Loans vs. Large Banks

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

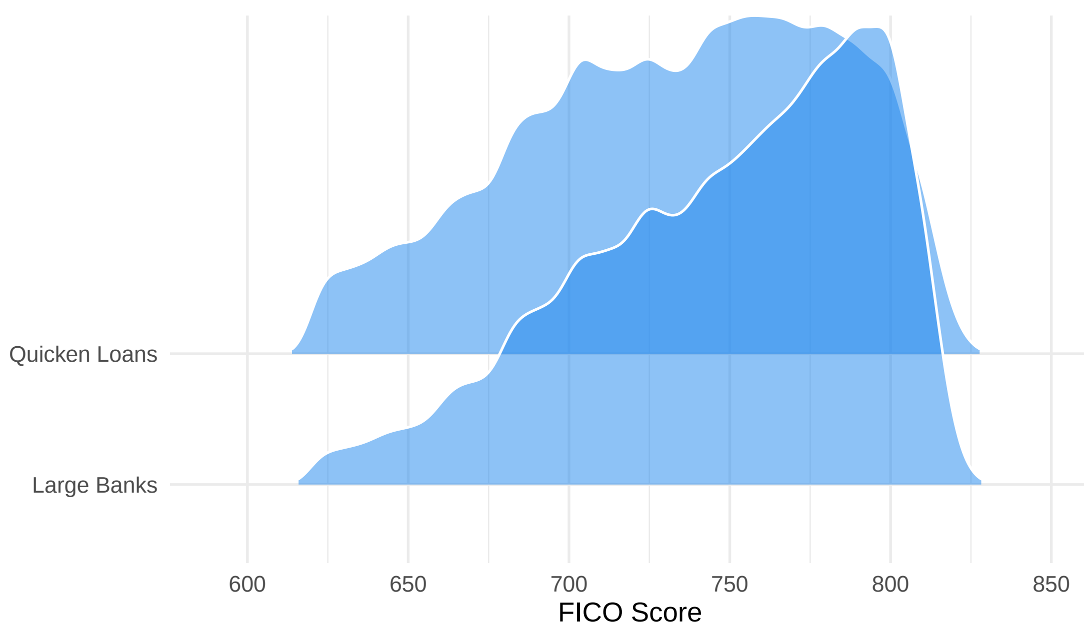


Table A1: **Descriptive Statistics - Mortgage Samples**

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 (1)	FICO < 725 (2)	725 < FICO < 750 (3)	750 < FICO < 775 (4)	FICO >775 (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