THE OHIO STATE UNIVERSITY
FISHER COLLEGE OF BUSINESS

Fisher College of Business
Working Paper Series

Charles A. Dice Center for Research in Financial Economics

Borrowers in the Shadows: The Promise and Pitfalls of Alternative Credit Data

Mark Jansen,
University of Utah

Sam Kruger,
University of Texas at Austin

Gonzalo Maturana,
Emory University

Amin Shams,
The Ohio State University

Dice Center WP 2025-13
Fisher College of Business WP 2025-03-013

Borrowers in the Shadows:

The Promise and Pitfalls of Alternative Credit Data

Mark Jansen[*] Sam Kruger[†] Gonzalo Maturana[‡] Amin Shams[§]

May 2025

ABSTRACT

More than 45 million U.S. adults lack traditional credit histories, creating a gap that alternative financial service data, such as payday lending records, could potentially fill. Using the staggered adoption of the largest alternative credit database, we examine the data's impact on automotive lenders in the subprime auto loan market. Alternative credit scores predict loan performance, leading lenders to offer better loan terms to higher-scoring borrowers. However, a history of using alternative financial services, even with relatively high alternative credit scores, comes with significant downsides: borrowers with payday loan histories experience higher delinquency rates, face higher interest rates, and reduced loan origination rates after the adoption of alternative credit data. A flexible machine learning model indicates that only 3.28% of alternative financial service users possess sufficiently strong credit histories to offset the stigma of using these services. Consequently, use of alternative credit data limits credit availability and raises traditional loan costs for most users of alternative financial services. Alternative financial services are used more frequently in lower-income areas and communities with higher shares of black residents, raising concerns that the adoption of alternative credit data may have disproportionate negative impacts on these populations. Our results contribute to the policy debate on credit data, consumer privacy, and financial inclusion.

JEL Classification: D12, D14, G23, G51.

[*]University of Utah, Eccles School of Business. E-mail:mark.jansen@eccles.utah.edu

[†]University of Texas at Austin, McCombs School of Business. E-mail: sam.kruger@mccombs.utexas.edu

[‡]Emory University, Goizueta Business School. E-mail: gonzalo.maturana@emory.edu

[§]Ohio State University, Fisher College of Business. E-mail: shams.22@osu.edu

Does alternative credit data enable financially underserved households to establish credit with traditional lenders? The United States has one of the most sophisticated and developed financial services industries in the world, but a significant fraction of Americans still lack access to traditional credit, such as bank loans or credit cards. For these financially underserved consumers, alternative financial services (AFS), such as payday lending, have become a significant source of financing. One of the greatest barriers to financial inclusion is the lack of available credit information about non-prime borrowers (Experian, 2022). In particular, more than 45 million Americans are unscorable because their financial history lacks sufficient detail or is missing altogether (Consumer Financial Protection Bureau, 2017). Rapid growth in the AFS industry has spurred parallel growth in alternative credit data. This paper studies how the adoption of alternative credit data affects automotive lending decisions and consumer access to credit.

Historically, credit transactions for AFS products were not tracked by the three major consumer credit bureaus (i.e., Experian, Equifax, and TransUnion). The main reason for this lack of coverage is that AFS lenders do not report to them. AFS lenders do not rely on traditional credit scores (e.g., FICO and VantageScore) because they have little predictive power in explaining defaults in the AFS space.[1] Since AFS borrowers often access short-term credit from multiple AFS sources, the typical three-week reporting time for a new loan to appear on a credit report is insufficient. This contrasts with the real-time credit updates used to assess AFS borrowers.

With the rise in demand for AFS products came the need for alternative data to evaluate loan requests by non-prime applicants. Usually, these data are constructed from three sources: (1) credit data from AFS lenders, (2) information regarding the timeliness of household payments such as rent, cell phone bills, utility bills, and insurance, and (3) public records and social media. Our focus in this paper is on the first of these data sources, credit data from AFS lenders. These data function similarly to traditional credit data and are used by AFS lenders to assess credit risk. In

[1] Similarly, Demyanyk and Van Hemert (2011) show that traditional credit scores had significantly less predictive power to explain mortgage defaults for subprime borrowers than for prime borrowers during the Great Recession.

the last 10 years, traditional lenders have begun to acquire these data to help assess credit risk for traditional loan products such as mortgages and automobile loans.[2]

AFS data has been welcomed by many policymakers as a way for people with limited traditional credit history to build a credit history that can be used to support access to traditional credit from lenders (U.S. Government Accountability Office, 2021). Moreover, many advocates for AFS data argue that these data are especially beneficial for low-income and minority borrowers, since these are the groups that have traditionally been excluded from the credit mainstream (Turner et al., 2006). Although AFS data offer potential benefits, using alternative credit sources such as payday loans can also create a stigma that could persist even with a positive repayment history. By revealing the usage of alternative credit products, AFS data may hurt more AFS users than it helps. As alternative data are incorporated into more lending decisions, understanding the implications of using such data becomes critically important.

We use proprietary data from one of the largest alternative credit data providers in the United States. These data include more than 250 million consumer loan transactions and are used for predictive modeling and risk management by peer-to-peer lenders, AFS lenders, and automotive lenders from which the data are primarily drawn. Most loan transactions found in the data are generated by individuals with non-prime credit scores. This setting serves as an ideal laboratory to answer our three main research questions: (1) To what extent do alternative credit data predict credit risk for automobile loans, and how are the data used by lenders once they become available? (2) How are financially underserved consumers affected when lenders use alternative data in their credit decisions? (3) Are some consumer segments affected differently by the adoption of alternative credit scoring? Comprehensive and conclusive answers to these questions are important for the financial services industry, academics, and the general public.

Despite the rapid adoption of alternative credit data in lending, little is empirically known about

[2]For example, Fannie Mae incorporated rental payments (a form of alternative data) into its automated underwriting system in September 2021 (U.S. Government Accountability Office, 2021).

the direction and magnitude of the impact of alternative credit data on lenders and consumers. The theory that motivates the use of credit data is decidedly mixed. For example, classic information theory informs us that more data leads to more informative decisions and less adverse selection, achieving more efficient outcomes for both lenders and consumers (Athreya et al., 2012; Livshits et al., 2016). In contrast, moral hazard is reduced by partial sharing of information rather than a policy of full sharing of information (Vercammen, 1995; Padillaa and Pagano, 2000). Furthermore, the segmentation between the three major credit bureaus and alternative credit data providers may hinder non-prime consumers from transitioning to traditional, lower-cost financial services, limiting their ability to leverage positive credit outcomes in the alternative credit market as a pathway to mainstream credit access.

We find that there is significant predictive power in alternative credit scores for auto loan performance, demonstrating that higher alternative credit scores are associated with significantly lower delinquency probabilities, even after controlling for traditional credit scores, loan characteristics, and borrower demographics. Moreover, this relationship holds when we separate our sample of borrowers into terciles of traditional credit scores, indicating that alternative credit data provide incremental predictive value beyond traditional credit models. These findings underscore the utility of alternative credit scores in improving risk assessment for lenders, particularly for borrowers with low or moderate traditional credit scores.

Next, we investigate the effect of alternative credit data on consumers. We show that lenders' use of alternative credit data significantly impacts loan terms. Using a difference-in-differences framework, we examine changes in auto loan pricing when lenders use alternative credit data. The treatment group comprises lenders who adopted alternative credit data, whereas the control group comprises lenders who have not used the data. A critical feature of the data is that when lenders sign up for the service with the data provider, they are required to provide previous loan data. This allows us to measure portfolio changes around the adoption of the new credit data. After accessing alternative credit data, lenders decrease interest rates for borrowers with high alternative credit scores. In particular, a one-standard-deviation increase in alternative credit score reduces interest rates by 14.7 basis points (bp), even when accounting for traditional credit scores. Focusing on the post-adoption period, in which we observe both loans and loan applications, higher alternative credit scores are also associated with higher loan origination probability, suggesting that alternative credit data impacts both loan approval and loan pricing. These effects are particularly pronounced for borrowers in the highest VantageScore tercile, suggesting that alternative credit scores complement traditional measures to aid in credit decisions.

Our results also reveal the downside of being included in alternative credit data. On average, borrowers with a history of using alternative financial products, as indicated by a "hit" in the alternative data, have higher delinquency rates. Because using alternative credit products is a negative signal of credit quality, these borrowers face higher interest rates and lower loan origination probabilities after lenders adopt the alternative credit data compared to individuals without a hit in the data. Although the alternative score partially offsets this negative signal by improving loan approval probabilities and reducing interest rates for higher scores, the overall effect of being in the alternative credit data remains significant, and having a hit in the data is generally associated with higher interest rates and a lower probability of loan origination, even for alternative credit users with the highest possible alternative credit scores.

Individuals who lack traditional credit scores (also known as "ghosts") represent a significant segment of the population with limited access to affordable credit, making them highly dependent on alternative financial services. Providing more data on ghosts is crucial to improving lenders ability to assess their creditworthiness, expand financial inclusion, and reduce borrowing costs for this underserved group. For ghosts, inclusion in alternative credit data has similar effects. Ghosts with a hit in alternative data experience higher delinquency rates and face higher borrowing costs than those without a hit. High alternative scores do not significantly offset this effect. Overall, there is no evidence that higher alternative credit scores reduce interest rates even for borrowers who are ghosts in traditional credit data.

These findings highlight the dual role of alternative credit data: While they enhance lenders ability to predict credit risk and improve credit access for high-scoring borrowers, they also signal increased risk, leading to adverse outcomes for borrowers included in the data. For ghosts, alternative credit data offer some benefit in terms of loan access, but the lack of impact on borrowing costs suggests limited advantages for this group.

Does alternative credit data allow at least some borrowers to build a positive credit history that can help them obtain traditional credit? To assess this question, we estimate flexible machine learning models using gradient boosting to assess the impact of alternative credit data on delinquency and interest rates. Consistent with our previous results, most loan applicants receive higher interest rates compared to the counterfactual estimates if they had not used alternative credit. Although some AFS users have high enough alternative credit scores to overcome this stigma, this is only true for 3.28% of borrowers we have used AFS. A larger share of AFS users (15.48%) have lower estimated delinquency probabilities as a result of their AFS usage. In general, the usage of alternative credit products is a negative signal that is difficult to overcome.

We also consider the extensive margin and investigate whether the adoption of alternative credit data affects the likelihood of loan origination. The evidence suggests that alternative credit data generally expands credit access for borrowers without a hit in the data with a flat to slightly declining number of loans for borrowers with a history of using alternative credit products.

Finally, we investigate whether lenders' use of a more data-driven underwriting process inadvertently has a disparate impact on certain consumer groups. Additional credit data can lead to more impartial and less discriminatory decision-making based on hard information, which leaves less discretion to loan officers. However, the adoption of alternative credit data adds a negative signal for AFS users. Because AFS usage is not randomly distributed in the population, it could impact some consumer groups more than others. In particular, we find that zip codes with lower incomes and higher minority population shares tend to have higher hit rates and lower credit scores in the alternative credit data. This result is particularly pronounced in zip codes with high black population shares. As a result, the adoption of alternative credit data is likely to reduce credit supply and increase interest rates for low-income black borrowers. To the extent that policymakers are concerned about the distributional impacts of changing lending technologies, this could be a concern.

A potential challenge to our identification strategy is that before the adoption of alternative credit data, we only observe approved loans, not loan applications. However, since auto dealerships submitting applications are unaware of the data adoption, it is reasonable to assume that the applicant pool remains largely consistent within a six-month window around the data adoption.

Our paper provides policymakers with important information relevant to regulating the use and reporting of credit information. This is particularly important because technological advances in the last decade have dramatically expanded the amount of information collected about consumers and the ability of financial institutions to process it. For example, this has had important implications in terms of dispersion in access to credit (e.g., Fuster et al., 2022; Bartlett et al., 2022).

Our paper contributes to a growing body of literature studying the implications of using alternative credit data in lending decisions. Di Maggio et al. (2022) show that fintech models that integrate alternative data approve borrowers with low traditional credit scores but a low propensity to default, improving economic outcomes for these borrowers and returns for fintech lenders. Agarwal et al. (2023) examine data from a large fintech lender in India and emphasize the role of digital footprint data in improving credit access for underserved borrowers. Similarly, Berg et al. (2020) analyze purchases from a German e-commerce company and find that the digital footprint that individuals leave online after registering on a website predicts consumer default and complements credit bureau information. We add to this research by showing that alternative credit scores complement traditional credit scores in aiding lending decisions in the U.S. auto market. Importantly, we show that alternative credit data do not necessarily increase access to credit for all borrowers, as indicated by the lower access to credit of borrowers with a history of using alternative financial products.

Recent research reveals a nuanced interaction between alternative and mainstream credit markets in the high-cost credit space. Fonseca (2023) find that restricting debt collection reduces credit access and increases payday loan borrowing, suggesting that regulatory restrictions may inadvertently push consumers towards higher-cost products that may subsequently affect their creditworthiness.[3] Liberman et al. (2021) discuss how high-cost debt, such as payday loans, can create a self-reinforcing stigma that negatively impacts future access to credit, even if they do not default on their high-cost loans. Consistent with this idea, we show that the use of payday loans acts as a negative signal to lenders, who restrict credit to these borrowers regardless of whether they repaid their payday loans. Our results are also consistent with the broader literature that highlights the importance of asymmetric information in credit markets (e.g., Stiglitz and Weiss, 1981).

Finally, our paper contributes to the recent and growing literature on auto lending. A major area of focus in this literature has been understanding the lending decisions of auto lenders. For example, Melzer and Schroeder (2017) find that auto lenders change loan terms to stay below the usury limit and Jansen et al. (2022) find that dealership sales personnel face incentives to shift weak borrowers from used to new cars, increasing loan defaults.[4] Jansen et al. (2024) examine finance and vehicle margins and find that deep subprime borrowers with the lowest credit scores generally fare better than subprime borrowers.[5] Closest to our paper, Einav et al. (2013) study

[3] Allcott et al. (2022) show, however, that barring payday lending would negatively impact welfare compared to the existing regulation and Miller and Soo (2020) show that improving access to traditional credit does not reduce payday loan use among subprime borrowers.

[4] Additional evidence that sales representatives have incentives that are not aligned with the interests of consumers can be found in Pierce (2012).

[5] Grunewald et al. (2023) show that auto borrowers are more sensitive to vehicle charges than finance charges. In addition, subprime auto borrowers are sensitive to down payments (Einav et al., 2012) and loan maturity (Argyle et al., 2020)

how automated credit scoring in automotive finance increases profits for lenders. They show that lender profits improve by screening high-risk borrowers more effectively and marketing to low-risk borrowers.

1 Data and Sample Selection

The total market for auto loans increased to a record $1.626 trillion in Q1 of 2024 (Federal Reserve Bank of New York, August 06, 2024). The market for auto loans to non-prime borrowers, which is the setting for our study, constitutes approximately 40% of this total (Zabritski, 2020). As the market for used cars has grown, so has the number of auto lenders competing to finance non-prime customers. Approximately 65,000 financial institutions are competing to finance vehicle purchases to non-prime consumers (Baines and Courchane, 2014). The auto finance industry is relatively fragmented with a combined market share of 37.7% for the top 10 originators as of 2013 (the starting point of our analysis), compared to a combined top-10 market share of 52% in the mortgage industry (Baines and Courchane, 2014). The Herfindahl-Hirschman index (HHI) for used car origination is less than 100.

Our analysis uses alternative credit data that were adopted by subprime automobile lenders on a staggered basis. The alternative credit data come from a company that collects and maintains individual credit information and provides reports and scores to lenders and other institutions to assess creditworthiness and financial behavior. In total, the database contains more than 250 million loan records from more than 20 million US adults. Data are sourced from alternative financial service providers, such as payday lenders, title lenders, and pawn shops. The database includes information on loan inquiries, employment data, wages, bank account information, and loan payment data. Non-prime consumers found in our data use on average 28 different types of financial services to spend, save, borrow, and plan during their financial lives (Network, 2017). In aggregate, the fee and interest revenue for these alternative financial products was $415 billion in 2023 (Gdalman et al., 2024). Significantly, over one quarter of all fees and interest from financial services were generated by subprime auto loans ($108.8 billion), which is the nexus of our study.[6] Our analysis primarily focuses on two variables from the alternative credit data: (1) an indicator variable for whether or not a person is in the alternative credit data (i.e., hit) and (2) the individual's alternative credit score, which the data provider calculates to predict loan default probabilities using all of the underlying data.

Our auto loan data come from individual loans and loan applications from six auto lenders serving mostly subprime borrowers that started using alternative credit data for loan underwriting on a staggered basis between 2013 and 2016. Because subprime auto lenders often serve borrowers with limited or poor credit histories, they potentially have more to gain by using alternative credit data to assess risk, making this a natural setting to study its impact. The use of alternative credit data has grown significantly in this market over the past 10 years. The data provide a snapshot of consumer finances at the time of application between 2014 and 2018 and again in June 2022. For each individual, we observe auto loan application information, such as the prospective borrowers' monthly income, length of employment, age, and zip code. For loans that are originated, we observe interest rates and other loan terms such as loan size and maturity, as well as loan performance data such as subsequent delinquencies. The data also include traditional credit scores (VantageScore 4.0, hereafter Vantage score) and alternative credit scores, both as of the date of the loan application and as of June 2022.

The lenders we study began using alternative credit data between 2013 and 2016. Before that, they assessed prospective borrowers using the traditional credit score, income, employment, scheduled debt payments, and downpayment capacity. Loan origination decisions were made based on this reported information plus any "soft information" they could gather at the time of the loan re-

[6]This excludes $9.6 billion of fees and interest from Buy Here, Pay Here (BHPH) auto loans, which are financed in-house by automobile dealers at high interest rates for subprime or no-credit borrowers.

quest. This approach to lending was typical before the advent of alternative data to assess prospective borrowers in subprime automotive finance. We observe all originated loans and most loan applications for each lender after they adopt alternative credit data until the end of the study in 2017. Before adoption, we observe loans that were active as of the date the lender adopted alternative credit data for five of the six lenders.

Figure 1 illustrates the geographic coverage of the data for auto loan applications in our sample. The shading indicates the density of loan applications, highlighting areas with varying levels of loan activity. The figure shows that the sample includes borrowers and loan applicants from a wide range of geographic areas.

Table I presents summary statistics for the sample, covering key metrics related to credit scores, loan characteristics, and demographic factors. The data include 7.2 million loan applications. These loan applications consist of originated loans before and after adoption of alternative credit data plus loan applications after adoption for which the lender queried the alternative data. Approximately 27% of the individuals in the sample are marked with a recorded hit (*Hit* = 1), indicating prior use of alternative financial services such as payday loans. The alternative credit score (*Alt Score*) has a mean of 703.38 with a standard deviation of 30.59 and ranges from 500 to 838. The Vantage score, measured for approximately 7 million observations, shows a mean of 555.34 and a standard deviation of 63.20, with values ranging from 300 to 850. The sample includes 243,000 individuals who do not have a Vantage score.

Most of our analysis focuses on originated loans, which include 536 thousand loan-level observations. Usage of alternative credit for borrowers in the loan sample is similar to the application sample, with a hit rate in the alternative data of 22.9%. Average alternative scores and Vantage scores are also similar to the overall sample of applications. The average loan interest rate (*Rate*) is 22.34% with a standard deviation of 5.59 ppt. The average loan amount is $16,263 with a standard deviation of $6,649. The loan terms have a mean of 66.69 months. The mean delinquency rate (*Delinqueny*) is 0.38, indicating that 38% of the borrowers were at some point delinquent by at least 30 days past due on a payment.

Based on matching the zip codes of loan application to census data, median household income at the zip-code level (*ZipIncome*) has a mean of $68,213 with a standard deviation of $29,458 and median zip code income levels in the data range from $5,704 to $441,278. The table also includes several other demographic variables at the zip code level: the proportion of black individuals in the zip codes in the sample is 0.092 on average (*Black*), and in the zip codes in the sample, an average of 7.3% of individuals live below the poverty line (*Poverty100*) and 13.5% receive food stamps (*Foodstamp*). Payday loan shops (*PDshops*) are prevalent in the samples regions, with an average of 0.42 locations per zip code.

Finally, the data include indicators for different lenders. The majority of loans (67%) come from Lender A, while smaller proportions are associated with Lenders B (17%), C (6%), D (5%), E (4%), and F (1%).[7] These summary statistics provide a comprehensive overview of the dataset, illustrating the distribution of credit scores, loan characteristics, and borrower demographics.

Figure 2 shows the distribution of alternative credit scores for auto loan applicants with hits in alternative credit data. The majority of applicants have scores clustered around 700. The distribution is relatively narrow, with a concentration of scores in the upper 600s to lower 700s and a sharp peak at a score of 708, which is frequently associated with AFS users who submitted one application but did not receive an AFS loan. A vertical red dashed line is plotted at 744, indicating the alternative credit score that the data assigns to individuals who were not previously present in the alternative database, that is, individuals with no prior interactions with any AFS lenders. Our analysis treats these individuals separately and does not use the score, but we include it in the figure as a reference. A clear feature of the data is that borrowers who enter the database by transacting with an AFS lender typically have lower alternative scores than borrowers without an AFS credit

[7] Table A.1 presents the average hit, Alt Score, Vantage Score, and interest rate for each of the six lenders, and Tables A.2 to A.4 repeat our main results excluding Lender A.

history.

Figure 3 examines the relation between alternative credit data and traditional Vantage scores. The binned scatter plot shows how alternative scores and hits vary with the Vantage score. The yellow squares in the plot indicate the probability of having a hit in the alternative credit data within each Vantage score bin. In the full sample, the average hit probability is 26.47%. Loan applicants without a Vantage score have a hit probability of 9.24%. For applicants with a Vantage score, hit probability decreases with Vantage score, consistent with the usage of alternative credit products being the highest for deep subprime borrowers. In particular, borrowers in the lowest tercile of Vantage scores (i.e., with scores below 525), have a hit probability of 38.48%, compared to a hit probability of 27.71% for borrowers with Vantage scores in the 525 to 574 range and a hit probability of 16.37% for borrowers in the top decile of the Vantage score spectrum with scores 575 and above. The dots and box plots in Figure 3 plot the distribution of alternative credit scores within each Vantage score bin, conditional on having a hit in the AFS data. In general, alternative scores are positively correlated with Vantage scores, but the relation is somewhat muted. In particular, the median alternative score is fairly stable across the Vantage score bins and there is a large variation in alternative scores within the Vantage score bins, indicating that the alternative data contain significant information that is not captured by Vantage scores.

2 Main Results

In this section, we assess how well alternative credit scores predict auto loan defaults and to what extent lenders use alternative credit scores to set interest rates. We start by analyzing alternative credit scores conditional on being in the alternative credit data and then add borrowers who are not in the alternative credit data to assess the potential negative signal of being in the data.

2.1 Does Alternative Credit Data Predict Auto Loan Performance?

For alternative credit data to be useful to lenders, they must demonstrate predictive value to assess loan default probabilities, even after controlling for other observables such as income and traditional credit scores. We start by assessing whether or not this is the case using panel data on auto loan performance. The loan data includes loans that were originated both before and after the adoption of the alternative credit data. For this set of results, we limit the sample to borrowers with a hit in the alternative credit data.

Table II estimates regressions of the form:

$$\text{Delinquency}_{i,j,l,t} = \beta_1 \text{AltScore}_{i,j,l,t} + \beta_2 \text{Vantage4}_{i,j,l,t} + \beta_3 \text{LoanAmount}_{i,j,l,t} + \beta_4 \text{LoanTerm}_{i,j,l,t}$$

$$+ \beta_5 \text{ZipIncome}_{j,t} + \alpha_j + \delta_{l,t} + \varepsilon_{i,j,l,t},$$

where i indexes the loan, j indexes the county, l indexes the lender, and t indexes the application month. The dependent variable is an indicator for whether a loan becomes 30-day delinquent. All regressions in Table II include lender-by-month fixed effects to control for differences in delinquency across lenders and over time, as well as county fixed effects. Standard errors are double-clustered by state and loan application month. Column (1) regresses delinquency on alternative credit score with no control variables except these fixed effects. The estimated coefficient of -0.0259 is highly significant and indicates that a one-standard-deviation increase in alternative credit score is associated with a 2.6 ppt decrease in delinquency probability, which is equivalent to a decrease of 5.7% relative to the mean delinquency rate of 45.5%.

Is this new information incremental to other data already available to lenders? In particular, do alternative credit scores have additional predictive power for delinquency after controlling for traditional credit scores? To answer this question, column (2) includes both alternative and traditional credit scores. In this regression, both the alternative score and the Vantage score are included as independent variables to assess whether the alternative score adds predictive value beyond traditional credit scores. The coefficient for the alternative score remains negative and statistically significant with a magnitude of -0.0228, which is only slightly muted compared to column (1). This indicates that while traditional credit scores (measured by the Vantage score) explain part of the variation in delinquency, the alternative score provides significant incremental predictive power. The coefficient on the Vantage score is -0.0568, indicating that a one-standard-deviation increase in the Vantage score reduces delinquency probability by 5.7 ppt, which is highly significant.

Column (3) adds additional controls for the loan amount, loan term, and median income in the borrower's zip code. The coefficient on the alternative score remains stable, showing that the alternative score continues to significantly predict delinquency even after accounting for these loan-specific factors. The traditional credit score (Vantage score) also retains its strong negative association with delinquency. Loan amount and the income of the zip code have small but statistically significant negative effects on delinquency, and longer loan terms are associated with a higher probability of delinquency.

Columns (4) through (6) restrict the sample to different terciles of the Vantage score distribution to assess whether the predictive power of the alternative score varies across traditional credit score groups. In all three Vantage score terciles, alternative score has a negative and significant relation to delinquency, with a slightly stronger effect for the high Vantage tercile (column (6)). This indicates that even among borrowers with high traditional credit scores, the alternative score provides additional information to predict delinquency. Across all columns, the inclusion of lender by month and county fixed effects controls for time-invariant differences across lenders and counties.

Overall, the results in Table II demonstrate that the Alternative Score provides significant predictive power for delinquency, both on its own and in conjunction with traditional credit scores. These findings highlight the value of alternative credit data for lenders seeking to assess credit risk, even for borrowers with relatively high credit scores according to traditional scoring models.

Figure 4 graphically illustrates the relation between delinquency and alternative credit score with binned scatter plots of delinquency rates by alternative score broken out by Vantage score tercile. The figure presents the relation between delinquency and alternative credit score after controlling for loan amount, loan terms, average zip code income, lender-month fixed effects, and county fixed effects. The downward sloping lines in the plot show a clear negative correlation between alternative credit scores and delinquency rates, meaning that as credit scores increase, the likelihood of delinquency decreases. This relation is monotonic and appears to be close to linear. The three-color groups represent the different Vantage score terciles (low, medium, and high). The consistent negative slopes across all three Vantage score groups indicate that alternative credit scores consistently predict delinquencies among a wide range of applicants with different traditional credit profiles. The horizontal dashed lines plot the delinquency rates for loan applicants without a history of alternative credit usage in each Vantage score group.

Figure A.1 displays a similar relationship while additionally controlling for Vantage score, which is why there is little vertical separation between the three lines. The pattern remains consistent, with a negative correlation between delinquency and alternative credit score, reinforcing the conclusion that higher alternative credit scores are associated with lower delinquency rates. The different Vantage score terciles remain close together, showing that the overall relationship between credit score and delinquency is robust across regions and Vantage score terciles.

2.2 How do Lenders Use the Alternative Credit Data?

In this section, we evaluate how lenders use alternative credit data, exploiting the fact that lenders gained access to the data on a staggered basis. As in the previous section, we restrict the sample to applicants with activity in the alternative credit score data to assess how lenders use the credit scores.

The availability of detailed loan origination and performance data from the time around the adoption of alternative credit data by lenders allows us to test the impact of such data on both lenders and borrowers. Specifically, we estimate the following OLS model:

$$y_{i,t} = \beta_1 \text{AltScore}_i \times (1 - \text{Post}_t) + \beta_2 \text{AltScore}_i \times \text{Post}_t + \gamma \text{Post}_t + X_i \delta + \alpha_g + \lambda_c + u_{i,t} \qquad (1)$$

where $y_{i,t}$ represents the outcome for loan i at time t. AltScore_i is the standardized alternative credit score, and Post_t indicates the period after data adoption. The model includes control variables X_i, lender-month fixed effects α_g, and county fixed effects λ_c. The coefficient β_1 represents the effect of alternative scores before adoption, while β_2 represents this effect after adoption. The difference between these coefficients indicates how the relationship between alternative scores and outcomes changes after lenders gain access to the data. As outcome variables, we consider the interest rate of the loan and a delinquency indicator variable.

Table III reports the results. First, we assess how loan delinquency is related to alternative credit scores before and after lenders adopt the data. We expect that alternative scores should have a similar predictive value regardless of whether lenders observe the data, and this is what we find. Column (1) of Table III reports regressions of delinquency on alternative credit score separately before and after the adoption of the data. In both cases, delinquency rates decrease with alternative score, and the magnitudes are similar to the total magnitude estimated in Table II. Specifically, a one standard deviation increase in alternative score is associated with a 1.85 ppt decrease in delinquency probability before adoption and a 2.80 ppt decrease in delinquency after adoption. The larger post-adoption delinquency coefficient could be due to higher interest rates for low-alternative-score borrowers after adoption. Lenders could also make different loan approval decisions post-adoption, but this would likely have the opposite effect on delinquency rates by increasing screening for low-alternative-score borrowers.

We next consider how alternative credit scores affect interest rates by regressing the loan interest rate on alternative credit score before and after the adoption of the data. Column (2) of Table III shows no relation between alternative credit score and loan interest rates before the adoption of alternative credit data. In contrast, once lenders have access to alternative credit scores, we find that a one-standard-deviation increase in alternative credit score is associated with a decrease in interest rate of 14.7 bp. This decrease represents 32.3% of the decrease in interest rates associated to a one-standard-deviation increase in Vantage score of 45.5 bp.

In Appendix Table A.5, we repeat our regressions of loan interest rate in subsamples of loans split by Vantage score, after adopting alternative credit data. The alternative score is associated with lower interest rates in all samples, and its effect on interest rates is especially pronounced in the subsample of high-Vantage score loans.

2.3 Are Alternative Credit Users Penalized?

So far, we have examined the impact of alternative credit scores conditional on borrowers being in the alternative credit data. We now add borrowers who are not in the alternative credit data to assess the potential negative signal of using alternative financial services.

Table IV regresses delinquency and interest rates on an indicator variable, Hit, which takes a value of 1 if someone is in the alternative credit score data. Column (1) reports the results of regressing delinquency on Hit, controlling for Vantage score and other loan characteristics. Borrowers with a history of using alternative financial products have a delinquency rate that is 6.1 ppt higher than those without such a history, indicating that inclusion in alternative credit data serves as a signal of lower creditworthiness. This effect is large relative to the mean delinquency rate of 39% in this sample.

Column (2) adds alternative credit score in addition to the Hit indicator. To separate the effect of Hit from the effect of alternative score, we assign a standardized alternative score of zero to borrowers without a hit in the alternative credit score data.[8] The Hit coefficient of 0.0614 indicates that a borrower with the mean alternative credit score has a delinquency rate that is 6.1 ppt higher compared to an equivalent borrower with no history of using alternative credit. Compared to the mean alternative credit user, a one-standard-deviation higher alternative score is associated with a 3.1 ppt lower delinquency rate.

Column (3) shifts the focus to loan interest rates and shows that borrowers included in the alternative credit data pay interest rates that are 39.8 bp higher than those not in the data. This large and significant effect suggests that inclusion in alternative credit data is strongly associated with higher borrowing costs, potentially reflecting the perceived higher risk of these borrowers. Importantly, this coefficient is substantial compared to the mean interest rate of 22.5%. Column (4) includes the standardized alternative score alongside the hit indicator to account for the effect of the alternative score on interest rates. A one-standard-deviation increase in the alternative score reduces interest rates by 5.1 bp, indicating that the alternative score provides lenders with additional information that mitigates some of the risk reflected in the hit indicator. These findings further support the role of alternative scores in providing predictive value beyond the signal of inclusion in the data. However, to offset the 39.8 bp effect of Hit, a borrower would need to have an alternative score that is more than seven standard deviations above the mean.

Figure 5 shows a negative relationship between the alternative credit score and the interest rate on auto loans. We observe that higher alternative credit scores for each Vantage score tercile are associated with lower interest rates. However, borrowers not included in the alternative data (plotted as the dashed horizontal lines in the plot) appear to benefit from systematically lower interest rates, suggesting that lenders may perceive them as less risky. Overall, the empirical evidence thus far highlights the dual role of alternative credit data in providing predictive value while also signaling increased risk, leading to divergent loan outcomes for borrowers included in

[8]This differs from the data vendor's practice of assigning a score of 744 for individuals without a hit in the data.

versus excluded from the data.

Next, in Figure 6, we examine the effect of being included in alternative credit data (*Hit* = 1) on loan interest rates over time, with month 0 marking the adoption of the data by lenders. The point estimates show the coefficients for regressions of interest rates on the hit indicator for each month relative to adoption, controlling for loan amount, loan terms, average zip code income, lender fixed effects, and county fixed effects. Before adoption, the coefficients hover around zero, indicating that the alternative credit data was not priced by the lenders. The absence of a pre-trend suggests that the observed post-adoption pricing effects are not driven by endogenous market factors. After adoption, the coefficients increase steadily, suggesting that lenders progressively incorporate the information from alternative credit data, leading to higher interest rates for borrowers included in the data. As a result, our baseline analysis focused on the entire sample after the adoption of alternative credit data probably underestimates the impact of the data, which appears to be largest two to three years after adoption.

2.4 Ghosts

Our analysis thus far indicates that information on past use of alternative credit hurts most users of alternative credit. We next turn to a population that may have the most to gain from alternative credit data: individuals who lack traditional credit scores, frequently referred to as "ghosts" in the lending industry. One might expect that alternative data are more valuable when lenders have greater uncertainty about the borrower's type. As a result, borrowers with fewer records from traditional credit bureaus could benefit the most from alternative data. In this subsection, we examine whether the effects of using alternative credit data are particularly important for these types of consumers.

Table V examines how alternative data relate to delinquency and interest rates for a sample restricted to borrowers who lack a traditional credit score, frequently known as "ghosts." This table replicates the analysis from Table IV assessing whether the inclusion in alternative credit data disproportionately affects individuals lacking traditional credit scores.

Column (1) explores the relationship between delinquency and the *Hit* indicator. Borrowers included in the alternative credit data are 12.1 ppt more likely to experience delinquency compared to those excluded from the data. This statistically significant and economically important result highlights that even among individuals who lack traditional credit scores, inclusion in alternative credit data signals higher credit risk. The effect is substantial relative to the mean delinquency rate of 34.2% for this sample. Column (2) adds the standardized alternative credit score, which is statistically insignificant in the ghost subsample, likely due to lack of power.

Columns (3) and (4) of Table V focus on interest rate as the dependent variable. Borrowers included in alternative credit data pay significantly higher interest rates, with a coefficient of 62.4 bp in column (3), compared to those not included. Standardized alternative credit score is statistically insignificant when added in column (4), again likely due to a lack of power. These findings indicate that, for ghosts, inclusion in alternative credit data significantly increases borrowing costs, even when controlling for additional loan characteristics. The alternative credit score does not affect loan interest rates in a meaningful way and its coefficient remains insignificant in the interest rate regression.

Overall, the results in Table V indicate that alternative credit data have an effect on ghosts that is similar to their effect on borrowers more generally. Ghosts who have used alternative credit are generally riskier and pay higher interest rates compared to borrowers without a hit in the alternative credit data, and there is no evidence that high alternative credit scores lead to lower interest rates for ghosts.

3 Building Credit with Alternative Financial Services

One of the promises of alternative data is that it may also provide a path for non-traditional borrowers to establish a positive credit history. In this section, we assess how often this is the case. In general, using alternative credit is a negative signal. How often do loan applicants have strong enough alternative credit scores to overcome the negative signal of having used alternative financial services in the past?

We take two approaches to address this question. First, we use a regression setting to estimate loan delinquency and interest rates for borrowers with a hit in the AFS data relative to a counterfactual in which the borrowers did not hit in the alternative credit data. Second, we implement a machine learning algorithm to examine the impact of hits and alternative scores in a more flexible framework that better accommodates nonlinearities and interaction effects.

For OLS estimates, we regress delinquency and interest rate on Hit, alternative score, Vantage score, and zip code income with fixed effects for lender-month and applicant county. Estimates for interest rate and delinquency also control for loan amount and loan term. For each borrower with a hit in the alternative credit data, we then use the model to estimate the combined effect of Hit and the borrower's alternative score relative to a counterfactual in which the borrower has no hit in the alternative credit data and has an alternative score that is the same as other borrowers with no hit in the data.

Figure 7 plots histograms of the OLS results. Panel (a) shows that predicted delinquency rates for people with a history of using alternative credit are higher than those for their counterparts 94.76% of the time (the blue area of the histogram). To offset the negative effect of showing in the alternative credit data on delinquency rates, borrowers need an extremely high alternative score, and this only occurs 5.24% of the time, as plotted in green in the figure. The bottom figure shows that the negative effect of a hit is even stronger for interest rates, with predicted interest rates for borrowers in the alternative credit data being higher 100% of the time. On average, the usage of alternative credit is associated with interest rates that are 39 bp higher.

These results suggest that a hit in the alternative financial services data is typically a negative signal even for borrowers whose alternative credit profile suggest a lower default likelihood in our sample. However, a limitation of these linear models is that they might miss out on important interactions and nonlinearities that could affect lending decisions. To estimate more flexible models, we implement machine learning models to predict interest rates and delinquency based on alternative credit scores and other borrower and loan characteristics.

We evaluate over 200 model configurations in four classes of machine learning models: (1) gradient boosting (XGBoost), (2) random forests, (3) neural networks, and (4) elastic net regression. Each model is trained using 5-fold cross-validation to prevent overfitting, with the best model selected based on out-of-sample root mean square error (RMSE). The models incorporate borrower characteristics, including traditional and alternative credit scores, loan characteristics (amount and term length), borrower zip code demographics (income and race), and fixed effects for lender and month. XGBoost emerges as the best-performing model, with optimal parameters of maximum depth of 11, learning rate of 0.011, minimum child weight of 7, and L2 regularization parameter of 0.11 for delinquency estimations, and maximum depth of 11, learning rate of 0.096, minimum child weight of 3, and L2 regularization parameter of 0.98 for the rate model. For delinquency, the model yields a cross-validated area under the receiver operating characteristic curve (ROC-AUC) of 0.679. For interest rates, this configuration achieves a cross-validation R^2 of 0.37, capturing important nonlinear patterns while maintaining strong out-of-sample performance. For consistency, we also employ the same model to estimate delinquency.

To understand how the models use alternative credit information, we employ SHAP (SHapley Additive exPlanations) values. SHAP values provide a unified measure of the importance of the characteristics that shows how each variable contributes to individual predictions. For each prediction, SHAP values decompose the difference between the model's prediction and the average prediction into contributions from each feature. This allows us to understand not just which features are important, but also how they affect predictions across their range of values. In our context, SHAP values help quantify how much having a hit in the alternative data and different alternative credit scores contribute to predicted interest rates relative to baseline predictions. By analyzing the distribution of these effects among borrowers, we can identify when alternative credit information helps or hurts different types of borrowers.

Panel A of Figure 8 shows the combined effect on delinquency of having a hit in the alternative credit data along with the borrower's alternative credit score. The y-axis plots SHAP values representing the total impact of alternative credit information (both having a hit and the alternative score), and the x-axis shows the alternative credit score. The figure shows a strong non-linear relationship. For borrowers with alternative scores below around 750, the combined effect is positive but diminishing, suggesting that borrowers with higher alternative scores show a lower likelihood of default. Mean SHAP values continue to decrease after 750,, but the rate of decline slows considerably. The rightmost data points show the impact on delinquency likelihood of not having a hit in the alternative data (labeled "No Hit"), which serves as a reference level. In particular, borrowers with very high alternative scores (above 800) achieve SHAP values lower than the SHAP values associated with having no hit in the data, suggesting that borrowers with a hit, but also with a high enough alternative score, can offset the negative impact on the likelihood of delinquency typically associated with appearing in the alternative data. The vertical differences in SHAP values for borrowers with the same alternative score come from interactions with different covariates, which the model flexibly estimates.

Panel B of Figure 8 plots SHAP values for other covariates. The plots indicate that delinquency rates generally decrease with Vantage score and increase with loan amount. We also consider how delinquency varies across zip codes with different characteristics. The likelihood of delinquency decreases moving from low income to high income zip codes and from zip codes with a low percentage of black population to zip codes with high black populations.

Next, we examine a similar model that estimates the impact of alternative credit information on interest rates. Panel A of Figure 9 shows the combined effect on interest rates of having a hit in the alternative credit data along with the borrower's alternative credit score. The y-axis plots the SHAP values representing the total impact of alternative credit information (both having a hit and the alternative score), and the x-axis shows the alternative credit score. The figure shows a strong non-linear relationship. For borrowers with alternative scores below 700, the combined effect is positive but diminishing, suggesting that higher alternative scores partially offset the negative impact that being in the data has on interest rates. There is a sharp decline in the combined effect for scores between 700 and 750, after which the impact stabilizes near zero. The rightmost data points show the impact of not having a hit in the alternative data (labeled "No Hit"), which serves as a reference level. Notably, even borrowers with very high alternative scores (above 800) rarely achieve SHAP values as low as the SHAP values associated with having no hit in the data, indicating that the negative signal of using alternative financial services is difficult to overcome even with extremely high alternative scores. The vertical differences in SHAP values for borrowers with the same alternative score come from interactions with different covariates, which the model flexibly estimates.

Panel B of Figure 9 plots SHAP values for other covariates. The plots indicate that interest rates generally decrease with Vantage score and loan amount. We also consider how interest rates vary across zip codes with different characteristics. Moving from low-income to high-income zip codes is associated with a decrease in interest rates of approximately 0.3 ppt. The share of the zip codes population that is black has an even stronger relation with interest rates of close to 1.0 ppt when moving from a zip code that is 0% black to a zip code that is nearly 100% Black. In addition to estimating the direction and magnitude of these relations, the models account for nonlinearities and interactions, represented by a range of SHAP values for the same covariate value.

We now turn back to the question of how often borrowers have strong enough alternative credit scores to overcome the negative signal that they have used alternative financial services using our SHAP analysis. Figure 10 plots the distribution of SHAP values for both hit and no-hit populations. With more flexible models, alternative credit usage has a greater ability to improve credit outcomes. Panel (a) shows that alternative credit usage is associated with a lower delinquency probability 15.48% of the time compared to 5.24% of the time using the linear model. The average SHAP values for delinquency are -0.041 for no-hit population and 0.141 for the hit population. This suggests that having no hit decreases the odds of delinquency by about 4% (exp(-0.041) $\approx$ 0.96) relative to the baseline while having a hit increases the odds of delinquency by about 15% (exp(0.141) $\approx$ 1.15). For interest rates, the SHAP values indicate that alternative credit usage improves credit outcomes for only 3.28% of borrowers, while 96.72% of alternative financial service users do not have high enough alternative scores that offset the stigma of using alternative credit. The average SHAP values for rates are -0.038 for no-hit borrowers and 0.122 for the hit borrowers, meaning that having no hit lowers the interest rate by 0.038 percentage points relative to the baseline while having a hit raises it by 0.122 ppt.

4 Differences Across Demographic Groups

The previous section indicates that the vast majority of borrowers with alternative credit usage are hurt by the revelation of this information. To the extent that some demographic groups are more likely to use alternative credit products, this could have differential effects on different groups.

To evaluate this possibility, we regress Hit and the alternative score (conditional on having a hit) on zip-code-level demographic variables. The results are reported in Table VI. In column (1), we regress Hit on the average income of zip codes with a standardized coefficient of -0.0015, indicating that zip codes with one-standard-deviation higher income are associated with 0.15 ppt lower probability of having a hit in the alternative credit data. Column (2) adds race to the regression with a positive and statistically significant result for minority population share.[9] The coefficient estimate indicates that a one-standard-deviation increase in minority population share is associated with a 1.6 ppt higher probability of being a user of alternative credit products. In column (3), we add more detailed racial characteristics with a large positive relation between Hit and black population share and smaller effects for other races. A one-standard-deviation increase in a zip code's black population share is associated with a 1.7 ppt increase in the likelihood of being in the alternative credit data, whereas this magnitude is 0.4 ppt, 0.2 ppt, and -0.4 ppt for population shares of Asians, Hispanics, and American Indians, respectively.

In columns (4) to (6), we repeat the same regressions with alternative score as the dependent variable, conditional on having a hit in the alternative data. Across the three specifications, zip code income has a positive relation to the alternative score. Alternative scores are also positively correlated with minority population share, as well as with the shares of Asians, Hispanics, and American Indians. In contrast, alternative score is negatively related with the share of the black population.

The results in Table VI show a clear pattern of more frequent alternative credit usage and lower alternative scores in zip codes with lower incomes and a higher share of the black population. As a result, these areas are likely to have the largest negative effects when lenders adopt alternative credit data.

[9] We define minority population share as the fraction of the population in a zip code that is identified as Black, American Indian, Asian, or Hispanic.

5 Loan Originations

The data we have analyzed thus far are conditional on borrowers receiving a loan. It is also possible that alternative credit data could affect the extensive margin of whether an applicant is approved for a loan. Analyzing the extensive margin is challenging for three reasons. First, we do not observe whether a loan application is approved. We only observe applications and whether they result in an origination. Second, we do not observe loan applications before the alternative credit data is adopted, so we cannot conduct difference-in-differences analysis around the data adoption dates like we do for interest rates and delinquencies. Finally, even after adoption of the alternative credit data, we only see applications for which the lender queried the alternative credit data, and we do not observe how many loan applications they received without querying the data.

Nonetheless, we can get some sense of the external margin from total numbers of loan originations before and after data adoption and from changes in the percent of originated loans that have a hit in the alternative credit data. Panel A of Figure 11 plots the total number of loans that are originated before and after data adoption in event time. Loans with and without hits in the alternative data are displayed separately. The general pattern is that originations to borrowers without a hit in the data (plotted as black squares) trended upward after data adoption, while loans to borrowers with a hit (plotted as blue squares) remained flat. This suggests that the use of the alternative data mainly benefits applicants who are not in the data, consistent with our findings for interest rates.

In Panel B of Figure 11, the blue squares plot the percentage of loans with a hit in the alternative data before and after data adoption. Consistent with the trends in Panel A, the percentage of loans with a hit drops considerably after data adoption. Post-adoption, hits are also more common in applications that do not result in an originations, as shown by the red squares in the plot. Finally, Table A.6, presents OLS regression results showing a negative relation between the average hit rate per county and indicator variables for data adoption and loan origination. Overall, the results suggest that lenders use the data to shift their lending toward applicants who have not previously used alternative financial services.

6 Conclusion

This study highlights the transformative potential of alternative credit data to improve credit risk assessment and loan pricing, which has important implications for financial inclusion. By providing incremental predictive power for delinquency beyond traditional credit metrics, alternative credit scores enable lenders to make more informed decisions, particularly for underserved borrowers. However, the dual role of these data—as both a predictor of risk and a potential negative signal—underscores the complexity of their impact on borrower outcomes.

The findings in this paper reveal a dual dynamic in the use of alternative credit data, where its positive impact on borrowers is offset by the negative signal associated with being included in the data. On the one hand, alternative credit data improve lenders ability to make informed decisions by reducing interest rates and increasing originations for borrowers with higher alternative scores. This highlights the potential of alternative data to improve financial inclusion by expanding credit access and offering better loan terms to underserved populations. On the other hand, inclusion in the alternative credit data itself acts as a negative signal and is associated with higher delinquency rates. In response to this credit risk, lenders set higher interest rates for borrowers with a hit in the alternative credit data. This negative signal is partially offset for borrowers with high alternative credit scores, but the net effect is typically negative even for borrowers with relatively high alternative credit scores. As a result, very few borrowers build a positive credit history with alternative credit data, and the increased adoption of these data is likely a hindrance to financial inclusion for users of alternative financial services, even for individuals who lack any traditional credit history.

The adoption of alternative credit data also has significant distributional implications. Because the usage of alternative financial services is higher in areas with lower incomes and higher proportions of black residents, the availability of alternative credit data is likely to increase average interest rates and decrease credit availability for these populations.

References

Agarwal, Sumit, Shashwat Alok, Pulak Ghosh, and Sudip Gupta, 2023, Financial inclusion and alternate credit scoring: Role of big data and machine learning in fintech, *Working Paper* .

Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman, 2022, Are high-interest loans predatory? theory and evidence from payday lending, *The Review of Economic Studies* 89, 1041–1084.

Argyle, Bronson S, Taylor D Nadauld, and Christopher J Palmer, 2020, Monthly payment targeting and the demand for maturity, *Review of Financial Studies* 33, 5416–5462.

Athreya, Kartik, Xuan S Tam, and Eric R Young, 2012, A quantitative theory of information and unsecured credit, *American Economic Journal: Macroeconomics* 4, 153–183.

Baines, Arthur P, and Marsha J Courchane, 2014, Fair lending: Implications for the indirect auto finance market, *Study Prepared for the American Financial Services Association* .

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, 2022, Consumer-lending discrimination in the fintech era, *Journal of Financial Economics* 143, 30–56.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri, 2020, On the rise of fintechs: Credit scoring using digital footprints, *The Review of Financial Studies* 33, 2845–2897.

Consumer Financial Protection Bureau, 2017, Who are the Credit Invisible?, [Accessed November 6, 2024].

Demyanyk, Yuliya, and Otto Van Hemert, 2011, Understanding the subprime mortgage crisis, *The Review of Financial Studies* 24, 1848–1880.

Di Maggio, Marco, Dimuthu Ratnadiwakara, and Don Carmichael, 2022, Invisible primes: Fintech lending with alternative data, Technical report, National Bureau of Economic Research.

Einav, Liran, Mark Jenkins, and Jonathan Levin, 2012, Contract pricing in consumer credit markets, *Econometrica* 80, 1387–1432.

Einav, Liran, Mark Jenkins, and Jonathan Levin, 2013, The impact of credit scoring on consumer lending, *The RAND Journal of Economics* 44, 249–274.

Experian, 2022, More than a score: The case for financial inclusion, Blog available at https://www.experian.com/blogs/insights/the-case-for-financial-inclusion/.

Federal Reserve Bank of New York, August 06, 2024, Household debt increased moderately in q2 2024; auto and credit card delinquency rates remain elevated, *News Release* www.newyorkfed.org/newsevents/news/research/2024/20240806.

Fonseca, Julia, 2023, Less mainstream credit, more payday borrowing? evidence from debt collection restrictions, *Journal of Finance* 78, 63–103.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther, 2022, Predictably unequal? The effects of machine learning on credit markets, *Journal of Finance* 77, 5–47.

Gdalman, Hannah, MK Falgout, Necati Celik, and Meghan Greene, 2024, Finhealth spend report 2024: Record spending on credit services puts pressure on vulnerable households, *FinHealth Spend* finhealthnetwork.org/wp-content/uploads/2024/08/FinHealth-Spend-Report-2024-FHN.pdf.

Grunewald, Andreas, Jonathan Lanning, David Low, and Tobias Salz, 2023, Auto dealer loan intermediation: Consumer behavior and competitive effects, Working paper.

Jansen, Mark, Samuel Kruger, and Gonzalo Maturana, 2024, Dealer financing in the subprime auto market: Markups and implicit subsidies, *Management Science, Forthcoming* .

Jansen, Mark, Hieu Nguyen, Lamar Pierce, and Jason Snyder, 2022, Product sales incentive spillovers to the lending market, *Management Science, Forthcoming* .

Liberman, Andres, Daniel Paravisini, and Vikram Pathania, 2021, High-cost debt and perceived creditworthiness: Evidence from the uk, *Journal of Financial Economics* 142, 719–736.

Livshits, Igor, James C Macgee, and Michele Tertilt, 2016, The democratization of credit and the rise in consumer bankruptcies, *Review of Economic Studies* rdw011.

Melzer, Brian, and Aaron Schroeder, 2017, Loan contracting in the presence of usury limits: Evidence from automobile lending, *Working Paper* .

Miller, Sarah, and Cindy K Soo, 2020, Does increasing access to formal credit reduce payday borrowing?, Technical report, National Bureau of Economic Research.

Network, Financial Health, 2017, 2017 financially underserved market size study, Available at https://finhealthnetwork.org/research/2017-financially-underserved-market-size-study/.

Padillaa, A. Jorge, and Marco Pagano, 2000, Sharing default information as a borrower discipline device, *European Economic Review* 44, 1951–1980.

Pierce, Lamar, 2012, Organizational structure and the limits of knowledge sharing: Incentive conflict and agency in car leasing, *Management Science* 58, 1106–1121.

Stiglitz, Joseph E, and Andrew Weiss, 1981, Credit rationing in markets with imperfect information, *American Economic Review* 71, 393–410.

Turner, Michael, Alyssa Stewart Lee, Ann Schnare, Robin Varghese, and Patrick D Walker, 2006, Give credit where credit is due: Increasing access to affordable mainstream credit using alternative data, *The Brookings Institution, Washington, DC* .

U.S. Government Accountability Office, 2021, Mortgage Lending: Use of Alternative Data Is Limited but Has Potential Benefits.

Vercammen, James A., 1995, Credit bureau policy and sustainable reputation effects in credit markets, *Economica* 62, 461–478.

Zabritski, Melinda, 2020, State of the automotive finance market. Fourth Quarter 2019. Experian.

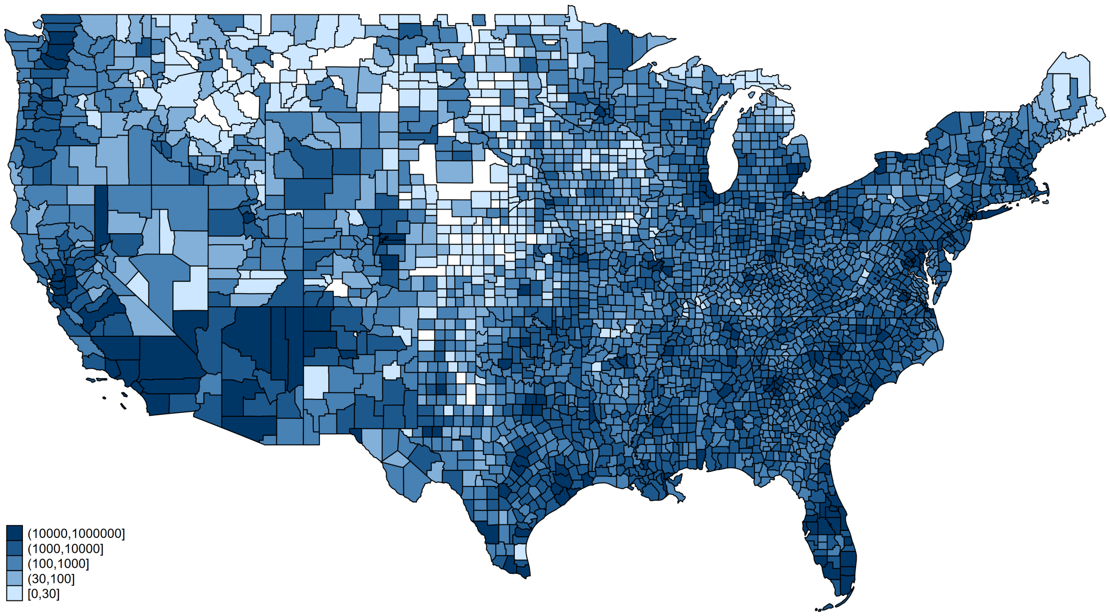


Figure 1. Geographic Coverage. This figure shows the geographic coverage of all loan applications in our data across different counties in the US. Each county is shaded according to the number of loan applications, with darker blue indicating higher number of applications and lighter shades representing fewer applications.

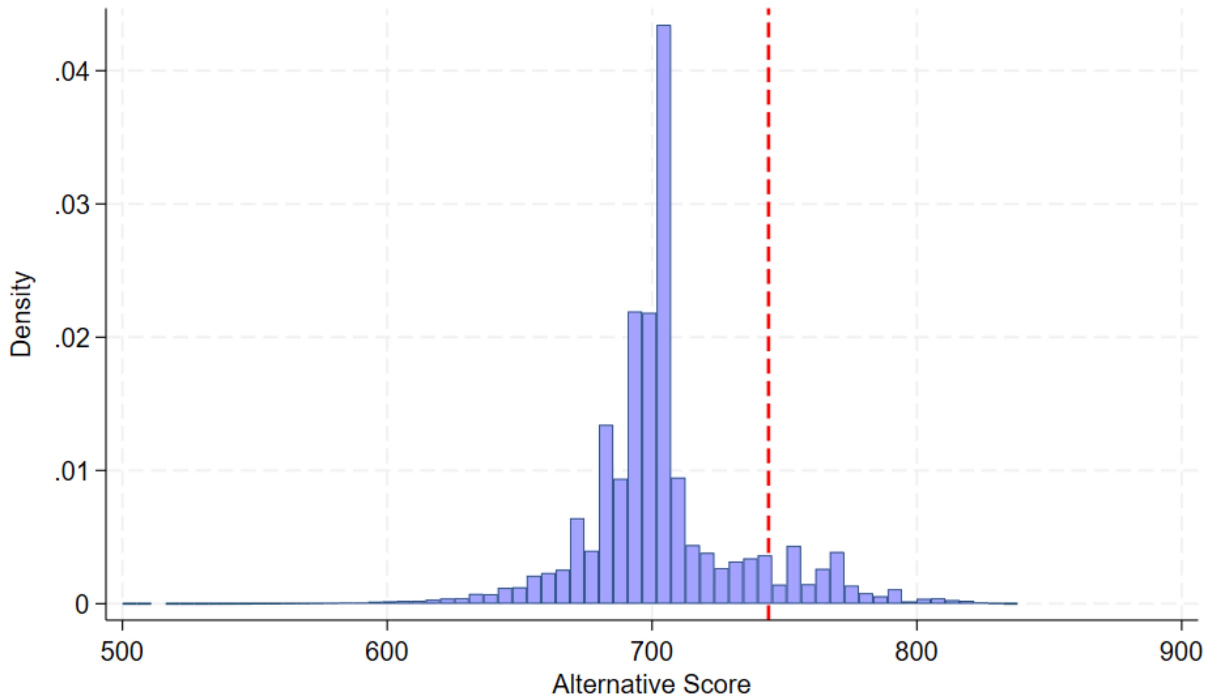


Figure 2. Distribution of the Alternative Score for Rated Loan Applicants. This figure shows the distribution of the alternative credit score for all applicants in our sample who have a hit in the alternative data, and as a result, receive an alternative credit score. The vertical red dashed line at 744 indicates the alternative credit score assigned to individuals who were not previously present in the alternative database.

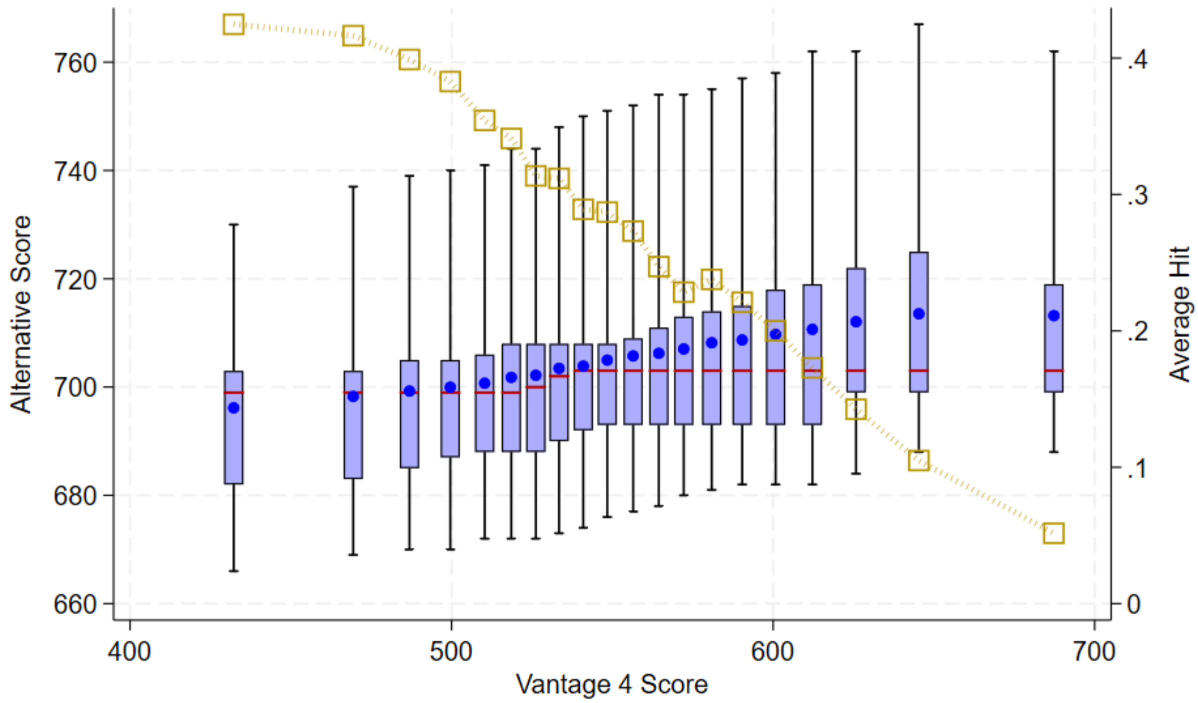


Figure 3. The Relationship between the Alternative Score and Vantage4 Score. Panel A shows the 10- to 90-percentile range of borrowers' alternative scores for 20 quantiles of Vantage 4 score for all loan applicants in our sample. Boxes represent the interquartile range of the alternative score for applicants who hit the alternative data. The horizontal red line in each box represents the median value, and the blue circles show the mean value. The gold hollow squares show the average ratio of the hit population for each quantile of the Vantage 4 score.

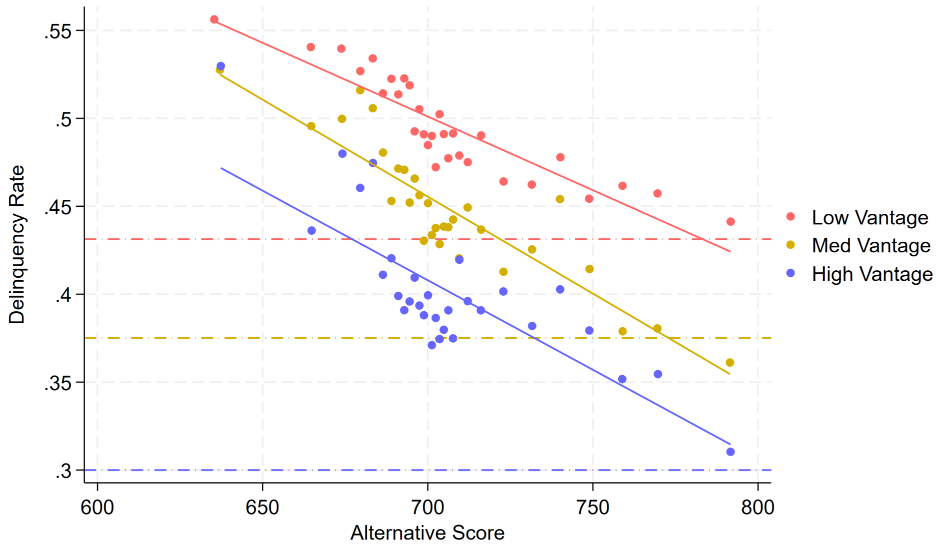


Figure 4. Relationship between Loan Delinquency and the Alternative Score for Different Terciles of Vantage Score. This figure shows the binscatter of loan delinquency over the alternative score for all loans in our sample. The red dots and lines correspond to the lowest tercile of the Vantage score, gold to the middle tercile, and blue to the highest tercile. The results control for the loan amount, loan terms, average zip code income, and month-lender and county fixed effects. The horizontal dashed line shows the average value for applicants in each Vantage group who do not hit the alternative credit data.

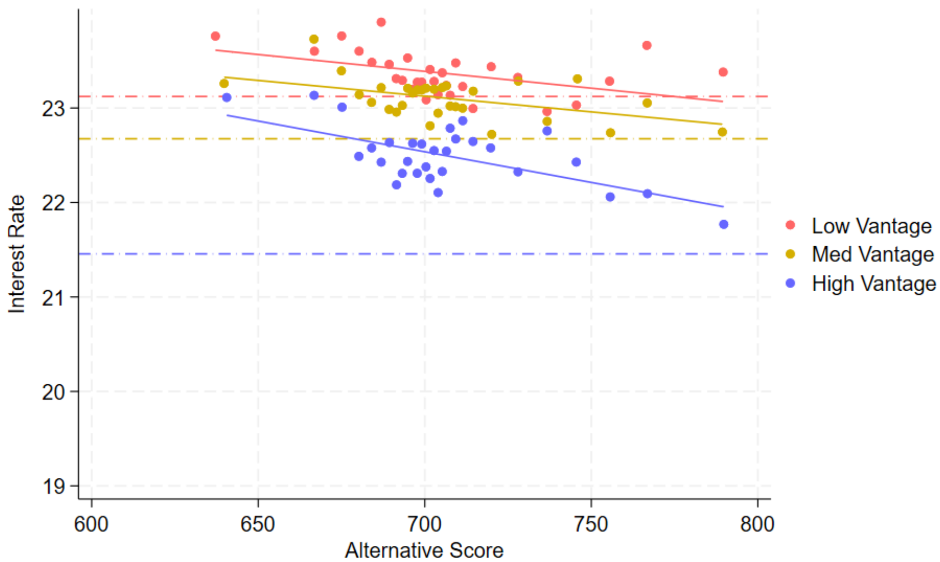


Figure 5. Relationship between Loan Rate and the Alternative Score for Different Terciles of Vantage Score. This figure shows the binscatter of loan interest rate and the alternative score for loans originated in the post-adoption period. The red dots and lines correspond to the lowest tercile of the Vantage score, gold to the middle tercile, and blue to the highest tercile. Both panels control for the vantage score, average zip code income, and month-lender and county fixed effects. The horizontal dashed line shows the average value for applicants in each Vantage group who do not hit the alternative credit data.

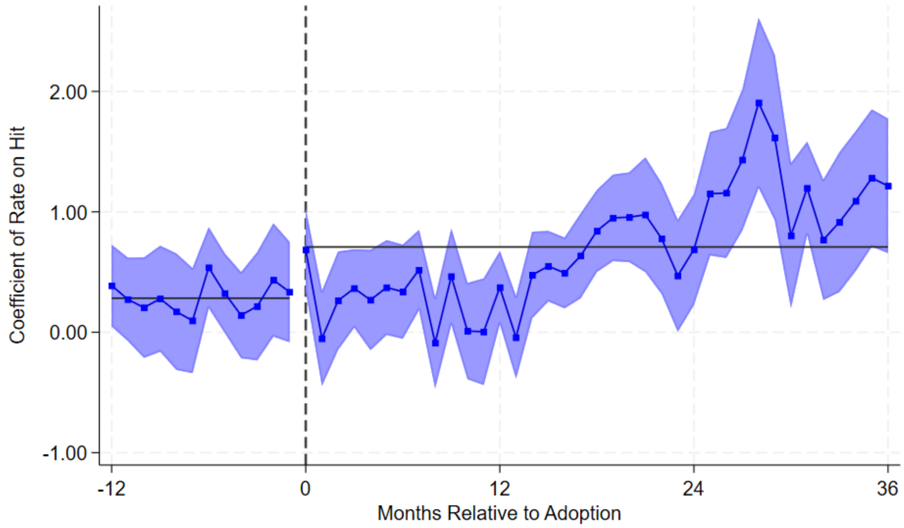
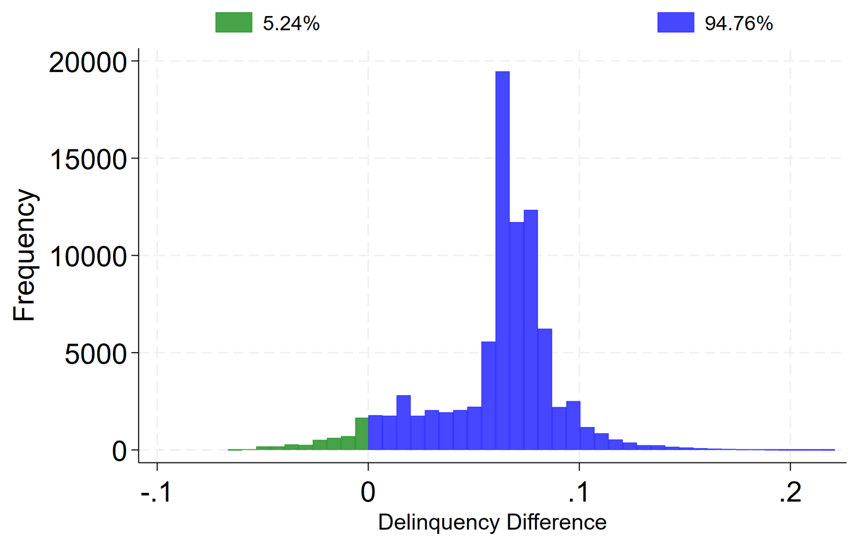


Figure 6. The Effect of Hit on Loan Rates Over Time. This figure plots the coefficients of a regression of loan rates on hit, control variables including vantage score, loan amount, loan terms, zip code income, and lender-month and county fixed effects. The regression is run for each month relative to the time of adoption separately and then graphed over time. Lender B is excluded due to lack of data before adoption. The blue area shows the 95% confidence interval for the coefficients, and the gray lines show the average values for pre and post period.

(a) Delinquency



(b) Rate

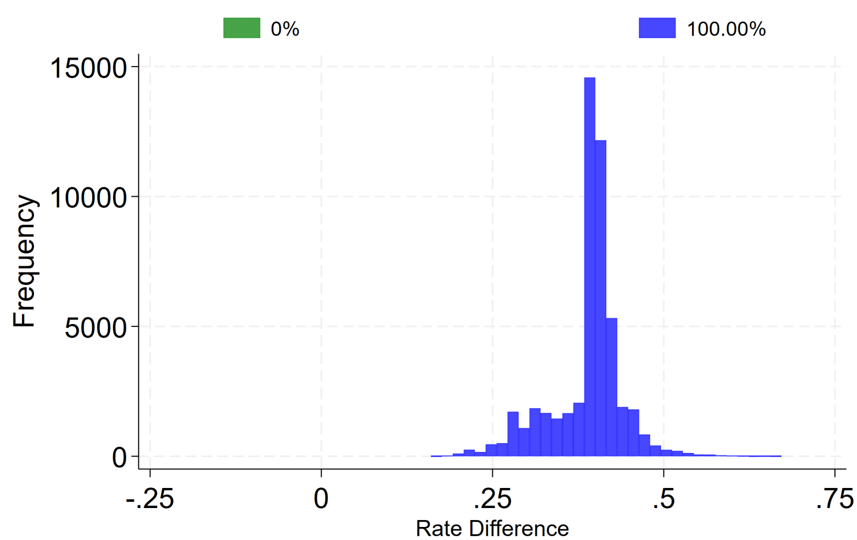
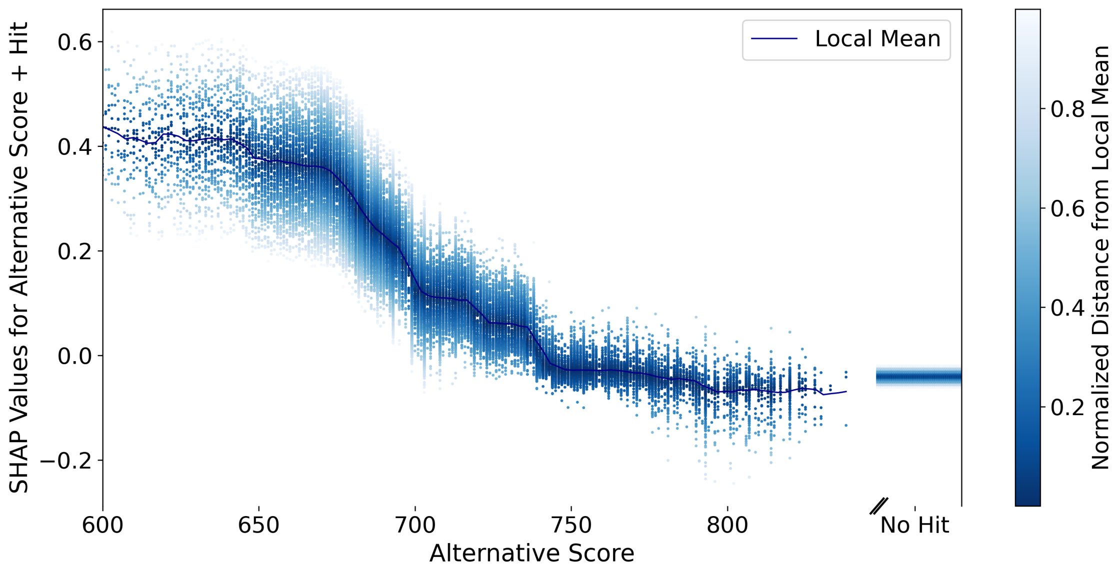


Figure 7. Estimated Impact of Hit on Delinquency and Interest Rate Using OLS. Panel (a) plots the distribution of differences between predicted delinquency rates for applicants with a hit in the alternative credit data and their corresponding predicted values if they had not been in the data. Panel (b) displays the same analysis for loan interest rates. Both analyses are based on OLS models using delinquency or interest rate as the dependent variable, with hit indicator and alternative credit score as independent variables, controlling for Vantage score, loan characteristics, zip code income, and lender-month and zipcode fixed effects.

(a) SHAP values for Alt Score + Hit



(b) SHAP values for other variables

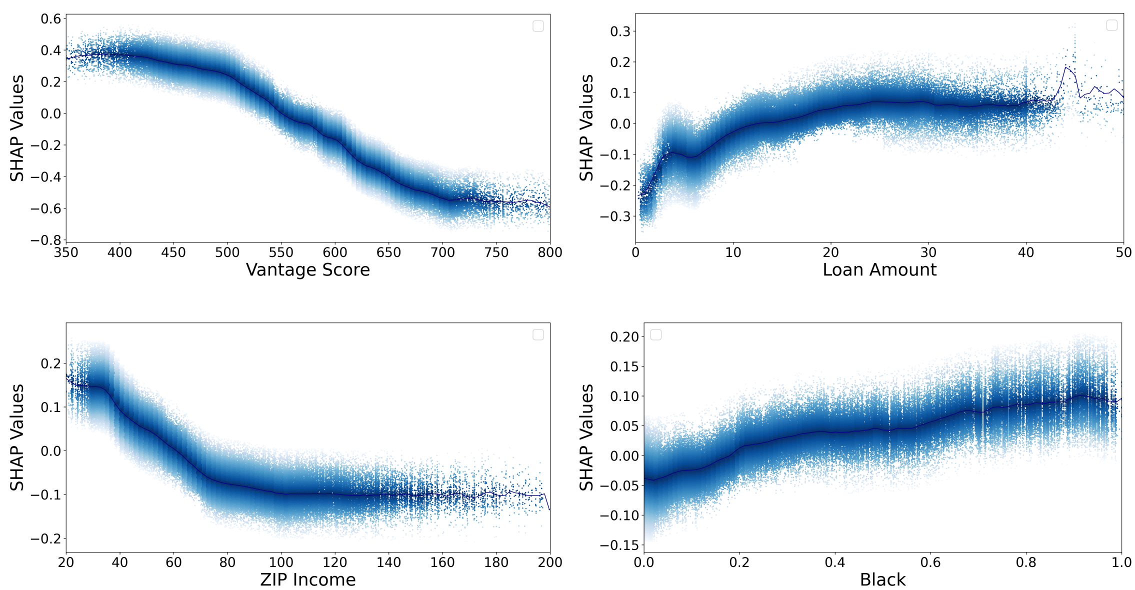
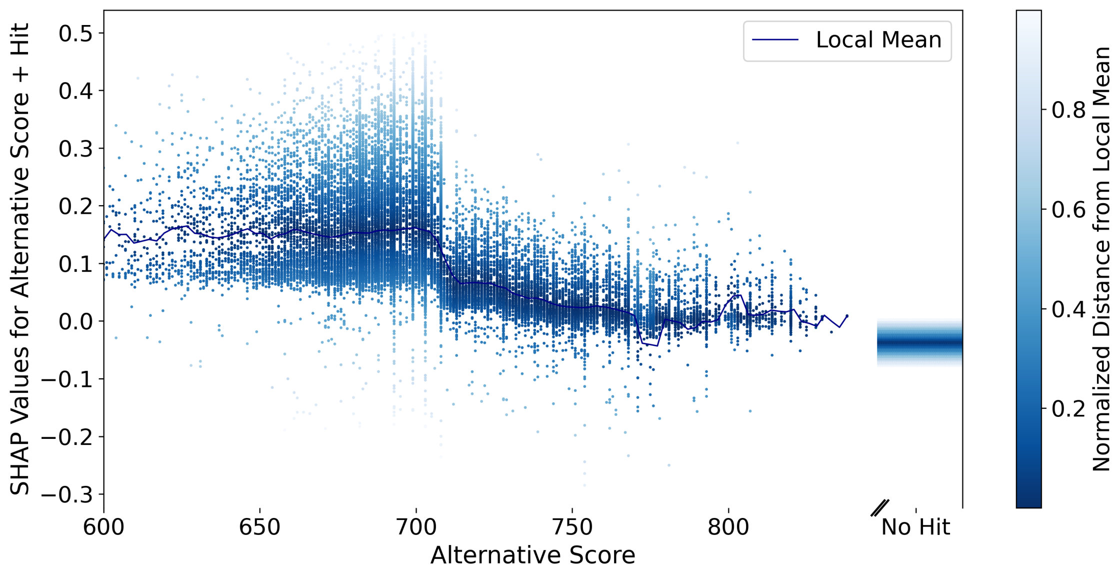


Figure 8. The Relationship between Alternative Score and Delinquency Predictions. This figure shows how the alternative score, hitting the alternative data, and other borrower characteristics influence loan delinquency likelihood predictions from the selected XGBoost model. In panel A, the Y-axis shows the sum of SHAP values for hit and alternative score. The color intensity indicates the distance from the local mean SHAP value, with darker blues representing closer proximity to the mean. The dark blue line shows the local mean SHAP values. The figure displays values for borrowers without alternative data hits after the axis break. Panel B plots the SHAP values for Vantage score, loan amount, zip code income, and percentage of Black population in zip codes against their respective values.

(a) SHAP values for Alt Score + Hit



(b) SHAP values for other variables

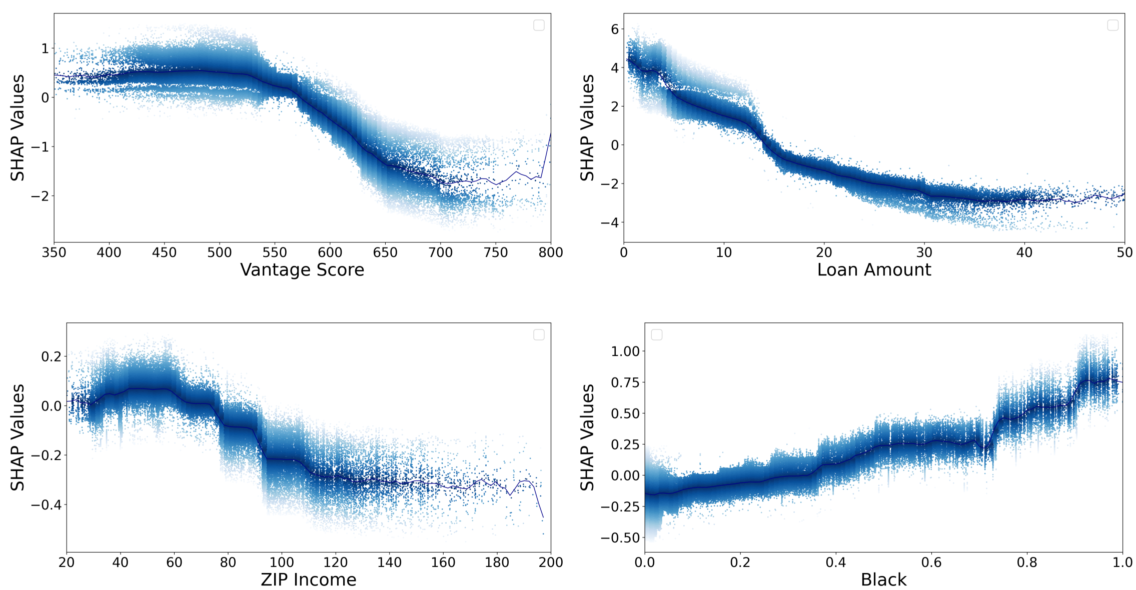
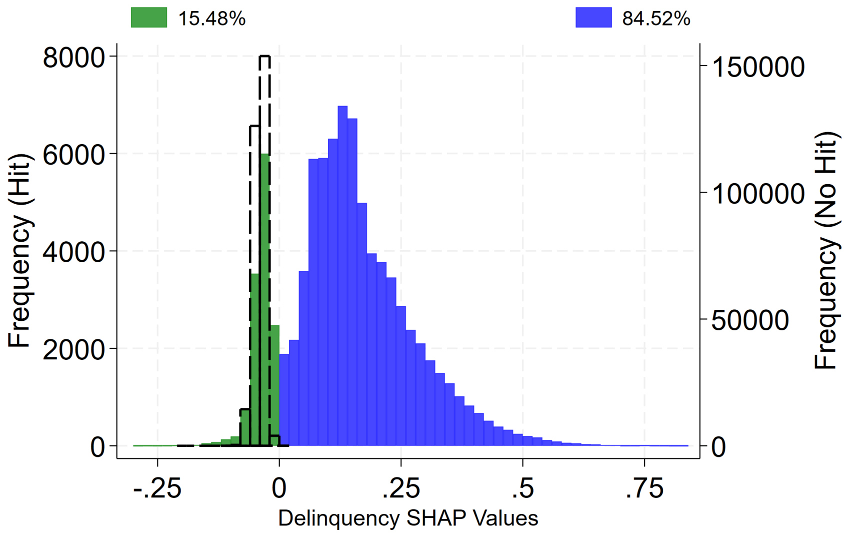


Figure 9. The Relationship between Alternative Score and Loan Rate Predictions. This figure shows how the alternative score, hitting the alternative data, and other borrower characteristics influence loan rate predictions from the selected XGBoost model. In panel A, the Y-axis shows the sum of SHAP values for hit and alternative score. The color intensity indicates the distance from the local mean SHAP value, with darker blues representing closer proximity to the mean. The dark blue line shows the local mean SHAP values. The figure displays values for borrowers without alternative data hits after the axis break. Panel B plots the SHAP values for Vantage score, loan amount, zip code income, and percentage of Black population in zip codes against their respective values.

(a) Delinquency



(b) Rate

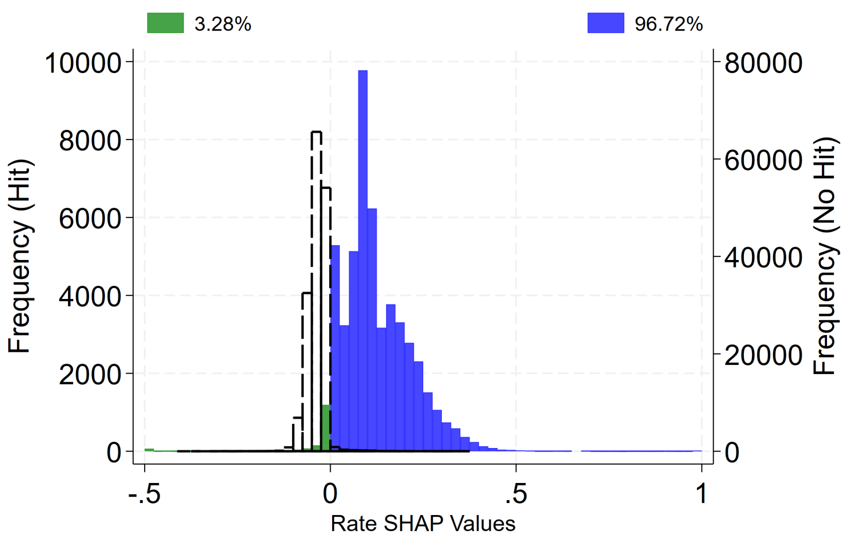
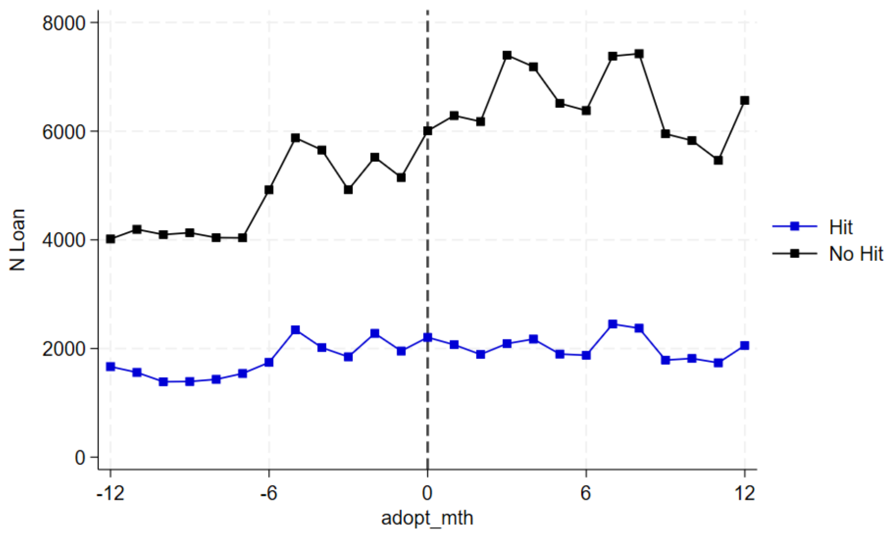


Figure 10. Distribution of SHAP Values for Alternative Credit Data Impact. Panel (a) plots the distribution of the sum of SHAP values for hit indicator and alternative score derived from the XGBoost model. The dashed black lines display the distribution for borrowers without alternative data hits, and the dashed vertical red line indicates their average SHAP value. Blue and green bars show the distribution for borrowers with alternative data hits - blue bars represent hit borrowers with SHAP values above the average for no-hit borrowers, and green bars represent those below this average. Panel (b) presents a similar analysis for loan interest rates.

(a) Number of loans by hit and no hit



(b) Hit ratio

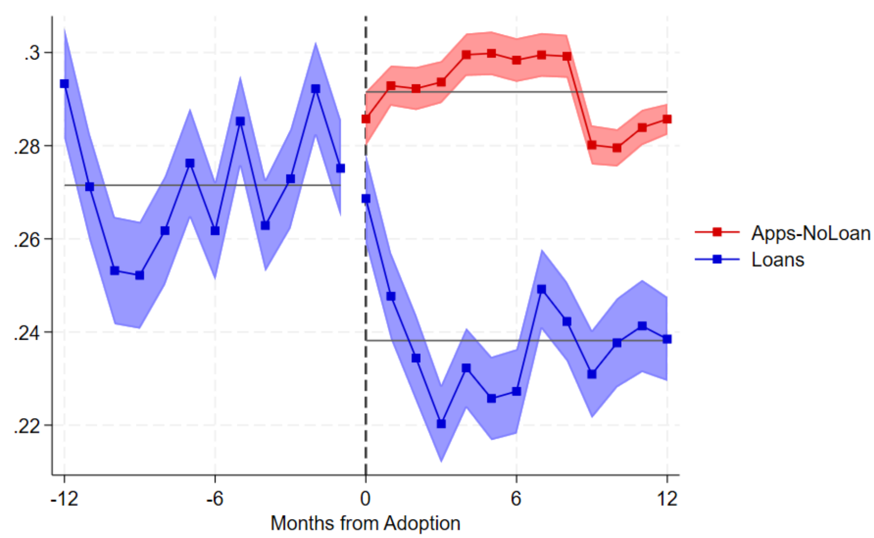


Figure 11. The Relationship Between Hit, Number of Loans, and Origination. Panel A shows the number of originated loans over time. Panel B plots the hit ratio for originated loans in blue and for applications that did not result in a loan in red. The hit ratio is calculated as the ratio of borrowers with a hit for each month relative to adoption. Lender B is excluded due to lack of data before adoption. The light blue and red areas shows the 95% confidence interval for the hit ratio, and the gray lines show the average values for pre and post period.

Table I. Summary Statistics.

This table reports summary statistics for loan applications and originated loans as well as zip code level demographics in our sample starting from April 2009 to April 2017. The number of observations, mean, standard deviation, and the minimum and maximum value for each variable are reported for the applicants.

	N Obs	Mean	SD	Min	Max
Application Data					
Hit	7,202,291	0.265	0.441	0	1
Alt Score	1,906,246	703.376	30.585	500	838
Vantage 4	6,959,134	555.343	63.197	300	850
Loan Data					
Hit	535,717	0.229	0.420	0	1
Alt Score	122,680	707.044	31.945	543	838
Vantage 4	529,237	555.388	56.792	300	848
Rate	308,779	22.338	5.587	0.010	42
Loan Amount ($K)	535,717	16.263	6.649	0.001	76.526
Loan Term	535,686	66.693	12.160	1	136
Delinquency	535,717	0.382	0.486	0	1
Zip Code Demographic Data					
ZipIncome (K)	25,417	68.213	29.458	5.704	441.278
Black	25,417	0.092	0.170	0	1
Poverty100	25,417	0.073	0.064	0	1
Foodstamp	25,417	0.135	0.099	0	1
PDshops	25,417	0.415	1.159	0	14

Table II. Alternative Score and Delinquency.

This table estimates panel regressions of loan delinquency on loan characteristics **for the hit population**. Columns 4-6 restrict the sample to three different terciles of Vantage 4 score, Column 4 representing the lowest and Column 6 representing the highest Vantage score. All regressions include lender-month and county-fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample	Full Sample	Full Sample	Low Vantage	Medium Vantage	High Vantage
AltScore STD	-0.0259***	-0.0228***	-0.0227***	-0.0188***	-0.0251***	-0.0255***
	(0.0013)	(0.0013)	(0.0013)	(0.0016)	(0.0019)	(0.0023)
Vantage 4 STD		-0.0568***	-0.0548***	-0.0392***	-0.0725***	-0.0592***
		(0.0021)	(0.0021)	(0.0044)	(0.0100)	(0.0064)
Loan Amount			-0.0017***	-0.0011*	-0.0014**	-0.0020***
			(0.0004)	(0.0006)	(0.0005)	(0.0006)
Loan Term			0.0052***	0.0047***	0.0057***	0.0053***
			(0.0003)	(0.0004)	(0.0004)	(0.0005)
ZipIncome			-0.0008***	-0.0008***	-0.0008***	-0.0010***
			(0.0001)	(0.0001)	(0.0002)	(0.0002)
Dep Var Mean	.455	.455	.455	.511	.444	.375
Lender-Month FE	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Observations	122283	121942	121929	48898	43603	28197
Adjusted R^2	.0863	.0942	.103	.11	.0825	.0793

Table III. Alternative Score and Loan Terms Before and After Adoption.

This table estimates OLS regressions **for the hit population**, where the dependent variables are the loan delinquency in Column (1) and loan interest rate in Column (2). **Lender B is excluded** due to lack of data in pre-adoption period. All regressions include lender-month and county fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1)	(2)
	Delinquency	Rate
AltScore-Pre	-0.0185***	0.0317
	(0.0020)	(0.0422)
AltScore-Post	-0.0280***	-0.1468***
	(0.0017)	(0.0192)
Post	0.0027	-0.0530
	(0.0142)	(0.2193)
Vantage 4 STD	-0.0582***	-0.4549***
	(0.0023)	(0.0725)
Loan Amount	-0.0020***	-0.1902***
	(0.0005)	(0.0086)
Loan Term	0.0054***	-0.0087
	(0.0003)	(0.0100)
ZipIncome	-0.0009***	0.0010
	(0.0001)	(0.0011)
Dep Var Mean	.458	22
Lender-Month FE	YES	YES
County FE	YES	YES
Observations	97921	51055
Adjusted R^2	.107	.4

Table IV. Presence in the Alternative Data, Delinquency, and Interest Rates.
This table estimates OLS regressions in the **post-adoption period** where the dependent variables are the loan delinquency in Columns 1 and 2 and loan interest rate in Columns 3 and 4. All regressions include lender-month and county fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1)	(2)	(3)	(4)
	Delinquency	Delinquency	Rate	Rate
Hit	0.0612***	0.0614***	0.3980***	0.3985***
	(0.0042)	(0.0041)	(0.0383)	(0.0386)
AltScore STD		-0.0309***		-0.0507***
		(0.0022)		(0.0123)
Vantage 4 STD	-0.0684***	-0.0677***	-0.6589***	-0.6577***
	(0.0012)	(0.0012)	(0.0348)	(0.0349)
Loan Amount	0.0001	0.0001	-0.1504***	-0.1504***
	(0.0004)	(0.0004)	(0.0057)	(0.0057)
Loan Term	0.0053***	0.0053***	-0.0042	-0.0042
	(0.0003)	(0.0003)	(0.0048)	(0.0048)
ZipIncome	-0.0010***	-0.0010***	-0.0056***	-0.0055***
	(0.0000)	(0.0000)	(0.0007)	(0.0007)
Dep Var Mean	.388	.388	22.5	22.5
Lender-Month FE	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Observations	384461	384461	213574	213574
Adjusted R^2	.0707	.0716	.457	.457

Table V. Alternative Data, Delinquency, and Loan Terms for the Ghost Sample.
This table estimates OLS regressions for borrowers without a Vantage 4 score in the **post-adoption period**, where the dependent variable are the loan delinquency in Columns 1 and 2 and loan interest rate in Columns 3 and 4. All regressions include lender-month and county fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1)	(2)	(3)	(4)
	Delinquency	Delinquency	Rate	Rate
Hit	0.1207***	0.1213***	0.6237***	0.6212***
	(0.0251)	(0.0254)	(0.2130)	(0.2137)
AltScore STD		-0.0281		0.2920
		(0.0275)		(0.1823)
Loan Amount	0.0061**	0.0062**	-0.1498***	-0.1499***
	(0.0024)	(0.0024)	(0.0222)	(0.0222)
Loan Term	0.0029***	0.0029***	-0.0158	-0.0160
	(0.0009)	(0.0009)	(0.0103)	(0.0103)
ZipIncome	-0.0012***	-0.0012***	-0.0062**	-0.0063**
	(0.0003)	(0.0003)	(0.0031)	(0.0031)
Dep Var Mean	.342	.342	24.2	24.2
Lender-Month FE	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Observations	5968	5968	3603	3603
Adjusted R^2	.0333	.0333	.335	.335

Table VI. Alternative Score and Demographic Characteristics.

This table reports OLS regression estimates **at the zip code level**, where the dependent variables are hitting the alternative data in Columns 1-3, and the alternative score for the hit population in Columns 4-6. All independent variables are standardized to have zero mean and unit standard deviation. 'Minority' denotes the standardized share of the non-white population, while race-specific variables represent the standardized share of each racial group. Standard errors are reported in parentheses. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1)	(2)	(3)	(4)	(5)	(6)
	Hit	Hit	Hit	AltScore	AltScore	AltScore
ZipIncome	-0.0015**	0.0007	0.0006	1.3692***	1.3976***	1.1822***
	(0.0007)	(0.0007)	(0.0008)	(0.0906)	(0.0923)	(0.1045)
Minority		0.0163***			0.2140***	
		(0.0007)			(0.0639)	
Black			0.0172***			-0.3971***
			(0.0007)			(0.0554)
American Indian			-0.0041***			0.2755***
			(0.0007)			(0.0749)
Asian			0.0043***			0.4361***
			(0.0007)			(0.0770)
Hispanic			0.0021**			0.0084
			(0.0008)			(0.1090)
Other			0.0058***			0.6228***
			(0.0007)			(0.0672)
Dep Var Mean	.228	.228	.228	703	703	703
Observations	23785	23785	23785	23785	23785	23785
Adjusted R^2	.000139	.0222	.0338	.0139	.0142	.0205

Appendix for:

"Borrowers in the Shadows: The Promise and Pitfalls of Alternative Credit Data"

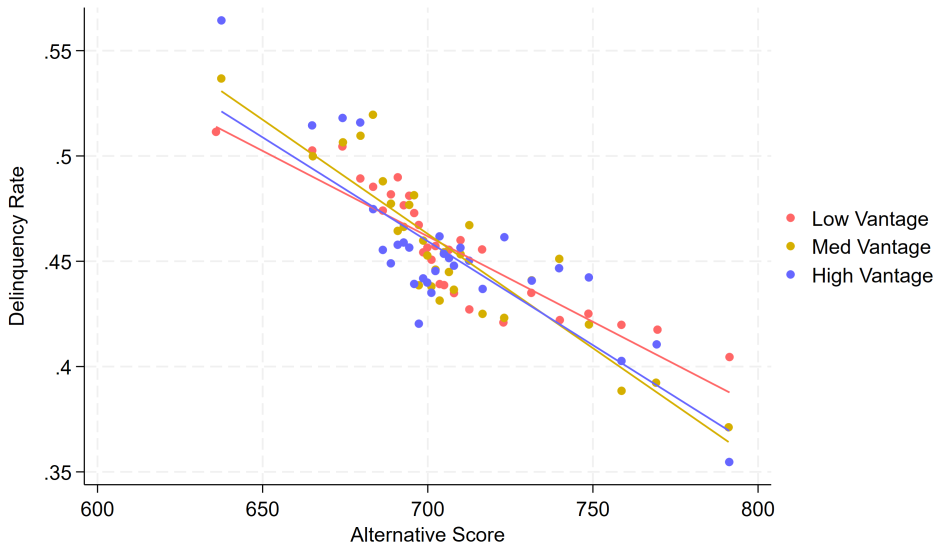


Figure A.1. Relationship between Loan Delinquency and the Alternative Score for Different Terciles of Vantage Score After Controlling for Vantage Score. This figure shows results similar to those in Figure 4 with the only difference being that the results also control for Vantage score. The figure shows the binscatter of loan delinquency over the alternative score for all loans in our sample. The red dots and lines correspond to the lowest tercile of the Vantage score, gold to the middle tercile, and blue to the highest tercile. The results control for the loan amount, loan terms, average zip code income, Vantage score and month-lender and county fixed effects. The horizontal dashed line shows the average value for applicants in each Vantage group who do not hit the alternative credit data.

Table A.1. Loan Characteristics Across Lenders. This table shows loan characteristics across different lenders in our sample. The table shows number of loans, hit ratio, average alternative score for borrowers who hit the alternative data, Vantage 4 score, and interest rate.

Lender	N	Hit	AltScore	Vantage4	Rate
A	358,893	0.21	708.52	560.36	20.57
B	92,974	0.26	704.75	539.29	24.98
C	29,786	0.21	707.74	574.90	17.86
D	24,307	0.36	702.55	528.99	22.19
E	22,738	0.30	706.68	551.97	31.27
F	7,019	0.29	698.55	530.01	22.68

Table A.2. Alternative Score and Delinquency - Excluding Lender A.

This table estimates panel regressions of loan delinquency on loan characteristics **for the hit population**. Lender A is excluded. All regressions include lender-month and county-fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1)	(2)	(3)
	Delinquency	Delinquency	Delinquency
AltScore STD	-0.0248***	-0.0214***	-0.0213***
	(0.0021)	(0.0021)	(0.0021)
Vantage 4 STD		-0.0465***	-0.0464***
		(0.0037)	(0.0037)
Loan Amount			0.0006
			(0.0005)
Loan Term			0.0030***
			(0.0004)
ZipIncome			-0.0006***
			(0.0001)
Dep Var Mean	.464	.465	.465
Lender-Month FE	YES	YES	YES
County FE	YES	YES	YES
Observations	47443	47257	47246
Adjusted R^2	.166	.172	.177

Table A.3. Alt Score and Loan Terms Before and After Adoption - Excluding Lender A. This table estimates OLS regressions **for the hit population**, where the dependent variables are the loan delinquency in Column (1) and loan interest rate in Column (2). Lenders A and B are excluded due to lack of data in pre-adoption period. All regressions include lender-month and county fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1) Delinquency	(2) Rate
AltScore-Pre	-0.0292***	0.0750
	(0.0033)	(0.0668)
AltScore-Post	-0.0242***	-0.2266***
	(0.0032)	(0.0602)
Post	-0.0222	-0.2339
	(0.0443)	(0.3836)
Vantage 4 STD	-0.0516***	-0.0563
	(0.0046)	(0.1255)
Loan Amount	0.0024***	-0.2573***
	(0.0008)	(0.0149)
Loan Term	-0.0001	-0.0150
	(0.0005)	(0.0213)
ZipIncome	-0.0006***	0.0041**
	(0.0002)	(0.0021)
Dep Var Mean	.487	23.3
Lender-Month FE	YES	YES
County FE	YES	YES
Observations	23313	20025
Adjusted R^2	.272	.403

Table A.4. Alternative Data, Delinquency, and Loan Terms - Excluding Lender A.
This table estimates OLS regressions in the **post-adoption period** where the dependent variables are the loan delinquency in Columns 1 and 2 and loan interest rate in Columns 3 and 4. Lender A is excluded. All regressions include lender-month and county fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1) Delinquency	(2) Delinquency	(3) Rate	(4) Rate
Hit	0.0531***	0.0533***	0.1452***	0.1460***
	(0.0044)	(0.0044)	(0.0433)	(0.0432)
AltScore STD		-0.0200***		-0.0650***
		(0.0035)		(0.0239)
Vantage 4 STD	-0.0593***	-0.0587***	-0.2398***	-0.2378***
	(0.0033)	(0.0032)	(0.0220)	(0.0222)
Loan Amount	0.0012***	0.0012***	-0.1422***	-0.1422***
	(0.0004)	(0.0004)	(0.0094)	(0.0094)
Loan Term	0.0035***	0.0035***	-0.0179***	-0.0179***
	(0.0003)	(0.0003)	(0.0060)	(0.0059)
ZipIncome	-0.0007***	-0.0007***	-0.0031***	-0.0030***
	(0.0001)	(0.0001)	(0.0006)	(0.0006)
Dep Var Mean	.348	.348	24.3	24.3
Lender-Month FE	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Observations	117698	117698	111419	111419
Adjusted R^2	.112	.112	.448	.448

Table A.5. Relationship between Alternative Score and Loan Rates for Terciles of Vantage Score. This table estimates OLS regressions in the post-adoption period, for three different terciles of Vantage 4 score, where the dependent variable is loan interest rate. Column 1 represents the lowest and Columns 3 the highest Vantage score. All regressions include lender-month and county fixed effects. Standard errors, reported in parentheses, are double-clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	(1)	(2)	(3)
	Low Vantage	Medium Vantage	High Vantage
AltScore STD	-0.0756***	-0.0757***	-0.1671***
	(0.0205)	(0.0215)	(0.0330)
Vantage 4 STD	-0.1898***	-0.5267***	-0.9174***
	(0.0509)	(0.1144)	(0.1227)
Loan Amount	-0.1390***	-0.1406***	-0.1520***
	(0.0110)	(0.0088)	(0.0103)
Loan Term	-0.0105	-0.0097	0.0071
	(0.0080)	(0.0081)	(0.0126)
ZipIncome	-0.0001	-0.0022	-0.0028
	(0.0017)	(0.0015)	(0.0019)
Dep Var Mean	24	22.9	21.4
Lender-Month FE	YES	YES	YES
County FE	YES	YES	YES
Observations	21995	17140	10833
Adjusted R^2	.342	.404	.439

Table A.6. County-Level Regression of Hit Ratio on Post Adoption.

This table estimates OLS regressions corresponding to the results in Figure 11 where the dependent variable is the average hit per county. The sample includes from 12 months before to 12 months after the adoptions. Column 1 includes county fixed effects, and Column 2 includes month and county fixed effects. Standard errors, reported in parentheses, are clustered by state and loan application year-month. * indicates 10% significance, ** indicates 5% significance, and *** indicates 1% significance.

	Avg Hit	Avg Hit
Post Adoption	-0.0192***	
	(0.0051)	
GotLoan		-0.0422***
		(0.0059)
Dep Var Mean	.211	.22
AdoptMonth FE	No	YES
County FE	YES	YES
Observations	33080	43494