NBER WORKING PAPER SERIES

MODERNIZING ACCESS TO CREDIT FOR YOUNGER ENTREPRENEURS: FROM FICO TO CASH FLOW

Christopher M. Hair
Sabrina T. Howell
Mark J. Johnson
Siena Matsumoto

Working Paper 33367
http://www.nber.org/papers/w33367

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025

We are grateful to Tetyana Balyuk, Brock Blake, Greg Buchak, Kelly Cochran, Brigham Frandsen, Ryan Gilland, Khasi-marc Jamieson, Melissa Koide, Jason Kotter, Dan O'Malley, Sandip Nayak, Sarah Mazurek, Ryan Metcalf, Pradeep Mittal, David Snitkof, Abby Sleight, Sam Taussig, Nishant Vats, and Jing Xue for support and comments. We also thank participants at Oxford, the NBER Summer Institute 2025, UNC Kenan- Flagler, Georgia Tech, the IMF-WIFPR Conference on Financial and Real Implications of Technologies, the GT - Atlanta Fed Household Finance Conference, the U.S. SBA, the GSU-MS AI and FinTech Conference, and UChicago Symposium on Private Firms. This project was partially conducted with data gathered for a project using Federal funds under award MB22OBD8020266-T1-01 to FinRegLab from the U.S. Department of Commerce. The conclusions are those of the authors and do not necessarily reflect the U.S. Department of Commerce or the National Bureau of Economic Research.

Modernizing Access to Credit for Younger Entrepreneurs: From FICO to Cash Flow
Christopher M. Hair, Sabrina T. Howell, Mark J. Johnson, and Siena Matsumoto
NBER Working Paper No. 33367
January 2025
JEL No. G21, G23, G32, J14, J16

ABSTRACT

This paper studies financial constraints over the entrepreneur's lifecycle in the context of small business lending. Underwriting relies on personal credit scores, which are backward-looking and favor long histories. In causal analysis, we show that younger entrepreneurs' loan applications have higher approval rates when the lender incorporates recent cash flows from business checking accounts. The result does not reflect lender risk-taking and is concentrated among lower-FICO applicants and the youngest entrepreneurs. Personal credit scores are both noisier and mechanically biased downward for younger entrepreneurs. Cash flow data provide an independent signal that reclassifies these entrepreneurs as less likely to default.

Christopher M. Hair
Brigham Young University
chris_hair@byu.edu

Sabrina T. Howell
Harvard University
Harvard Business School
and NBER
showell@hbs.edu

Mark J. Johnson
Brigham Young University
Department of Finance
markjjohnson@byu.edu

Siena Matsumoto
University of California, Berkeley
Haas School of Business
siena_matsumoto@berkeley.edu

1 Introduction

Personal and small business lending in the U.S. relies heavily on personal credit scores. While credit registries increase efficiency, they are imperfect (Einav et al., 2013; Blattner et al., 2022). This paper begins with the observation that credit scores are strongly correlated with age, far more so than with other protected class characteristics, such as gender or race. Credit scores tend to favor older people because they reward a long history of successful debt repayment, placing greater weight on the length of one's credit history than on individual repayment events. Fair lending laws discourage the direct use of age in underwriting, but permit favorable treatment for the elderly. These features suggest the possibility of a structural disadvantage for younger entrepreneurs. To our knowledge, we offer the first focused, empirical study exploring financial constraints by age.

A disadvantage for younger people in U.S. lending markets may help explain why entrepreneurs are most commonly in their 40s when they found their firms, despite evidence that latent demand is higher among younger people (Azoulay et al., 2020; Harris, 2024). We show that incorporating recent cash flow information from bank statements ("cash flow data") substantially reduces the mechanical bias against younger entrepreneurs imposed by credit score-based underwriting. Specifically, younger entrepreneurs applying for small business loans are more likely to be approved when the lender uses cash flow data alongside conventional metrics. A distributional analysis shows that this is because cash flow data disproportionately reclassifies younger entrepreneurs toward lower predicted default—not because young borrowers have stronger cash flows on average but because credit scores are mechanically biased downward and are noisier signals of repayment ability for younger applicants.

The context for our analysis is that in the 1990s, underwriting shifted from local loan officers using soft information to arm's length models that relied heavily on personal credit scores.[1] After the Financial Crisis, new fintech lenders exploited digitization and automation to incorporate cash flow information from bank statements into credit scoring models for both personal and small

[1] See Black and Strahan (2002); Petersen and Rajan (2002); Carter and McNulty (2005); Berger et al. (2005b). From the industry press, see Goldstein and Lauckner (2022).

business lending (FinRegLab, 2019). Unlike credit scores, which emphasize history and delinquency, cash flow data highlight recent account activity and repayment capacity. These new data are now being adopted more widely in conventional lending.[2] We hypothesize that incorporating cash flow data into underwriting models could mitigate the mechanical disadvantage facing younger entrepreneurs from lower or noisier personal credit scores.

While some banks have used transaction account data to monitor existing borrowers and predict delinquency (Mester et al., 2007; Norden and Weber, 2010; Khandani et al., 2010; Frost et al., 2019), they have rarely used it for arm's length underwriting of new small business applicants. Conversations with practitioners point to three reasons. First, transaction data was siloed across business units within banks. Second, the technology to systematically score bank statements only matured after banks had retreated from small business lending. And third, translating bank statement variables into compliance-approved default models is operationally costly.[3] New fintech lenders introduced cash flow variables because they did not face these frictions, they brought IT capability from Silicon Valley, and they needed new data to find creditworthy borrowers among those without strong bank relationships.

Small business offers an important setting. These firms depend heavily on debt, employ about half of private-sector workers, and account for nearly all U.S. firms (Robb and Robinson, 2014). Existing evidence that alternative data can improve lending technologies motivates our analytical strategy. Examples include the presence of digital (i.e., cashless) payments in China, India, and Mexico (Ouyang, 2026; Chioda et al., 2024; Ghosh et al., 2025), and alternative consumer characteristics such as education or phone device type in the U.S. and Germany (Di Maggio et al., 2022; Berg et al., 2020). Our setting differs in that we study arm's-length fintech lenders obtaining standardized bank statement data from third-party providers on applicants with whom they have no prior relationship. This closely resembles the Open Banking model in which consumer data on bank statements in particular flows to competing lenders (Babina et al., 2025).

[2]See Crosman (2024), and Lawler (2024). Also, in 2024 the big three credit bureaus introduced consumer scores that integrate bank statement attributes (Experian, 2024; VantageScore, 2024).

[3]As a former head of the U.S. Small Business Administration observes, "Surprisingly, large banks such as J.P. Morgan and Wells Fargo did not historically use bank account transactions in their underwriting and in many cases did not even have easy access to the information due to the siloed nature of the bank information systems" (Mills, 2024).

To our knowledge, this is the first paper to study this model at scale in the U.S.

We use data from three fintech companies serving small businesses between 2013 and 2024: two lenders (Lenders A and B) and one platform that routes applications to participating lenders.[4] From all three companies, we obtain standard underwriting inputs, including the owner's credit score as well as the firm's location, industry, size, and age. We also obtain measures about cash inflows and outflows constructed from three or six recent business checking account bank statements. These cash flow variables are standard among fintech lenders. They capture revenues, withdrawals, balances, volatility, and distress signals such as negative balances or insufficient funds transactions. We also obtain information they collect but which is not used in underwriting, such as the business owner's age, gender, and race. Importantly, we observe loan approval and origination outcomes for all three sources.

Our primary causal analysis uses the platform's data, covering about 900,000 applications-forwards routed to 111 lenders. Each application is routed to a subset of participating lenders based on lender-specified criteria observable to the platform, with a mean of five lenders per application and a median of three. Lenders A and B contribute loan performance data for the distributional analysis, and Lender B additionally provides a setting with true random assignment of applications to loan officers.

To establish the relevance of the cash flow variables, we examine whether and how they predict loan default and approval, after controlling for the credit score (FICO), industry, firm age, firm size, time, and state. Consistent with intuition, measures of business health and revenue—more inflows (i.e., credits) and higher balances—are associated with lower chances of default and higher chances of approval, while measures of distress—overdrafts, insufficient-funds events, and cash flow volatility—predict higher chances of default and lower chances of approval.[5] These relationships motivate our subsequent classification of lenders by their reliance on cash flow data and our distributional analysis of default reclassification.

[4]These companies have requested anonymity. There are no Paycheck Protection Program loans or otherwise subsidized or mission-driven loans in the sample.

[5]The relationships between cash flow variables and default are robust to controlling for the interest rate, and cash flow variables do not strongly predict interest rates.

Our main analysis causally identifies the effect of cash flow-based underwriting on access to credit and how it varies by entrepreneur age. We classify Platform lenders as cash flow-intensive (CFI) underwriters using a machine learning (ML) procedure: for each lender, we compare a Baseline model relying on FICO to a Cash Flow model that augments it with bank statement variables, and designate lenders for whom the Cash Flow model improves approval prediction as CFI (70% of lenders). We manually verify that CFI-classified lenders require bank statements as part of an application. Although applicants are not randomly assigned to lenders, applications on the Platform are routed to multiple lenders simultaneously, allowing us to use application fixed effects to absorb all observed and unobserved applicant characteristics (following Khwaja and Mian (2008)). Identification then comes from comparing how the same applicant is treated by CFI versus non-CFI lenders.

The potential concern with this empirical strategy is that the set of lenders to which an applicant is forwarded could be correlated with applicant characteristics in ways that interact with age. However, the results are quite similar when we restrict the sample to applicants forwarded to both CFI and non-CFI lenders, where forwarding-stage variation in lender type is fully within-application. Also, we observe all the variables that the Platform employs in its forwarding algorithm, which do not include age or other protected class characteristics, and we verify this empirically with formal balance tests. More broadly, a benefit of our within-application empirical strategy (where applications are forwarded by a third party to multiple lenders) is that it removes the powerful element of search from the approval process, in which borrowers who are riskier select into different search behavior (Agarwal et al., 2024).

We find that assignment to a CFI lender raises the chance of approval for younger entrepreneurs by 2.4 percentage points (12% of the mean) for those up to 40, with larger effects for the youngest entrepreneurs. Younger applicants assigned to CFI lenders also receive lower interest rates conditional on approval. These effects disappear among applicants with very high FICO scores (above 799), and are concentrated among those with poor or low scores (below 739). In this group, the effect is 16-18% of the mean.

In interpreting these results, one possibility is that CFI lenders approve younger applicants at higher rates because they take more risk, not because they can identify "diamonds in the rough" using alternative cash flow data. CFI lenders in our sample have higher APRs, consistent with this possibility. To assess it, we conduct three analyses. First, we show that approval by CFI, non-CFI, and both are equally predictive of business survival, and that being young and assigned to a CFI lender does not predict survival, in contrast to the negative relationship we would expect if risk preferences explained the main finding. To conduct this analysis, we collect data on business survival for a random subset of about 46,000 applications, since we do not observe loan performance on the Platform. Second, we show that the main effects are *larger* among especially conservative lenders, filtered either on approval rate or by the FICO score distribution of approved loans. Third, we replicate the main result using data from Lender B, where applications are randomly assigned to loan officers who vary in their reliance on cash flow information. This provides a setting with true random assignment that is free of any concerns one might have with the lender matching process. In sum, differential risk-taking does not explain the results.

We next ask why cash flow data benefits younger entrepreneurs. If lenders minimize default risk, approving more young applicants under a Cash Flow model implies that cash flow variables reclassify young borrowers toward lower predicted default probabilities. Two distinct mechanisms could produce this pattern. First, FICO may be a noisier signal of repayment ability for young borrowers, in which case cash flow data adds information where FICO is least informative. Second, cash flow data may alleviate a bias induced by FICO by revealing that young borrowers are, on average, less risky than their FICO scores suggest. We find evidence consistent with both.

To assess the noise channel, we train models of default on data from Lenders A and B and compare predictive performance. A model using FICO and standard controls performs substantially worse for young applicants than for older applicants. Adding cash flow variables closes this gap entirely for young applicants while leaving performance for older applicants essentially unchanged. For example, adding cash flow variables increases the ROC-AUC by 28% relative to the random-guess benchmark for entrepreneurs up to 35, but by only 4.7% for owners over 35. This is the pattern we would expect if FICO were noisier for the young and cash flow data provided orthogonal information. Under information frictions of the kind described by Stiglitz and Weiss (1981), less noisy screening at the margin translates into expanded credit access for marginal applicants who would otherwise be rationed out, consistent with our finding that effects are largest among low-FICO applicants and conservative lenders, where rationing is most likely to bind.

To explore the bias channel, we follow Fuster et al. (2022) and compute, for each applicant, the change in predicted default probability when moving from the Baseline to the Cash Flow model. Plotting the cumulative distribution of these changes separately by age, we find that younger entrepreneurs are disproportionately represented among those reclassified toward lower predicted default. Within low-FICO groups, younger and older applicants exhibit similar cash flow profiles, which means the reclassification is not driven by young borrowers having objectively stronger cash flows. Instead, the patterns are consistent with FICO assigning lower scores to young applicants than their repayment behavior warrants, a mechanical age bias that cash flow data partially corrects. On average, this correction lowers predicted default by 0.48 percentage points (3% of the mean) for entrepreneurs younger than 35. To put the magnitude in perspective, Dobbie et al. (2020) attribute a 9% decline in predicted default to the removal of bankruptcy flags from credit reports.

In sum, we show that incorporating cash flow data into U.S. small business underwriting expands access to credit for younger entrepreneurs because it compensates for the noise and mechanical bias they face in conventional credit scoring. Overall, our results suggest that adding cash flow data is not zero sum; while there is a benefit for younger entrepreneurs, there is no apparent negative impact for older entrepreneurs. This is consistent with Stiglitz and Weiss (1981)-style rationing among the young population in the absence of cash flow data, while credit scores offer more accurate information for older entrepreneurs.

Our paper contributes to literature and policy along several dimensions. The first dimension is the role of age in entrepreneurship. On the one hand, theory and cross-country demographic studies suggest that younger populations should yield more entrepreneurship (Levesque and

Minniti, 2011; Liang et al., 2018; Hopenhayn et al., 2022). Yet there is somewhat contradictory evidence that in the U.S., entrepreneurs skew older (Zhang and Acs, 2018; Azoulay et al., 2020). At the same time, personal credit scores are known to correlate with age and, theoretically, may understate the creditworthiness of young consumers (Albanesi and Vamossy, 2024; Bakker et al., 2025; Chatterjee et al., 2025). We help to link together existing findings by providing the first focused, causal empirical analysis of financial constraints by age, and are the first to study the issue among business owners.

The second dimension is around the use of alternative data in lending technologies. As noted above, Ouyang (2026), Ghosh et al. (2025), and Chioda et al. (2024) show in settings without well-developed credit registries how digital (i.e., cashless) payments can improve access to credit. Other papers focus on alternative underwriting inputs such as device type and other web-based footprints, or characteristics such as education (Berg et al., 2020; Di Maggio et al., 2022). A larger literature studies how fintech lenders have expanded access to credit, especially to underserved geographic areas.[6] We build on these papers in two ways. First, we identify causal effects on access to credit using within-application variation across competing lenders. That is, we provide causal evidence on how alternative data affects access rather than focusing on changes in predictive performance. Second, this is the first paper to study how cash flow data from *bank statements* affects access to credit, and we do so in the U.S. context, where bank statement data is the central source of alternative data in lending.

This point leads to our final contribution, which is to inform ongoing policy debates about Open Banking. Open Banking policies enable consumer-authorized bank data to flow to third-party lenders (He et al., 2023; Alok et al., 2024; Babina et al., 2025). Prior cross-country evidence suggests Open Banking can improve credit access (Rishabh, 2024; Agarwal et al., 2021), but evidence on which populations benefit is limited and context-dependent. We document a specific benefit to young business owners who have lower FICO scores.

[6]See, among others, Jagtiani and Lemieux (2018); Agarwal et al. (2021); Bartlett et al. (2022); Johnson et al. (2023). Berg et al. (2022) provide a literature review.

2 Institutional Background

This section provides institutional context for our analysis. We first describe how U.S. small business underwriting came to rely heavily on personal credit scores. We then describe the emergence of cash flow–based underwriting and the reasons traditional lenders did not adopt it earlier. Finally, we discuss the relationship between personal credit scores, entrepreneur age, and the fair lending regulatory environment.

2.1 Small Business Underwriting and the Rise of Personal Credit Scores

Small business lending is important for job creation and economic mobility. However, it is difficult to underwrite (i.e., assess credit quality) in this population because small firms are heterogeneous even within narrow sectors, often rely on informal accounting practices, and the loans are too small to merit substantial firm-specific due diligence (Mills and McCarthy, 2016). Traditionally, small and regional banks were the primary source of small business loans, overcoming information asymmetry through soft information gathered by loan officers in long-term relationships (Petersen and Rajan, 1994; Berger and Udell, 1995). As banks consolidated in the 1990s and 2000s, large banks expanded their market share and increasingly employed standardized risk scoring (Black and Strahan, 2002; Berger et al., 2005a; Strahan, 2017). This shift made lending more arm's-length and made personal credit scores a key underwriting input (Arora, 2023; Carter and McNulty, 2005).

Following the Financial Crisis, higher capital requirements led banks to retreat from small business lending (Chen et al., 2017). Remaining bank lending was held to high underwriting standards, requiring income measures from annual financial statements and tax documents.[7] Firms lacking strong historical financials were left underserved (Cole and Damm, 2020), and small business credit cards became the most common form of financing for small firms (Fed, 2024).[8] The retreat of large banks from arm's-length, non-credit-card small business lending

[7] Based on conversations with industry experts, including Dan O'Malley, the Founder and CEO of Numerated.
[8] Credit card underwriting also typically does not consider bank account information; see NerdWallet (2024).

created a gap that new fintech companies entered in the early 2010s, exploiting digitization and automation to lend at arm's length.

2.2 Cash Flow–Based Underwriting

The most important new input that fintech lenders introduced was information from recent bank statements, which we refer to as *cash flow data*. Bank statements offer a direct measure of repayment ability; that is, whether the applicant can service debt based on recent inflows, outflows, and signs of distress. Three features make cash flow data particularly well-suited to small business underwriting at scale. First, bank statements are nearly universal: practically every business in developed economies has a checking account. Second, statements are relatively standardized, offering comparable information across businesses that are otherwise *sui generis*. Third, simple inflows and outflows are less correlated with socioeconomic factors than other alternative inputs such as education or spending habits.

Fintech lenders typically rely on a small set of variables derived from bank statements. While the specific inputs and modeling strategies vary across lenders, conversations with industry practitioners suggest that most use fewer than 10 basic measures: credits and debits to capture inflows and outflows, balance levels and volatility to capture financial stability, and indicator variables flagging events that signal distress, such as insufficient-funds transactions or overdrafts.[9] Fintechs initially obtained these data by pulling them electronically from PDFs provided by applicants or by "screen-scraping" bank websites. These processes have since been digitized and commodified by data aggregators such as Plaid, Yodlee, and Ocrolus.

A natural question is why traditional lenders, who often already held the borrower's transaction account, did not adopt cash flow–based underwriting earlier. Banks have long used client cash flow information for relationship-based loans, and a separate literature shows that transaction account data improves monitoring and default prediction for existing borrowers

[9] Based on conversations with industry experts, including Venkatesh Bala (Biz2Credit), David Snitkof (Ocrolus), and Sam Taussig (formerly Kabbage and AmEx).

(Mester et al., 2007; Norden and Weber, 2010; Khandani et al., 2010; Frost et al., 2019). But systematic use of these data for *arm's-length* underwriting of new applicants has been rare. Conversations with practitioners point to three reasons.[10] First, transaction data was siloed across business units. Retail deposit operations maintained data that was separate from underwriting decisions made in commercial lending, and bank information systems were not designed for cross-unit access.[11] Second, the technology to digitize and systematically score bank statements only matured after banks had already retreated from small business lending. Third, translating bank statement variables into compliance-approved default models is operationally costly, since existing scoring infrastructure was built around credit bureau inputs and validating new variables for fair-lending review adds model-risk and a burden on documentation. Fintechs, building from scratch and serving applicants without prior commercial lending relationships, faced none of these frictions.

Cash flow data are now beginning to move into conventional lending. In May 2024, both Experian and VantageScore launched consumer credit score products that incorporate bank statement–based attributes, marketed as targeting thin-file consumers.[12] More broadly, fintech lenders' demand for customer-permissioned data sharing has driven the Open Banking movement: Babina et al. (2025) document that by 2021, 80 countries had taken steps toward Open Banking, and in 2024 the U.S. established a regulatory structure under CFPB Section 1033 set to take effect between 2026 and 2030.

2.3 Personal Credit Scores, Age, and Fair Lending

Personal credit scores rise systematically with age. In our data, FICO scores rise almost linearly from below 670 for entrepreneurs under 30 to over 720 for entrepreneurs over 70 (Figure 1).

[10] Based on conversations with industry experts including Sam Taussig (formerly Kabbage and AmEx), David Snitkof (Ocrolus), Kelly Cochran (FinRegLab), Dan O'Malley (Numerated), and Karen Mills.

[11] As a former head of the U.S. Small Business Administration observes, "Surprisingly, large banks such as J.P. Morgan and Wells Fargo did not historically use bank account transactions in their underwriting and in many cases did not even have easy access to the information due to the siloed nature of the bank information systems" (Mills, 2024).

[12] Experian's Cashflow Attributes leverages checking and savings account information; VantageScore's product is marketed as using "consumer-permissioned Open Banking data." See Experian (2024) and VantageScore (2024).

National data exhibit the same pattern. Average FICO scores peak near 757 for individuals in their 80s, with the steepest gains in the 50s and 60s (Appendix Figure A.1). This relationship is mechanical. FICO assigns 15% of its score to length of credit history and an additional 10% to credit mix which are categories that, by construction, take time to accumulate.[13] Younger applicants therefore receive lower scores not because they have demonstrated worse repayment behavior, but because they have had less time to demonstrate creditworthiness. Chatterjee et al. (2025) provide a theoretical foundation for this relationship. In a model with hidden information about borrower type, credit scores rise with age because the market gradually learns each borrower's repayment propensity.

There is some evidence from the literature on the implications for consumer debt. For example, Agarwal et al. (2009) show that among prime borrowers, middle-aged individuals borrow at lower interest rates than younger people after controlling for observable risk characteristics. Albanesi and Vamossy (2024) show that using the same inputs, ML models outperform traditional credit scores particularly for young, low-income, and minority populations. Regarding entrepreneurship, Levesque and Minniti (2011) theorize that while younger individuals have substantial entrepreneurial potential, they face systematically tighter capital constraints than older entrepreneurs. These constraints may be reinforced by other obligations. Krishnan and Wang (2019) document that outstanding student loan balances raise the cost of business failure and reduce access to credit, a channel that weighs particularly heavily on younger entrepreneurs. Survey evidence indicates that latent entrepreneurial demand is concentrated among the young; 71% of 18–34-year-olds report wanting to start a business, compared to 53% of those aged 45–54 (Harris, 2024). Yet, entrepreneurs in practice are most commonly in their 40s when they found firms (Azoulay et al., 2020). The gap between latent demand and observed entry is consistent with binding financial constraints among the young.

Fair lending regulation reinforces this dynamic by treating age asymmetrically. The Equal Credit Opportunity Act prohibits discrimination on the basis of age, but the implementing

[13] The remaining categories are payment history (35%), amounts owed (30%), and new credit (10%, which penalizes recent applications). See `https://www.myfico.com/credit-education/whats-in-your-credit-score`.

regulation explicitly permits scoring systems that favor older applicants. The CFPB's official commentary states: "Favoring the elderly: Any system of evaluating creditworthiness may favor a credit applicant who is age 62 or older" (CFPB, 2024). Lenders may use age directly in underwriting models only if they meet additional disclosure and oversight requirements. The result is an environment in which credit scoring methodology can produce age-correlated outcomes that are permitted by design, while adjustments that would benefit younger applicants face substantial regulatory frictions. This asymmetry is the institutional foundation for the empirical patterns we document.

3 Data Sources and Summary Statistics

3.1 Data Sources

We use data from three fintech companies. Two are non-bank, online lenders and one is a platform. These companies have requested anonymity. All three exclusively serve U.S. small businesses and request cash flow variables from bank statements as part of the application process.

The Platform. The Platform is a fintech company that connects prospective small business borrowers with lenders. Businesses apply through the Platform's website, which routes them to one or more of its 111 partner lenders. All applicants provide the same, standardized set of data as part of the initial application process, and all of these data are forwarded to the lenders. The lenders include fintechs and conventional banks. In our baseline data for the Platform, an observation is a forward from the platform to a lender. Each application is forwarded to an average (median) of 5 (3) lenders. Lenders provide the Platform with simple criteria, such as FICO or revenue thresholds. Conditional on meeting the lender's requirements, the Platform sends applicants to a set of lenders based on an internal matching algorithm. While we do not observe the algorithm, we know that it does not use age or other protected class indicators. The lenders then decide whether to approve the loan and if so, on what terms. We observe these offers, which we describe as "approval" decisions. An applicant may be approved by multiple lenders and may ultimately accept at most one of the offers, which we also observe.

The Lenders. We obtain comprehensive loan application, approval, origination, and performance data from two lenders, which are reasonably similar in the applicant characteristics they consider and the loans they underwrite. We call one "Lender A" and the other "Lender B." Lender A provides both term loans and lines of credit, while Lender B only provides term loans.[14] Lender A's underwriting process is highly automated and it uses only a risk scoring model. Lender B incorporates discretionary judgment by loan officers. Specifically, Lender B's underwriting process includes a manual component, with each case assigned to an individual underwriter. Lender B also screens out subprime applicants (who have a FICO score below 660) in an initial step that we do not observe, contributing to a higher approval rate in our Lender B data.[15]

Data Types and Collection. We obtained three types of data from all three companies: applications, third party underwriting inputs, and loan characteristics. Information about the applicant firm and owner is collected from the firm directly, such as owner name, address, owner age, firm founding date, and industry.[16] The owner first name is important for us to construct a gender indicator. Note that not all of this application data is used in underwriting (e.g., it would be illegal to use gender). We also observe underwriting inputs collected from third parties, specifically the FICO score from a credit bureau and bank statement information, provided by Plaid (Lender A) and Ocrolus (Lender B and the Platform).[17]

[14] We control for loan type in our analysis but do not find it has any material effects on the results.

[15] Lenders A and B as well as most lenders on the platform require a personal guarantee and take out a blanket UCC lien against business assets. Their loans are described as collateralized with non-revenue, non-real estate business assets, but in practice they do not verify collateral or secure the loan with specific collateral.

[16] Industry is in the form of NAICS codes for both Lenders A and B, and industry categories for the Platform. These are aggregated to 17 NAICS industries appearing in our data. We group sectors "Management of Companies and Enterprises" and "Administrative and Support and Waste Management and Remediation Services" with "Professional, Scientific, and Technical Services".

[17] In the latter case, we employ the original JSON files containing bank statement variables that are gathered from the Ocrolus API. Lender A and the Platform acquire three months of bank statements. Lender B acquires six months. Industry practice is typically three months, and rarely more than six.

For all three companies, we observe whether a loan was approved, information on the offered loan terms—amount, maturity, and interest rate—and whether the loan was ultimately originated. Since we observe performance for the two Lenders but not for the platform, we collect firm survival data for a subset of the Platform applicants (introduced below). There are 104,150 applications and 18,434 loans for Lender A, from November 2013 to June 2024. For Lender B, there are a total of 58,668 applications and 25,762 loans, spanning August 2015 to May 2024. Finally, we observe 199,300 applications, 904,471 application-forwards, 191,421 offers, and 35,821 originated loans on the Platform, spanning January 2022 to December 2023. Across all lenders, there are 262,723 unique firms.

3.2 Summary Statistics

We present summary statistics about the full analysis sample of applicants and borrowers in Table 1. We use the applicant data from all three companies to analyze approval decisions, and borrower data from the two lenders to analyze default. The first set of columns combines applicants from all three companies. Since the Platform sends each application to multiple lenders, the unit of observation is an application-forward, so a Platform applicant can appear in this sample more than once. The second set of columns reports the borrower-level data we use to analyze default and is restricted to Lenders A and B, where we observe loan performance. Variables are defined in Appendix B.

Loan Approval and Performance. The first set of variables focus on underwriting activity. The average requested and originated amounts are between $88,000 and $144,000 across the different samples. The overall approval rate is 25% and the origination rate is 7.5%. Originated loans average $88,000 with maturities of 2.4 years in the full sample, and $115,000 with maturities of 3.2 years among Lenders A & B. We construct an indicator variable for a non-performing loan (which we interchangeably call "default"). It takes a value of one if the loan is charged off, more than 60 days past due, or the borrower has received forbearance and a modified loan. Lenders A and B have a 17% default rate.

The average offered APR in the pooled sample is 66%, which is much higher than the 16% among originated loans. The higher rates for approved applicants reflect both differences in the types of lenders doing business on the platform as well as offers that are rejected.[18] While some lenders on the platform provide term loans comparable to those offered by Lenders A and B, others focus on short-term loans with payments via automatic ACH transfers, which have very high annualized interest rates.[19] Figure A.2 shows the number of applications over time for each of the three data sources.

FICO and Bank Statements. The average applicant has a FICO score of 687, which qualifies as Prime (CFPB, 2020). Borrowers from Lenders A and B have a higher average FICO of 728, reflecting the screening from underwriting and the selection of applicants into these loans. Bank statement variables are averaged across the three months prior to application. Credits (i.e., inflows) average $100,000 for applicants and $131,000 for borrowers. They decline slightly when adjusted for new debt. Withdrawals are higher than overall credits. The average balance in the account is $33,000 for applicants and $42,000 for borrowers. There are three variables that may signal distress. One is the number of overdrafts or insufficient funds (NSF) transactions, which occur when the balance falls below zero.[20] The second is the number of low or negative balances, which occur when the balance goes below zero or below $1,000. The third is the number of daily pay loans, which are merchant cash advances (MCA) that typically have very high interest rates and are reflected in the bank statement by daily withdrawals to the MCA lender. We also observe the standard deviation of credits and balances for Lenders A and B, which reflect operational volatility.

[18]There are two types of loans: Lines of credit and term loans. These compose 41% and 59% of the loans, respectively. In our analysis, we combine them and include a control for loan type.

[19]ACH loans are often quoted in factors. For example, a 1.3 factor means that the borrower must repay $1.30 for each dollar borrowed. When payments are deducted daily over a 3 month window, this would amount to an APR above 120%.

[20]Some banks impose an insufficient funds fee and reject the transaction, while others permit the account to go negative and impose an overdraft fee.

Business Characteristics. We observe key dimensions from an underwriting perspective beyond financial information. Our proxy for firm size is employment. Applicants average about seven employees and borrowers average about nine. The firms operate in diverse industries, with the highest concentrations in Construction and Retail Trade (see Table A.1). The average age of applicant firms is seven years, while it is 10 years among borrowers.

Entrepreneur Age. We measure owner age using the date of birth of the primary owner. The median age on the Platform is 44, close to the 42 years reported by Azoulay et al. (2020) for all new companies in the U.S. For this reason, we focus on a threshold of 40 years to define “young" entrepreneurs, but also present results using thresholds of 35, 45, and 50.[21] Figure 2 shows that younger entrepreneurs appear to be at a disadvantage. They have a 21.5% loan approval rate, compared to 25% for older owners (Panel A), yet among borrowers they have identical default rates (Panel B).[22] Figure 3 shows that the chance of approval is strongly increasing in entrepreneur age and has a much steeper slope for younger applicants.

The disadvantage is largely tied to FICO. Younger applicants score 17 points lower on average—or 25% of a standard deviation—yet once they borrow, default rates are similar (Figure 2 Panel C). By contrast, cash flow metrics differ little by age. As one example, Figure 2 Panel D shows the number of insufficient funds transactions. We see that while approval clearly selects on this variable, there is little difference across age groups. Similarly, Table 2 shows that the disparity in credits is just 9% of the standard deviation, and the distress proxies are essentially the same across the age groups.[23] Table 2 also indicates that young owners are more likely to have younger firms (4.6 vs. 9 years) and smaller firms (6.3 vs 7.8 employees), but are otherwise similar across gender, socioeconomic status (determined by location), and business survival outcomes.

Finally, among low-FICO applicants (< 740), younger entrepreneurs have comparable or

[21] Figure A.3 shows histograms of age in the pooled applicant (Panel A) and originated loan (Panel B) samples.

[22] Panel A uses data from all three sources; Table A.2 shows a similar pattern in the Lender A & B data, where 35% of young applicants are approved, compared to 43% of older owners.

[23] In Table A.2, we present these statistics for the Lender A and B subset. We see similar patterns across young and old, even though the applicant pool is overall lower risk. The average FICO is 714 for young owners vs. 731 for mature owners, while the cash flow variables are generally similar.

stronger cash flow profiles, which we document in Table A.3. For example, younger owners have fewer overdrafts and MCA loans, which are negative signals. They have similar balances, and while they have somewhat lower inflows, older owners have much higher withdrawals. This suggests that cash flow underwriting may be especially valuable in identifying creditworthy but FICO-constrained young owners.

Selection into the Data. The applicants in our data are selected relative to the overall population of U.S. small businesses, in that they chose to apply to an online fintech lender. However, they are observably representative along important dimensions. For example, the average small firm in the U.S. has 11.7 employees as of 2020, which is similar to 10 employees in the Lender data and 7 in the Platform data.[24] Women make up 21.7% of employer firm owners, which is very close to the figures in our data. Table A.1 compares the industry composition with the national data on employer firms with 1-19 employees from the U.S. Census Bureau.[25] The data are overall fairly representative. As mentioned above, the distribution of entrepreneur age is also representative.

4 Effect of Cash Flow Information on Credit Access

In this section, we study how exposure to cash flow-based underwriting affects access to credit. We proceed in three steps. First, we establish that cash flow variables predict approval decisions on average and that the intensity of their use varies substantially across lenders, motivating our classification of lenders as cash-flow intensive (CFI) or not (Section 4.1). Second, leveraging this observed variation in lenders' cash flow intensity, we use a within-application design to provide causal evidence on the relationship between cash flow-based underwriting and access to credit for younger vs. older entrepreneurs (Section 4.2). Third, we address potential challenges to causal interpretation of these results, including presenting a second causal design using data from Lender B, where applications are randomly assigned to individual loan officers who vary in their use of

[24] See `https://advocacy.sba.gov/wp-content/uploads/2023/03/Frequently-Asked-Questions-About-Small-Business-March-2023-508c.pdf`.

[25] Available here `https://www.census.gov/programs-surveys/susb/data.html`.

cash flow information.

4.1 Cash-Flow Variables and Loan Approval

We first examine the degree to which cash flow variables explain loan approval decisions by constructing basic underwriting models with and without these variables. We do this for the overall sample and then for each Platform lender, forming the basis for our empirical design. The Baseline model predicts a loan approval indicator using FICO, firm size (employees), firm age, and fixed effects for industry, state, loan type, and calendar quarters. The Cash Flow model is identical, but adds a vector of bank statement metrics: average credits, debits, balances, volatility, and distress flags for overdrafts and MCA withdrawals. Table A.4 contains a list of features included in each model.

The random forest ML algorithm that we implement (Ho, 1995), like OLS and other classical models, aims to minimize prediction error.[26] There are three steps to the estimation process. First is the selection of "hyperparameters," which set higher-order model behavior.[27] To select hyperparameters, we follow the standard technique in the ML literature of "tuning" (Probst et al., 2019). This is a Bayesian optimization procedure where hyperparameters are selected to maximize model performance (AUC-PR) across random subsamples of the data using k-folds cross-validation.[28] In the second step, we construct training, validation, and testing samples by randomly selecting loans from the entire dataset.[29] The training dataset is comprised of 80% of the sample, while the remaining 20% is used for a testing dataset (or "holdout sample"). Only the

[26]Like other recent research on ML models in finance (e.g., Fuster et al. (2022)), we focus on random forest models. In our setting, we find that random forests performs similarly to other modern ML methods like XGBoost and are much faster to train and considerably easier to implement, with a smaller number of hyperparameters and less tendency to overfit.

[27]This is similar to the way an econometrician must select the bandwidth in a regression discontinuity design or the number of lagged periods in a vector autoregressive model. Examples of hyperparameters for random forests are the total number of trees in the forest, how many variables each tree considers simultaneously, and the number of samples required to split an internal node of a tree.

[28]Research in the field of ML presents this technique as a robust hyperparameter selection strategy. See, e.g., Kohavi et al. (1995); Snoek et al. (2012); Turner et al. (2021)

[29]We split the data with stratification on the outcome variable, which ensures the same proportion of each outcome class in the training dataset and testing dataset.

training dataset is used to fit the model, while the holdout sample is used for prediction. Within the training dataset, 20% of observations are used for a validation process that fine-tunes the model and helps to reduce overfitting. The third step is to estimate the model many times; "bootstrapping" obtains average estimates and standard errors for the model. This estimation method is explained in more detail in Appendix C.

Across all three datasets, we find convincing evidence that on average, cash flow variables are used in underwriting. In Table 3, we show the OLS regressions in Panel A and the performance metrics from the ML models in Panel B. Moving from the Baseline to Cash Flow model improves the OLS R^2 by 9.4%. Using the more flexible ML model, the ROC-AUC increases from 0.803 to 0.816, an improvement of 4.3% relative to the random-guess benchmark.[30] Column 2 in Panel A shows that cash flow variables predict approval in expected directions. For example, more credits to the account and higher bank balances are associated with higher chances of approval, whereas more withdrawals and higher volatility in revenues are associated with lower chances of approval. Daily pay loans and the number of days with low or negative balances also predict lower approval. Importantly, the coefficients and R^2 are similar when we re-estimate the model separately by age (Columns 3 and 4), indicating that the CF-Young interactions we exploit later do not arise from lenders running age-specific models.

While these results indicate that the lenders on average use cash flow variables, there is significant heterogeneity. To quantify it, we re-estimate the random forest ML model for each lender, and examine the improvement in ROC-AUC between the Baseline and Cash Flow models. Figure A.4 depicts the distribution of ROC-AUC improvement. The improvement spans a broad range and includes some negative values, which we interpret as evidence that for those lenders, cash flow variables add no genuine predictive power beyond the Baseline—the Cash Flow model overfits the training data and underperforms out of sample. Some lenders may not use cash flow variables because their training samples are too small to support high-dimensional models without overfitting, they lack the sophistication to recognize their importance, or they rely on

[30] The calculation is (0.816-0.50)/(0.803-.50) where both models are benchmarked against an ROC-AUC of 0.5–the random-guess model.

alternative technologies. Our results are consistent with the Platform's assertion that it works with both conventional and fintech lenders.

4.2 Within-Application Assignment to Cash Flow-Intensive Lenders

Having established that lenders vary in how intensively they use cash flow data in underwriting, we turn to estimating the causal effect of assignment to a relatively more cash flow intensive (CFI) lender. Our identification strategy exploits within-application variation, following Khwaja and Mian (2008). The Platform typically forwards the same application to multiple lenders—an average of five and median of three at the unique application level—which allows us to compare how the same applicant is treated across CFI and non-CFI lenders. By including application fixed effects in our regressions, we control for all observed and unobserved applicant risk characteristics, effectively isolating lender-side differences in underwriting practices. A lender that emphasizes cash flow metrics will estimate a different level of risk for the very same applicant than a lender who does not. The within-applicant design tests whether the variation translates into different approval outcomes for younger vs. older business owners.

CFI Classification. As described in Section 4.1, we classify lenders as CFI if the lender-specific Cash Flow model produces a positive ROC-AUC improvement over the Baseline; 70% of lenders meet this threshold.[31] The 30% of lenders that do not meet the threshold serve as the comparison group—the non-CFI lenders. Summary statistics on the random forest model performance by CFI status are included in Table A.5 at the lender and application-forward levels. We see higher improvements in all performance metrics for CFI lenders.

To verify that CFI-classified lenders actually use cash flow data in underwriting, we clerically researched the application and underwriting processes for the subset of lenders whose names the Platform disclosed rather than identifying only by ID number (53% of lenders). All but one of

[31]We exclude lenders receiving fewer than 50 application forwards or with an approval rate above 95%, since the random forest classifier requires sufficient within-lender variation. Further details about the methodology for constructing the cash flow–intensity measure are in Appendix C.

these lenders required bank statements as part of the application, either by submitting PDFs or by integrating with bank-account aggregators such as Plaid. The lender that did not require statements no longer served small businesses at the time of our search.[32] We also verified that some of the non-CFI lenders explicitly do not require bank statements, though the Platform sends all the lenders the same cash flow variables, which they may or may not use.

Summary statistics at the lender level by CFI status are in Table 4. Columns 1-4 report statistics for all Platform lenders, Columns 5-8 separate by CFI status, and the final column reports the difference in means. CFI lenders extend offers with substantially higher interest rates and shorter maturities, consistent with the Platform serving both fintech lenders and more conventional lenders that focus on longer-term, lower-rate credit. Applicants are similar across the two groups: those forwarded to CFI lenders have slightly lower FICO scores (683 vs. 697), but the difference is significant only at the 10% level, and other applicant characteristics are statistically indistinguishable. We expand on this further below, where we explore whether the results reflect risk-taking, but it is worth noting that this difference is consistent with fintech lenders targeting borrowers with positive cash-flow attributes that mitigate the risk implied by their credit scores. We include statistics that confirm this pattern at the Application level in Appendix Table A.6. Here, the differences between the columns are almost all statistically significant because of the large sample. From an economic perspective, the borrower characteristics and cash flow variables are very similar. We show this in the final column, where we scale the difference by the pooled standard deviation. The scaled differences are small for borrower risk measures. For example, the difference in FICO scores is just 0.02 of a standard deviation. However, the CFI lenders have higher approval rates and interest rates. Below, we ensure that the results do not reflect endogeneity in the set of lenders to which an applicant is forwarded.

[32] Examples of evidence are statements such as the following from lender websites or application portals: “A representative will contact you and request three to four months business bank statements along with standard identification;" “Simply submit a one-page application and 3 months’ business bank statements to get funding within as little as 24 hours"; “Last, we will ask for a secure connection to your business bank account or for bank statements for the most recent 3 months;" and “bank statements are absolutely central to your loan application."

Empirical Approach. We estimate the effect of CFI assignment on young owners by regressing approval decisions for application i at time t from lender l on the interaction between the lender-level CFI_l indicator and the Young Owner_i indicator:

$$\mathbb{1}(\text{Approved}_{ilt}) = \beta_1 \mathbb{1}(\text{Young Owner}_i) \times \mathbb{1}(\text{CFI}_l) + \text{Lender}_l + \text{Application}_i + \varepsilon_{ilt} \tag{1}$$

The model includes lender fixed effects (Lender_l) and application fixed effects (Application_i), which absorb all observable and unobservable attributes of the applicant that may influence approval decisions. That is, the applicant's risk characteristics are constant across lenders, so we do not require firm- or owner-level controls for identification. Any differences in approval across lenders must arise from lender-side behavior, such as how intensively a lender uses cash flow data. Standard errors are clustered by application.

Three standard assumptions support a causal interpretation. *Overlap* requires that applicants in our sample be exposed to lenders of both types. Of the 184,503 unique applications (comprising 880,885 observations), 69,109 (563,453 observations) are forwarded to both CFI and Non-CFI lenders. *SUTVA* requires that one lender's decision not be influenced by the presence or decision of others. This holds in our setting because the Platform forwards applications without conveying real-time information between lenders, and lenders typically make offers roughly simultaneously. *Conditional independence* requires that, given application fixed effects, the assignment of an applicant to a particular lender be uncorrelated with potential outcomes. We discuss the most plausible threat to this assumption—systematic forwarding patterns that interact with applicant age—after presenting the main results.

Main Results. We present estimates of Equation 1 in the full sample in Table 5. The dependent variable in Panel A is an indicator for loan approval (i.e., an offer), which is our main outcome of interest. We use four age thresholds across the columns: 35, 40, 45, and 50. For the below-50 thresholds, we omit from the sample borrowers between that threshold and 50, so that we are not comparing very young and young owners. For example, in column 1, we compare up to-35 owners with over-50 owners. Table 5 Panel A Column 1 indicates that applicants up to 35 have a 2.5 percentage point higher chance of approval at CFI lenders relative to non-CFI lenders. This represents 13% of the mean approval rate, which is 20%. The effect is similar for applicants up to 40 (12% of mean) and attenuates monotonically as the age threshold rises, falling to 10.5% at the up to-50 threshold (Column 4). The monotonic attenuation indicates that the benefits of cash flow underwriting are concentrated among the youngest applicants for whom credit history is shortest and FICO information is thinnest.

Panel B reports effects on the interest rate (APR) of the offer, restricted to applicants who receive at least two offers (which is necessary to compare APRs across lenders, but also restricts attention to a more creditworthy subsample of the applicant pool). Column 1 indicates that younger entrepreneurs (< 35) enjoy a 1.8 pp lower APR from CFI lenders, which is 2.2% of the mean. The effect is similar across the columns. The contrast between the approval and APR magnitudes is consistent with the screening dynamics in Stiglitz and Weiss (1981). A better screening technology primarily expands access for borrowers who were previously rationed out, with secondary effects on price as the newly-approved are no longer pooled with riskier types.

In the Appendix, we report alternative specifications. First, we include all applicants regardless of age in Table A.7. The results remain robust, albeit with lower magnitudes in the initial columns as we would expect, since the youngest are compared with all older applicants, including those even modestly older. Next, in Table A.8, we omit application fixed effects, instead employing industry, quarter, and lender fixed effects. The approval results are generally similar to those in Table 5.

Heterogeneity in FICO Score. If FICO is a noisier signal of repayment ability for younger applicants, the benefit of CFI assignment should be largest among applicants with lower FICO scores, where credit-history information is thinnest and FICO is most noisy. We test this in Table 6. In Panel A, we split the sample into four groups using industry-standard breakpoints and examine effects on approval (columns 1-4) and APR (columns 5-8). We focus here on the 40-year threshold for "Young." The effect is concentrated among young applicants with Poor or Low FICO scores.

Young applicants with Poor FICO scores are 2.4 pp more likely to be approved by CFI lenders, which represents 16% of their mean approval rate of 15%. For those with Low FICO scores, the effect is 3.91 percentage points, or 18% of their mean approval rate of 21.5%. The effect is much smaller for High-FICO and is statistically insignificant for Super FICO applicants.

In Panel B, we present results for all the age groups in triple interaction models. We focus here on the FICO segment where we see the greatest differentiation, in the Low to High range (670-799). This is where we expect cash flow data to matter the most, since extremely poor scores more often indicate weaknesses that cash flow data cannot mitigate (column 1 of Panel A). In the odd columns of Panel B, we interact $\mathbb{1}(\text{Young Owner}_i) \times \mathbb{1}(\text{CFI}_l)$ with the continuous FICO score within this range. In even columns, we interact with an indicator for Low FICO. Here, we fully interact the fixed effects so that the difference in coefficients replicates the split sample. The coefficients in columns 1 and 3 indicate that a single point higher FICO in this range is associated with a 0.03 pp lower chance of approval for young applicants at CFI lenders, using the 35-year and 40-year definitions of young, respectively. Column 2 shows that being Low vs. High FICO has a 2.1 pp effect, or almost 10% of the mean. The results attenuate using 45 as the young definition, and disappear when we use 50 (columns 5-8).

The effects on APR in Panel A are directionally similar but noisier. Poor-FICO applicants receive interest rates about 5.5 pp lower at CFI lenders (column 5), which is 5% of the high baseline APR for this group. We do not find significant interaction effects for APR. While this may reflect the smaller sample, it could also indicate that young people with lower or less informative FICO scores benefit from cash flow-based lending primarily via more access to credit rather than the intensive margin decision of interest rate.

In Table A.9 we present alternative results where we split the entire FICO distribution into two groups, Low and High, and consider each of the age thresholds from 35 (columns 1-2) to 50 (columns 7-8). We report results for approval (Panel A) and APR (Panel B). Across all models, we see greater benefits among lower-FICO applicants, with the effects attenuating as we increase the age threshold.

4.3 Forwarding Endogeneity and Risk Taking

In this section, we rule out alternative explanations for the effect of CFI assignment on approval. First, we discuss implications of non-random assignment to lenders (i.e., endogenous forwarding). Next, we examine whether greater risk taking by CFI lenders could explain the results. Third, we replicate our results with an entirely different empirical strategy using random assignment to loan officers at Lender B.

Forwarding Endogeneity. A concern with the within-application design is that the set of lenders to which an applicant is forwarded is not random. The Platform routes applications based on criteria specified by the lenders, and the routing decision could correlate with applicant characteristics in ways that are, in turn, correlated with age. We observe all variables employed by the Platform's forwarding algorithm, which do not include age or other protected class characteristics. However, it could be that younger people might tend to be more price insensitive and are forwarded to lenders who have low approval standards and higher prices.

Two tests address this concern. First, the results in Table 5 are robust to restricting the sample to applicants forwarded to both CFI and non-CFI lenders (Appendix Table A.10), where forwarding-stage variation in lender type is fully within-application. This restriction differences out any selection on whether an applicant sees a CFI lender at all, and the estimates are nearly identical to the full-sample results.

Second, in Tables A.11 and A.12 we show that applications forwarded to high-APR versus low-APR lenders, and to high- versus low-approval-rate lenders, are statistically indistinguishable on owner age, owner gender, firm age, firm size, and other observable characteristics. There is a small FICO difference across high- and low-APR lenders. Table A.13 shows that applications forwarded *only* to CFI lenders have somewhat lower FICO and a slightly higher share of young owners than applications forwarded only to non-CFI lenders, but applications forwarded to both lender types, the sample from which we get identification, are well-balanced on observables.

Risk-taking as an Alternative Explanation. A related concern is that CFI lenders might approve more young applicants because they take more risk in general, not because they identify creditworthy young borrowers that FICO misses. We address this concern in three ways: (1) examining business survival outcomes; (2) subsample analyses on the most conservative lenders in the Platform; and (3) in the next subsection, a separate identification strategy using Lender B.

If CFI lenders take more risk, their loans should default at higher rates. To test this—since we do not observe loan performance in the Platform data—we proxy for default with business failure. We collected survival outcomes for a random subset of 46,000 unique applications.[33] Then, in Table 7, we present a series of regressions estimated in this sample with business survival as the dependent variable. Columns 1 and 2 test whether approval by a CFI lender predicts lower survival than approval by a non-CFI lender. The coefficients should be interpreted relative to the outcome of rejection, which is the omitted group. The three indicators for approval by CFI, non-CFI, and both fully saturate the approved set. In Column 1, we show that approval is similarly associated with subsequent survival for *approved by CFI* (7%) and *approved by non-CFI* (6%). The remaining concern is that applicants forwarded to CFI lenders may differ systematically from those forwarded to non-CFI lenders in ways correlated with survival, as discussed above. Column 2 addresses this by restricting the sample to applicants forwarded to both lender types. The coefficients look almost identical even after eliminating forwarding-stage selection of applicants, confirming that concerns about the Platform's forwarding algorithm are limited.

Next, we focus on the approved subsample and test whether young applicants approved by CFI lenders survive at different rates than older applicants approved by CFI lenders. The results are in columns 3-6 of Table 7. The interaction of young with CFI lenders is statistically and economically zero across all four columns (0.00 to 0.02). Together, these results show that the higher approval rate for young applicants at CFI lenders is not driven by CFI lenders systematically extending approvals to borrowers who fail at higher rates.

[33] The clerical workers checked whether the business remained open or appeared closed online as of September 2024. The clerical workers assessed whether the business appears to be open based on a Google search of the business name and state, with a check on the industry for common names. They also examined whether the business is identified as active based on a Google sidebar. The 46,000 unique applications correspond to 250,000 application-forwards and 43,000 approvals.

The second test confirms this finding by focusing on a subsample of lenders who are relatively conservative in their loan offers. In Table A.14 Panel A, we restrict the sample to lenders who approve no more than 10% of applicants, which is the average rate for non-CFI lenders when the unit of observation is the application-forward. Even though the sample is only about one quarter the size used in Table 5 Panel A, the effects are considerably *larger*. For example, using the 40 year age threshold, the effect is 4.6 pp relative to a mean of just 4.4%. In other words, among conservative lenders, assignment to a CFI lender doubles a young entrepreneur's chance of an approval. Panel B repeats the exercise using an alternative definition of conservative lender–those for whom the 25th percentile FICO score for approved applications is above 669.[34] We see the results are similar or slightly larger than in Table 5 Panel A.

Random Assignment to Cash Flow-Intensive Loan Officers Our main analysis in Section 4.2 identifies the effect of cash flow–intensive underwriting using within-application variation across competing lenders. Here, we provide a second causal design at Lender B, where applications are randomly assigned to loan officers who differ in their reliance on cash flow information. Unlike the Platform analysis, where identification comes from comparing how the same applicant is treated by different lenders, identification here comes from random assignment of applicants to officers with different underwriting approaches. The two designs use different sources of variation and the complementary results here, albeit in a small sample, show that the main results are not a spurious artifact of something unique about the Platform.

Similar to our classification of lenders on the Platform, we identify individual loan officers at Lender B as CFI based on the predictive power of cash flow variables in their approval decisions. We observe 60 loan officers (by first and last name), of which 15 have sufficient observations where owner age is observed to construct a CFI measure. These loan officers considered about 11,500 applicants. Using a logit model, we estimate the incremental predictive power of cash flow variables (measured by ROC AUC improvement) for approval decisions made by each officer with

[34]This is the industry threshold separating Poor/Fair from Low, and is also very close to the average 25th percentile approved FICO score across lenders (which is 660 for CFI lenders, and 671 for non-CFI lenders).

sufficient observations. We use logit because there are not enough within-officer observations for an ML model to be useful. Since Lender B is generally cash-flow intensive, all officers show some improvement in ROC AUC; we classify loan officers in the top-tercile of this AUC improvement as CFI. The results are similar using other thresholds.

We present summary statistics about these data in Table A.15. Panel A reports data at the application level, while Panel B reports data at the unique loan officer level. Random assignment predicts that key ex-ante cross-sectional applicant characteristics should be roughly evenly split across CFI and non-CFI loan officers, though we expect some noise due to the small number of officers. As above, the final column of Table A.15 reports t-test results for all variables. Importantly, FICO, cash flow variables, and age are the same across both types of loan officers, consistent with random assignment. Additionally, there is no difference in default rate between CFI and non-CFI loan officers; in fact, the rate is slightly lower for CFI loan officers. Furthermore, CFI loan officers approve a somewhat lower share of applicants. Therefore, CFI loan officers do not take higher risk.

The estimation approach is as follows:

$$\begin{aligned}\mathbb{1}(\text{Approved}_i) = \beta_1 \mathbb{1}(\text{Young Owner}_i) + \beta_2 \mathbb{1}(\text{Young Owner}_i) \times \mathbb{1}(\text{CFI}_k) \\ + \text{Loan Officer}_k + \text{Industry}_i + \alpha_t + \varepsilon_{ik}\end{aligned} \tag{2}$$

Here, the coefficient of interest is β_2 on the interaction of younger owner and a CFI loan officer. We include loan officer, applicant industry, and application quarter (α_t) fixed effects. Standard errors are clustered jointly by quarter and approver. The loan officer and time fixed effects ensure that any variation in the composition of applicants should not bias the results. However, CFI loan officers could be different from their counterparts, raising the concern that the type of officer who is CFI also favors certain age groups. While we cannot exclude this possibility, we believe it is unlikely for two reasons: first, age is not directly used in underwriting; second, such a preference would be illegal. The industry controls rule out a channel where the preference comes via industry heuristics. Note also that as mentioned below, the default rates are the same across CFI and non-

CFI loan officers, so there is no difference in risk preferences.

We present the results in Table 8, using the same structure as the main analysis.[35] Assignment to a CFI loan officer has a strong positive effect on approval, which declines as we raise the age threshold, consistent with the strongest benefits accruing to the youngest applicants. For example, the effect is 8.75 pp using the 35 year threshold (17% of the mean), and 7 pp using the 40 year old threshold (13.5% of the mean), and 3.6 pp using the 50 year old threshold (7% of the mean). Next, in Panel B, we divide the sample around median FICO.[36] Lower FICO scores explain the large effects among younger applicants. In sum, while this sample is smaller and limited to a single lender, it is comforting that we are able to essentially replicate the main finding.

5 Mechanism: Why Cash Flow Data Helps Young Borrowers

Why does cash flow data benefit younger entrepreneurs specifically? If lenders minimize default risk, higher approval rates for young applicants with cash flow–based underwriting must reflect young borrowers being reclassified toward lower predicted default when cash flow data are added to a FICO-based model. Two distinct channels could produce this reclassification. First, FICO may be a noisier signal of repayment ability for young borrowers. This *noise channel* could lead to higher approval because cash flow data better separates the good and bad types. Second, FICO may be systematically biased downward for younger entrepreneurs. This *bias channel* could lead to higher approval because cash flow data causes the algorithm to reclassify younger people as having lower predicted default risk—not necessarily because their cash flow profiles are objectively better than older applicants', but because FICO assigns them lower scores than their underlying repayment behavior warrants.

We find evidence consistent with both channels. Cash flow data closes a substantial gap in predictive performance for young borrowers (Section 5.1), and adding cash flow variables

[35]As above, we also adjust the sample to compare a group of young owners to over-50 owners in each column. The full sample results are similar, reported in Table A.16.

[36]We divide FICO around the median due to the smaller sample and higher FICO scores in this population.

disproportionately reclassifies young applicants toward lower predicted default in ways that cannot be explained by stronger underlying cash flow profiles (Section 5.2).

5.1 FICO is Noisier for Young Borrowers

The noise channel can produce higher approval rates on its own. Interpreted through the lens of the Stiglitz and Weiss (1981) model, better information separates creditworthy from non-creditworthy applicants, allowing lenders to extend credit to marginal young applicants who were previously rationed out even if young applicants are no less risky on average. We conduct two tests to explore whether this noise channel is at play. First, we explore whether adding cash flow variables to a default-prediction model improves performance more for young borrowers than for older borrowers. If FICO is noisier for the young, then the Cash Flow model that augments FICO with bank statement variables should add more orthogonal information for young borrowers than for older ones. To test this, we re-estimate the random forest models from Section 4.1, but now with default as the outcome variable, separately for younger and older entrepreneurs.

The results in Table 9 reveal exactly the pattern we expect. For owners up to 35, the FICO-based Baseline model has an out-of-sample ROC-AUC of 0.608. Adding cash flow variables raises ROC-AUC by 0.03, which is a 28% improvement relative to the random-guess benchmark. For owners over 35, the Baseline AUC is higher (0.671) and the improvement from adding cash flow data is much smaller (0.008, or 4.7%). The pattern is robust across alternative measures of predictive performance (area under the precision-recall curve and the H-measure) and the differences in AUC improvement are statistically significant at every age threshold we examine. Figure A.5 plots the ROC curves by owner age group.

Second, we predict default using OLS models in Table 10. The first two columns use the full sample, while the remaining four consider younger owners (columns 3-4) and older owners (columns 5-6). The odd columns present the Baseline (FICO-only) model, while the even columns add the cash flow variables. In the full sample, the cash flow variables improve the R-squared by 13.8%. The improvement is much larger within the young population, at 23%, than within the older population, at 11.1%. This confirms the conclusion from the ML model that the cash flow variables are more informative for younger entrepreneurs.

Finally, we explore whether cash flow data can potentially replace FICO altogether. This is particularly relevant for thin-file applicants who may have no usable credit score at all. We re-estimate the default models after removing FICO entirely and present the results in Table A.17. This enables us to compare a Cash Flow model that excludes FICO to the FICO-based Baseline. Across the full sample the two perform almost identically, with an AUC of 0.642 each (Tables 9 and A.17). The comparison is more telling for the youngest owners, where the FICO-free CF model modestly outperforms the FICO Baseline (AUC of 0.617 versus 0.608 up to 35, and 0.637 versus 0.633 up to 40). At older thresholds the two are comparable. We do not interpret this result to imply that cash flow data dominate FICO, but rather that FICO predicts default well for older borrowers but poorly for younger ones, and therefore recent cash flow variables can substitute for FICO among the young.

5.2 Cash Flow Data Indicates Lower Default Risk for Young

While the noise channel can explain the access result in Section 4.2, a second mechanism may also be at play. If incorporating recent cash flow data makes young applicants appear more creditworthy on average, then young applicants are not just being sorted more accurately, but are being reclassified upward. This is the bias channel: FICO and other backward-looking inputs are not timely enough to capture the current creditworthiness of young business owners, and incorporating recent cash flow data corrects a systematic downward bias in their scores.

We test the bias channel using a distributional analysis in the spirit of Fuster et al. (2022). For each applicant in the Lender A and Lender B samples, we compute the change in predicted default probability when moving from the Baseline (FICO-based) model to the Cash Flow model. Figure 4 plots the cumulative distribution of these changes separately for younger and older entrepreneurs. The young CDF is shifted leftward relative to the older CDF, indicating that young applicants are disproportionately reclassified toward *lower* predicted default when cash flow data is added. More than half of young applicants see a downward revision in predicted default; older applicants are more evenly split.

To gauge the economic magnitude of this reclassification, Table A.18 reports the average change in predicted default when moving from the Baseline to the Cash Flow model. The reduction is largest for the youngest owners and shrinks monotonically with age. Predicted default falls by 0.48 percentage points for owners up to 35 (a 3.0% reduction relative to their 15.8% baseline rate), by 0.40 points for those up to 40 (2.6%), and by 0.31 points for those up to 45 (2.0%), each significant at the 1% level. For older owners there is no comparable reduction; the average change is small and turns slightly positive only for owners over 45. As a benchmark, Dobbie et al. (2020) find that removing a bankruptcy flag from credit reports lowers predicted default by roughly 3 percentage points or 9%.

Two interpretations of this leftward shift are possible. The first is that young applicants have objectively stronger cash flow profiles than older applicants, so adding cash flow data favorably reclassifies them. The second is that FICO assigns young applicants lower scores than their underlying creditworthiness warrants, and the Cash Flow model provides more timely repayment information that corrects this by reducing the predictive weight placed on FICO. Cash flow data provide an independent signal of repayment ability and help to correct the bias. The two interpretations differ in what young and older applicants look like *within* a FICO band. The first predicts that young applicants in a given FICO band have stronger cash flows than older applicants in the same band while the second does not.

The data favor the second interpretation. Within FICO bands, younger and older applicants exhibit similar cash flow profiles. For example, among the lowest FICO applicants the average levels and volatility of credits, balances, and distress signals are qualitatively similar or better for younger borrowers than older ones (Table A.3). The Fuster-style reclassification toward lower predicted default for young applicants is therefore not driven by superior underlying cash flows. Instead, this pattern is consistent with FICO assigning lower scores to young applicants than the rest of their financial profile warrants. Cash flow data partially corrects this mechanical age bias by giving lenders a signal that does not depend on the length of credit history.

Finally, a different perspective on our analysis asks whether adding cash flow data to the underwriting model has a negative impact on apparent creditworthiness and thus access to credit for older people. Figure 4 suggests that at least for the median older borrower, it has little effect. While our data do not permit us to establish the absolute impact on access to credit, since this would require an understanding of the market lending equilibrium, it is possible that the effect is not zero sum; that is, there may be a benefit for younger entrepreneurs but no negative impact on older ones. This could occur, for example, if young, thin credit file entrepreneurs pool together and lenders ration within that group. In contrast, older individuals have a different signal—long credit histories that are more indicative of type—so they may never pool with the young. Overall, our results are consistent with this possibility.

6 Conclusion

Interest in cash flow-based underwriting has grown among policymakers and practitioners in recent years, especially in the debate about Open Banking mandates. Advocates argue that incorporating timely information about ability to repay from bank statements can help democratize access to credit and foster innovation and competition in the financial services industry (e.g., Chopra (2023)). In this paper, we offer the first analysis of variables drawn directly and transparently from bank statements, which any lender could replicate. We compare these with FICO, the traditional central input for credit scoring. Unlike much existing work on alternative inputs in lending, we focus on the U.S., where there is an active policy debate about access to data (Mills and McCarthy, 2016).

We document that younger entrepreneurs are mechanically disadvantaged by underwriters' reliance on personal credit scores, which favor a long history of repaying debt. In the U.S., personal credit scores rise precipitously for people in their 40s and 50s, on average, which is precisely when we see the steepest increases in entrepreneurship. In this paper, we explore the connection between these facts by studying whether younger entrepreneurs benefit from cash flow-based underwriting, in which lenders incorporate recent information from business checking accounts. Using two complementary identification strategies we find that younger entrepreneurs are substantially more likely to be approved at lenders that use cash flow data. The benefits are concentrated among the youngest applicants and those with low FICO scores. We focus on cash flow information because of its rising prevalence in underwriting and obvious connection to business health, but our findings likely generalize to other underwriting inputs that predict default and capture information orthogonal to traditional credit scores.

Two mechanisms appear to drive our results. First, FICO is a noisier signal of repayment ability for young borrowers than for older ones, so adding cash flow variables to a FICO-based default prediction model reduces that differential noise. Second, FICO appears to be systematically biased downward for young borrowers, because cash flow-based models disproportionately reclassify young borrowers toward lower predicted default, but young borrowers do not have superior cash flow metrics within a given FICO band. Cash flow data can mitigate FICO's mechanical age bias by providing lenders a signal that does not depend on the length of credit history.

References

Agarwal, S, S Alok, P Ghosh, and S Gupta, 2021, Financial inclusion and alternate credit scoring: Role of big data and machine learning in fintech, *SSRN, Rochester, NY* .

Agarwal, Sumit, John C Driscoll, Xavier Gabaix, and David Laibson, 2009, The age of reason: Financial decisions over the life cycle and implications for regulation, *Brookings Papers on Economic Activity* .

Agarwal, Sumit, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao, 2024, Searching for approval, *Econometrica* 92, 1195–1231.

Albanesi, Stefania, and Domonkos F Vamossy, 2024, Credit scores: Performance and equity, Technical report, National Bureau of Economic Research.

Alok, Shashwat, Pulak Ghosh, Nirupama Kulkarni, and Manju Puri, 2024, Open banking and digital payments: Implications for credit access, Technical report, National Bureau of Economic Research.

Arora, Rohit, 2023, How to improve your credit score and secure a business loan, Technical report, Forbes.

Azoulay, Pierre, Benjamin F Jones, J Daniel Kim, and Javier Miranda, 2020, Age and high-growth entrepreneurship, *American Economic Review: Insights* 2, 65–82.

Babina, Tania, Saleem Bahaj, Greg Buchak, Filippo De Marco, Angus Foulis, Will Gornall, Francesco Mazzola, and Tong Yu, 2025, Customer data access and fintech entry: Early evidence from open banking, *Journal of Financial Economics* 169, 103950.

Bakker, Trevor J, Stefanie DeLuca, Eric A English, James S Fogel, Nathaniel Hendren, and Daniel Herbst, 2025, Credit access in the united states, Technical report, National Bureau of Economic Research.

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.

Berg, Tobias, Andreas Fuster, and Manju Puri, 2022, Fintech lending, *Annual Review of Financial Economics* 14, 187–207.

Berger, Allen N, W Scott Frame, and Nathan H Miller, 2005a, Credit scoring and the availability, price, and risk of small business credit, *Journal of money, credit and banking* 191–222.

Berger, Allen N, Nathan H Miller, Mitchell A Petersen, Raghuram G Rajan, and Jeremy C Stein, 2005b, Does function follow organizational form? evidence from the lending practices of large and small banks, *Journal of Financial Economics* 76, 237–269.

Berger, Allen N, and Gregory F Udell, 1995, Relationship lending and lines of credit in small firm finance, *Journal of business* 351–381.

Black, Sandra E, and Philip E Strahan, 2002, Entrepreneurship and bank credit availability, *The Journal of Finance* 57, 2807–2833.

Blattner, Laura, Jacob Hartwig, and Scott Nelson, 2022, Information design in consumer credit markets, *American Economic Review* .

Breiman, Leo, 2001, Random forests, *Machine learning* 45, 5–32.

Carter, David A, and James E McNulty, 2005, Deregulation, technological change, and the business-lending performance of large and small banks, *Journal of Banking & Finance* 29, 1113–1130.

CFPB, 2020, Borrower risk profiles, Technical report, Consumer Financial Protection Bureau.

CFPB, 2024, Regulation b (equal credit opportunity act): Official interpretations of § 1002.6, Provides official interpretations of § 1002.6 regarding rules for creditworthiness evaluations under the Equal Credit Opportunity Act.

Chatterjee, Satyajit, Dean Corbae, Kyle P Dempsey, and José-Víctor Ríos-Rull, 2025, Credit scores and inequality across the life cycle, Technical report, National Bureau of Economic Research.

Chen, Brian S, Samuel G Hanson, and Jeremy C Stein, 2017, The decline of big-bank lending to small business: Dynamic impacts on local credit and labor markets, Technical report, National Bureau of Economic Research.

Chioda, Laura, Paul Gertler, Sean Higgins, and Paolina C Medina, 2024, Fintech lending to borrowers with no credit history, Technical report, National Bureau of Economic Research.

Chopra, Rohit, 2023, Prepared remarks of cfpb director rohit chopra on the proposed personal financial data rights rule, Accessed: 2024-12-10.

Cole, Rebel A, and Jason Damm, 2020, How did the financial crisis affect small-business lending in the united states?, *Journal of Financial Research* 43, 767–820.

Crosman, Penny, 2024, How banks can overcome challenges of cash flow underwriting technology, Technical report, American Banker.

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

Dobbie, Will, Paul Goldsmith-Pinkham, Neale Mahoney, and Jae Song, 2020, Bad credit, no problem? credit and labor market consequences of bad credit reports, *The Journal of Finance* 75, 2377–2419.

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, 2024, New experian tool empowers financial inclusion through open banking insights, Technical report, Experian Press Release, May 9.

Fed, 2024, 2024 report on employer firms: Findings from the 2023 small business credit survey, Technical report, Federal Reserve Banks.

FinRegLab, 2019, The use of cash flow data in underwriting credit: Empirical research findings, Technical report, FinRegLab, Accessed: 2024-12-10.

Frost, Jon, Leonardo Gambacorta, Yi Huang, Hyun Song Shin, and Pablo Zbinden, 2019, Bigtech and the changing structure of financial intermediation, *Economic policy* 34, 761–799.

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

Ghosh, Pulak, Boris Vallee, and Yao Zeng, 2025, Fintech lending and cashless payments, *The Journal of Finance* .

Goldstein, Caroline, and Sally Lauckner, 2022, Is there a minimum credit score for business loan eligibility?, Online article, Updated June 30, 2022.

Hand, David J, 2009, Measuring classifier performance: a coherent alternative to the area under the roc curve, *Machine learning* 77, 103–123.

Harris, 2024, Rising entrepreneurs: Pathways to small business formation, *Justworks and The Harris Poll* .

He, Zhiguo, Jing Huang, and Jidong Zhou, 2023, Open banking: Credit market competition when borrowers own the data, *Journal of financial economics* 147, 449–474.

Ho, Tin Kam, 1995, Random decision forests, in *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, 278–282, IEEE.

Hopenhayn, Hugo, Julian Neira, and Rish Singhania, 2022, From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share, *Econometrica* 90, 1879–1914.

Jagtiani, Julapa, and Catharine Lemieux, 2018, Do fintech lenders penetrate areas that are underserved by traditional banks?, *Journal of Economics and Business* 100, 43–54.

Johnson, Mark J, Itzhak Ben-David, Jason Lee, and Vincent Yao, 2023, Fintech lending with lowtech pricing, Technical report, National Bureau of Economic Research.

Khandani, Amir E, Adlar J Kim, and Andrew W Lo, 2010, Consumer credit-risk models via machine-learning algorithms, *Journal of Banking & Finance* 34, 2767–2787.

Khwaja, Asim Ijaz, and Atif Mian, 2008, Tracing the impact of bank liquidity shocks: Evidence from an emerging market, *American Economic Review* 98, 1413–1442.

Kohavi, Ron, et al., 1995, A study of cross-validation and bootstrap for accuracy estimation and model selection, in *Ijcai*, volume 14, 1137–1145, Montreal, Canada.

Krishnan, Karthik, and Pinshuo Wang, 2019, The cost of financing education: Can student debt hinder entrepreneurship?, *Management Science* 65, 4522–4554.

Lawler, Ryan, 2024, Plaid introduces cash flow data-based credit report, Technical report, Axios.

Levesque, Moren, and Maria Minniti, 2011, Age matters: How demographics influence aggregate entrepreneurship, *Strategic Entrepreneurship Journal* 5, 269–284.

Liang, James, Hui Wang, and Edward P Lazear, 2018, Demographics and entrepreneurship, *Journal of Political Economy* 126, S140–S196.

Mester, Loretta J, Leonard I Nakamura, and Micheline Renault, 2007, Transactions accounts and loan monitoring, *The Review of Financial Studies* 20, 529–556.

Mills, Karen, and Brayden McCarthy, 2016, The state of small business lending: Innovation and technology and the implications for regulation, *Harvard Business School Entrepreneurial Management Working Paper* 17–042.

Mills, Karen G, 2024, *Fintech, Small Business & the American Dream How Technology Is Transforming Lending and Shaping a New Era of Small Business Opportunity* (Palgrave Macmillan; Second Edition).

Nadeau, Claude, and Yoshua Bengio, 1999, Inference for the generalization error, *Advances in neural information processing systems* 12.

NerdWallet, 2024, Business credit card requirements: What you need to apply, Published 7 months ago; provides an overview of the necessary information and qualifications for applying for a business credit card.

Norden, Lars, and Martin Weber, 2010, Credit line usage, checking account activity, and default risk of bank borrowers, *The Review of Financial Studies* 23, 3665–3699.

Ouyang, Shumiao, 2026, Cashless payment and financial inclusion, *Journal of Financial Economics* 180, 104277.

Petersen, Mitchell A, and Raghuram G Rajan, 1994, The benefits of lending relationships: Evidence from small business data, *The journal of finance* 49, 3–37.

Petersen, Mitchell A, and Raghuram G Rajan, 2002, Does distance still matter? the information revolution in small business lending, *The journal of Finance* 57, 2533–2570.

Probst, Philipp, Anne-Laure Boulesteix, and Bernd Bischl, 2019, Tunability: Importance of hyperparameters of machine learning algorithms, *Journal of Machine Learning Research* 20, 1–32.

Rishabh, Kumar, 2024, Beyond the bureau: Interoperable payment data for loan screening and monitoring, *Available at SSRN 4782597* .

Robb, Alicia M, and David T Robinson, 2014, The capital structure decisions of new firms, *The Review of Financial Studies* 27, 153–179.

Snoek, Jasper, Hugo Larochelle, and Ryan P Adams, 2012, Practical bayesian optimization of machine learning algorithms, *Advances in neural information processing systems* 25.

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

Strahan, Philip E, 2017, Are concentrated banks better informed than diversified ones?, *Journal of Accounting and Economics* 64, 278–283.

Turner, Ryan, David Eriksson, Michael McCourt, Juha Kiili, Eero Laaksonen, Zhen Xu, and Isabelle Guyon, 2021, Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020, in *NeurIPS 2020 Competition and Demonstration Track*, 3–26, PMLR.

VantageScore, 2024, Credit bureaus + bank data = vantagescore 4 plus, Technical report, VantageScore Website.

Zhang, Ting, and Zoltan Acs, 2018, Age and entrepreneurship: nuances from entrepreneur types and generation effects, *Small Business Economics* 51, 773–809.

Figure 1: FICO and Age Relationship

This figure uses applicant data from Lender A, Lender B, and the Platform to show the relationship between FICO score and age (N = 1,027,837). This is a binscatter with 20 equal-sized age bins and a line of best fit.

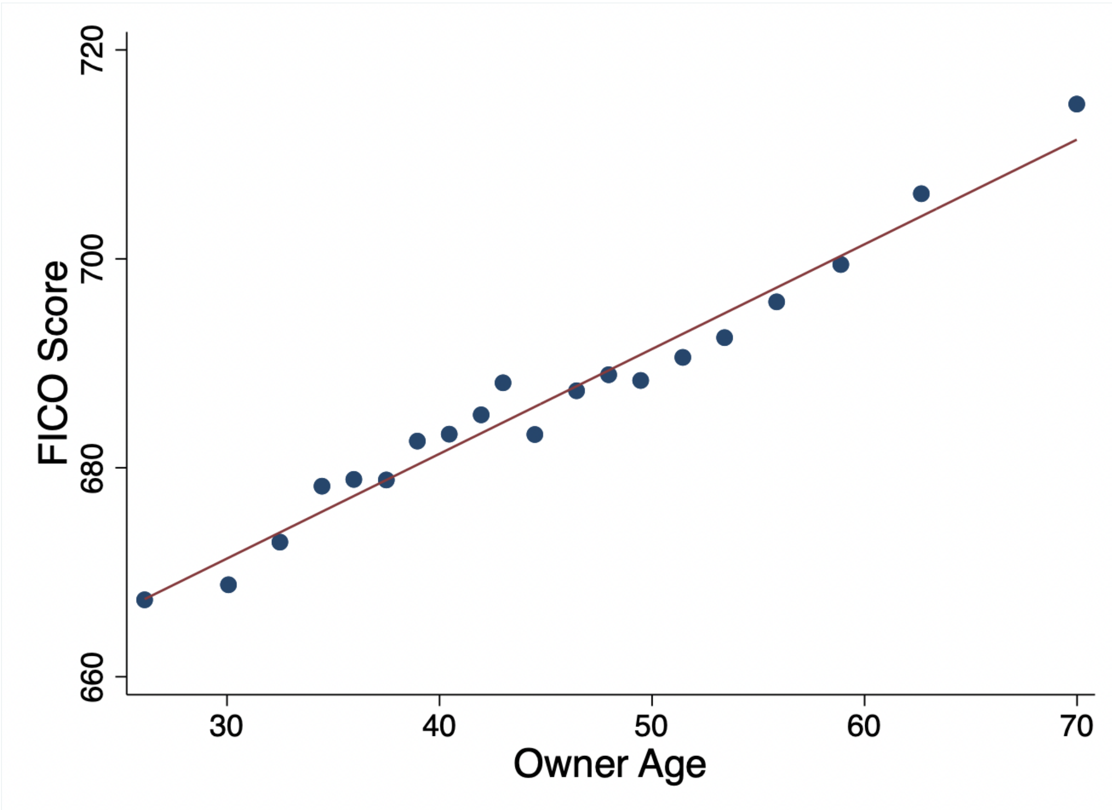


Figure 2: Variation in Loan Outcomes and Inputs by Entrepreneur Age

This figure uses applicant data from Lender A, Lender B, and the Platform to show the differences in key analysis variables for youngand mature owners (N = 1,027,837). Panel A shows the shares of younger and older owners whose loan applications are approved. Panel B shows, within the sample of borrowers, the shares of younger and older owners whose loans default (N = 20,190). Panels C and D present the mean FICO score and number of insufficient funds transactions in the bank statement for four groups: younger applicants, older applicants, younger borrowers, and older borrowers. Young owners are 40 or less, while mature owners are at over 40. Each bar includes 95% confidence intervals using standard errors.

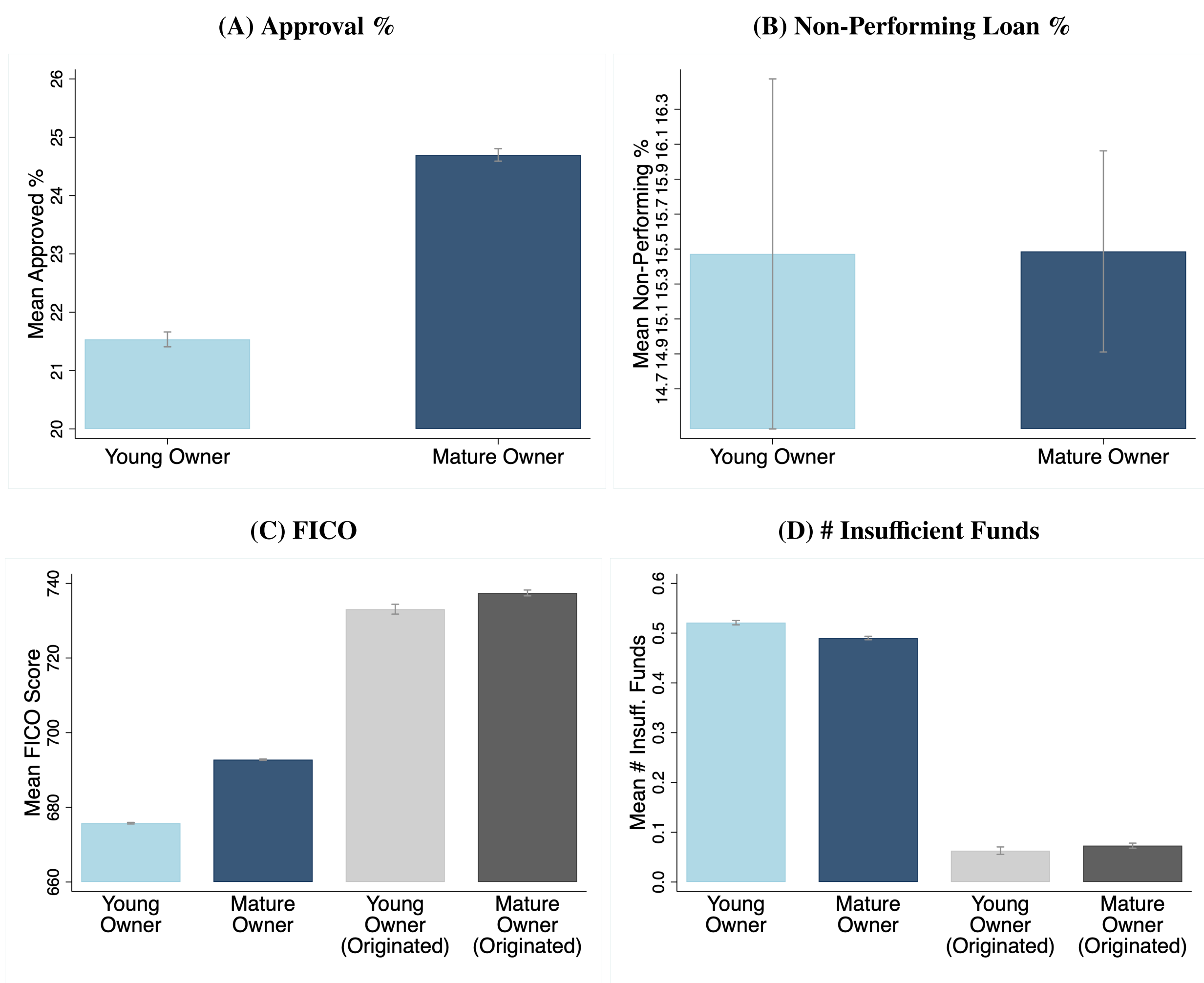


Figure 3: Approval and Owner Age

This figure uses applicant data from Lender A, Lender B, and the Platform to show how approval percentages change with age (N = 1,027,837). This is a binscatter up to and above the age 40 cutoff with 20 equal-sized age bins. The coefficients and standard errors for the owner age coefficient are reported for either side of the cutoff.

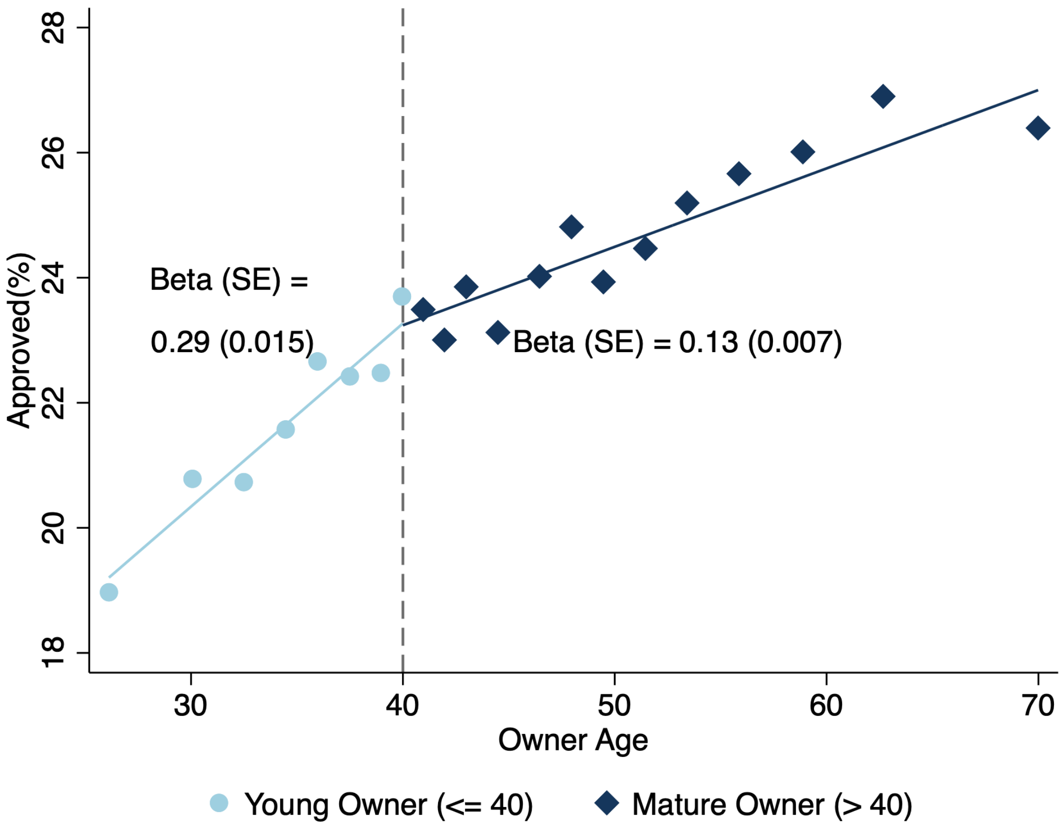


Figure 4: Cumulative Distribution of Change in Predicted Default

This figure shows cumulative distribution of change in predicted default between the cash flow model and the baseline model across all applicants. The underwriting models are first trained using loan default outcomes from loans originated by Lender A and Lender B (N=38,021). Then, the default probabilities of the hold-out samples of applicants to Lender A and Lender B with non-missing ages are used to create these cumulative distributions (N=4,037). We conduct this procedure for the two samples of entrepreneurs up to 40 years old and those over 40.

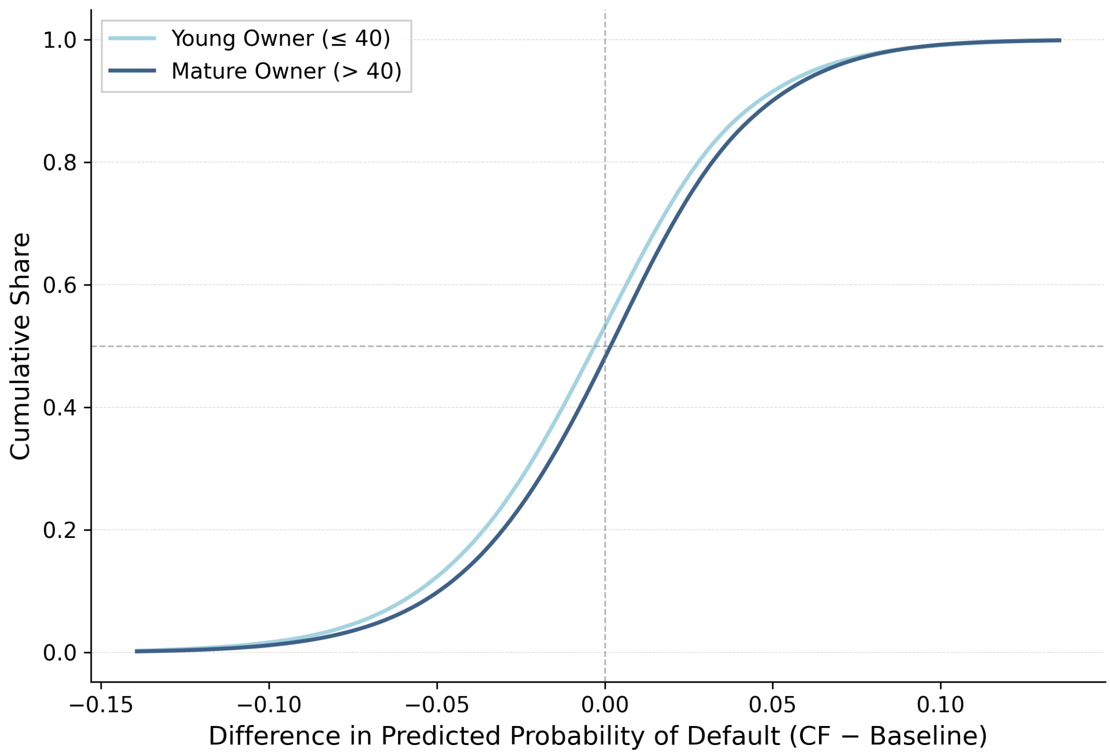


Table 1: Summary Statistics

This table contains summary statistics about loan applications and originated loans. The Loan Applicants sample uses data from Lender A, Lender B, and the Platform (N = 1,067,289). The Originated Loans sample uses data from Lender A and Lender B (N = 38,021). Platform statistics are at the application-forward level so that an application often enters the sample more than once since it is being evaluated by multiple lenders.

	Loan Applicants (Platform, Lender A & B)				Originated Loans (Lender A & B)			
	N	Mean	Median	SD	N	Mean	Median	SD
Loan Variables:								
Approved (%)	1,067,289	25			38,021	100		
Requested Loan Amount (Th$)	1,066,817	144	94	183	38,021	109	80	98
APR (%)	215,796	66	52	57	36,252	16	15	5.11
Originated (%)	1,067,289	7.50			38,021	100		
Non-Performing Loan (%)	38,021	17			38,021	17		
Originated Loan Amount (Th$)	79,858	88	59	98	38,021	115	78	104
Loan Maturity (Years)	80,030	2.42	2.00	1.90	38,021	3.15	3.00	1.41
Credit Score & Cash Flow (Bank Statement) Variables:								
FICO	1,067,289	687	686	67	38,021	728	726	49
Credits (Th$)	1,067,289	100	48	155	38,021	131	73	170
Balance (Th$)	1,067,289	33	11	64	38,021	42	20	64
(#) Insuff. Funds	1,067,289	0.49	0.00	1.38	38,021	0.05	0.00	0.25
(#) Low or Neg. Bal.	1,067,289	0.74	0.00	1.89	38,021	0.42	0.00	1.06
Withdrawals (Th$)	974,677	103	49	159	20,353	168	93	209
Credits (less new debt) (Th$)	962,571	87	41	136	19,370	169	97	195
$\mathbb{1}$(Daily Pay Loan)	1,035,177	0.33			23,950	0.06		
S.D. Credits (Th$)	162,818	12	6.42	14	38,021	13	7.63	14
S.D. Balance (Th$)	162,818	5.18	2.55	6.21	38,021	5.86	3.39	6.15
Borrower Age:								
Owner Age	1,027,837	45	44	11	20,190	49	48	11
Young Owner ($\leq$ 35)	1,027,837	0.22			20,190	0.12		
Young Owner ($\leq$ 40)	1,027,837	0.39			20,190	0.25		
Young Owner ($\leq$ 45)	1,027,837	0.56			20,190	0.40		
Young Owner ($\leq$ 50)	1,027,837	0.70			20,190	0.55		
Other Borrower Characteristics:								
Female	1,042,145	0.28			37,820	0.23		
Business Age (Years)	1,067,289	7.45	5.00	7.19	38,021	11	7.71	7.64
Young Firm (< 5)	1,067,289	0.49			38,021	0.26		
Number of Employees	1,067,289	7.36	4.00	11	38,021	10	6.00	13
Business Survival	250,695	0.35			0			
Pct Black Pop (%)	1,067,289	14	6.30	18	38,021	11	5.30	16
High Pct Black Pop (> 6%)	1,067,289	0.51			38,021	0.46		

Table 2: Summary Statistics by Owner Age

This table compares business owners who are young (≤ 40 years old) with those who are more mature (> 40). The Loan Applicants sample uses data from Lender A, Lender B, and the Platform (N = 1,067,289). The Originated Loans sample uses data from Lender A and Lender B (N = 38,021).

	Loan Applicants								Originated Loans							
	Young Owner (≤ 40)				Mature Owner (> 40)				Young Owner (≤ 40)				Mature Owner (> 40)			
	N	Mean	Median	SD	N	Mean	Median	SD	N	Mean	Median	SD	N	Mean	Median	SD
Loan Variables:																
Approved (%)	399,103	22			628,734	25			5,009	100			15,181	100		
Requested Loan Amount (Th$)	398,915	134	75	178	628,460	149	100	188	5,009	77	60	60	15,181	79	70	62
APR (%)	63,931	86	78	61	118,570	68	58	55	4,532	15	15	4.96	13,889	15	15	4.84
Originated (%)	399,103	5.06			628,734	6.54			5,009	100			15,181	100		
Non-Performing Loan (%)									5,009	15.5			15,181	15.5		
Originated Loan Amount (Th$)									5,009	72	59	56	15,181	77	60	59
Loan Maturity (Years)									5,009	2.45	2.08	1.19	15,181	2.51	2.08	1.21
Credit Score & Cash Flow (Bank Statement) Variables:																
FICO	399,103	676	675	66	628,734	693	691	67	5,009	733	731	48	15,181	737	734	49
Credits (Th$)	399,103	89	42	142	628,734	103	50	157	5,009	103	58	136	15,181	113	68	142
Balance (Th$)	399,103	29	9	59	628,734	35	12	66	5,009	36	17	59	15,181	37	17	58
(#) Insuff. Funds	399,103	0.52	0.00	1.42	628,734	0.49	0.00	1.40	5,009	0.06	0.00	0.27	15,181	0.07	0.00	0.32
(#) Low or Neg. Bal.	399,103	0.78	0.00	1.94	628,734	0.73	0.00	1.90	5,009	0.48	0.00	1.19	15,181	0.46	0.00	1.14
Withdrawals (Th$)	372,660	90	43	145	562,914	106	51	161	677	118	57	183	2,062	128	64	186
Credits (less new debt) (Th$)	368,556	76	36	123	554,912	89	43	136	398	161	88	194	1,358	161	90	188
$\mathbb{1}$(Daily Pay Loan)	397,970	0.32			625,658	0.34			4,932	0.04			14,887	0.05		
S.D. Credits (Th$)	34,609	10	5.10	13	89,039	12	6.28	14	5,009	13	7.36	13	15,181	13	7.86	14
S.D. Balance (Th$)	34,609	4.26	1.85	5.72	89,039	4.94	2.35	6.11	5,009	5.40	3.02	5.93	15,181	5.52	3.15	5.99
Borrower Characteristics:																
Owner Age	399,103	34	35	4.35	628,734	52	50	8.27	5,009	35	35	3.94	15,181	53	52	8.63
Female	387,110	0.27			615,946	0.28			4,993	0.23			15,148	0.23		
Business Age (Years)	399,103	4.62	3.42	4.30	628,734	9	6.33	7.99	5,009	6.40	4.69	5.41	15,181	12	9	7.83
Young Firm (< 5)	399,103	0.67			628,734	0.39			5,009	0.51			15,181	0.24		
Number of Employees	399,103	6.30	4.00	9	628,734	7.82	4.00	11	5,009	8.38	4.00	12	15,181	10	7.00	12
Business Survival	95,083	0.34			155,494	0.36										
Pct Black Pop (%)	399,103	15	7.10	19	628,734	13	6.00	18	5,009	12	5.90	16	15,181	12	5.50	16
Per Capita Income	398,280	36	33	16	627,373	38	34	18	5,006	38	34	18	15,163	39	35	18

Table 3: Predicting Loan Approval in OLS and ML

This table shows how credit score, cash flow, and borrower characteristics predict loan approval using data from Lender A, Lender B, and the Platform (N = 1,067,289). Panel A presents results using an OLS model. Columns 1-2 show results for the full population and columns 3-4 show the results split by owner age. All bank variables and FICO score are standardized to z-scores (in the full sample) and can be interpreted as the change in the dependent variable from 1 standard deviation of change. # Low or Neg. Bal. is the number of low or negative ending balances across the statements. # Insuff. Funds is the number of insufficient funds transactions. Missing values are replaced with median values. Standard errors are clustered by industry and quarter. R-squared improvement represents the percent improvement in unrounded R-squared from the baseline model (column 1). Panel B presents our performance evaluation of the Baseline and Cash Flow random forest models for predicting loan approval. Additional details on the model are found in Appendix Section C. Performance metrics are calculated as the mean of 1,000 bootstrap iterations. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: OLS

Dependent Variable:	Approved (%)			
	All		Owner (≤ 40)	Owner (> 40)
	(1)	(2)	(3)	(4)
FICO	7.14***	5.61***	5.74***	5.46***
	(0.50)	(0.33)	(0.40)	(0.30)
Credits (Th$)		4.40***	3.14**	4.75***
		(0.81)	(1.09)	(0.88)
Withdrawals (Th$)		-4.58***	-3.12**	-5.02***
		(0.86)	(1.21)	(1.00)
Balance (Th$)		0.45*	0.17	0.52*
		(0.24)	(0.21)	(0.25)
$\mathbb{1}$(Daily Pay Loan)		-2.39***	-1.96***	-2.67***
		(0.40)	(0.34)	(0.45)
(#) Low or Neg. Bal.		-5.43***	-4.55***	-5.77***
		(0.58)	(0.44)	(0.60)
(#) Insuff. Funds		-0.12	-0.26	-0.02
		(0.17)	(0.17)	(0.18)
S.D. Credits (Th$)		-1.03***	-0.60	-0.89**
		(0.23)	(0.40)	(0.31)
S.D. Balance (Th$)		0.57**	0.84**	0.77***
		(0.21)	(0.31)	(0.23)
Business Age (Years)	0.20***	0.21***	0.35***	0.17***
	(0.02)	(0.02)	(0.04)	(0.02)
Number of Employees	0.05***	0.04**	0.06***	0.04**
	(0.01)	(0.02)	(0.02)	(0.02)
Observations	1,067,272	1,067,272	399,091	628,718
Industry FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
Adj. R-squared	0.204	0.224	0.211	0.205
R-squared Improv.		9.4		
Y-mean	24.93	24.93	21.53	24.70

Panel B: Machine Learning

	ROC AUC	**AUC-PR**	**H-Measure**
FICO Model	0.803	0.615	0.295
Cash-Flow Model	0.816	0.639	0.323
Difference	0.013***	0.023***	0.027***

Table 4: Summary Statistics by Cash Flow-Intensive Lender Status

This table reports summary statistics from the Platform (N = 880,885) comparing lenders categorized by their intensity of cash flow-based underwriting (CFI). CFI lenders are those for which including bank statement-based cash flow variables significantly enhances the predictive accuracy of loan approval models relative to the Baseline FICO-driven model. The classification relies on improvements in model performance metrics such as ROC-AUC from a random forest classifier. We compare lender-level data by CFI intensity using a two-sample t-test. P-values are based on a two-sample t-test of the CFI and non-CFI lender means. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	All				CF Intensive Lender		Not CF Intensive Lender		
	N	Mean	Median	SD	N	Mean	N	Mean	Difference
Loan Variables:									
Approved (%)	76	21.2	20	16	53	21.4	23	20.7	0.71
# Forwards	76	6.53	7.14	3.32	53	7.33	23	4.69	2.64***
Requested Loan Amount (Th$)	76	171	159	75	53	177	23	158	19
APR (%)	74	59.0	53	45	52	69.3	22	34.7	35***
Originated (%)	76	4.19	3.22	4.60	53	3.87	23	4.90	-1.03
Originated Loan Amount (Th$)	75	84	62	78	52	89	23	72	18
Loan Maturity (Years)	75	2.11	1.00	2.44	52	1.66	23	3.12	-1.46**
Credit Score & Cash Flow (Bank Statement) Variables:									
FICO	76	687	686	33	53	683	23	697	-13.65*
Credits (Th$)	76	120	114	65	53	124	23	111	13
Balance (Th$)	76	39.8	37	20	53	40.4	23	38.5	1.88
(#) Insuff. Funds	76	0.61	0.54	0.34	53	0.59	23	0.65	-0.07
(#) Low or Neg. Bal.	76	0.64	0.55	0.39	53	0.62	23	0.71	-0.10
Withdrawals (Th$)	76	121	115	65	53	125	23	111	13
Credits (less new debt) (Th$)	76	100	96	55	53	104	23	92	11
$\mathbb{1}$(Daily Pay Loan)	76	0.38	0.38	0.11	53	0.39	23	0.35	0.04
Borrower Age:									
Owner Age	76	44.4	44	1.99	53	44.4	23	44.3	0.15
Young Owner (≤ 35)	76	0.23	0.22	0.07	53	0.22	23	0.23	-0.00
Young Owner (≤ 40)	76	0.41	0.40	0.08	53	0.40	23	0.41	-0.01
Young Owner (≤ 45)	76	0.58	0.57	0.07	53	0.58	23	0.58	-0.01
Young Owner (≤ 50)	76	0.72	0.72	0.05	53	0.72	23	0.72	-0.00
Other Borrower Characteristics:									
Female	76	0.29	0.28	0.06	53	0.29	23	0.28	0.00
Business Age (Years)	76	7.08	7.07	1.61	53	7.14	23	6.93	0.22
Young Firm (< 5)	76	0.51	0.52	0.13	53	0.51	23	0.53	-0.02
Number of Employees	76	8.07	7.58	3.24	53	8.28	23	7.57	0.71
Business Survival	72	0.37	0.41	0.14	51	0.39	21	0.33	0.07*
Pct Black Pop (%)	76	13.3	13	2.79	53	13.4	23	13.3	0.07
High Pct Black Pop (> 6%)	76	0.50	0.51	0.06	53	0.51	23	0.50	0.01

Table 5: Within-Application Effect of Assignment to Cash Flow-Intensive Lender

This table uses data from the Platform to test the effect of assignment to a cash flow-intensive lender on approval and offered interest rate (APR) for young entrepreneurs. The level of observation is the application-lender, so an applicant may appear multiple times as the application is forwarded to multiple lenders. Singleton applications are excluded because identification comes from within-application comparisons across lenders. Young is defined according to the column header and only applicants that fall below that threshold or are 50 years or older are included in the regressions. The dependent variable in Panel A is approval and in Panel B it is the APR. The coefficient on the interaction term "Young=1 × CFI=1" represents the difference in approval likelihood or APR for young applicants forwarded to cash flow-intensive lenders relative to applicants over 50 and those forwarded to non-cash flow-intensive lenders. Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Effect on Loan Approval

Dependent Variable:	Approved (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	2.53***	2.36***	2.19***	2.08***
	(0.31)	(0.27)	(0.25)	(0.25)
Observations	418,454	564,934	710,357	829,167
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.314	0.316	0.316	0.315
Y-mean	19.56	19.61	19.73	19.83

Panel B: Effect on the Interest Rate

Dependent Variable:	APR (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	-1.79**	-2.03***	-1.84***	-1.44***
	(0.84)	(0.59)	(0.50)	(0.46)
Observations	51,934	70,435	89,285	105,129
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.878	0.878	0.879	0.878
Y-mean	81.16	82.31	82.20	81.78

Table 6: Within-Application Effect of Assignment to Cash Flow-Intensive Lender by FICO Range

This table uses data from the Platform to test the effect of assignment to a cash flow-intensive lender on approval and percent interest rate (APR) for young entrepreneurs. The level of observation is the application-lender, so an applicant may appear multiple times as the application is forwarded to multiple lenders. In Panel A columns 1-4, the outcome is approval, which is an indicator multiplied by 100, so the coefficient represents a percentage point change in the chance of approval. In Panel A columns 5-8, the outcome is the APR within loan offers, conditional on the applicant receiving at least two loan offers. The FICO ranges reflect industry standard thresholds, which we call Poor (<670), Low (670-739), High (740-799), and Super (>799). In Panel B, we present triple interactions with the continuous FICO score, and then with Low FICO as defined in Panel A, restricting to the sample with Low and High FICO scores. Lender FE are interacted with Low FICO so that the difference between coefficients is the same as in Panel A. Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Split Sample

Dependent Variable:	Approved (%)				APR (%)			
FICO Range:	Poor	Low	High	Super	Poor	Low	High	Super
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Young Owner (< 40)=1 × CFI=1	2.44***	3.91***	1.46**	1.98	-5.52***	-1.66*	-0.99	-2.20
	(0.40)	(0.43)	(0.62)	(1.39)	(1.62)	(0.86)	(0.88)	(1.71)
Observations	251,610	208,532	104,792	24,799	22,159	30,002	18,259	4,771
Application FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.330	0.304	0.337	0.357	0.849	0.868	0.874	0.865
Y-mean	15.55	21.47	25.68	28.42	108.25	76.94	59.69	53.85

Panel B: Triple Interaction within Low-High FICO Range

Dependent Variable:	Approved (%)							
Young Owner (Young) Defined As:	≤ 35		≤ 40		≤ 45		≤ 50	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Young=1 × CFI=1 × FICO	-0.03**		-0.03***		-0.02*		-0.01	
	(0.01)		(0.01)		(0.01)		(0.01)	
Young=1 × CFI=1 × Low FICO=1		2.13**		2.61***		1.61**		1.35*
		(0.95)		(0.82)		(0.76)		(0.73)
Observations	216,991	216,991	288,520	288,520	362,228	362,228	424,268	424,268
Application FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.312	0.312	0.313	0.313	0.314	0.314	0.314	0.313
Y-mean	22.283	22.283	22.400	22.400	22.606	22.606	22.719	22.719

Table 7: Role of Risk in Cash Flow-Intensive Lender Approval

This table uses data from the Platform to examine whether CFI lenders tend to approve riskier businesses. We proxy for risk with business survival as of September 2024, observed for a random subset of applicants. Columns 1 and 2 use the full sample of application-forwards and report estimates from a regression of survival on indicators for approval status (approved by a CFI lender, approved by a non-CFI lender, or approved by both). Column 1 includes all forwards; Column 2 restricts to applicants forwarded to both CFI and non-CFI lenders. Columns 3–6 restrict to the approved subsample and report estimates of the Young × CFI interaction across age thresholds of 35, 40, 45, and 50. Lender FE absorb time-invariant lender characteristics; Quarter FE control for the time of application. Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Dependent Variable:	Business Survival					
Application Sample:	All	Fwd CFI & Non-CFI	Approved			
			Young ($\leq$ 35)	Young ($\leq$ 40)	Young ($\leq$ 45)	Young ($\leq$ 50)
	(1)	(2)	(3)	(4)	(5)	(6)
Approved by CFI	0.07***	0.07***				
	(0.01)	(0.01)				
Approved by Non-CFI	0.06***	0.06**				
	(0.02)	(0.03)				
Approved by Both	0.05**	0.05*				
	(0.02)	(0.03)				
CFI=1 × Young Owner			0.00	0.00	0.01	0.02
			(0.03)	(0.02)	(0.02)	(0.02)
Young Owner			0.01	0.00	-0.01	-0.01
			(0.03)	(0.02)	(0.02)	(0.02)
Observations	240,532	164,048	43,923	43,919	43,919	43,919
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.013	0.011	0.001	0.008	0.008	0.008
Y-mean	0.36	0.37	0.40	0.40	0.40	0.40

Table 8: Effect of Random Assignment to Cash Flow-Intensive Loan Officer

This table uses data from Lender B to test whether young entrepreneurs are more likely to have their loan application approved when randomly assigned to a cash flow-intensive Loan Officer (N = 11,535). The level of observation is an application. Young is defined according to the column header and only applicants that fall below that threshold or are 50 years or older are included in the regressions. Panel A uses the full sample and Panel B splits by FICO score. In odd (even) columns, the sample is restricted to applicants with a FICO score above (below) the median. The interaction term "Young=1 × CFI=1" represents the difference in approval likelihood for young applicants assigned to cash flow-intensive loan officers relative to applicants over 50 and those forwarded to non-cash flow-intensive loan officers. Standard errors are clustered by quarter and approver. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Effect on Loan Approval

Dependent Variable:	Approved (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	8.75***	6.98***	4.97**	3.63*
	(3.10)	(2.43)	(2.05)	(1.83)
Young=1	-8.53***	-6.14***	-4.78***	-3.63**
	(2.26)	(1.71)	(1.55)	(1.39)
Observations	6,550	7,970	9,726	11,535
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Loan Officer FE	Yes	Yes	Yes	Yes
R-squared	0.060	0.057	0.056	0.054
Y-mean	52.09	52.06	51.90	51.99

Panel B: Effect on Loan Approval by FICO Score

Dependent Variable:	Approved (%)							
	Young (≤ 35)		Young (≤ 40)		Young (≤ 45)		Young (≤ 50)	
FICO:	High	Low	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Young=1 × CFI=1	3.80	13.34**	4.69	8.70**	3.94	5.39*	3.29	3.28
	(5.35)	(5.60)	(3.35)	(3.75)	(2.94)	(3.05)	(2.49)	(2.75)
Young=1	-6.74*	-9.47**	-5.82***	-5.79**	-4.55**	-4.57*	-3.51**	-3.27
	(4.02)	(4.03)	(2.22)	(2.72)	(2.01)	(2.31)	(1.64)	(2.11)
Observations	3,319	3,231	4,023	3,947	4,884	4,842	5,782	5,753
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Officer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.075	0.067	0.070	0.065	0.070	0.062	0.068	0.059
Y-mean	55.80	48.28	55.65	48.39	55.63	48.14	55.76	48.20

Table 9: Machine Learning Model Performance

This table presents our performance evaluation of the Baseline and Cash Flow random forest models for predicting loan default overall and for different owner age groups based on the birth date of the primary owner or CEO. We present results from the Preferred Specification, for variables in this model, see Table A.4. This table uses data from Lender A and Lender B (N = 38,021) on originated loans where loan performance is available. Performance metrics are calculated as the mean of 1,000 bootstrap iterations. Definitions of the performance metrics *ROC AUC*, *AUC-PR* and *H-Measure* are provided in Appendix Section C; larger numbers indicate better predictive performance. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	ROC AUC	AUC-PR	H-Measure
Overall			
FICO Model	0.642	0.255	0.071
Cash Flow Model	0.657	0.272	0.086
Difference	0.015***	0.017***	0.016***
Up to 35			
FICO Model	0.608	0.220	0.074
Cash Flow Model	0.638	0.254	0.106
Difference	0.030***	0.034***	0.033***
Over 35			
FICO Model	0.671	0.264	0.104
Cash Flow Model	0.679	0.261	0.108
Difference	0.008***	-0.003***	0.004***
Up to 40			
FICO Model	0.633	0.219	0.076
Cash Flow Model	0.648	0.229	0.090
Difference	0.015***	0.011***	0.014***
Over 40			
FICO Model	0.663	0.263	0.098
Cash Flow Model	0.663	0.269	0.101
Difference	-0.000	0.005***	0.004***
Up to 45			
FICO Model	0.651	0.239	0.087
Cash Flow Model	0.652	0.232	0.087
Difference	0.001***	-0.006***	0.000
Over 45			
FICO Model	0.662	0.270	0.102
Cash Flow Model	0.659	0.267	0.098
Difference	-0.003***	-0.003***	-0.004***
Up to 50			
FICO Model	0.641	0.243	0.079
Cash Flow Model	0.662	0.258	0.098
Difference	0.020***	0.015***	0.019***
Over 50			
FICO Model	0.642	0.236	0.083
Cash Flow Model	0.651	0.245	0.090
Difference	0.009***	0.008***	0.007***

Table 10: Predicting Loan Outcomes in Regressions

This table shows how credit score, cash flow, and borrower characteristics predict loan default conditional on origination. We use data on originated loans from Lenders A and B. All bank variables and FICO score are standardized to z-scores (in the originated sample) and can be interpreted as the change in the dependent variable from 1 standard deviation of change. Standard errors are clustered by industry and quarter. R-squared improvement represents the percent improvement in unrounded R-squared from the baseline model (columns 1, 3, and 5). ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Dependent Variable:	Is Non-Performing (%)					
	All		Owner (≤ 40)		Owner (> 40)	
	(1)	(2)	(3)	(4)	(5)	(6)
FICO	-3.94***	-3.67***	-4.14***	-3.81***	-3.84***	-3.60***
	(0.25)	(0.25)	(0.52)	(0.52)	(0.29)	(0.30)
Credits (Th$)		-1.94***		-2.54***		-1.78***
		(0.43)		(0.87)		(0.51)
Withdrawals (Th$)		3.33***		4.21***		3.16***
		(0.86)		(1.50)		(0.95)
Balance (Th$)		-1.56***		-2.65***		-1.20***
		(0.36)		(0.73)		(0.43)
$\mathbb{1}$(Daily Pay Loan)		0.71***		0.27		0.82***
		(0.24)		(0.46)		(0.28)
(#) Low or Neg. Bal.		1.49***		2.04***		1.32***
		(0.31)		(0.62)		(0.37)
(#) Insuff. Funds		-0.15		-0.33		-0.12
		(0.24)		(0.54)		(0.28)
S.D. Credits (Th$)		1.67***		2.12**		1.52***
		(0.38)		(0.82)		(0.43)
S.D. Balance (Th$)		0.91**		1.36*		0.83*
		(0.38)		(0.75)		(0.46)
Business Age (Years)	-0.14***	-0.14***	-0.06	-0.06	-0.18***	-0.18***
	(0.03)	(0.03)	(0.09)	(0.09)	(0.04)	(0.04)
Number of Employees	-0.07***	-0.06**	-0.05	-0.02	-0.08***	-0.07***
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)
Observations	20,186	20,186	5,006	5,006	15,178	15,178
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Loan Type FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.040	0.045	0.037	0.045	0.041	0.045
R-squared Improv.		13.8		23.0		11.1
Y-mean	15.49	15.49	15.48	15.48	15.49	15.49

Appendix (For Online Publication)

A Supplemental Tables and Figures

Figure A.1: Relationship between FICO and Age Nationally

This figure uses National data from Experian as of 2019 Q2 on the average FICO for each reported age.

Credit Scores by Age

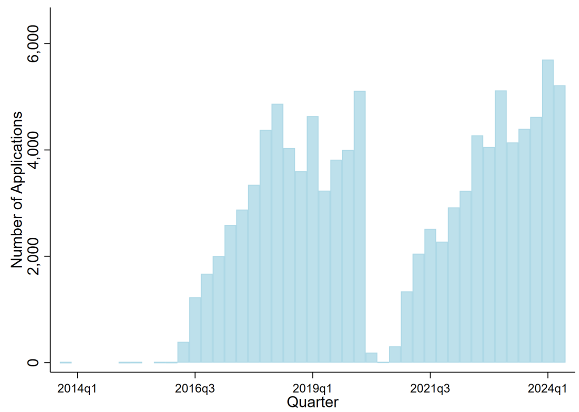
Age Range	Average FICO® Score Q2 2015	Average FICO® Score Q2 2019
23 to 29*	660	660
30 to 39	652	672
40 to 49	667	683
50 to 59	688	703
60 to 69	718	733
70 to 79	745	754
80 to 89	755	757
90 to 99	754	753

Source: Experian. *In Q2 2015, data was only available for ages 23 to 29; we compared to the same age range in Q2 2019.

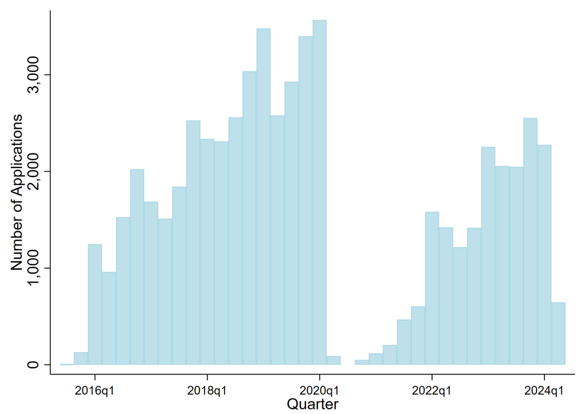
Figure A.2: Applications by Source

This figure shows the number of applications-forwards over time for each of the three data sources. Panel A uses data from Lender A (N = 104,150), Panel B uses data from Lender B (N = 58,668), and Panel C uses data from the Platform (N = 904,471).

(A) Lender A



(B) Lender B



(C) Platform

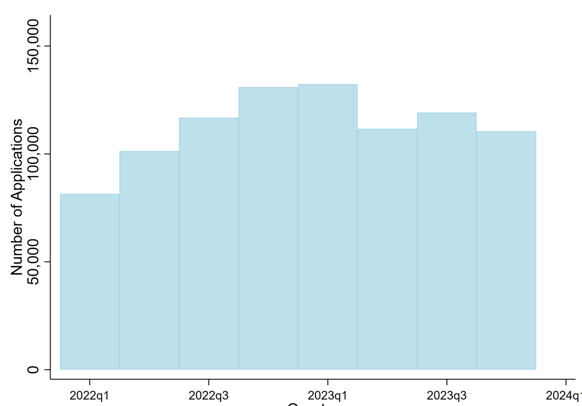
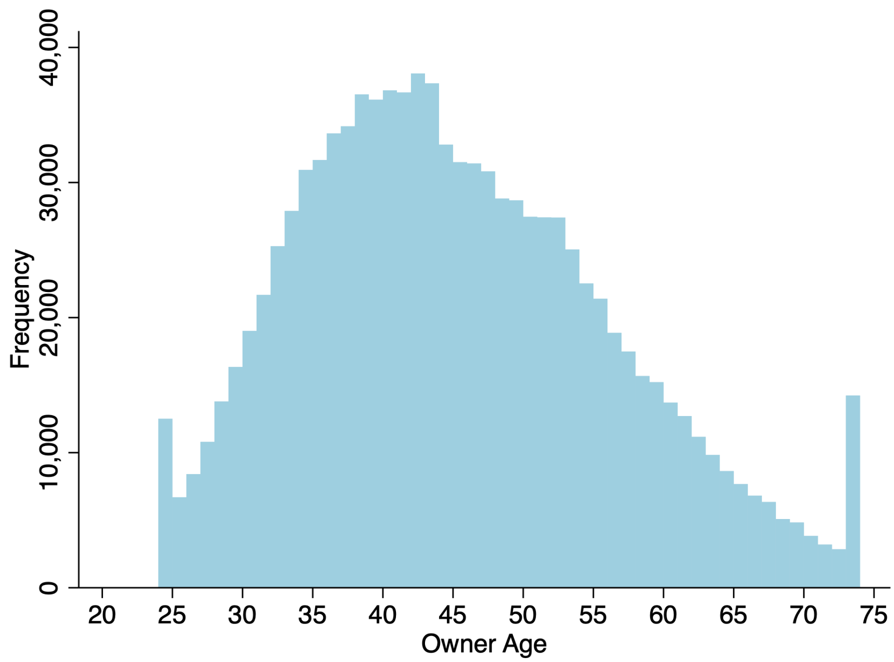


Figure A.3: Entrepreneur Age Distributions

This figure shows the distribution of age among owners in the data (each firm is identified as having one primary CEO or owner). Panel A uses data from Lender A, Lender B, and the Platform (N = 1,027,837) on loan applications without missing age. Panel B uses data from Lender A and Lender B (N = 20,190) on originated loans without missing age.

(A) Full Sample



(B) Lender A and B Sample

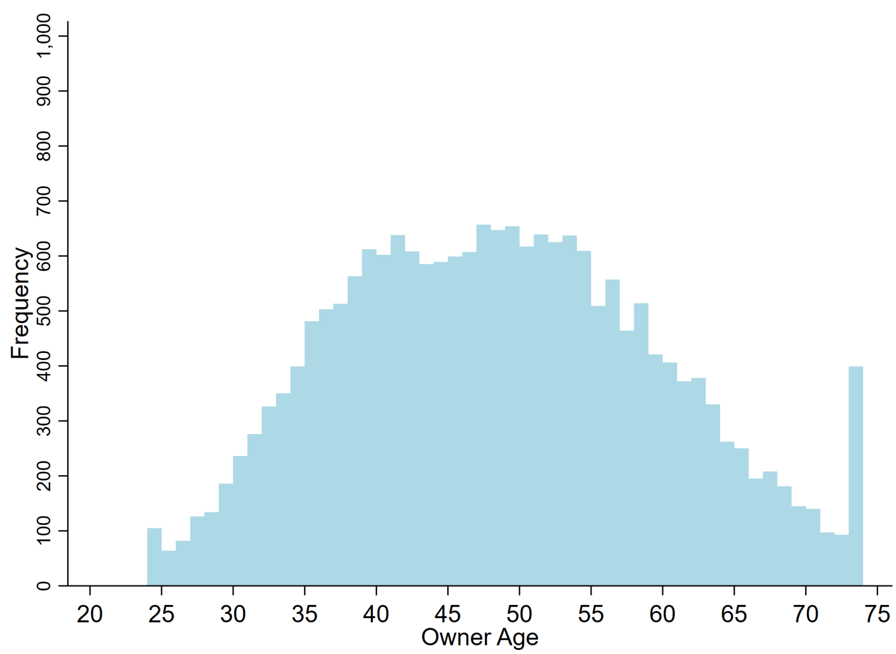


Figure A.4: AUC Improvement Distribution

This figure shows the distribution of AUC improvement from the FICO to the Cash Flow model for the 76 lenders in the Platform data. Lenders described as cash flow intensive have an improvement greater than 0 and are to the right of the red line.

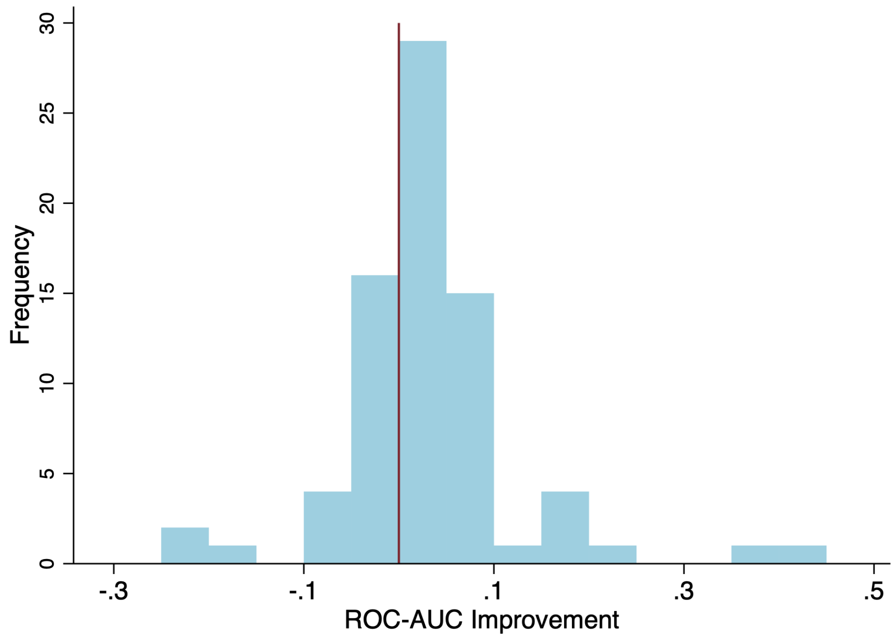
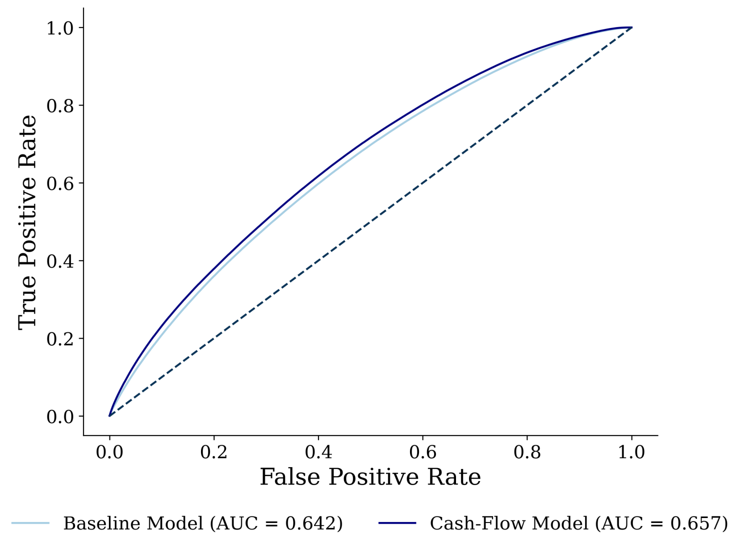


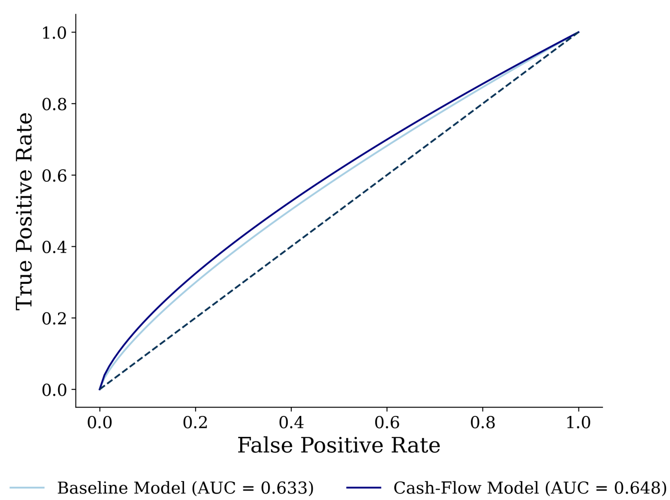
Figure A.5: Performance of Loan Default Prediction Models

This figure uses data from Lender A and Lender B on originated loans (N = 38,021) to report the performance of the Baseline (FICO) and Cash Flow (including bank statement variables) random forest models in predicting loan defaults on the test dataset, presented in Table 9. In Panel A, we plot the receiver operating characteristic (ROC) curve for the full sample. In Panel B and Panel C, we plot the ROC curve for Young (≤ 40) and Mature (> 40) Owners individually.

(A) Overall ROC Curves



(B) Young Owners ROC Curves



(C) Mature Owners ROC Curves

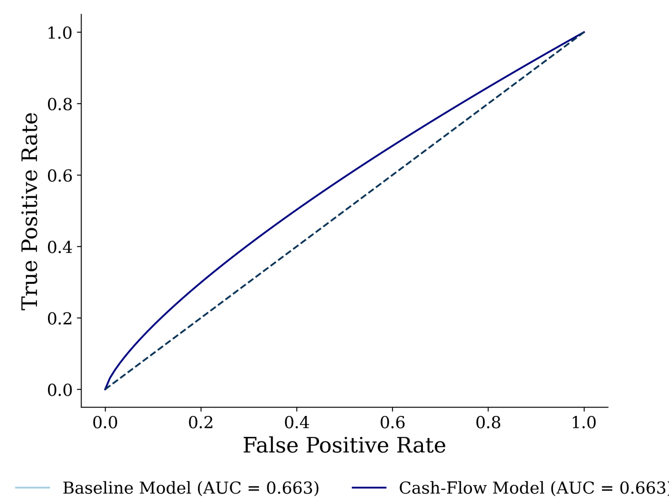
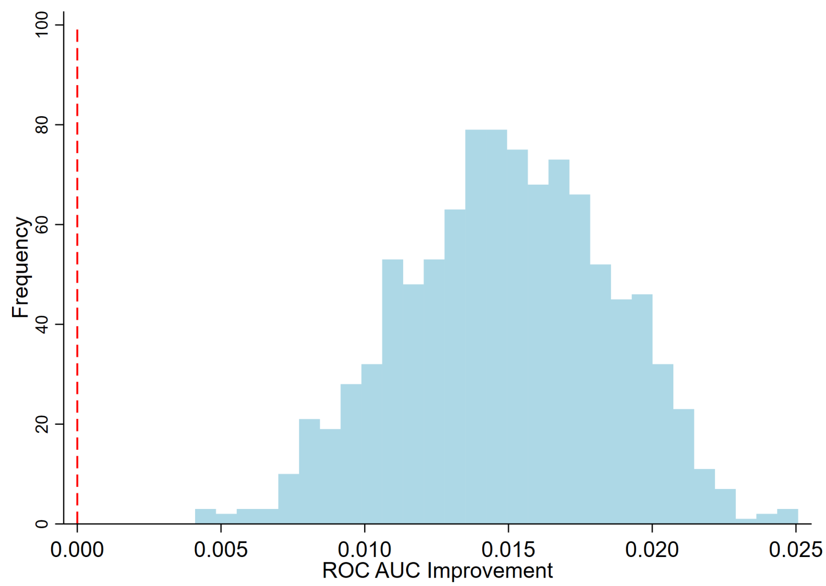


Figure A.6: Bootstrap Estimates of Differences in Predicting Default

This figure presents differences in performance metrics between the Baseline and Cash Flow random forest models for predicting loan default. We present results from the Preferred Specification; for variables in this model, see Table A.4. This figure uses data from Lender A and Lender B (N = 38,021) on originated loans where loan performance is available. Paired performance metrics are calculated for 1,000 bootstrapped iterations and plotted below. Definitions of the performance metrics *ROC AUC* and *H-Measure* are provided in Appendix Section C; larger numbers indicate better predictive performance. In the figure, positive numbers (bars above the red line) represent that the CF model has better performance.

(A) Differences in ROC AUC



(B) Differences in H-Measure

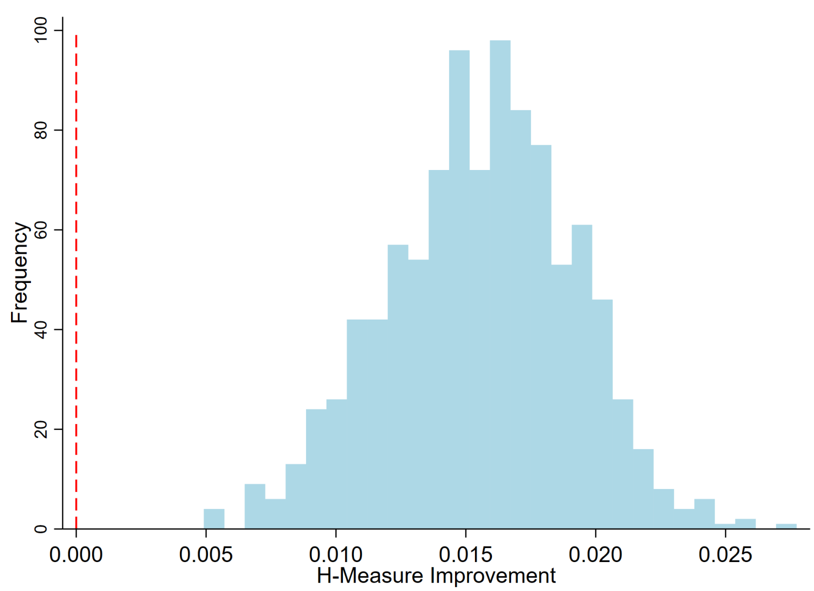


Table A.1: Industry Distribution

This table describes the industry distribution among small businesses nationwide (first column), loan applicants in our data (second set of columns) and borrowers in our data (third set of columns). We also report the share in each industry with young owners (Up to 40).

		Loan Applicants		Originated Loans	
Industry	% National	% Applications	% ≤ 40	% Loans	% ≤ 40
Accommodation and Food Services	7.8	10.7	0.5	8.5	0.4
Agriculture, Forestry, Fishing and Hunting	0.4	1.0	0.6	0.7	0.3
Arts, Entertainment, and Recreation	2.2	3.3	0.6	2.2	0.5
Construction	12.3	20.9	0.6	16.2	0.4
Educational Services	1.4	1.9	0.5	1.2	0.4
Finance and Insurance	4.1	2.1	0.5	1.4	0.4
Health Care and Social Assistance	10.5	8.9	0.5	13.9	0.4
Information	1.3	4.0	0.6	1.3	0.4
Manufacturing	3.3	5.0	0.5	5.0	0.3
Mining, Quarrying, and Oil and Gas Extraction	0.3	0.2	0.5	0.1	0.4
Other Services (except Public Administration)	12.1	0.2	0.4	0.8	0.4
Professional, Scientific, and Technical Services	20.0	3.9	0.4	22.5	0.4
Real Estate and Rental and Leasing	5.8	2.9	0.5	2.3	0.3
Retail Trade	10.6	16.1	0.6	13.3	0.4
Transportation and Warehousing	3.2	14.8	0.6	7.2	0.4
Utilities	0.1	0.6	0.6	0.1	0.2
Wholesale Trade	4.5	3.4	0.5	3.5	0.3

Table A.2: Summary Statistics by Age about Applicants to Lenders A & B

This table prepares summary statistics for Lender A and Lender B (N = 162,818) among loan applicants. It compares business owners who are young (≤ 40 years old) with those who are more mature (> 40).

	Young Owner (≤ 40)		Mature Owner (> 40)		
	N	Mean	N	Mean	Difference
Loan Variables:					
Approved (%)	34,609	35	89,039	43	-7.994***
Requested Loan Amount (Th$)	34,609	80	89,037	89	-8.107***
APR (%)	8,156	19	27,605	18	0.809***
Originated (%)	34,609	17	89,039	22	-4.962***
Non-Performing Loan (%)	5,009	15	15,181	15	-0.014
Originated Loan Amount (Th$)	5,900	80	19,597	89	-8.372***
Loan Maturity (Years)	5,900	2.71	19,597	2.94	-0.225***
Credit Score & Cash Flow (Bank Statement) Variables:					
FICO	34,609	714	89,039	731	-17.070***
Credits (Th$)	34,609	79	89,039	97	-17.973***
Balance (Th$)	34,609	26	89,039	32	-6.033***
(#) Insuff. Funds	34,609	0.38	89,039	0.30	0.083***
(#) Low or Neg. Bal.	34,609	2.70	89,039	1.95	0.749***
Withdrawals (Th$)	8,168	90	23,219	118	-27.659***
Credits (less new debt) (Th$)	4,062	140	15,217	145	-4.541
$\mathbb{1}$(Daily Pay Loan)	33,960	0.14	86,722	0.14	-0.004*
S.D. Credits (Th$)	34,609	10	89,039	12	-1.536***
S.D. Balance (Th$)	34,609	4.26	89,039	4.94	-0.672***
Borrower Characteristics:					
Owner Age	34,609	34	89,039	53	-19.183***
Female	33,919	0.24	87,996	0.24	0.006**
Business Age (Years)	34,609	5.68	89,039	11	-5.652***
Young Firm (< 5)	34,609	0.56	89,039	0.26	0.304***
Number of Employees	34,609	7.18	89,039	8.70	-1.521***
Pct Black Pop (%)	34,609	14	89,039	13	1.188***
High Pct Black Pop (> 6%)	34,609	0.52	89,039	0.48	0.037***

Table A.3: Summary Statistics on Low-FICO Originated Loans (Lender A & B)

This table reports summary statistics on originated loans from Lender A and Lender B, restricted to business owners with a FICO score below 740 (N = 10,962). It compares young owners (≤ 40 years old) with those who are more mature (> 40), and the final column reports the difference in means.

	All				Young Owner (≤ 40)		Mature Owner (> 40)		
	N	Mean	Median	SD	N	Mean	N	Mean	Difference
Credit Score & Cash Flow (Bank Statement) Variables:									
FICO	10,962	699	702	26	2,772	697	8,190	700	-2.24***
Credits (Th$)	10,962	102	61	128	2,772	94	8,190	104	-9.69***
Balance (Th$)	10,962	30	14	50	2,772	29	8,190	30	-1.66
(#) Insuff. Funds	10,962	0.09	0.00	0.34	2,772	0.08	8,190	0.09	-0.01
(#) Low or Neg. Bal.	10,962	0.55	0.00	1.23	2,772	0.58	8,190	0.54	0.04
Withdrawals (Th$)	1,354	115	57	170	342	98	1,012	121	-22.56**
Credits (less new debt) (Th$)	856	150	83	177	180	141	676	153	-12.02
$\mathbb{1}$(Daily Pay Loan)	10,770	0.05	0.00	0.22	2,730	0.04	8,040	0.05	-0.01**
S.D. Credits (Th$)	10,962	13	7.29	13	2,772	12	8,190	13	-0.77***
S.D. Balance (Th$)	10,962	5.00	2.79	5.65	2,772	4.79	8,190	5.08	-0.28**

Table A.4: Approval and Default Machine Learning Features

This table lists the features included in the ML analysis. The Minimal Specification consists of the features listed under Baseline Minimal Model. The Preferred Model includes all of the features in the Minimal Model column plus the additional columns listed. The Cash Flow Model consists of all of the Baseline Features plus the additional features listed under Cash Flow Model. Note that this includes transformations of the cash flow variables included in the OLS regressions.

Baseline Minimal Model	**Baseline Preferred Model**
FICO Score	Industry
Business Age (Years)	Quarter Number (1-4)
Number of Employees	Late Quarter (After Median)
Lender ID	Region (NE, Midwest, South, West)
Requested Loan Amount (Log)	Loan Type
State	
Cash Flow Model	
Credits (Log)	Balance (Log)
Withdrawals (Log)	(#) Insuff. Funds
(#) Low or Neg. Balance	**1**(Daily Pay Loan)
S.D. Credits	S.D. Balance
Missing Withdrawals	Missing Daily Pay Loans
Credits (less new debt) (Log)	Missing Credits (less new debt)
Debits to Credits Ratio	Balance × Credits
(#) Insuff. Funds × (#) Low or Neg. Balance	Low Credit Utilization
Coeff. Variation Balance	Coeff. Variation Credits
$\mathbb{1}$(Daily Pay Loan) to Balance Ratio	Never Low or Neg. Balance
Never Insuff. Funds	(#) Insuff. Funds > 5
Balance Volatility Ratio	Credits to Balance Ratio

Table A.5: Summary Statistics on Machine Learning Variables by Cash Flow-Intensive Lender Status

This table reports summary statistics on the Random Forest Model Performance definitions of (Platform) cash flow-intensity (CFI) (N = 880,885). Application Level statistics are limited to applications sent to both cash flow-intensive and not cash flow-intensive lenders. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	All				CF Intensive Lender		Not CF Intensive Lender		
	N	Mean	Median	SD	N	Mean	N	Mean	Difference
Lender Level:									
AUC Improvement	76	0.04	0.01	0.15	53	0.08	23	-0.05	0.13***
AUC FICO Model	76	0.68	0.70	0.11	53	0.68	23	0.68	0.00
AUC Cash Flow Model	76	0.70	0.72	0.10	53	0.73	23	0.64	0.08***
AUC P-Value (Difference)	73	0.04	0.00	0.15	53	0.03	20	0.07	-0.03
AUC-PR FICO Model	76	0.36	0.36	0.18	53	0.37	23	0.33	0.03
AUC-PR Cash Flow Model	76	0.38	0.38	0.19	53	0.41	23	0.31	0.10**
AUC-PR P-Value (Difference)	73	0.06	0.00	0.17	53	0.04	20	0.10	-0.06
H-Measure FICO Model	76	0.20	0.16	0.14	53	0.18	23	0.23	-0.05
H-Measure Cash Flow Model	76	0.22	0.22	0.11	53	0.23	23	0.18	0.05*
H-Measure P-Value (Difference)	73	0.04	0.00	0.13	53	0.04	20	0.04	0.00
Application-Forward Level:									
AUC Improvement	563,453	0.04	0.03	0.04	459,667	0.05	103,786	-0.01	0.06***
AUC FICO Model	563,453	0.73	0.73	0.05	459,667	0.72	103,786	0.76	-0.04***
AUC Cash Flow Model	563,453	0.75	0.76	0.04	459,667	0.75	103,786	0.75	0.01***
AUC P-Value (Difference)	563,354	0.00	0.00	0.06	459,667	0.00	103,687	0.02	-0.02***
AUC-PR FICO Model	563,453	0.36	0.41	0.14	459,667	0.38	103,786	0.26	0.12***
AUC-PR Cash Flow Model	563,453	0.39	0.42	0.16	459,667	0.42	103,786	0.25	0.17***
AUC-PR P-Value (Difference)	563,354	0.01	0.00	0.04	459,667	0.01	103,687	0.00	0.00***
H-Measure FICO Model	563,453	0.19	0.17	0.08	459,667	0.18	103,786	0.23	-0.05***
H-Measure Cash Flow Model	563,453	0.22	0.22	0.07	459,667	0.22	103,786	0.21	0.01***
H-Measure P-Value (Difference)	563,354	0.01	0.00	0.04	459,667	0.01	103,687	0.00	0.00***

Table A.6: Summary Statistics by Cash Flow-Intensive Lender Status

This table reports summary statistics by Lender (Platform) cash flow-intensity (CFI) (N = 880,885). This table is limited to applications sent to both cash flow-intensive and not cash flow-intensive lenders. We compare application-forward level data by CFI Lender using a two-sample t-test. Survival is measured by whether the business is open according to Google as of data collection in September, 2024. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	All				CF Intensive Lender		Not CF Intensive Lender			
	N	Mean	Median	SD	N	Mean	N	Mean	Diff.	Std. Diff.
Loan Variables:										
Approved (%)	563,453	21.0	0.00	41	459,667	23.6	103,786	9.18	14***	0.40
# Forwards	563,453	10.8	10.00	4.97	459,667	11.0	103,786	9.65	1.37***	0.27
Requested Loan Amount (Th$)	563,453	166	100	203	459,667	168	103,786	158	9.76***	0.05
APR (%)	99,280	78	71	49	91,267	80	8,013	54.9	25***	0.64
Originated (%)	563,453	3.48	0.00	18	459,667	3.55	103,786	3.19	0.37***	0.02
Originated Loan Amount (Th$)	19,630	57.4	34	68	16,324	55.7	3,306	65.7	-10***	-0.15
Loan Maturity (Years)	19,630	1.18	1.00	1.16	16,324	1.04	3,306	1.87	-0.83***	-0.57
Credit Score & Cash Flow (Bank Statement) Variables:										
FICO	563,453	696	690	57	459,667	696	103,786	697	-1.08***	-0.02
Credits (Th$)	563,453	113	56	164	459,667	115	103,786	104	11***	0.07
Balance (Th$)	563,453	38.5	14	69	459,667	39.0	103,786	36.3	2.69***	0.04
(#) Insuff. Funds	563,453	0.38	0.00	1.13	459,667	0.37	103,786	0.41	-0.05***	-0.04
(#) Low or Neg. Bal.	563,453	0.37	0.00	1.03	459,667	0.36	103,786	0.42	-0.06***	-0.05
Withdrawals (Th$)	563,453	114	57	166	459,667	116	103,786	105	11***	0.06
Credits (less new debt) (Th$)	563,453	95	47	139	459,667	97	103,786	87	9.44***	0.07
$\mathbb{1}$(Daily Pay Loan)	563,453	0.35	0.00	0.48	459,667	0.35	103,786	0.35	-0.00	-0.00
Borrower Age:										
Owner Age	563,453	45.2	44	11	459,667	45.2	103,786	45.3	-0.14***	-0.01
Young Owner (≤ 35)	563,453	0.20	0.00	0.40	459,667	0.20	103,786	0.20	0.00***	0.01
Young Owner (≤ 40)	563,453	0.37	0.00	0.48	459,667	0.37	103,786	0.37	0.01***	0.01
Young Owner (≤ 45)	563,453	0.55	1.00	0.50	459,667	0.55	103,786	0.54	0.01***	0.01
Young Owner (≤ 50)	563,453	0.70	1.00	0.46	459,667	0.70	103,786	0.69	0.00***	0.01
Other Borrower Characteristics:										
Female	550,832	0.28	0.00	0.45	449,393	0.28	101,439	0.29	-0.01***	-0.02
Business Age (Years)	563,453	7.87	5.33	7.26	459,667	7.84	103,786	7.98	-0.13***	-0.02
Young Firm (< 5)	563,453	0.45	0.00	0.50	459,667	0.45	103,786	0.43	0.01***	0.03
Number of Employees	563,453	7.82	4.00	11	459,667	7.92	103,786	7.38	0.55***	0.05
Business Survival	32,048	0.42	0.00	0.49	28,808	0.42	3,240	0.42	-0.01	-0.02
Pct Black Pop (%)	563,453	12.9	6.00	17	459,667	12.9	103,786	13.1	-0.25***	-0.01
High Pct Black Pop (> 6%)	563,453	0.50	0.00	0.50	459,667	0.50	103,786	0.50	-0.01***	-0.01

Table A.7: Effect of Assignment to Cash Flow-Intensive Lender on Approval and APR (Within-Application), All Applicant Ages

This table uses data from the Platform to test the effect of assignment to a cash flow-intensive lender on approval and APR for young entrepreneurs (N = 880,885). The level of observation is the application-lender, so an applicant may appear multiple times as the application is forwarded to multiple lenders. Young is defined according to the column header. Panel A shows if cash flow-intensive lenders are more likely to approve young entrepreneurs and Panel B shows the impact on APR. The interaction term "Young=1 × CFI=1" represents the difference in approval likelihood or APR for young applicants forwarded to cash flow-intensive lenders relative to older applicants and those forwarded to non-cash flow-intensive lenders. Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Effect on Loan Approval

Dependent Variable:	Approved (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	1.31***	1.40***	1.65***	2.08***
	(0.25)	(0.22)	(0.22)	(0.25)
Observations	829,167	829,167	829,167	829,167
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.315	0.315	0.315	0.315
Y-mean	19.83	19.83	19.83	19.83

Panel B: Effect on the Interest Rate

Dependent Variable:	APR (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	-0.94	-1.55***	-1.70***	-1.44***
	(0.80)	(0.54)	(0.46)	(0.46)
Observations	105,129	105,129	105,129	105,129
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.878	0.878	0.878	0.878
Y-mean	81.78	81.78	81.78	81.78

Table A.8: Effect of Assignment to Cash Flow-Intensive Lender on Approval and APR, Without Application Fixed Effects

This table uses data from the Platform to test the effect of assignment to a cash flow-intensive lender on approval and APR for young entrepreneurs (N = 880,885). The level of observation is the application-lender, so an applicant may appear multiple times as the application is forwarded to multiple lenders. This is the same specification as Table 5 but without application fixed effects and instead including Industry, Quarter, and Lender fixed effects. Young is defined according to the column header and only applicants that fall below that threshold or are 50 years or older are included in the regressions. Panel A shows if cash flow-intensive lenders are more likely to approve young entrepreneurs and Panel B shows the impact on APR. The interaction term "Young=1 × CFI=1" represents the difference in approval likelihood or APR for young applicants forwarded to cash flow-intensive lenders relative to applicants over 50 and those forwarded to non-cash flow-intensive lenders. Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Effect on Loan Approval

Dependent Variable:	Approved (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	1.94***	1.85***	1.68***	1.60***
	(0.26)	(0.23)	(0.22)	(0.21)
Observations	444,774	600,472	754,764	880,614
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.091	0.091	0.091	0.091
Y-mean	19.38	19.44	19.55	19.66

Panel B: Effect on the Interest Rate

Dependent Variable:	APR (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	2.34***	1.40***	1.20***	1.22***
	(0.56)	(0.43)	(0.37)	(0.34)
Observations	72,926	98,810	124,941	146,545
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.800	0.797	0.797	0.797
Y-mean	87.86	88.92	88.73	88.22

Table A.9: Effect of Assignment to Cash Flow-Intensive Lender on Approval and APR by FICO Range, Supplementary Analysis (Within-Application)

This table uses data from the Platform to test the effect of assignment to a cash flow-intensive lender on approval (Panel A) and APR (Panel B) for young entrepreneurs. We split the sample into two FICO score groups, one below the threshold for "Very Good" of 739, and one above. Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Approval with Full Sample Split Across Age Groups

Dependent Variable:	Approved (%)							
Young Owner (Young) Defined As:	≤ 35		≤ 40		≤ 45		≤ 50	
FICO Range:	< 740	≥ 740	< 740	≥ 740	< 740	≥ 740	< 740	≥ 740
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Young=1 × CFI=1	3.14***	2.23***	3.02***	1.46**	2.62***	2.03***	2.43***	2.05***
	(0.33)	(0.74)	(0.30)	(0.62)	(0.28)	(0.57)	(0.27)	(0.54)
Observations	337,903	80,549	460,142	104,792	579,505	130,852	675,804	153,363
Application FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.311	0.335	0.313	0.337	0.312	0.340	0.312	0.339
Y-mean	18.13	25.53	18.23	25.68	18.31	25.99	18.39	26.18

Panel B: APR with Full Sample Split Across Age Groups

Dependent Variable:	APR (%)							
Young Owner (Young) Defined As:	≤ 35		≤ 40		≤ 45		≤ 50	
FICO Range:	< 740	≥ 740	< 740	≥ 740	< 740	≥ 740	< 740	≥ 740
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Young=1 × CFI=1	-1.85*	-1.82	-2.70***	-0.99	-2.75***	-0.44	-2.08***	-0.54
	(1.07)	(1.31)	(0.77)	(0.88)	(0.66)	(0.72)	(0.62)	(0.66)
Observations	37,981	13,945	52,166	18,259	66,192	23,083	77,762	27,358
Application FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.870	0.871	0.870	0.874	0.870	0.876	0.869	0.875
Y-mean	89.22	59.23	90.24	59.69	90.17	59.37	89.75	59.16

Table A.10: Effect of Assignment to Cash Flow-Intensive Lender on Approval and APR (Within-Application), Limited to Applications Sent to Both Types of Lender

This table uses data from the Platform to test the effect of assignment to a cash flow-intensive lender on approval and APR for young entrepreneurs (N = 880,885). The level of observation is the application-lender, so an applicant may appear multiple times as the application is forwarded to multiple lenders. This is the same specification as Table 5 but limited to applications sent to both cash flow-intensive and not cash flow-intensive lenders (N = 563,453). Young is defined according to the column header and only applicants that fall below that threshold or are 50 years or older are included in the regressions. Panel A shows if cash flow-intensive lenders are more likely to approve young entrepreneurs and Panel B shows the impact on APR. The interaction term "Young=1 × CFI=1" represents the difference in approval likelihood or APR for young applicants forwarded to cash flow-intensive lenders relative to applicants over 50 and those forwarded to non-cash flow-intensive lenders. Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Effect on Loan Approval

Dependent Variable:	Approved (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	2.56***	2.38***	2.21***	2.09***
	(0.31)	(0.27)	(0.25)	(0.25)
Observations	283,792	380,107	479,642	563,453
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.292	0.294	0.295	0.294
Y-mean	20.73	20.74	20.86	20.97

Panel B: Effect on the Interest Rate

Dependent Variable:	APR (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	-1.82**	-2.03***	-1.83***	-1.44***
	(0.84)	(0.59)	(0.50)	(0.46)
Observations	40,228	53,909	68,579	81,207
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.876	0.876	0.876	0.876
Y-mean	73.95	74.60	74.51	74.23

Table A.11: Summary Statistics by Lender Cost (APR)

This table reports summary statistics by Lender (Platform) cost (N = 880,885). A high cost lender is defined as above average APR (18%). We compare application-forward level data by Lender type using a two-sample t-test. Survival is measured by whether the business is open according to Google as of data collection in September, 2024. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	All				High Cost Lender		Low Cost Lender		
	N	Mean	Median	SD	N	Mean	N	Mean	Difference
Loan Variables:									
Approved (%)	77	20.7	18	17	38	20.1	39	21.2	-1.11
# Forwards	77	6.45	7.12	3.36	38	8.91	39	4.06	4.85***
Requested Loan Amount (Th$)	77	171	155	81	38	159	39	182	-22.67
APR (%)	77	58.2	53	44	38	94	39	22.9	71***
Originated (%)	77	4.17	3.12	4.58	38	2.98	39	5.33	-2.35**
Originated Loan Amount (Th$)	76	84	61	81	37	73	39	95	-22.00
Loan Maturity (Years)	76	2.15	1.00	2.45	37	0.80	39	3.42	-2.62***
Credit Score & Cash Flow (Bank Statement) Variables:									
FICO	77	686	686	35	38	674	39	699	-25.16***
Credits (Th$)	77	117	113	65	38	123	39	111	12
Balance (Th$)	77	38.8	36	20	38	36.4	39	41.2	-4.78
(#) Insuff. Funds	77	0.59	0.52	0.33	38	0.62	39	0.57	0.05
(#) Low or Neg. Bal.	77	0.63	0.55	0.36	38	0.64	39	0.61	0.03
Withdrawals (Th$)	77	117	114	65	38	124	39	111	12
Credits (less new debt) (Th$)	77	97	94	55	38	103	39	92	11
$\mathbb{1}$(Daily Pay Loan)	77	0.37	0.38	0.11	38	0.43	39	0.32	0.12***
Borrower Age:									
Owner Age	77	44.3	44	2.01	38	44.3	39	44.2	0.14
Young Owner (≤ 35)	77	0.23	0.22	0.07	38	0.23	39	0.23	-0.01
Young Owner (≤ 40)	77	0.41	0.40	0.08	38	0.41	39	0.41	-0.01
Young Owner (≤ 45)	77	0.58	0.58	0.07	38	0.58	39	0.59	-0.00
Young Owner (≤ 50)	77	0.72	0.72	0.05	38	0.72	39	0.72	-0.00
Other Borrower Characteristics:									
Female	77	0.28	0.28	0.06	38	0.28	39	0.29	-0.01
Business Age (Years)	77	7.04	7.06	1.64	38	7.11	39	6.97	0.14
Young Firm (< 5)	77	0.52	0.52	0.14	38	0.52	39	0.52	-0.01
Number of Employees	77	7.80	7.46	3.07	38	8.07	39	7.54	0.53
Business Survival	71	0.37	0.41	0.14	36	0.41	35	0.33	0.08**
Pct Black Pop (%)	77	13.5	13	3.11	38	13.2	39	13.7	-0.53
High Pct Black Pop (> 6%)	77	0.51	0.51	0.07	38	0.51	39	0.51	0.00

Table A.12: Summary Statistics by Lender Approval Rate

This table reports summary statistics by Lender (Platform) approval rate (N = 880,885). A high approval lender is defined as above average approval rate (53%). We compare application-forward level data by Lender type using a two-sample t-test. Survival is measured by whether the business is open according to Google as of data collection in September, 2024. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	All				High Approval Lender		Low Approval Lender		
	N	Mean	Median	SD	N	Mean	N	Mean	Difference
Loan Variables:									
Approved (%)	77	20.7	18	17	38	32.9	39	8.84	24***
# Forwards	77	6.45	7.12	3.36	38	6.75	39	6.16	0.59
Requested Loan Amount (Th$)	77	171	155	81	38	160	39	181	-21.21
APR (%)	77	58.2	53	44	38	61.9	39	54.6	7.28
Originated (%)	77	4.17	3.12	4.58	38	5.55	39	2.82	2.73***
Originated Loan Amount (Th$)	76	84	61	81	38	79	38	90	-11.64
Loan Maturity (Years)	76	2.15	1.00	2.45	38	2.09	38	2.21	-0.13
Credit Score & Cash Flow (Bank Statement) Variables:									
FICO	77	686	686	35	38	686	39	687	-0.58
Credits (Th$)	77	117	113	65	38	115	39	119	-3.64
Balance (Th$)	77	38.8	36	20	38	37.7	39	40.0	-2.29
(#) Insuff. Funds	77	0.59	0.52	0.33	38	0.54	39	0.64	-0.10
(#) Low or Neg. Bal.	77	0.63	0.55	0.36	38	0.54	39	0.71	-0.16**
Withdrawals (Th$)	77	117	114	65	38	116	39	119	-3.16
Credits (less new debt) (Th$)	77	97	94	55	38	96	39	99	-2.88
$\mathbb{1}$(Daily Pay Loan)	77	0.37	0.38	0.11	38	0.37	39	0.37	0.00
Borrower Age:									
Owner Age	77	44.3	44	2.01	38	44.4	39	44.2	0.21
Young Owner (≤ 35)	77	0.23	0.22	0.07	38	0.22	39	0.24	-0.01
Young Owner (≤ 40)	77	0.41	0.40	0.08	38	0.41	39	0.42	-0.01
Young Owner (≤ 45)	77	0.58	0.58	0.07	38	0.58	39	0.59	-0.01
Young Owner (≤ 50)	77	0.72	0.72	0.05	38	0.72	39	0.72	0.00
Other Borrower Characteristics:									
Female	77	0.28	0.28	0.06	38	0.29	39	0.28	0.01
Business Age (Years)	77	7.04	7.06	1.64	38	6.98	39	7.10	-0.12
Young Firm (< 5)	77	0.52	0.52	0.14	38	0.52	39	0.51	0.01
Number of Employees	77	7.80	7.46	3.07	38	7.71	39	7.89	-0.19
Business Survival	71	0.37	0.41	0.14	38	0.35	33	0.39	-0.03
Pct Black Pop (%)	77	13.5	13	3.11	38	13.4	39	13.5	-0.09
High Pct Black Pop (> 6%)	77	0.51	0.51	0.07	38	0.50	39	0.51	-0.01

Table A.13: Summary Statistics by Cash Flow-Intensive Lender Status

This table reports summary statistics by Lender (Platform) cash flow-intensity (CFI) (N = 880,885). We compare application-forward level data by CFI Lender using a two-sample t-test. Survival is measured by whether the business is open according to Google as of data collection in September, 2024. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	All				Only CF Intensive Lender		Never CF Intensive Lender		Both Types of Lender	
	N	Mean	Median	SD	N	Mean	N	Mean	N	Mean
Loan Variables:										
Approved (%)	184,503	18.0	0.00	27	99,848	17.2	15,546	13.5	69,109	20.3
# Forwards	184,503	4.77	3.00	4.27	99,848	2.98	15,546	1.24	69,109	8.15
Requested Loan Amount (Th$)	184,503	140	76	186	99,848	133	15,546	137	69,109	152
APR (%)	76,367	96	87	58	31,776	116	1,138	28.1	43,453	84
Originated (%)	184,503	5.10	0.00	16	99,848	5.48	15,546	7.67	69,109	3.96
Originated Loan Amount (Th$)	33,116	52.5	28	69	13,678	44.0	1,269	88	18,169	56.5
Loan Maturity (Years)	33,117	1.31	0.92	1.81	13,679	1.06	1,269	5.72	18,169	1.19
Credit Score & Cash Flow (Bank Statement) Variables:										
FICO	184,503	675	675	70	99,848	657	15,546	714	69,109	693
Credits (Th$)	184,503	85	38	142	99,848	79	15,546	60.4	69,109	98
Balance (Th$)	184,503	29.2	8.56	60	99,848	26.2	15,546	24.7	69,109	34.4
(#) Insuff. Funds	184,503	0.66	0.00	1.67	99,848	0.81	15,546	0.59	69,109	0.47
(#) Low or Neg. Bal.	184,503	0.70	0.00	1.65	99,848	0.84	15,546	0.83	69,109	0.46
Withdrawals (Th$)	184,503	85	38	143	99,848	80	15,546	61.1	69,109	99
Credits (less new debt) (Th$)	184,503	70	31	120	99,848	65.4	15,546	48.9	69,109	82
$\mathbb{1}$(Daily Pay Loan)	184,503	0.35	0.00	0.48	99,848	0.37	15,546	0.27	69,109	0.36
Borrower Age:										
Owner Age	184,503	44.0	43	11	99,848	43.2	15,546	44.6	69,109	44.9
Young Owner ($\leq$ 35)	184,503	0.24	0.00	0.43	99,848	0.26	15,546	0.24	69,109	0.21
Young Owner ($\leq$ 40)	184,503	0.42	0.00	0.49	99,848	0.45	15,546	0.40	69,109	0.38
Young Owner ($\leq$ 45)	184,503	0.59	1.00	0.49	99,848	0.62	15,546	0.57	69,109	0.56
Young Owner ($\leq$ 50)	184,503	0.73	1.00	0.44	99,848	0.76	15,546	0.70	69,109	0.71
Other Borrower Characteristics:										
Female	179,337	0.29	0.00	0.45	96,746	0.29	15,111	0.29	67,480	0.28
Business Age (Years)	184,503	6.52	4.33	6.76	99,848	5.88	15,546	6.67	69,109	7.42
Young Firm (< 5)	184,503	0.56	1.00	0.50	99,848	0.61	15,546	0.55	69,109	0.48
Number of Employees	184,503	6.34	4.00	9.40	99,848	6.06	15,546	5.26	69,109	6.99
Business Survival	18,519	0.37	0.00	0.48	6,896	0.35	580	0.40	11,043	0.38
Pct Black Pop (%)	184,503	14.7	7.00	19	99,848	15.5	15,546	14.7	69,109	13.5
High Pct Black Pop (> 6%)	184,503	0.54	1.00	0.50	99,848	0.56	15,546	0.53	69,109	0.51

Table A.14: Effect of Assignment to Cash Flow-Intensive Lender on Approval for Low-Risk Lenders (Within-Application)

This table uses data from the Platform to test the effect of assignment to a cash flow-intensive lender on approval. In Panel A, we restrict the sample to lenders who approve less than 10% of loan applications. In Panel B, we restrict the sample to lenders for whom the 25th percentile FICO score for approved applications is above 669 (this score is close to the average 25th percentile in the overall sample, and is also the industry threshold separating Poor/Fair from Good). Standard errors are clustered by applicant. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Within Lenders with Low Approval Rates

Dependent Variable:	Approved (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	5.13***	4.63***	4.09***	3.51***
	(0.32)	(0.29)	(0.27)	(0.27)
Observations	92,853	124,266	156,169	182,651
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.391	0.393	0.390	0.386
Y-mean	4.45	4.42	4.41	4.41

Panel B: Within Lenders with High Average FICO Scores Among Approved Applications

Dependent Variable:	Approved (%)			
	Young (≤ 35)	Young (≤ 40)	Young (≤ 45)	Young (≤ 50)
	(1)	(2)	(3)	(4)
Young=1 × CFI=1	3.23***	2.50***	2.15***	2.04***
	(0.66)	(0.56)	(0.51)	(0.49)
Observations	132,874	177,196	223,164	261,941
Application FE	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
R-squared	0.422	0.426	0.427	0.426
Y-mean	20.49	20.47	20.66	20.79

Table A.15: Summary Statistics by Assignment to Cash Flow-Intensive Loan Officer

This table reports summary statistics by Loan Officer (Lender B) cash flow-intensity (CFI) (N = 11,535). Panel A compares application level data by CFI Loan Officer using a two-sample t-test. Panel B compares loan officer level data by CFI intensity using a two-sample t-test. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Application Level Statistics

	All				CF Intensive Loan Officer		Not CF Intensive Loan Officer		
	N	Mean	Median	SD	N	Mean	N	Mean	Difference
Loan Variables:									
Approved (%)	11,535	52.0	100	50	5,776	49.9	5,759	54.1	-4.19***
Non-Performing Loan (%)	1,222	25.0	0.00	43	606	24.4	616	25.6	-1.23
Credit Score & Cash Flow (Bank Statement) Variables:									
FICO	11,535	744	747	42	5,776	744	5,759	745	-0.85
Credits (Th$)	11,535	163	80	222	5,776	160	5,759	165	-4.58
Balance (Th$)	11,535	68.4	32	93	5,776	69.6	5,759	67.2	2.39
(#) Insuff. Funds	11,535	0.03	0.00	0.29	5,776	0.03	5,759	0.03	0.00
(#) Low or Neg. Bal.	11,535	0.35	0.00	0.96	5,776	0.35	5,759	0.34	0.01
Withdrawals (Th$)	11,535	178	94	224	5,776	177	5,759	179	-1.67
Credits (less new debt) (Th$)	11,535	159	85	197	5,776	158	5,759	160	-1.54
$\mathbb{1}$(Daily Pay Loan)	11,535	0.02	0.00	0.15	5,776	0.02	5,759	0.02	-0.00
S.D. Credits (Th$)	11,535	16.7	10	16	5,776	16.8	5,759	16.6	0.14
S.D. Balance (Th$)	11,535	8.42	5.35	7.49	5,776	8.50	5,759	8.35	0.15
Borrower Age:									
Owner Age	11,535	50.3	50	11	5,776	50.3	5,759	50.4	-0.16
Young Owner (≤ 35)	11,535	0.09	0.00	0.28	5,776	0.09	5,759	0.08	0.01
Young Owner (≤ 40)	11,535	0.21	0.00	0.41	5,776	0.21	5,759	0.21	0.00
Young Owner (≤ 45)	11,535	0.36	0.00	0.48	5,776	0.36	5,759	0.36	0.00
Young Owner (≤ 50)	11,535	0.52	1.00	0.50	5,776	0.52	5,759	0.52	0.00
Other Borrower Characteristics:									
Female	11,408	0.23	0.00	0.42	5,710	0.22	5,698	0.23	-0.01
Business Age (Years)	11,535	11.4	9.29	7.77	5,776	11.3	5,759	11.4	-0.12
Young Firm (< 5)	11,535	0.18	0.00	0.39	5,776	0.19	5,759	0.18	0.01
Number of Employees	11,535	9.84	5.00	13	5,776	9.78	5,759	9.89	-0.11
Pct Black Pop (%)	11,535	11.0	5.00	16	5,776	11.0	5,759	11.0	-0.04
High Pct Black Pop (> 6%)	11,535	0.44	0.00	0.50	5,776	0.44	5,759	0.44	-0.00

Panel B: Loan Officer Level Statistics

	All				CF Intensive Loan Officer		Not CF Intensive Loan Officer		
	N	Mean	Median	SD	N	Mean	N	Mean	Difference
Loan Variables:									
Approved (%)	15	56.1	56	12	5	52.9	10	57.6	-4.75
Non-Performing Loan (%)	15	20.6	20	13	5	24.1	10	18.9	5.23
Credit Score & Cash Flow (Bank Statement) Variables:									
FICO	15	740	740	8.47	5	739	10	741	-1.80
Credits (Th$)	15	177	154	72	5	162	10	184	-22.36
Balance (Th$)	15	65.9	62	23	5	62.3	10	67.7	-5.36
(#) Insuff. Funds	15	0.05	0.03	0.06	5	0.04	10	0.06	-0.02
(#) Low or Neg. Bal.	15	0.32	0.34	0.12	5	0.35	10	0.31	0.04
Withdrawals (Th$)	15	189	182	68	5	179	10	194	-14.75
Credits (less new debt) (Th$)	15	173	155	64	5	165	10	177	-12.03
$\mathbb{1}$(Daily Pay Loan)	15	0.04	0.02	0.03	5	0.05	10	0.03	0.02
S.D. Credits (Th$)	15	16.7	16	4.47	5	16.3	10	17.0	-0.70
S.D. Balance (Th$)	15	8.26	7.98	2.13	5	8.09	10	8.35	-0.26
Borrower Age:									
Owner Age	15	49.7	50	1.42	5	49.7	10	49.7	-0.07
Young Owner ($\leq$ 35)	15	0.10	0.09	0.02	5	0.10	10	0.10	0.00
Young Owner ($\leq$ 40)	15	0.23	0.22	0.04	5	0.23	10	0.22	0.01
Young Owner ($\leq$ 45)	15	0.38	0.38	0.05	5	0.38	10	0.38	-0.00
Young Owner ($\leq$ 50)	15	0.55	0.53	0.08	5	0.54	10	0.55	-0.01
Other Borrower Characteristics:									
Female	15	0.23	0.23	0.03	5	0.23	10	0.22	0.01
Business Age (Years)	15	11.0	11	1.00	5	11.1	10	11.0	0.06
Young Firm (< 5)	15	0.20	0.20	0.05	5	0.20	10	0.19	0.00
Number of Employees	15	10.8	11	2.08	5	10.3	10	11.1	-0.75
Pct Black Pop (%)	15	11.3	11	2.45	5	11.2	10	11.3	-0.04
High Pct Black Pop (> 6%)	15	0.45	0.45	0.06	5	0.46	10	0.45	0.01

Table A.16: Effect of Random Assignment to Cash Flow-Intensive Loan Officer, All Applicant Ages

This table uses data from Lender B to test whether young entrepreneurs are more likely to have their loan application approved when randomly assigned to a cash flow-intensive Loan Officer (N = 11,535). The level of observation is an application. Young is defined according to the column header. The interaction term "Young=1 × CFI=1" represents the difference in approval likelihood for young applicants assigned to cash flow-intensive loan officers relative to older applicants and those assigned to non-cash flow-intensive loan officers. In odd (even) columns, the sample is restricted to applicants with a FICO score above (below) the median. Standard errors are clustered by quarter and approver. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Dependent Variable:	Approved (%)							
	Young (≤ 35)		Young (≤ 40)		Young (≤ 45)		Young (≤ 50)	
FICO:	High	Low	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Young=1 × CFI=1	2.11	12.80**	4.14	8.93***	3.45	5.87**	3.29	3.28
	(5.49)	(4.91)	(3.09)	(3.24)	(2.85)	(2.71)	(2.49)	(2.75)
Young=1	-5.35	-8.05**	-5.24**	-4.90**	-4.53**	-4.32**	-3.51**	-3.27
	(4.22)	(3.46)	(2.06)	(2.27)	(1.96)	(2.02)	(1.64)	(2.11)
Observations	5,782	5,753	5,782	5,753	5,782	5,753	5,782	5,753
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Officer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.067	0.060	0.068	0.060	0.068	0.060	0.068	0.059
Y-mean	55.76	48.20	55.76	48.20	55.76	48.20	55.76	48.20

Table A.17: Machine Learning Model Performance without FICO

This table presents our performance evaluation of the Baseline and Cash Flow random forest models for predicting loan default after removing the FICO variables from both models. Results are reported for different owner age groups based on the birth date of the primary owner or CEO and for the overall sample. This table uses data from Lender A and Lender B (N = 38,021) on originated loans where loan performance is available. Performance metrics are calculated as the mean of 1,000 bootstrap iterations. Definitions of the performance metrics *ROC AUC*, *AUC-PR* and *H-Measure* are provided in Appendix Section C; larger numbers indicate better predictive performance. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	ROC AUC	AUC-PR	H-Measure
Overall			
Baseline Model (no FICO)	0.614	0.233	0.049
Cash Flow Model (no FICO)	0.642	0.259	0.072
Difference	0.027***	0.026***	0.023***
Up to 35			
Baseline Model (no FICO)	0.548	0.186	0.040
Cash Flow Model (no FICO)	0.617	0.237	0.086
Difference	0.068***	0.052***	0.046***
Over 35			
Baseline Model (no FICO)	0.642	0.248	0.078
Cash Flow Model (no FICO)	0.669	0.254	0.098
Difference	0.027***	0.006*	0.020***
Up to 40			
Baseline Model (no FICO)	0.602	0.210	0.058
Cash Flow Model (no FICO)	0.637	0.222	0.080
Difference	0.036***	0.012*	0.022***
Over 40			
Baseline Model (no FICO)	0.625	0.235	0.065
Cash Flow Model (no FICO)	0.648	0.265	0.091
Difference	0.024***	0.030***	0.027***
Up to 45			
Baseline Model (no FICO)	0.615	0.226	0.064
Cash Flow Model (no FICO)	0.632	0.225	0.070
Difference	0.017***	-0.002	0.006
Over 45			
Baseline Model (no FICO)	0.624	0.225	0.063
Cash Flow Model (no FICO)	0.643	0.251	0.083
Difference	0.019***	0.026***	0.020***
Up to 50			
Baseline Model (no FICO)	0.624	0.233	0.066
Cash Flow Model (no FICO)	0.656	0.250	0.091
Difference	0.032***	0.018***	0.025***
Over 50			
Baseline Model (no FICO)	0.626	0.226	0.069
Cash Flow Model (no FICO)	0.641	0.240	0.082
Difference	0.016***	0.014**	0.014***

Table A.18: Default Probability Shifts Under Cash Flow Underwriting

This table presents the mean and median estimated shift in predicted probability of default under the cash flow underwriting model compared to the baseline model for young and mature owners. We additionally present the fraction of young or mature owners who see a decrease in predicted default probability. We calculate "Δ" for each observation in the holdout dataset as the predicted probability of default under the cash flow model minus the predicted probability of default under the baseline model. The columns "Mean Δ (pp)" denotes the average Δ for a group in percentage points. "Median Δ (pp)" denotes the median, and "Share with Δ < 0" denotes the proportion of observations with a decrease in their predicted probability of default under the cash flow model. These results are derived from the Preferred Specification; for variables in this model, see Table A.4. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	Baseline P(default)	**Mean Δ (pp)**	**Median Δ (pp)**	**Share with Δ < 0**
Overall	0.168	-0.017	0.056	0.494
	(0.000)	(0.036)	(0.069)	(0.007)
Up to 35	0.158	-0.482***	-0.393**	0.540**
	(0.002)	(0.117)	(0.180)	(0.018)
Over 35	0.148	0.012	0.113	0.487
	(0.001)	(0.061)	(0.093)	(0.010)
Up to 40	0.156	-0.401***	-0.300**	0.532**
	(0.002)	(0.089)	(0.139)	(0.014)
Over 40	0.147	0.073	0.174*	0.481*
	(0.001)	(0.064)	(0.095)	(0.011)
Up to 45	0.154	-0.309***	-0.236**	0.526**
	(0.002)	(0.078)	(0.120)	(0.013)
Over 45	0.146	0.134**	0.257**	0.471**
	(0.001)	(0.067)	(0.099)	(0.011)

B Variable Definitions

B.1 Lender A (Application Level)

Loan Variables:

- Approved (%): An indicator variable that takes the value of one if a loan is approved and zero otherwise (for easy interpretation, multiplied by 100).
- Requested Loan Amount (Th$): The loan amount requested at the time of application (in thousands of dollars).
- Interest Rate (%, APR): The interest rate of the loan (in percentage points).
- Originated (%): An indicator variable that takes the value of one if a loan is originated and zero otherwise (for easy interpretation, multiplied by 100).
- Non-Performing Loan (%): An indicator variable that takes the value of one if a loan is in forbearance or default and zero otherwise (for easy interpretation, multiplied by 100).
- Originated Loan Amount (Th$): The loan amount (in thousands of dollars).
- Loan Maturity (Years): The loan maturity (in years).

Credit Score & Cash Flow (Bank Statement) Variables:

- Applicant's Credit Score (FICO): The credit score of the primary applicant (330-850).
- Credits (Th$): The average monthly business revenue, implied from bank statements over the three months (in thousands of dollars).
- Balance (Th$): The average monthly balance in the bank account of the borrower (in thousands of dollars).

- (#) Insuff. Funds: The average monthly number of insufficient balance fees.
- (#) Low or Neg. Balance: The average monthly number of cases where the borrower's account balance is either low ($< \$1,000$) or negative.
- Withdrawals (Th$): The average monthly amount of debit transactions for applicants over the last three months (in thousands of dollars).
- $\mathbb{1}$(Daily Pay Loan): An indicator that takes the value of one if an applicant has daily pay loans and zero otherwise.
- S.D. Credits (Th$): The average monthly standard deviation of deposits of the borrower's account over the past three months (in thousands of dollars).
- S.D. Balance (Th$): The average monthly standard deviation of bank account balances of the borrower over the past three months (in thousands of dollars).

Borrower Characteristics:

- Owner Age (Years): The owner's age at the time of the loan application (in years).
- Female: An indicator variable that takes the value of one if the borrower is female and zero otherwise.
- Business Age (Years): The age of the borrower's business at the time of the loan application (in years).
- Young Firm: An indicator variable that takes the value of one if the loan's borrowing firm age is below 5 years (the 50th percentile in business age across all loans), and zero otherwise.
- Number of Employees: The number of employees in a firm.
- Business Survival: Survival is measured by whether the business is open according to Google as of data collection in September, 2024.

- Pct Black Pop: The fraction of the black population in the borrower's zip code. We collect the total population by race (Asian, Black, Native, Pacific Islander, White, Other) at the zip code level. The data is from the 2020 American Community Survey (ACS) 5-year tables. Since ACS data uses zip code tabulation areas (a Census-specific analogue to zip code), we use a crosswalk provided by Uniform Data System to map ACS data to our zip codes.
- High Pct Black Pop: An indicator variable that takes the value of one if the fraction of the black population in the borrower's zip code is above 6% (the 50th percentile in Pct Black Pop across all loans), and zero otherwise.

B.2 Lender B (Application Level)

Loan Variables:

- Approved (%): An indicator variable that takes the value of one if a loan is approved and zero otherwise (for easy interpretation, multiplied by 100).
- Requested Loan Amount (Th$): The loan amount requested at the time of application (in thousands of dollars).
- Interest Rate (%, APR): The interest rate of the loan (in percentage points).
- Originated (%): An indicator variable that takes the value of one if a loan is originated and zero otherwise (for easy interpretation, multiplied by 100).
- Non-Performing Loan (%): An indicator variable that takes the value of one if a loan is in forbearance or default and zero otherwise (for easy interpretation, multiplied by 100).
- Originated Loan Amount (Th$): The loan amount (in thousands of dollars).
- Loan Maturity (Years): The loan maturity (in years).

Credit Score & Cash Flow (Bank Statement) Variables:

- Applicant's Credit Score (FICO): The credit score of the borrower (450-850).
- Credits (Th$): The average monthly credits in the bank account of the borrower over the last six months (in thousands of dollars).
- Balance (Th$): The average monthly balance in the bank account of the borrower (in thousands of dollars).
- (#) Insuff. Funds: The average monthly number of insufficient balance fees over the past six months.
- (#) Low or Neg. Balance: The average monthly number of cases where the borrower's account balance is either low ($< \$1,000$) or negative over the past six months.
- Withdrawals (Th$): The average monthly amount of debit transactions for applicants (in thousands of dollars).
- Credits (less new debt)(Th$): The average monthly business revenue adjusted for loans, implied from bank statements over the last year (in thousands of dollars).
- $\mathbb{1}$(Daily Pay Loan): An indicator that takes the value of one if an applicant has daily pay loans in the last six months and zero otherwise.
- S.D. Credits (Th$): The average monthly standard deviation of deposits of the borrower's account over the past six months (in thousands of dollars).
- S.D. Balance (Th$): The average monthly standard deviation of bank account balances of the borrower over the past six months (in thousands of dollars).

Borrower Characteristics:

- Owner Age (Years): The owner's age at the time of the loan application (in years).

- Female: An indicator variable that takes the value of one if the borrower is female and zero otherwise.
- Business Age (Years): The age of the borrower's business at the time of the loan application (in years).
- Young Firm: An indicator variable that takes the value of one if the loan's borrowing firm age is below 5 years (the 50th percentile in business age across all loans), and zero otherwise.
- Number of Employees: The number of employees in a firm.
- Business Survival: Survival is measured by whether the business is open according to Google as of data collection in September, 2024.
- Pct Black Pop: The fraction of the black population in the borrower's zip code. We collect the total population by race (Asian, Black, Native, Pacific Islander, White, Other) at the zip code level. The data is from the 2020 American Community Survey (ACS) 5-year tables. Since ACS data uses zip code tabulation areas (a Census-specific analogue to zip code), we use a crosswalk provided by Uniform Data System to map ACS data to our zip codes.
- High Pct Black Pop: An indicator variable that takes the value of one if the fraction of the black population in the borrower's zip code is above 6% (relatively the 50th percentile in Pct Black Pop across all loans), and zero otherwise.

Loan Officer Variables

- CF Intensive: We identify CFI loan officers as those with top-tercile improvement in AUC ROC from the Baseline to the Cash Flow Logistic model.

B.3 The Platform (Application-Forward Level)

Loan Variables:

- Approved (%): An indicator variable that takes the value of one if a loan is approved and zero otherwise (for easy interpretation, multiplied by 100).
- Requested Loan Amount (Th$): The loan amount requested at the time of application (in thousands of dollars).
- Interest Rate (%, APR): The interest rate of the loan (in percentage points).
- Originated (%): An indicator variable that takes the value of one if a loan is originated and zero otherwise (for easy interpretation, multiplied by 100).

Credit Score & Cash Flow (Bank Statement) Variables:

- Applicant's Credit Score (FICO): The credit score of the borrower (400-850).
- Credits (Th$): The average monthly credits in the bank account of the borrower over the last three months (in thousands of dollars).
- Balance (Th$): The average monthly balance in the bank account of the borrower over the last three months (in thousands of dollars).
- (#) Insuff. Funds: The average monthly number of insufficient balance fees over the last three months.
- (#) Low or Neg. Balance: The average monthly number of cases where the borrower's account balance is either low ($< \$1,000$) or negative over the last three months.
- Withdrawals (Th$): The average monthly amount of debit transactions for applicants over the last three months (in thousands of dollars).
- Credits (less new debt)(Th$): The average monthly business revenue adjusted for loans, implied from bank statements over the last three months (in thousands of dollars).
- $\mathbb{1}$(Daily Pay Loan): An indicator that takes the value of one if an applicant has daily pay loans in the last three months and zero otherwise.

Borrower Characteristics:

- Owner Age (Years): The owner's age at the time of the loan application (in years).
- Female: An indicator variable that takes the value of one if the borrower is female and zero otherwise.
- Business Age (Years): The age of the borrower's business at the time of the loan application (in years).
- Young Firm: An indicator variable that takes the value of one if the loan's borrowing firm age is below 5 years (the 50th percentile in business age across all loans), and zero otherwise.
- Number of Employees: The number of employees in a firm.
- Pct Black Pop: The fraction of the black population in the borrower's zip code. We collect the total population by race (Asian, Black, Native, Pacific Islander, White, Other) at the zip code level. The data is from the 2020 American Community Survey (ACS) 5-year tables. Since ACS data uses zip code tabulation areas (a Census-specific analogue to zip code), we use a crosswalk provided by Uniform Data System to map ACS data to our zip codes.
- High Pct Black Pop: An indicator variable that takes the value of one if the fraction of the black population in the borrower's zip code is above 6% (relatively the 50th in Pct Black Pop across all loans), and zero otherwise.

Lender Variables

- CF Intensive: We identify CFI lenders as those for which the AUC ROC is higher in the Cash Flow ML model than in the Baseline ML model. We exclude lenders receiving under 50 application forwards, with an approval rate over 95%. This leaves us with a sample of 76 lenders.

B.4 Additional ML Variables

Credit Score & Cash Flow (Bank Statement) Variables:

- Debits to Credits Ratio: The ratio of the log of variables *Withdrawals (Th$)* over *Credits (Th$)*.
- Balance $\times$ Credits: An interaction term between variables *Balance (Th$)* and *Credits (Th$)*.
- (#) Insuff. Funds $\times$ (#) Low or Neg. Balance: An interaction term between variables *(#) Insuff. Funds* and *(#) Low or Neg. Balance*.
- Low Credit Utilization: An indicator variable that takes the value of 1 if the variable *Credits (Th$)* is below the median, and zero otherwise.
- Coeff. Variation Balance: The ratio of the log of variables *S.D. Balance (Th$)* over *Balance (Th$)*.
- Coeff. Variation Credits: The ratio of the log of variables *S.D. Credits (Th$)* over *Credits (Th$)*.
- $\mathbb{1}$(Daily Pay Loan) to Balance Ratio: The ratio of the log of variables $\mathbb{1}$*(Daily Pay Loan)* over *Balance (Th$)*.
- Never Low or Neg. Balance: An indicator variable that takes the value of 1 if *(#) Low or Neg. Balance* is equal to zero, and zero otherwise.
- Never Insuff. Funds: An indicator variable that takes the value of 1 if *(#) Insuff. Funds* is equal to zero, and zero otherwise.
- (#) Insuff. Funds > 5: An indicator variable that takes the value of 1 if *(#) Insuff. Funds* is greater than 5, and zero otherwise.
- Balance Volatility Ratio: The ratio of the log of variables *S.D. Balance (Th$)* over *Credits (Th$)*.

- Credits to Balance Ratio: The ratio of the log of variables *Credits (Th$)* over *Balance (Th$)*.

Borrower Characteristics:

- Quarter Number: The quarter number (1-4) of the loan application date.
- Late Quarter: The loan application is after the median loan application date in the combined data (post 10/1/2022).
- Region: The borrower's state is mapped to the following regions: Midwest, Northeast, South, West, and Other. This classification is based on the US Census designation by state.

C Machine Learning Methods

A random forest classifier is a ML method that uses a series of decision trees to classify or predict outcomes. Each tree is trained on a randomly selected subset of the data (bagging) and a random subset of features, ensuring diversity across the ensemble. The classifier determines its final prediction through majority voting in classification tasks. This aggregation process reduces overfitting and improves generalization compared to individual decision trees. Since the development of the method (Ho, 1995; Breiman, 2001), random forests have been increasingly applied to complex predictive tasks in the economics and finance fields.

In this study, the primary objective is to develop classifications and predictions of target values, and to measure how these predictions change across different model specifications.[37] Random forests are well-suited for such tasks, as they not only offer predictive accuracy but also provide insights into the relative importance of input features. Feature importance is calculated directly from the trained model's attributes, ranking features based on their contributions to reducing impurity across all trees in the forest. This analysis helps evaluate how the relevance of different variables shifts between models, offering valuable insights into the underlying data-generating process.

Hyperparameter Tuning We tune the hyperparameters for the random forest models using an optimization routine with a TPE Sampler that maximizes model AUC-PR through iterative Bayesian sampling. The hyperparameter tuning executes 100 trials of stratified K-fold cross-validation. Our implementation uses Optuna, a hyperparameter optimization framework for Python. These optimized hyperparameters are selected once for each model specification, and then used for all subsequent estimation and prediction.

In the first stage, the training dataset is used with 5-fold cross-validation to evaluate model performance. 5-fold cross-validation splits the data into five parts, with each part serving as the

[37]The analysis utilized a custom framework built in Python, with dependencies including scikit-learn, numpy, pandas, optuna and other libraries for data manipulation and visualization.

validation set once while the remaining four parts are used for training, repeating this process for each fold and preparing a mean AUC-PR score.[38] This cross-validation procedure is repeated 100 times, each with a different set of hyperparameters chosen through Bayesian optimization by Optuna. After 100 iterations, Optuna identifies the best hyperparameter configuration based on the highest mean AUC-PR score achieved across all cross-validation folds. This entire process is done once for the baseline model and once for the cash flow model, resulting in two sets of optimized hyperparameters.

- Estimators: specifies the number of trees that the model will train. A higher number of trees produces a more complex model at the risk of overfitting.
- Maximum Tree Depth: defines the maximum depth of each individual tree in the model. A higher depth produces a model able to measure more complex patterns at the risk of overfitting.
- Minimum Samples to Split: The minimum number of samples required to split an internal node.
- Feature Sample by Tree: the fraction of features (columns) that will be randomly selected to grow each tree. This prevents overfit by ensuring the model does not rely too heavily on any particular feature or combination of features.
- Maximum Leaf Nodes: Maximum number of leaf nodes in the trees.
- Minimum Child Weight: the minimum sum of instance weights (hessian) needed in a child node. A node is split only if the resulting child nodes have at least a certain amount of "weight."

Bootstrapped Model Evaluation In the second stage, the training dataset is randomly sampled with replacement and then used to train a model for each bootstrap iteration.[39] This mirrors the

[38] We tune our hyperparameters to maximize AUC-PR because AUC-PR can be more informative and stable for datasets like ours with unbalanced outcome classes. Choosing to maximize ROC AUC yields almost identical results.

[39] We stratify splits by the outcome class to maintain consistent proportions of the outcomes in the validation sample.

common bootstrapping approach used by, e.g., Fuster et al. (2022). Each model's performance is then evaluated on the holdout (testing) dataset, and mean performance metrics are collected and standard errors are calculated. To evaluate the differences between the baseline and cash flow models, paired t-tests are conducted on the differences in performance.

Performance Metrics We evaluate the performance of the models using three measures, which are all reported in Table 9. The first, known as the receiver-operating characteristic (ROC) area under the curve (AUC), is widely used to evaluate binary classification models.[40] The ROC curve describes the true positive rate (TP) and the false positive rate (FP) at various thresholds.[41] The ROC AUC is the integral of the ROC curve, where a value of 0.5 indicates no discriminative ability (equivalent to random guessing), and a value of 1.0 indicates perfect discrimination. Typically, an AUC of 0.6 is considered desirable in information-scarce environments, whereas AUCs of 0.7 or greater are the goal in information-rich environments.

The other measures of performance are better suited to our context where defaults are sparse. One is the Precision-Recall curve, where Precision $= \frac{TP}{TP+FP}$ (the share of predicted defaults that actually default) and Recall $= \frac{TP}{TP+FN} = TPR$ (the share of defaults that were predicted). The intuition is the same as in the above, where a larger area under the curve implies a better model. Next, we compute the H-Measure for each model, an alternative to AUC proposed by Hand (2009).[42] Here, the Cash Flow model outperforms the Baseline model by 28% in the minimal and 21% in the preferred specification. Finally, we establish that the Cash Flow models perform significantly better (at the 1% level) across all three measures. This is performed with 1,000

[40] We note that ROC AUC is known to be less informative in situations like ours, when positive outcomes are sparse and false positive rates are naturally low. Also, Hand (2009) demonstrates that the AUC has a potentially serious deficiency: it uses different weights when computing misclassification costs for different classifiers. This means that in some settings, AUC is incoherent. In our setting, these weights are unlikely to differ significantly.

[41] For any binary classification model of default, a decision rule (threshold) is needed to determine whether a particular observation with some non-zero probability of default (e.g., 4%) should be classified as a "0" (non-default) or a "1" (default). As this threshold is lowered, the TP approaches 1: virtually every person who eventually defaults is accurately predicted as a "1" (default). But simultaneously, the FP approaches 1: nearly everyone is categorized as a "1," even those who do not eventually default. Conversely, increasing the threshold leads to a lower FP (since almost every observation is predicted as a "0"), but also lowers the TP (observations likely to default are incorrectly predicted as a "0"). The ROC curve avoids the need to select one threshold as a decision rule, and instead maps out this relationship between the TP and the FP for a specific model over all possible thresholds.

[42] We use the hmeasure package for Python to compute the H-Measure.

bootstraps using the "Corrected Resampled t-Test" proposed by Nadeau and Bengio (1999). We graph the results for the ROC AUC and H-Measure in Figure A.6.

Lender Cash Flow Intensity Measure To construct a measure of cash-flow intensity (mentioned in Section 4.2) for each lender in our Platform data we first restrict the sample to those with at least 50 application forwards and approval rates below 95 percent. For each lender, we then perform 100 stratified random splits of its application-level data (80 percent training, 20 percent testing). On each split we run the Baseline and Cash Flow model in a Random Forest classifiers using identical tuning parameters (Estimators: 100, Maximum Tree Depth: 10, Minimum Samples to Split: 15, Minimum Child Weight: 20). We create the average ROC AUC for all lenders for whom the model was able to run and calculate the percent improvement from the baseline to the cash flow model. We then define a lender as "Cash Flow Intensive" if their improvement is positive. Further details about lender level performance metrics are contained in Table A.5.