Invisible Primes: Fintech Lending with Alternative Data

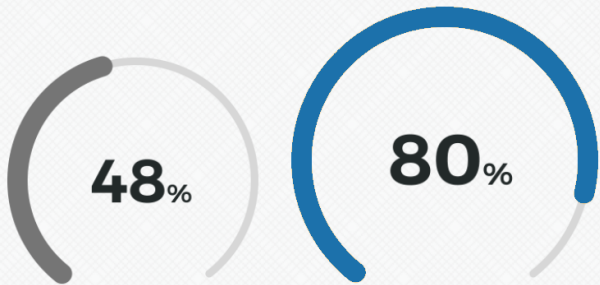
Marco Di Maggio (*HBS*; *NBER*) and **Dimuthu Ratnadiwakara** (*LSU*)

The Digital Future: Fintech, AI, and the Path to Financial Inclusion | December 2023

Lab for
Inclusive
FinTech

Four in five Americans

have never defaulted on a credit product, yet less than half have access to prime credit.* The implication is eye-opening. With a smarter credit model, lenders could approve almost twice as many borrowers, with fewer defaults.



* According to an Upstart retrospective study completed in December 2019.

Credit score is central to credit decisions

Lenders **primarily use credit scores** to evaluate the probability that an individual will repay loans

- GSE cutoff of 620; Marcus, SunTrust - 660, SoFi -680

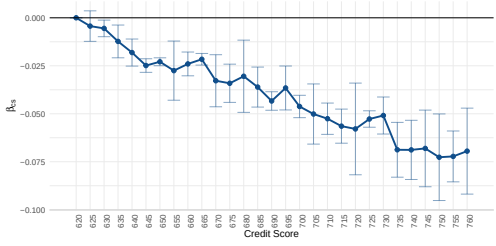
Credit score is central to credit decisions

Lenders **primarily use credit scores** to evaluate the probability that an individual will repay loans

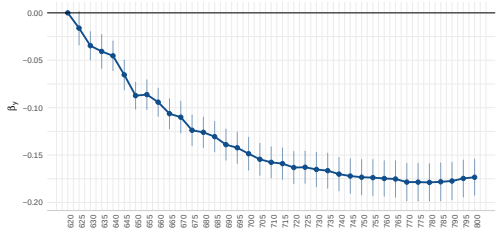
- GSE cutoff of 620; Marcus, SunTrust - 660, SoFi -680

Credit score is a **good predictor of default**, in general

Credit card defaults



Mortgage defaults



Large segment below acceptable credit score

*"FICO scores are good, but they're **not perfect**."— Roger Hochschild, Discover Financial Services CEO*

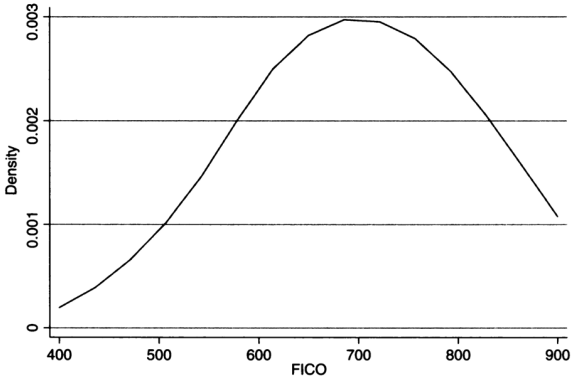


FIGURE I
FICO Distribution (U.S. Population)

Source: Keys, B. J., Mukherjee, T., Seru, A., Vig, V. (2010)

Large segment below acceptable credit score

*"FICO scores are good, but they're **not perfect**."— Roger Hochschild, Discover Financial Services CEO*

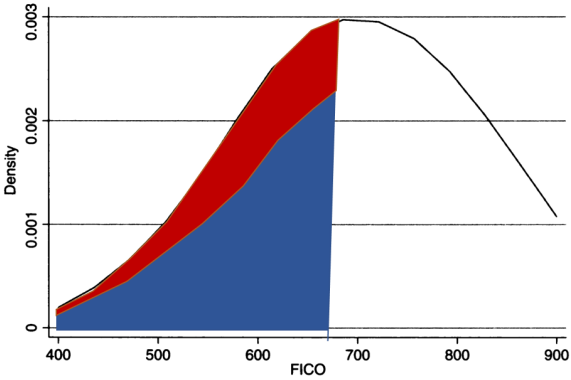


FIGURE I
FICO Distribution (U.S. Population)

Recent graduates

Recent immigrants

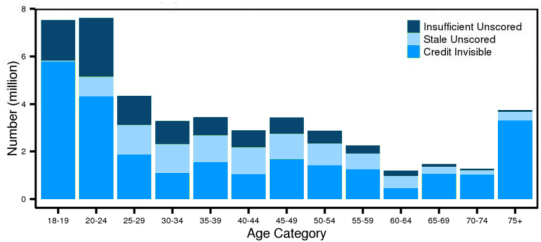
Self-employed individuals

...

Source: Keys, B. J., Mukherjee, T., Seru, A., Vig, V. (2010)

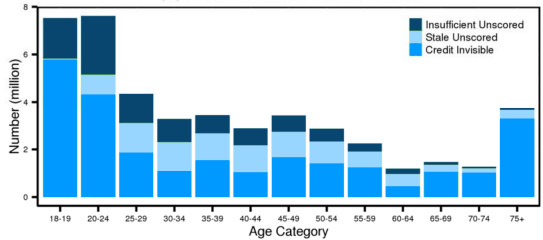
26 million adults lack a credit record - Credit Invisibles

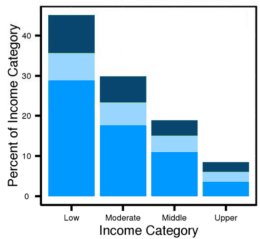
Younger, lower-income, minorities

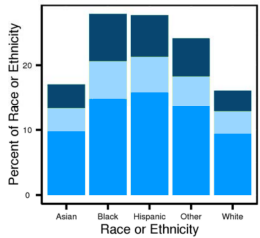


26 million adults lack a credit record - Credit Invisibles

Younger, lower-income, minorities







Research Questions

Winners and Losers

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Credit availability?

  "Adding this kind of alternative data into the mix thus holds out the promise of opening up credit for millions of additional consumers"–Richard Cordray, director of the CFPB

Research Questions

Winners and Losers

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Credit availability?

  "*Adding this kind of alternative data into the mix thus holds out the promise of opening up credit for millions of additional consumers*"–Richard Cordray, director of the CFPB

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Lower rates?

Research Questions

Winners and Losers

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Credit availability?

  "Adding this kind of alternative data into the mix thus holds out the promise of opening up credit for millions of additional consumers"–Richard Cordray, director of the CFPB

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Lower rates?

- Types of alternative data

Research Questions

Winners and Losers

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Credit availability?

  "Adding this kind of alternative data into the mix thus holds out the promise of opening up credit for millions of additional consumers"–Richard Cordray, director of the CFPB

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Lower rates?

- Types of alternative data

- **Impact** on borrowers $\Longrightarrow$ Better financial outcomes?

Research Questions

Winners and Losers

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Credit availability?

  "Adding this kind of alternative data into the mix thus holds out the promise of opening up credit for millions of additional consumers"–Richard Cordray, director of the CFPB

- Use of **alternative data** and **AI credit models** $\Longrightarrow$ Lower rates?

- Types of alternative data

- **Impact** on borrowers $\Longrightarrow$ Better financial outcomes?

- Do alternative data inadvertently **reduce** credit access for some households?

Roadmap

Data/Setting

The Platform's Underwriting Model

The Predictive Power of the Platform's Model

Data or Model?

The Effects of the Advanced Underwriting Model on Borrowers

Our Setting: Upstart

- Anonymized administrative data and the underwriting algorithm from Upstart, a major fintech lender

- Use alternative data and artificial intelligence in making credit underwriting and pricing decisions
  - → Variables from credit report, alternative variables, AI ⟶ prob. of default

- Attracted the attention of the CFPB in 2017; concerns about the potential violation of fair lending regulations; issued a **'no-action letter'**

Key Features of the Data

Close to the ideal setting

- **Credit decisions and interest rates** for all applicants
- Full **information set** of the Platform
- **Counterfactual credit decisions and interest rates** for all applicants
  - A traditional model developed with the CFPB; Used for regulatory reporting
  - Representative of the traditional lenders' credit decisions
  - An additional benchmark model provided by a large bank

Key Features of the Data

Close to the ideal setting

- **Credit decisions and interest rates** for all applicants
- Full **information set** of the Platform
- **Counterfactual credit decisions and interest rates** for all applicants
  - A traditional model developed with the CFPB; Used for regulatory reporting
  - Representative of the traditional lenders' credit decisions
  - An additional benchmark model provided by a large bank
- A **panel** of both the funded *and* rejected applicants
  - Credit report for an additional 12 months after application for all applicants

Data

Covers the 2014-2021 period

	Disqualified	Funded
Number of Obs	2,374,912	770,523
Credit report at origination	Yes	Yes
≈ 1400 variables		
Loan performance data	No	Yes
Credit report after 1 year	Yes	Yes
Upstart model outcome	Yes	Yes
Counterfactual model outcome	Yes	Yes
Alternative variables	Yes	Yes
Education, employment, device, digital footprint		

Funded vs. Disqualified Applications

- 10k - 3-5 year unsecured loans
- On average, the funded applicants exhibit a 70-point **higher credit score** and a $14,000 **higher annual income**.
- However, they also have **higher total liabilities and credit balances**.
- Borrowers who get funded are also more likely to be **college-educated**, **less likely to be hourly employees**, and more likely to **use a computer** and use the loan for debt consolidation

Roadmap

Data/Setting

The Platform's Underwriting Model

The Predictive Power of the Platform's Model

Data or Model?

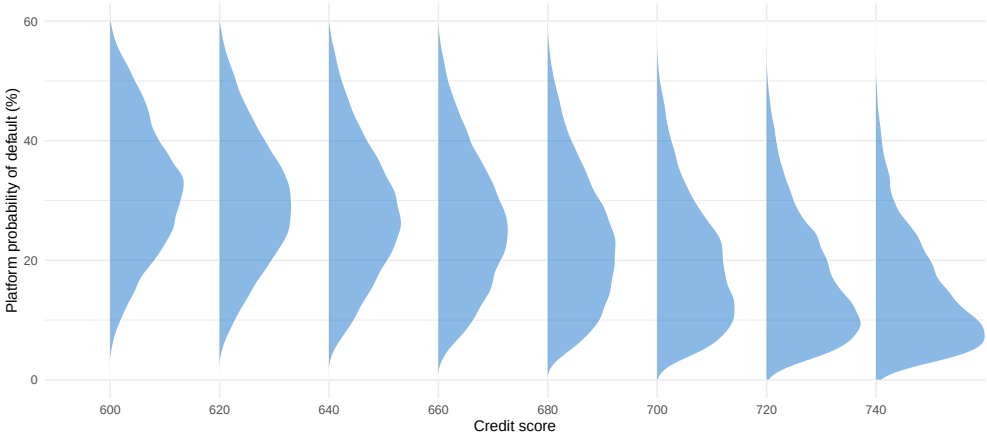
The Effects of the Advanced Underwriting Model on Borrowers

The Platform Model vs. Traditional Benchmarks

- **Platform's Underwriting Model:**
  - → Uses over 1,600 variables (traditional+alternative) for predicting the PD
  - → Utilizes sophisticated machine learning techniques.
- **Traditional Model (Benchmark 1):**
  - → Developed in coordination with the CFPB for regulatory reporting
  - → Logistic regression; traditional variables
- **Large Bank Model (Benchmark 2):**
  - → Used by one of the top 25 banks in the U.S.
  - → Approves based on credit scores, DTI, and loan amounts
- **Conventional Credit Scores (Benchmark 3)**

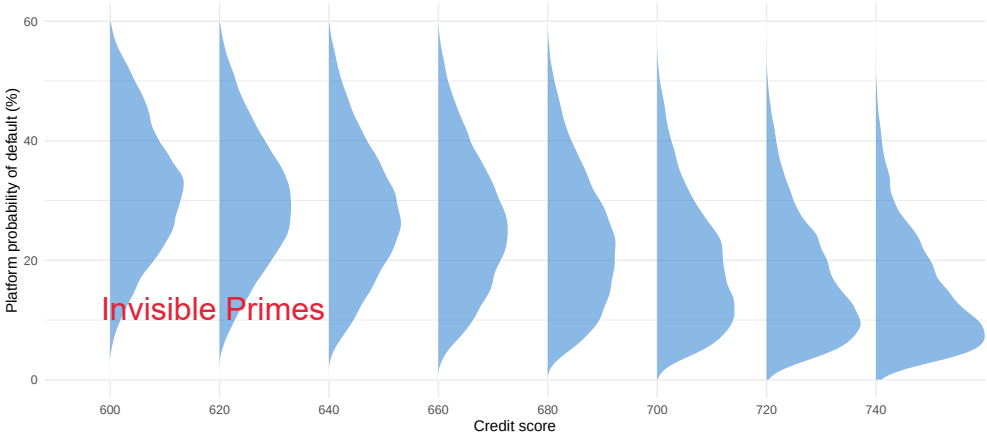
Platform PD vs. Credit Score

The Platform's model identifies risk factors not captured by the traditional models



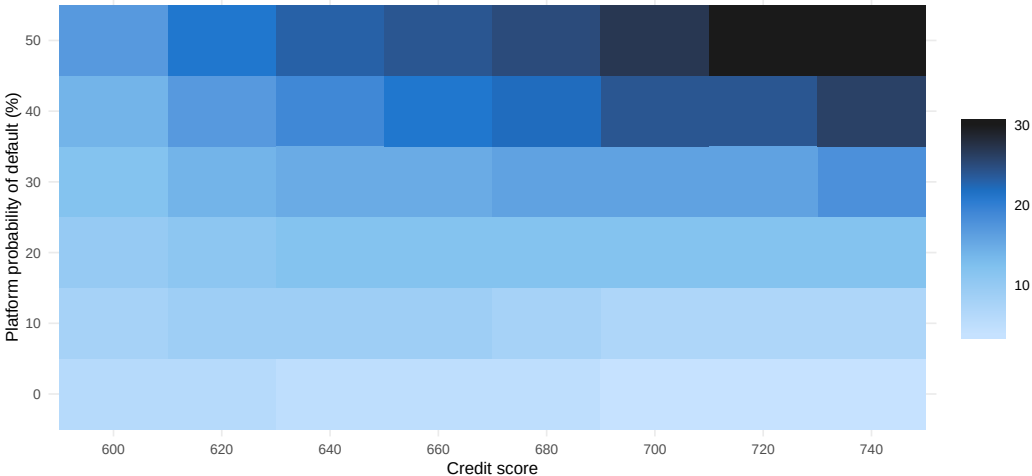
Platform PD vs. Credit Score

The Platform's model identifies risk factors not captured by the traditional models



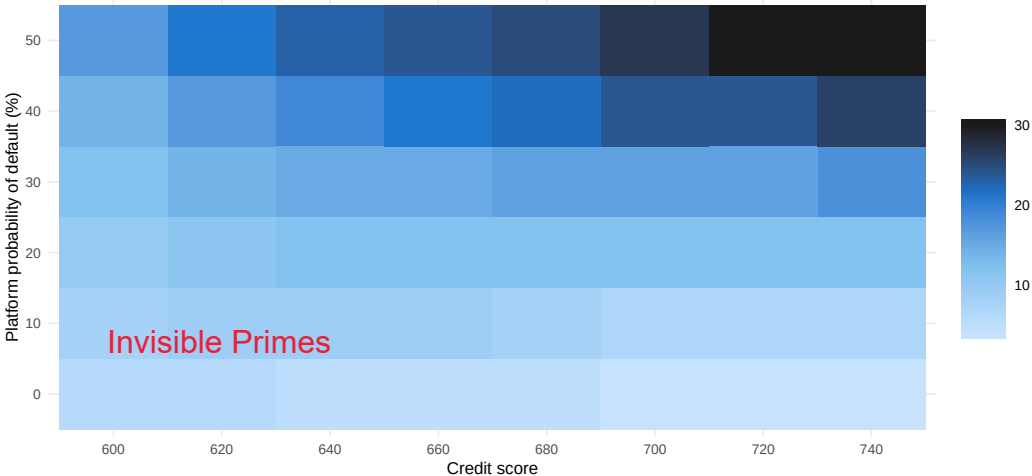
Model Performance: Platform PD vs. Credit Score

The platform model does a good job in identifying invisible primes



Model Performance: Platform PD vs. Credit Score

The platform model does a good job in identifying invisible primes



Roadmap

Data/Setting

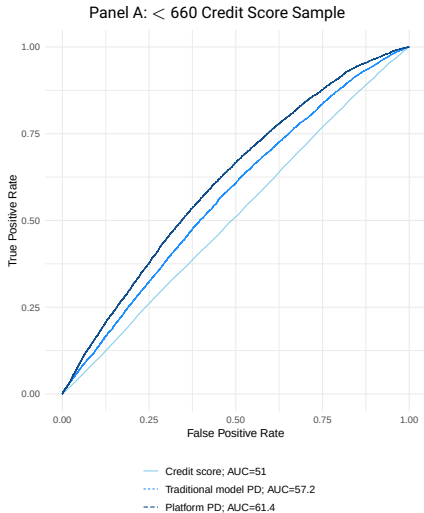
The Platform's Underwriting Model

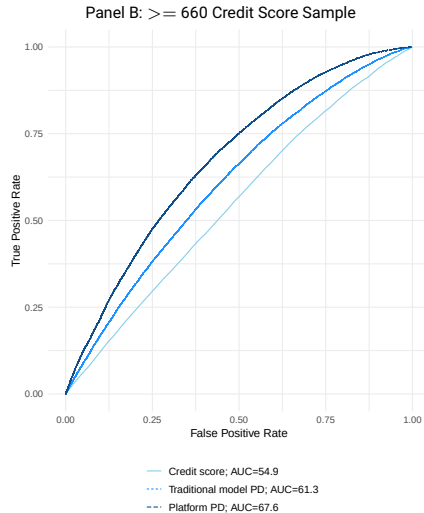
The Predictive Power of the Platform's Model

Data or Model?

The Effects of the Advanced Underwriting Model on Borrowers

Model Performance Comparison (AUC)





Roadmap

Data/Setting

The Platform's Underwriting Model

The Predictive Power of the Platform's Model

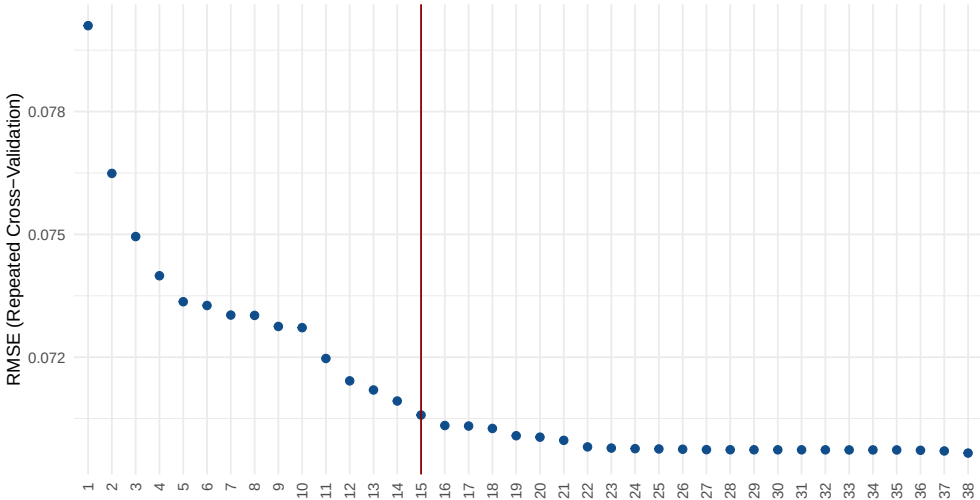
Data or Model?

The Effects of the Advanced Underwriting Model on Borrowers

What are the main features?

- We use Recursive Feature Elimination with Random Forests (RFE-RF) to select the most relevant variables in predicting the Upstart score.
- The RFE-RF procedure performs feature selection by iteratively training a random forest model, then ranking the different features, and finally removing the lowest ranking features, i.e. the ones not improving the predictive power of the model.
- It allows the different features to have non-linear effects and includes interactions.

Recursive feature elimination: Root-mean-square deviation



Education Exchange Rate

- Conditional on credit score, loan amount, and age:
  - To go from high school or less to an advanced degree: $107k
  - From associate to advanced: $114k
  - From college to advanced: $22k

Education Exchange Rate

- Conditional on credit score, loan amount, and age:
  - To go from high school or less to an advanced degree: $107k
  - From associate to advanced: $114k
  - From college to advanced: $22k
- Conditional on age, income and loan amount:
  - High school to Advanced: 37 points
  - Associate to Advanced: 23 points
  - College to Advanced: 4 points

Contribution of data vs. the sophisticated model

- **Decompose the AUC** difference between the traditional and the Platform model
  - → Recall, the traditional model is a logistic regression using all the traditional variables

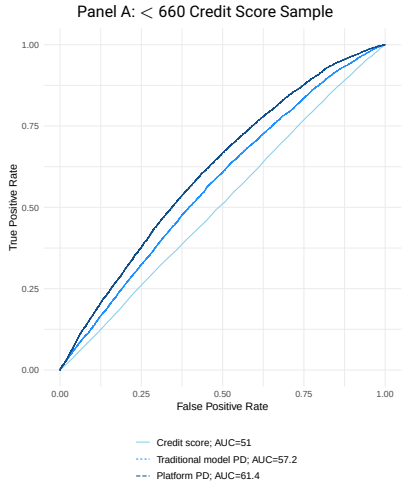
- A third logistic model (new model) that augments the traditional model's output with:
  - → Education, Employment type, Employment industry, Loan purpose, Device, Technology

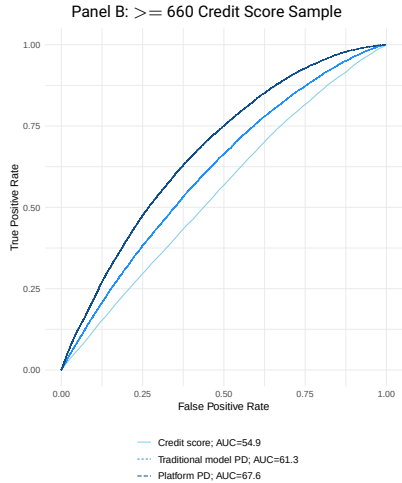
- Any difference in AUC between the traditional model and the new model can be attributed to the use of **alternative data**

- The difference in AUC between the new model and the Platform model can be attributed to the **increased sophistication** of the model

Contribution of data vs. the sophisticated model

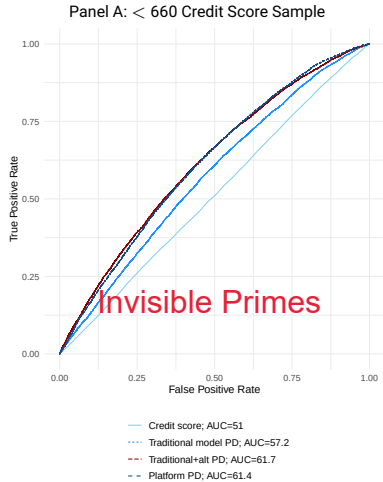
Superior performance is mostly due to alternative data

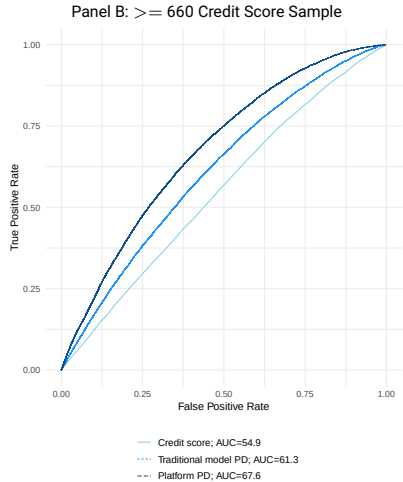




Contribution of data vs. the sophisticated model

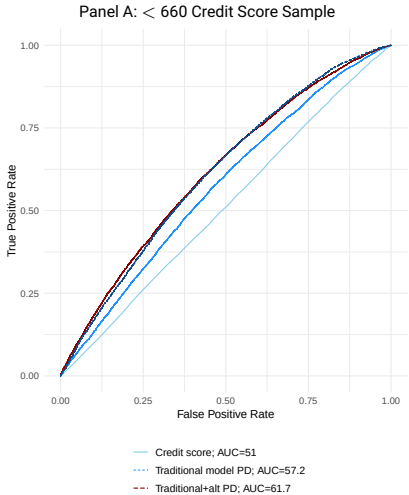
Superior performance is mostly due to alternative data

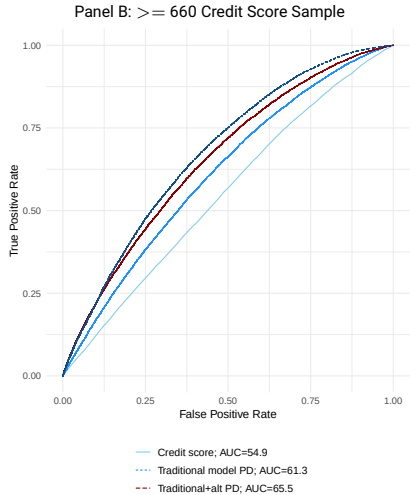




Contribution of data vs. the sophisticated model

Superior performance is mostly due to alternative data





Roadmap

Data/Setting

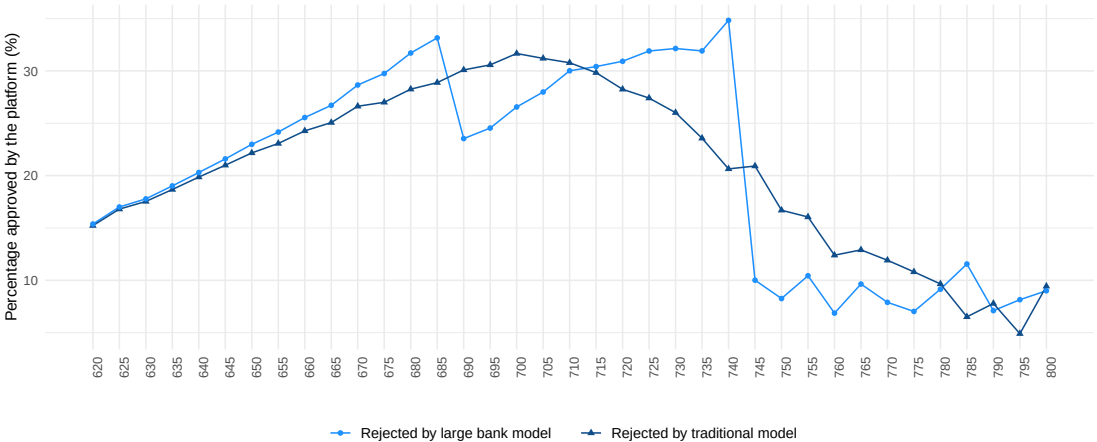
The Platform's Underwriting Model

The Predictive Power of the Platform's Model

Data or Model?

The Effects of the Advanced Underwriting Model on Borrowers

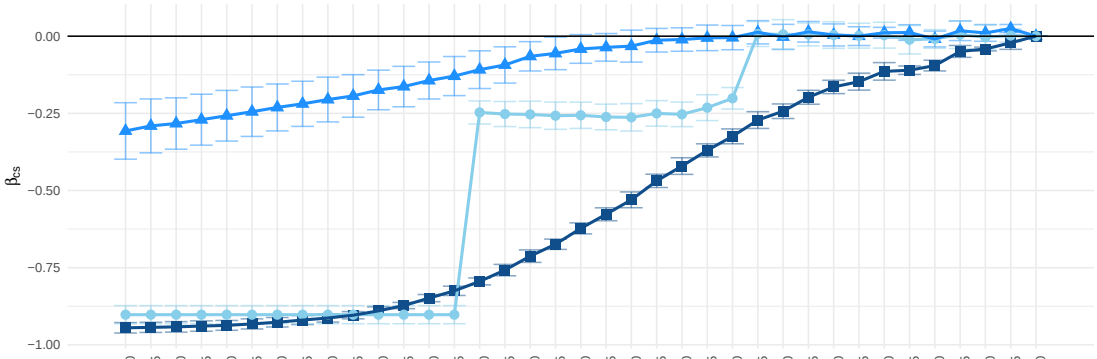
Platform approval rates for rejected loans by traditional models



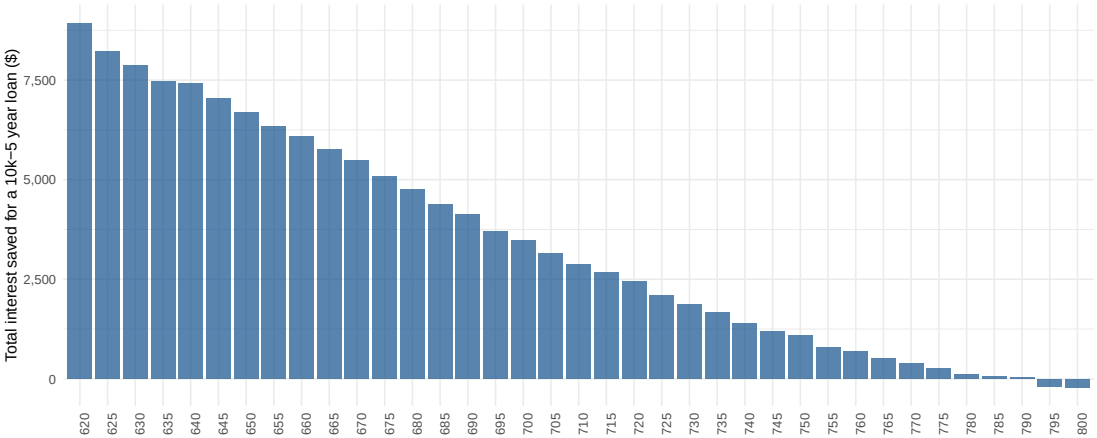
Loan approval by different models

$$approved_{i,z,y} = \sum_{cs} \beta_{cs} \times I(credit\ score_i \in cs) + \mu_{z,y} + \epsilon_{i,z,y}$$

i, z, y, and, cs represent the application, zip code, year, and credit score bin, respectively. Omitted category $cs = 800$



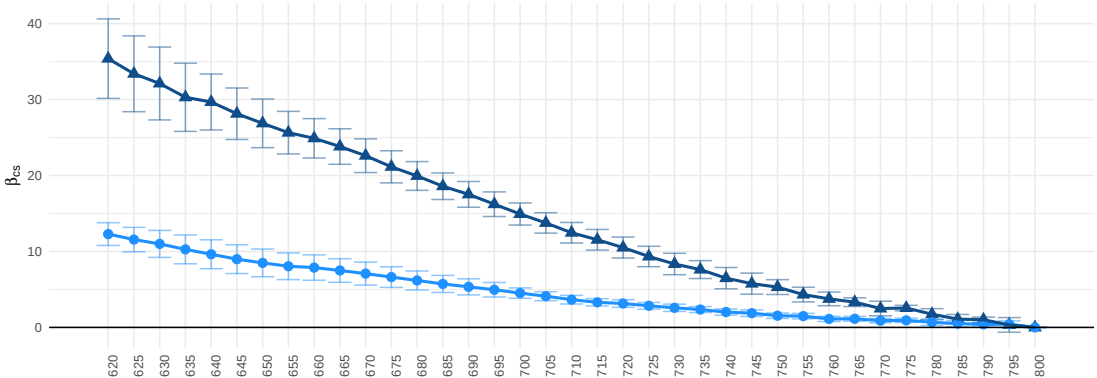
Savings due to the Platform model, compared to the traditional model



Regression evidence (Dependent variable: interest rate)

$$apr_{i,z,y} = \sum_{cs} \beta_{cs} \times I(credit\ score_i \in cs) + \mu_{z,y} + \epsilon_{i,z,y}$$

i, z, y, and, cs represent the application, zip code, year, and credit score bin, respectively. Omitted category $cs = 800$



Does better access to credit improve borrower outcomes?

Challenges:

- Omitted variable bias in OLS estimation
- Ability to observe longitudinal data for both approved and rejected applicants

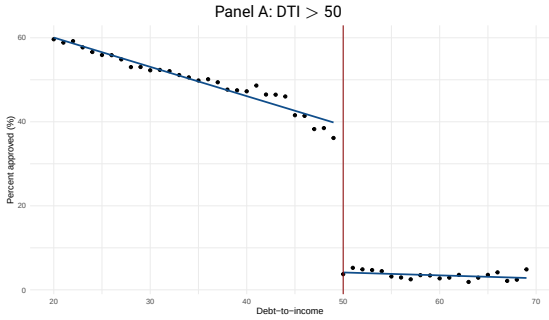
Our strategy: Utilizing two disqualifying criteria in an RDD framework

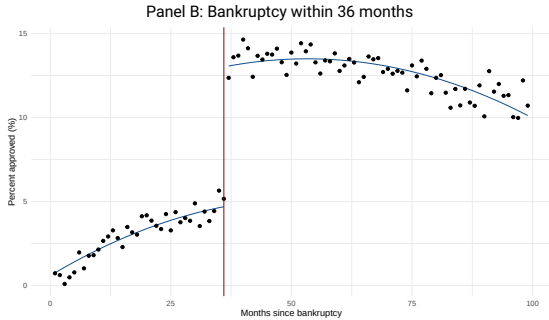
- DTI $>$ 50%
- Loan application post-bankruptcy within 36 months

Data: Platform data tracks both rejected and approved applicants for at least 12 months post-application

Measures of financial health: Change in credit score, missed credit card payments, and first-time home purchase.

RDD Setup





$$\text{Approved}_i = \beta_0 + \beta_1 I(\text{DTI} > 50\%) + \beta_2 I(\text{DTI} > 50\%) * \text{DTI} + \beta_3 \text{DTI} + \boldsymbol{\beta_4 X} + \mu_{zy} + \eta_i \quad (1)$$

$$Y_i = \gamma_0 + \gamma_1 \widehat{\text{Approved}}_i + \gamma_2 \text{DTI} + \boldsymbol{\Gamma_3 X} + \mu_{zy} + \mu_i \quad (2)$$

$$\text{Approved}_i = \beta_0 + \beta_1 I(MSB > 36) + \beta_2 \text{MSB} + \beta_3 \text{MSB}^2 + \boldsymbol{\beta_4 X} + \mu_{zy} + \eta_i \quad (3)$$

$$Y_i = \gamma_0 + \gamma_1 \widehat{\text{Approved}}_i + \gamma_2 \text{MSB} + \gamma_3 \text{MSB}^2 + \boldsymbol{\Gamma_4 X} + \mu_{zy} + \mu_i \quad (4)$$

The Effects of Credit Access: Regression Results

Panel A: Debt-to-income ratio Discontinuity

	Credit score $<$ 660			Credit score $>=$ 660		
	Credit card delinq	Credit score change	Mortgage	Credit card delinq	Credit score change	Mortgage
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{\text{Approved}}$	-0.176*	0.082***	0.132*	-0.028	-0.008	-0.004
	(0.099)	(0.025)	(0.073)	(0.074)	(0.014)	(0.059)
Controls	Y	Y	Y	Y	Y	Y
Zip code*Year	Y	Y	Y	Y	Y	Y
N	29,607	29,607	21,118	12,768	12,767	7,581
Adjusted R^2	0.302	0.300	0.053	0.339	0.496	0.293

Panel B: Months Since Bankruptcy Discontinuity

	Credit score $<$ 660			Credit score $>=$ 660		
	Credit card delinq	Credit score change	Mortgage	Credit card delinq	Credit score change	Mortgage
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{\text{Approved}}$	-0.223**	0.050***	0.077	0.057	-0.014	-0.077
	(0.096)	(0.018)	(0.055)	(0.309)	(0.067)	(0.356)
Controls	Y	Y	Y	Y	Y	Y
Zip code*Year	Y	Y	Y	Y	Y	Y
N	49,760	49,663	42,755	15,545	15,533	10,269
Adjusted R^2	0.159	0.151	0.124	0.116	0.142	0.172

Conclusion

- Credit scores are not reliable predictors for certain borrower groups, suggesting the need for alternative means of assessments

- By considering easy-to-collect non-traditional factors like education, employment, and digital footprints, lenders could identify invisible prime borrowers and provide them access to cheaper credit

- Invisible primes' financial health improves significantly as a result of access to cheaper credit