Invisible Primes: Fintech Lending with Alternative Data

Marco Di Maggio
Harvard Business School; NBER

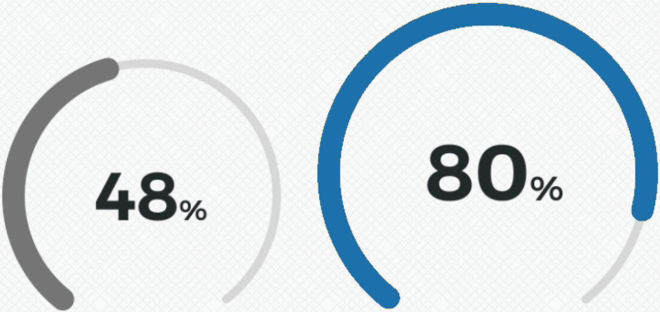
Dimuthu Ratnadiwakara
Louisiana State University

Don Carmichael
Upstart Network, Inc

11th FDIC Consumer Research Symposium

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.



Americans who have access to prime credit

Americans who never defaulted on a credit product

* 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

- GSEs have a minimum credit score of 620
- Marcus and SunTrust have a minimum credit score of 660 for personal loan, 680 for SoFi

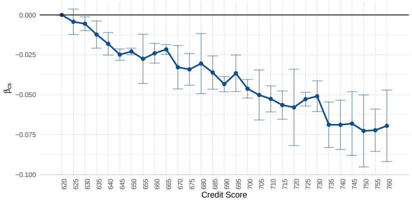
Credit score is central to credit decisions

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

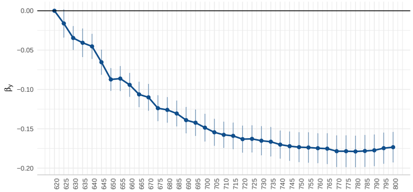
- GSEs have a minimum credit score of 620
- Marcus and SunTrust have a minimum credit score of 660 for personal loan, 680 for SoFi

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

Credit card defaults

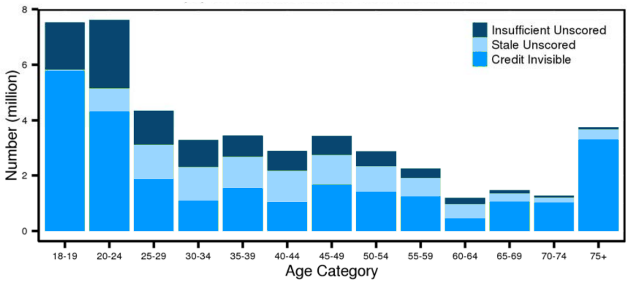


Mortgage defaults



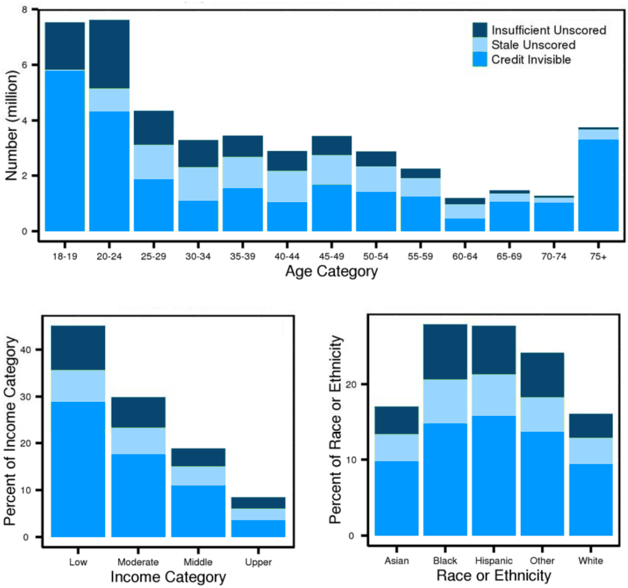
“FICO scores are good, but they’re ***not perfect****.”*
— Roger Hochschild, Discover Financial Services CEO

Credit Invisibles



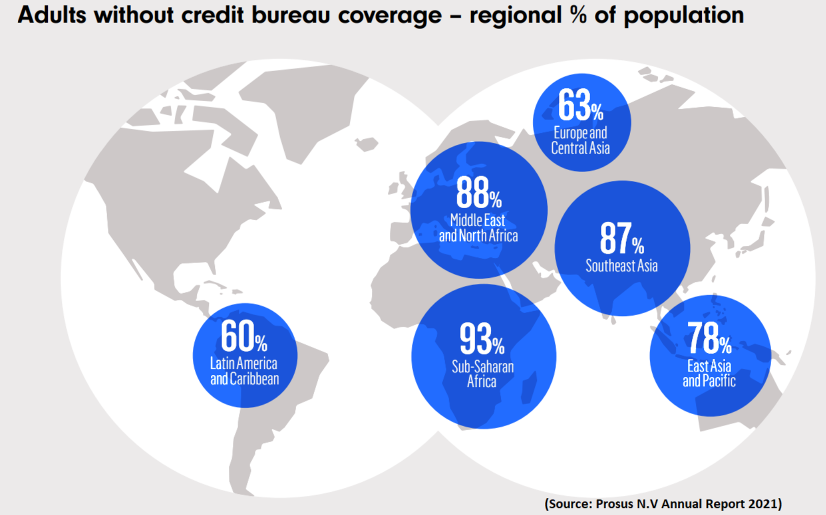
(Source: The CFPB Office of Research)

Credit Invisibles



(Source: The CFPB Office of Research)

Credit Invisibles



Invisible Primes?

The advent of fintech lenders has the potential to identify the prime borrowers among the credit invisibles

- Sophisticated algorithms
- Alternative data
- Online presence

Invisible Primes?

The advent of fintech lenders has the potential to identify the prime borrowers among the credit invisibles

- Sophisticated algorithms
- Alternative data
- Online presence

Research questions:

- Broader credit access?
- Types of data?
- Profitable to lend to invisible primes?
- Are borrowers better off?

Invisible Primes?

The advent of fintech lenders has the potential to identify the prime borrowers among the credit invisibles

- Sophisticated algorithms
- Alternative data
- Online presence

Research questions:

- Broader credit access?
- Types of data?
- Profitable to lend to invisible primes?
- Are borrowers better off?

Empirical Challenge:

- Observing the counterfactual

Our Setting: Upstart

- Anonymized administrative data from Upstart, a major fintech lender

- Why Upstart?
  - Use non-traditional variables such as education and job history for underwriting
  - CFPB granted a no-action letter (NAL) to Upstart regarding its automated model
  - Counterfactual based on a model developed by CFPB

Main Results

- Credit score is a poor predictor of default for Upstart loans
  - Particularly for low credit score borrowers

Main Results

- Credit score is a poor predictor of default for Upstart loans
  - Particularly for low credit score borrowers
- Upstart's model more accurately predicts the credit risk

Main Results

- Credit score is a poor predictor of default for Upstart loans
  - Particularly for low credit score borrowers
- Upstart's model more accurately predicts the credit risk
- $\approx$ 25% of borrowers with low credit scores would have been rejected by traditional models (extensive margin)

Main Results

- Credit score is a poor predictor of default for Upstart loans
  - Particularly for low credit score borrowers
- Upstart's model more accurately predicts the credit risk
- $\approx$ 25% of borrowers with low credit scores would have been rejected by traditional models (extensive margin)
- Significantly lower interest rates to low-credit score borrowers (intensive margin)

Main Results

- Credit score is a poor predictor of default for Upstart loans
  - Particularly for low credit score borrowers
- Upstart's model more accurately predicts the credit risk
- $\approx$ 25% of borrowers with low credit scores would have been rejected by traditional models (extensive margin)
- Significantly lower interest rates to low-credit score borrowers (intensive margin)
- Larger impact for individuals with thin credit files, younger borrowers, borrowers with advanced degrees

Main Results

- Credit score is a poor predictor of default for Upstart loans
  - Particularly for low credit score borrowers
- Upstart's model more accurately predicts the credit risk
- $\approx$ 25% of borrowers with low credit scores would have been rejected by traditional models (extensive margin)
- Significantly lower interest rates to low-credit score borrowers (intensive margin)
- Larger impact for individuals with thin credit files, younger borrowers, borrowers with advanced degrees
- **Overall benefits**:
  - Higher IRR for Upstart
  - Improved borrowers' financial health: borrowers are 20% less likely to default on their credit cards, their credit scores increase by about 9% and more likely to purchase a home

Main Results

- Credit score is a poor predictor of default for Upstart loans
  - Particularly for low credit score borrowers
- Upstart's model more accurately predicts the credit risk
- $\approx$ 25% of borrowers with low credit scores would have been rejected by traditional models (extensive margin)
- Significantly lower interest rates to low-credit score borrowers (intensive margin)
- Larger impact for individuals with thin credit files, younger borrowers, borrowers with advanced degrees
- **Overall benefits**:
  - Higher IRR for Upstart
  - Improved borrowers' financial health: borrowers are 20% less likely to default on their credit cards, their credit scores increase by about 9% and more likely to purchase a home
- Similar results from publicly available mortgage data

Data

- 770,523 loans and 2,374,912 disqualified applications from 2014 to 2021 (Q1)
- Access to all credit report variables and performance data
- Information on education, the type of the device (smartphone/computer) used by applicants, the operating system, the employment type and tenure, etc.

Data

- 770,523 loans and 2,374,912 disqualified applications from 2014 to 2021 (Q1)
- Access to all credit report variables and performance data
- Information on education, the type of the device (smartphone/computer) used by applicants, the operating system, the employment type and tenure, etc.
- **Upstart's probability of default estimate**

Data

- 770,523 loans and 2,374,912 disqualified applications from 2014 to 2021 (Q1)
- Access to all credit report variables and performance data
- Information on education, the type of the device (smartphone/computer) used by applicants, the operating system, the employment type and tenure, etc.

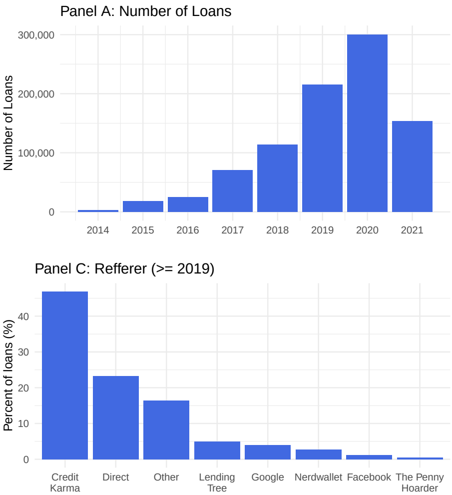
- **Upstart's probability of default estimate**
- **Traditional model** - the counterfactual

Data

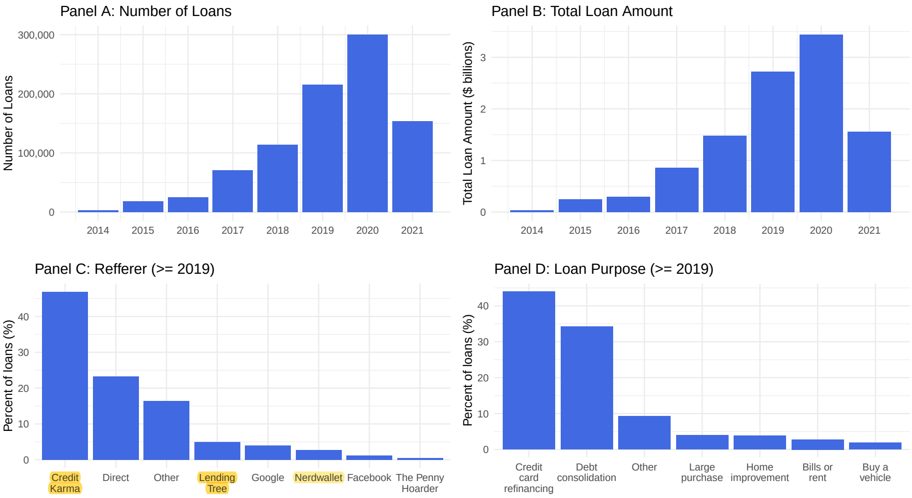
- 770,523 loans and 2,374,912 disqualified applications from 2014 to 2021 (Q1)
- Access to all credit report variables and performance data
- Information on education, the type of the device (smartphone/computer) used by applicants, the operating system, the employment type and tenure, etc.

- **Upstart's probability of default estimate**
- **Traditional model** - the counterfactual
- A **panel** of both the funded *and* rejected applicants

Upstart's Loan Growth and Source



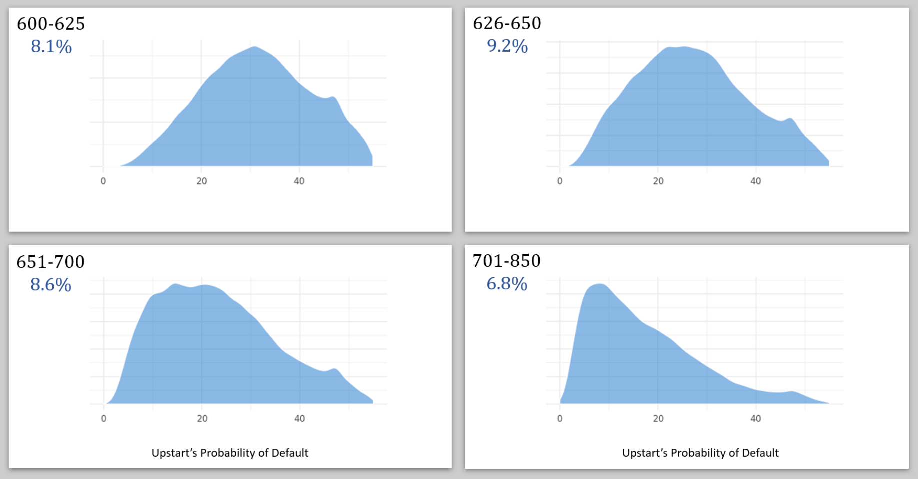
Upstart's Loan Growth and Source



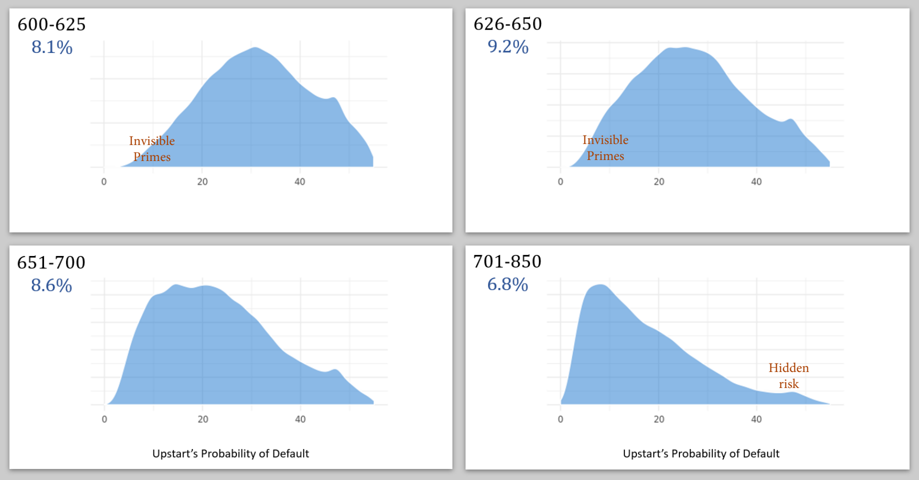
Descriptive Statistics: Funded vs. Disqualified Applications

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

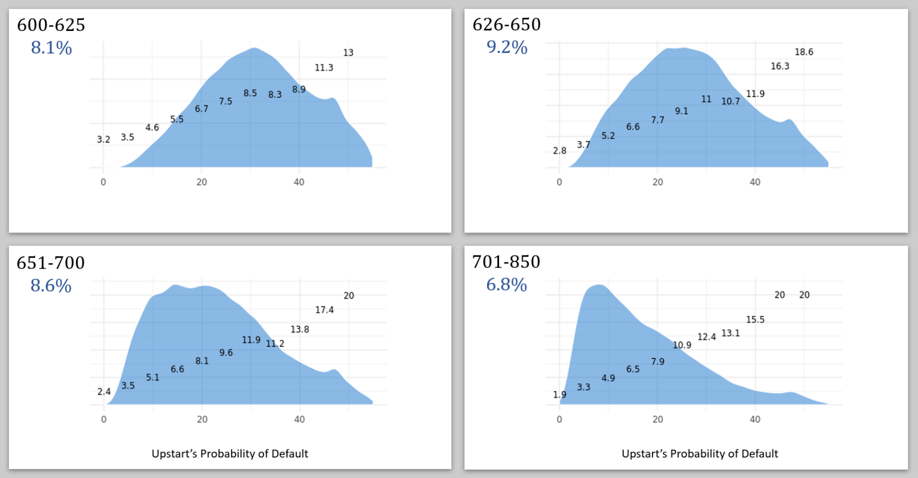
Predictability of Default



Predictability of Default



Predictability of Default



Alternative data and/or sophisticated model?

Main drivers of Upstart's credit model

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

Education Exchange Rates

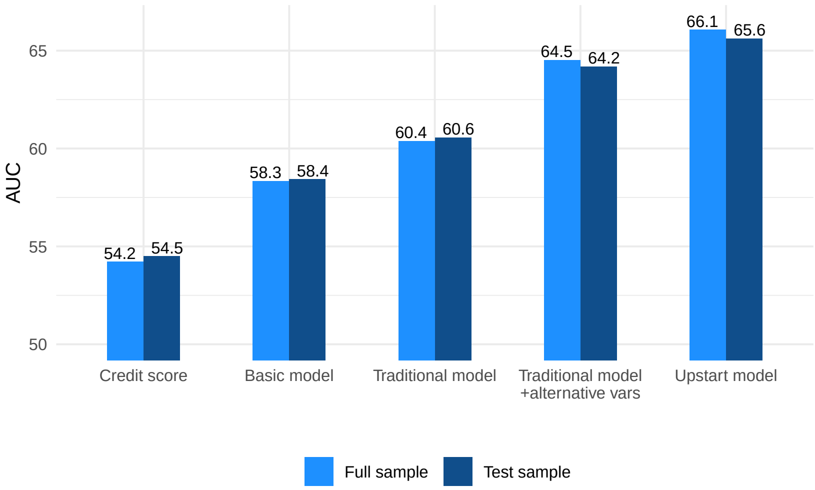
- Conditional on credit score, loan amount, and age:
  - To go from high school or less to an advanced degree: $107k
  - From associate to advanced: $114k
  - From college to advanced: $22k

Education Exchange Rates

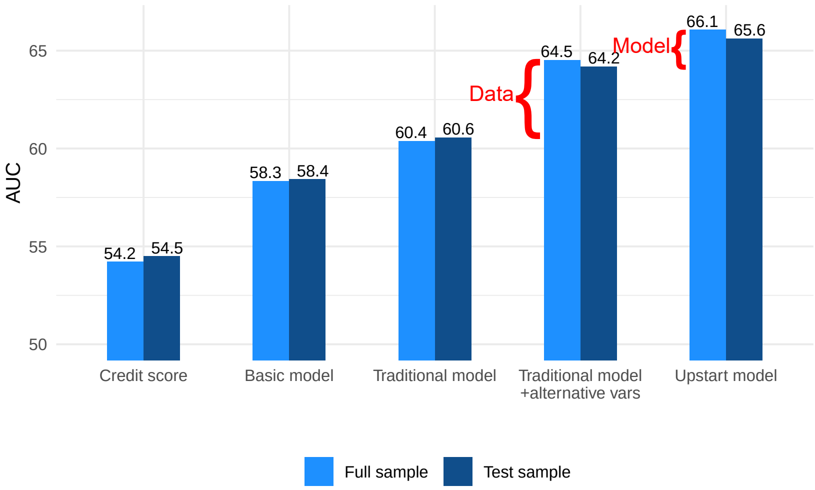
- Conditional on credit score, loan amount, and age:
  - To go from high school or less to an advanced degree: $107k
  - From associate to advanced: $114k
  - From college to advanced: $22k

- Conditional on age, income and loan amount:
  - High school to Advanced: 37 points
  - Associate to Advanced: 23 points
  - College to Advanced: 4 points

Alternative data significantly contributes to the predictive power

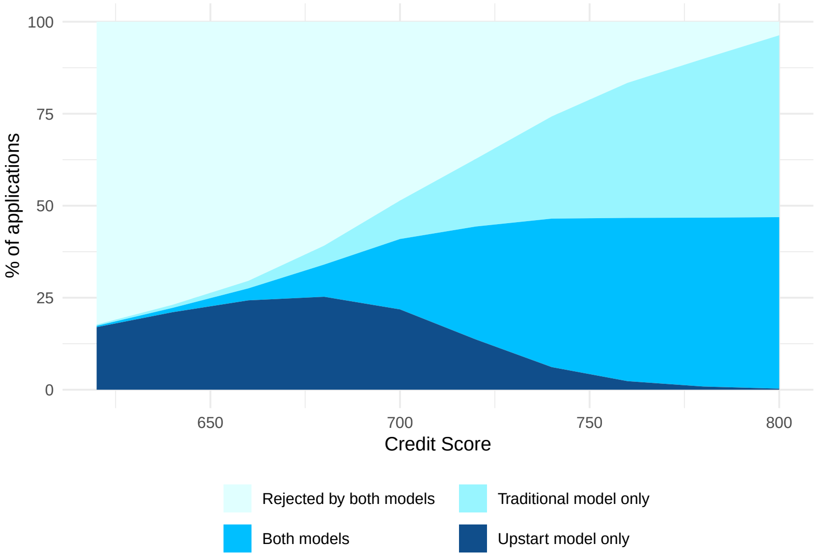


Alternative data significantly contributes to the predictive power

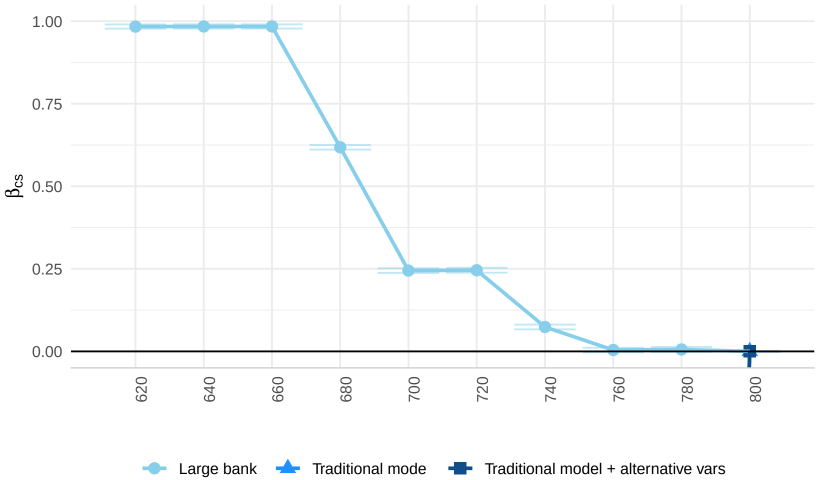


Does the use of alternative data improve credit access?

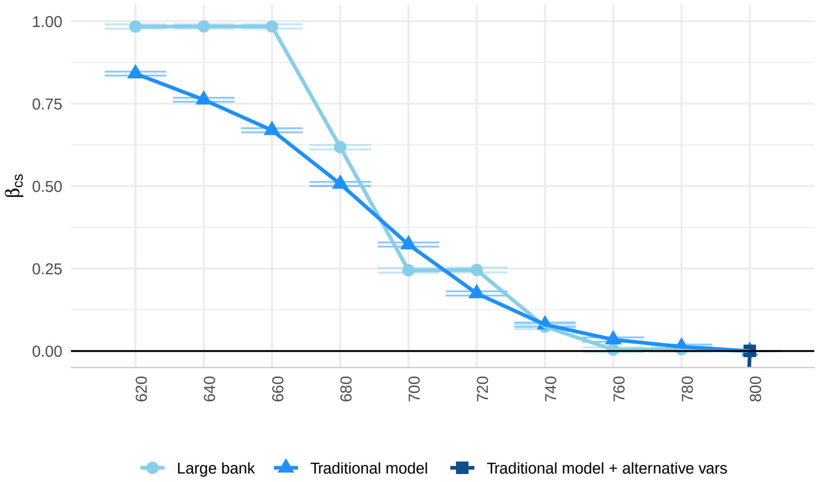
Upstart is more likely to approve applicants with low credit scores



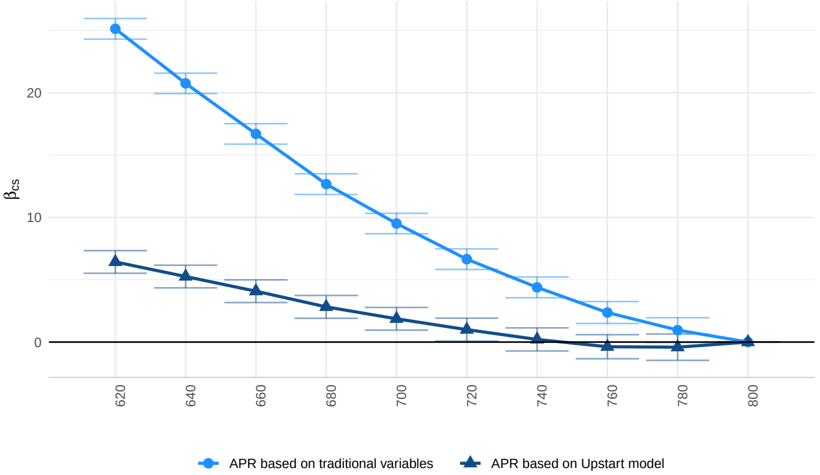
Traditional vs. Upstart: Rejection rate



Traditional vs. Upstart: Rejection rate



Traditional vs. Upstart: Interest rate difference



Who benefits the most due to the use of alternative data?

Borrower Differences

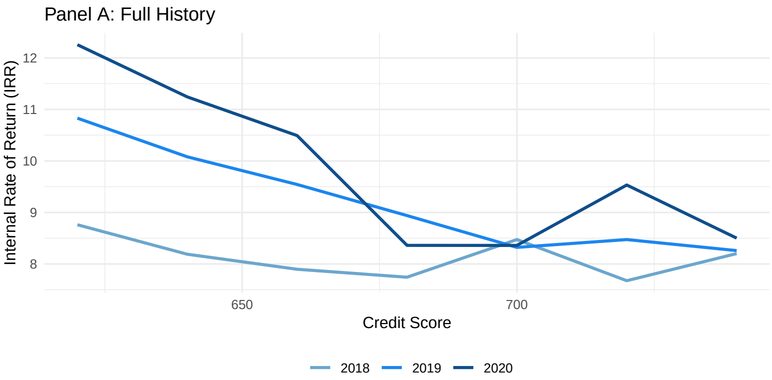
	Rejected by Trad. Model		Trad. APR - Upstart APR	
	(1)	(2)	(3)	(4)
Credit score/100	-0.660***		-15.229***	
	(0.001)		(0.033)	
Credit score $<$ 660		0.332***		9.067***
		(0.003)		(0.068)
Income $<$ 55k	-0.108***	-0.128***	-4.996***	-4.859***
	(0.001)	(0.002)	(0.028)	(0.033)
Advanced degree	0.067***	0.046***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

Regional Differences

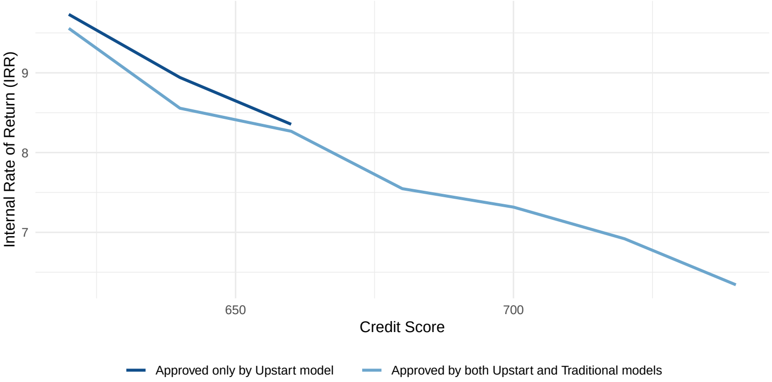
	Dep.var = Rejected by trad. model			Dep. var = Trad. APR - Upstart APR		
	(1)	(2)	(3)	(4)	(5)	(6)
Minority fraction in middle third	0.014***			0.569***		
	(0.003)			(0.065)		
Minority fraction in top third	0.023***			0.823***		
	(0.004)			(0.125)		
Fraction of renters in middle third		0.007**			0.228***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

Is lending to ‘invisible primes’ value enhancing for the lender?

Internal Rate of Return: 3-year loans, full history

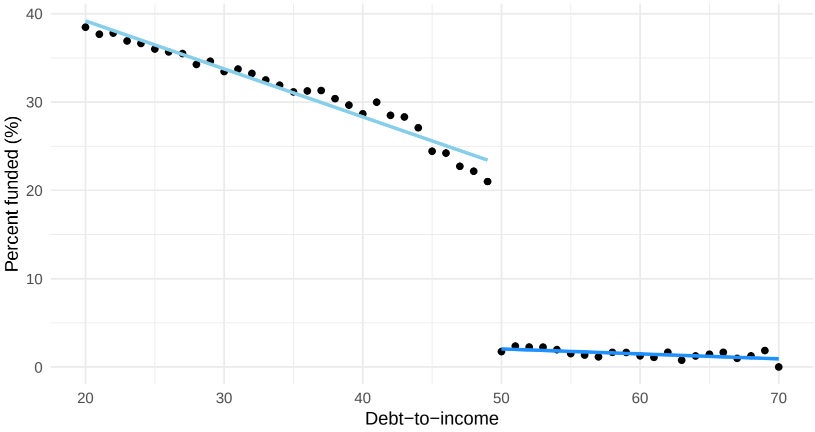


Internal Rate of Return: By Model Outcome



Are the borrowers better off?

We exploit a discontinuity in the probability of approval



Fuzzy RDD

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

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

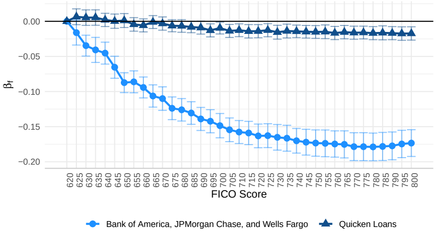
The Effects of Credit Access: RD

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

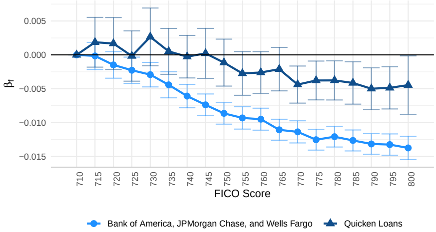
External Validity

- We compare the performance of mortgages originated by the three largest banks (Bank of America, Chase and Wells Fargo) to the performance of mortgages originated by Quicken Loans.
- FICO is not a good predictor of default for a fintech lender like Quicken borrowers

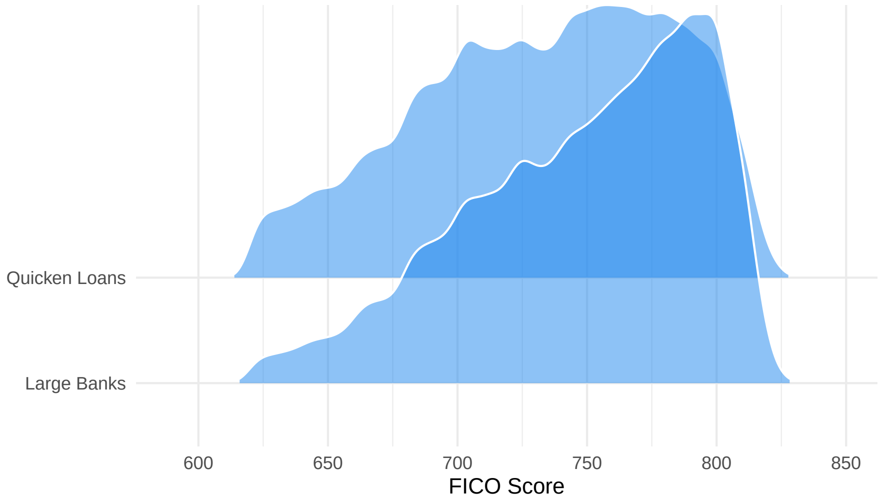
Panel A: Non-agency mortgages



Panel B: Mortgages sold to Freddie Mac



FICO Score Distribution



Conclusion

- A superior ability to predict default translates into broader access to credit
- Benefits for low-score borrowers on both the extensive and intensive margins
- Significant decrease in other liabilities defaults, increase in credit score and home purchase
- These positive findings do not refute arguments around some of the potential concerns around privacy and potential for statistical discrimination
- However, they do show that there are quantifiable benefits to both borrowers and lenders