20598 – Finance with Big Data

Week 8 Lecture: Banking in the age of AI

Clément Mazet-Sonilhac
clement.mazetsonilhac@unibocconi.it

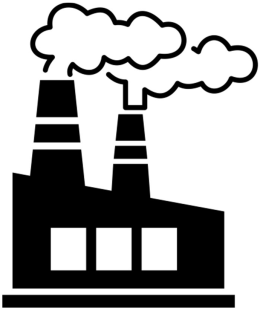
Department of Finance, Bocconi University

Outline

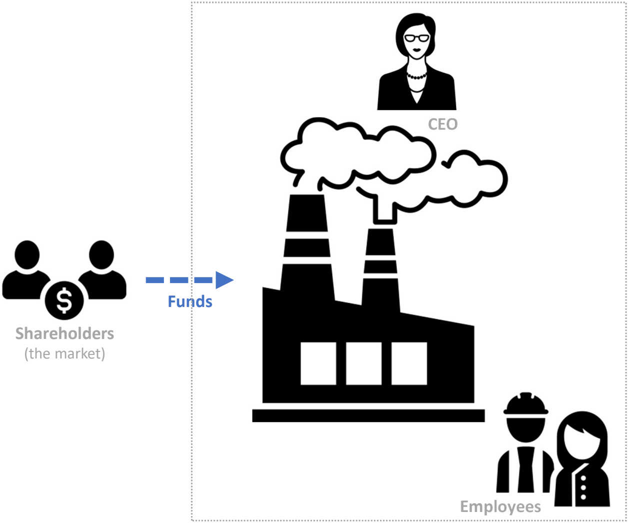
Hackathon

ML in Bank and FinTech Lending

The Hackathon : Great Job!




CEO
Employees


CEO
Funds
Shareholders
(the market)
Employees

The Hackathon: Great Job!

Let's discuss your work

Group 1, 5-9 and 10

Replication crisis?

RESEARCH ARTICLE | SOCIAL SCIENCES

Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty

Nate Breznau, Eike Mark Rinke, Alexander Wuttke, +162, and Tomasz Żółtak Authors Info & Affiliations

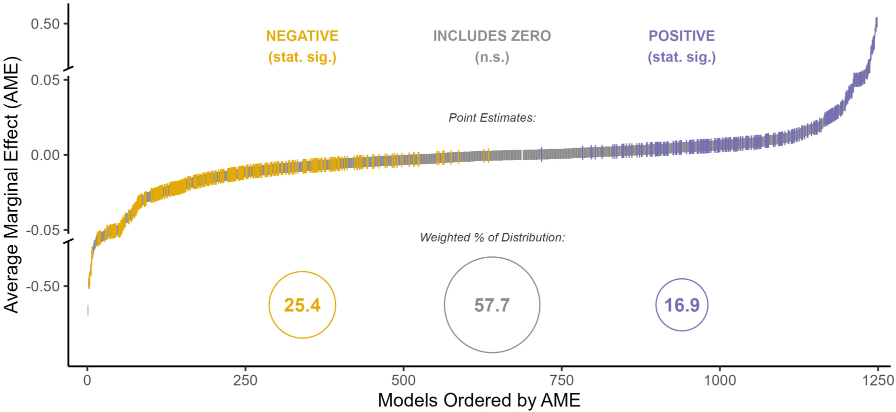
Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022

October 28, 2022 | 119 (44) e2203150119 | https://doi.org/10.1073/pnas.2203150119

THIS ARTICLE HAS BEEN UPDATED

- 161 researchers in 73 research teams using the same data independently
- Test the same prominent social science hypothesis: Does greater immigration reduces support for social policies among the public?

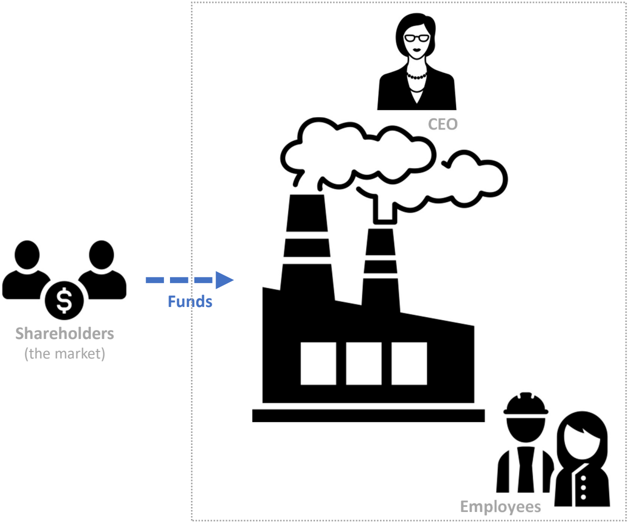
Replication crisis?

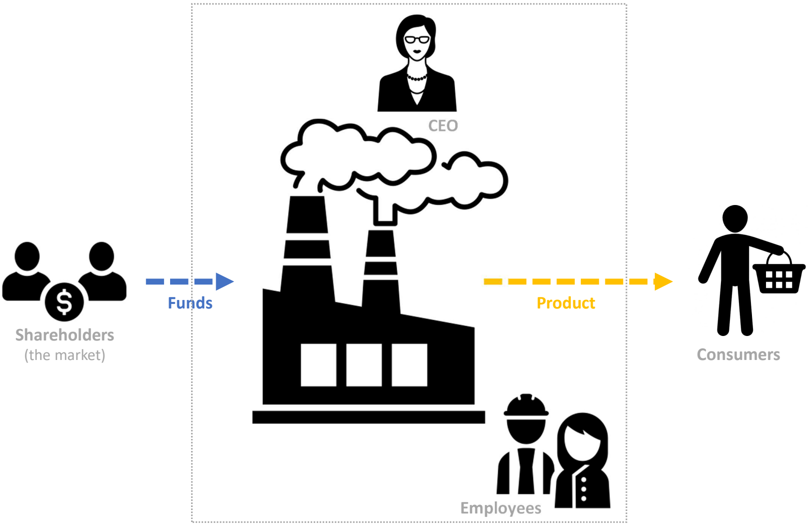


- Research teams reported both widely diverging numerical findings and substantive conclusions despite identical start conditions

The Hackathon : Great Job!

What about my work?

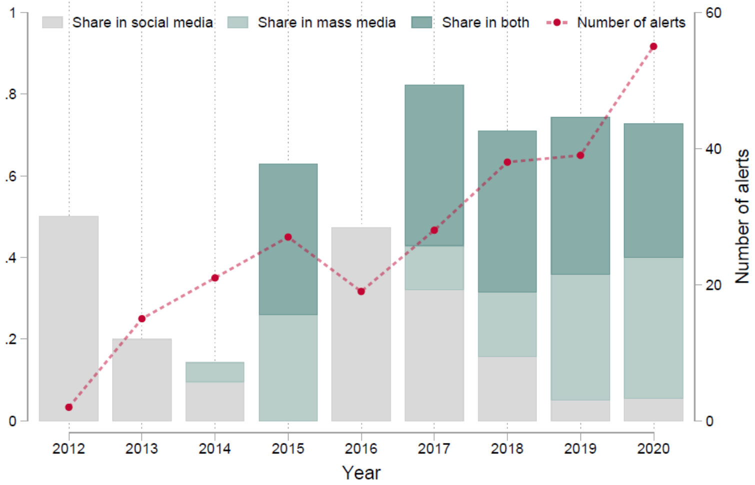

CEO
Funds
Shareholders
(the market)
Employees


CEO
Funds
Product
Shareholders
(the market)
Consumers
Employees

Some don't like it hot (Mazet-Sonilhac et al., 2024)

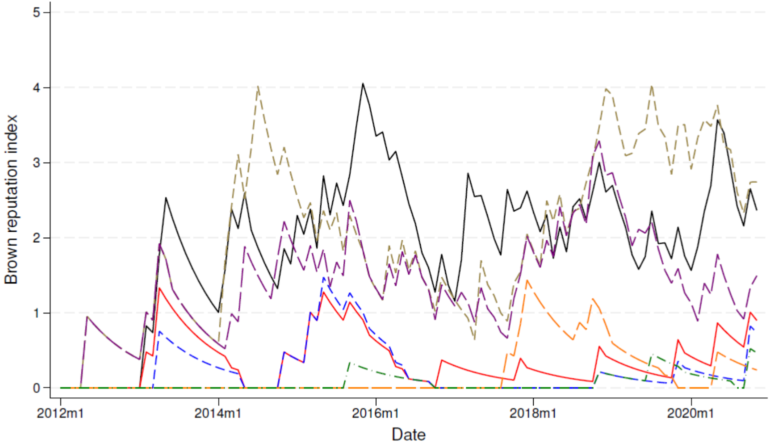
- How do bank's depositors react to NGO campaigns against brown banks?
- We merge several data sources :
  - Sigwatch data (you know it)
  - Deposits at the bank $\times$ city level (panel)
  - Interest rates for new loans (quarterly survey)
  - Green voting at city-level
  - Exposure to natural disasters

Some don't like it hot (Mazet-Sonilhac et al., 2024)



Some don't like it hot (Mazet-Sonilhac et al., 2024)

An Index for Brown Reputation with Fading Memory

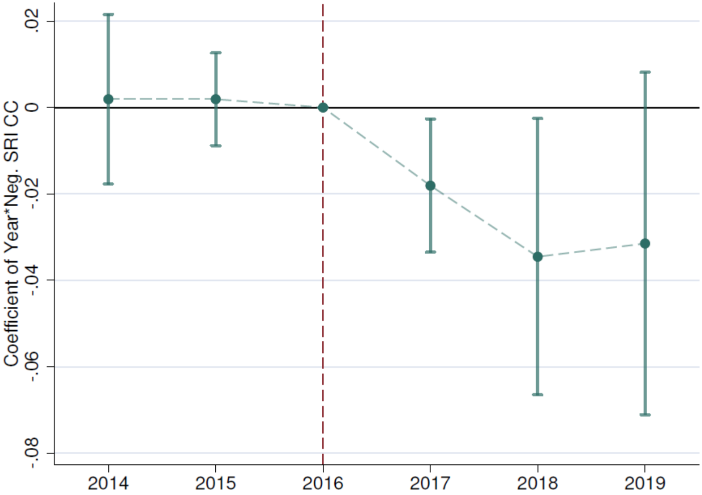


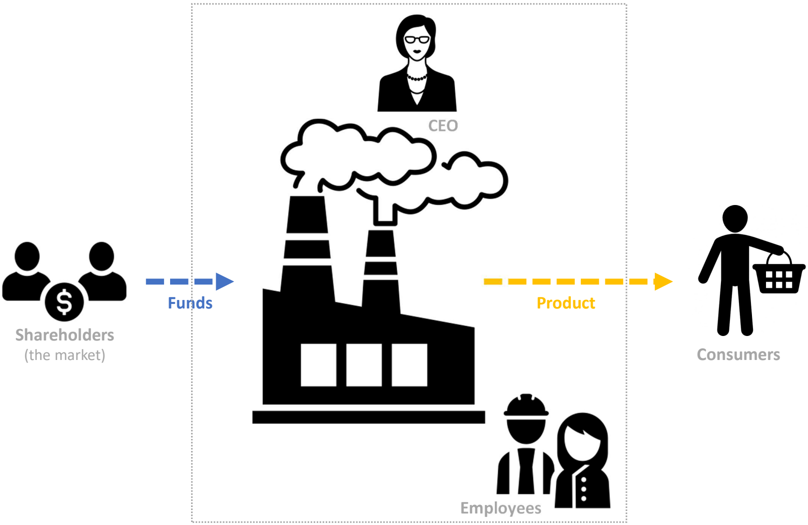
Some don't like it hot (Mazet-Sonilhac et al., 2024)

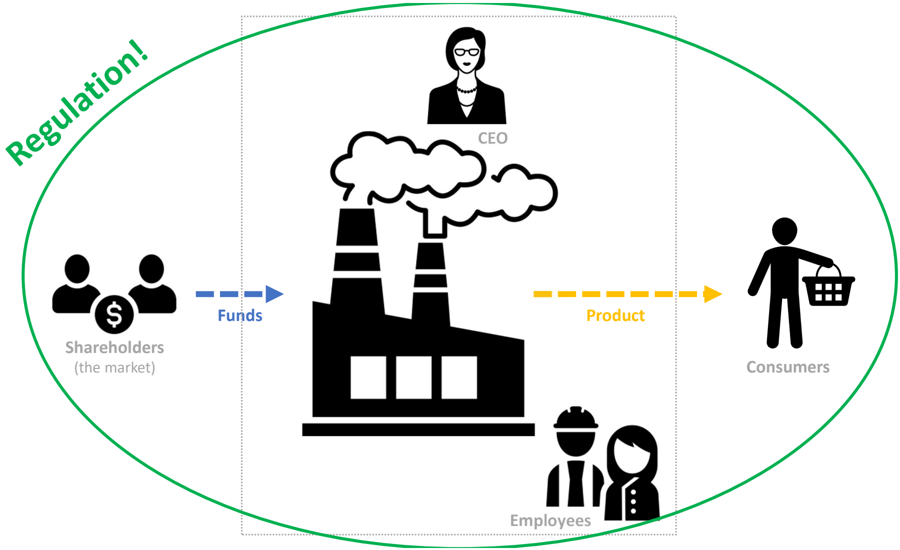
	$\text{Log(Deposits)}_{i,c,t}$			
	Large Banks		**All Banks**	
	(1)	(2)	(3)	(4)
Shock	-0.016***		-0.022***	
	[0.003]		[0.003]	
Shock × **Green vote intensity**	-0.004***		-0.003**	
	[0.001]		[0.001]	
Media Shock		-0.007		-0.014**
		[0.006]		[0.006]
Media shock × **Green vote intensity**		-0.008***		-0.006**
		[0.003]		[0.003]
Green vote intensity	0.001	-0.001	-0.006	-0.009
	[0.009]	[0.008]	[0.008]	[0.008]
Bank FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	131,849	131,849	138,727	138,727
Adj. R2	0.87	0.87	0.88	0.88

Some don't like it hot (Mazet-Sonilhac et al., 2024)

Impact of the Macron reform for Banking Flexibility




CEO
Funds
Product
Shareholders
(the market)
Consumers
Employees


Regulation!
CEO
Shareholders
(the market)
Funds
Product
Consumers
Employees

Questions?

Financial Intermediation with Big Data

Why do Banks Exist ?

Why do Banks Exist ?

1. Liquidity and Payment Services
2. Managing Risk
3. Reduce Information Frictions and Allocate Funds
4. Asset Transformation (Diamond and Dybvig, 1983)

Why do Banks Exist ?

1. Liquidity and Payment Services
2. Managing Risk
3. **Reducing Information Frictions and Allocate Funds**
4. Asset Transformation (Diamond and Dybvig, 1983)

Reducing Information Frictions and Allocate Funds

Asymmetric Information

- Banks determine which projects get financed. What is a good project?
- Banks can invest in technologies that allow them to screen loan applicants and monitor their projects $\rightarrow$ **informed lending**

Reducing Information Frictions and Allocate Funds

Asymmetric Information

- Banks determine which projects get financed. What is a good project?
- Banks can invest in technologies that allow them to screen loan applicants and monitor their projects → **informed lending**
- Low-tech informed lending :
  - Engage in long-term relationships
  - Geography and specialization matter
  - Screening : 1-on-1 meetings with borrowers, manual financial analysis

Reducing Information Frictions and Allocate Funds

Asymmetric Information

- Banks determine which projects get financed. What is a good project?
- Banks can invest in technologies that allow them to screen loan applicants and monitor their projects $\rightarrow$ **informed lending**

Today's lending: more data + machine learning methods

Machine Learning in Bank Lending

- A algorithm can be as good as a loan officer to evaluate the credit risk of an applicant?

Machine Learning in Bank Lending

- A algorithm can be as good as a loan officer to evaluate the credit risk of an applicant→ Evidence from PC Lab #6

- Banks are changing their industrial organization: less local branches, less loan officers (more data scientist), centralized loan approval process

Machine Learning in Bank Lending

Gambacorta et al., 2019 or e.g. Turkson et al., 2016

- Compares the predictive power of credit scoring models based on :
  - i machine learning techniques (decision tree approach)
  - ii traditional loss and default models (Linear model)

Machine Learning in Bank Lending

Gambacorta et al., 2019 or e.g. Turkson et al., 2016

- Compares the predictive power of credit scoring models based on :
  - i machine learning techniques (decision tree approach)
  - ii traditional loss and default models (Linear model)
- Loan-level data from a lending company in China for the period between May and Sep. 2017
  - Financial data, alternative data (phone calls, family), and ML-credit score
- Use an (exogenous) change in regulation policy on shadow banking in China that caused lending to decline and credit conditions to deteriorate.

Machine Learning in Bank Lending

Gambacorta et al., 2019 or e.g. Turkson et al., 2016

- Three OLS and Logit models :

$$\mathbb{P}\left(\text{Default}_{i,t}\right) = \phi\left(\alpha\, \mathbf{CS}_{i,t} + \mu_p + \mu_t + \epsilon_{i,t}\right) \tag{1}$$

$$\mathbb{P}\left(\text{Default}_{i,t}\right) = \phi\left(\beta\, \mathbf{X}_{i,t} + \mu_p + \mu_t + \epsilon_{i,t}\right) \tag{2}$$

$$\mathbb{P}\left(\text{Default}_{i,t}\right) = \phi\left(\beta\, \mathbf{X}_{i,t} + \delta\, \mathbf{Z}_{i,t} + \mu_p + \mu_t + \epsilon_{i,t}\right) \tag{3}$$

- **CS** is the ML-credit score, **X** is a set of financial variables, **Z** a vector of alternative variables

Machine Learning in Bank Lending

Gambacorta et al., 2019 or e.g. Turkson et al., 2016

ROC curves for different models — Figure 2

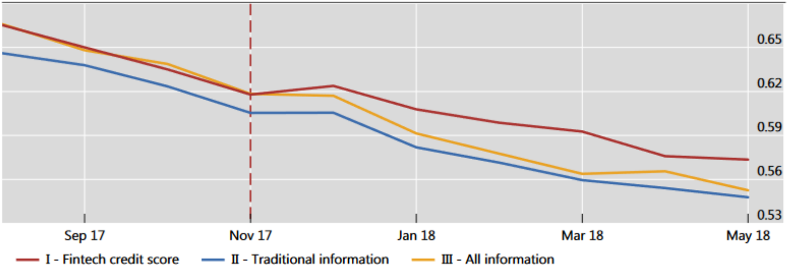
Source: Authors' calculations.

Machine Learning in Bank Lending

Gambacorta et al., 2019 or e.g. Turkson et al., 2016

Discriminatory powers of the models before and after the regulatory shock

Figure 6



The vertical axis reports the area under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores. The AUROC ranges from 50% (purely random prediction) to 100% (perfect prediction). The vertical dashed line indicates the date of the shock. In particular, it refers to a largely unexpected regulatory change that occurred in China on 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. This regulatory policy has led many financial intermediaries to increase their lending requirements, resulting in deteriorating credit conditions for borrowers.

Source: Authors' calculations.

New players: FinTech Lending

The Rise of FinTech Lending

The Rise of FinTech Lending

- Entry of new players = FinTech lending
- In the US, online lenders account for about 8–12% of new mortgage loans in 2016 (Source: Gambacorta et al., 2019)

The Rise of FinTech Lending

2011 Market Share	
Wells Fargo	24.20%
Bank of America	10.58%
JP Morgan Chase	9.95%
U.S. Bank Home Mortgage	4.38%
Citigroup	4.29%
Ally-GMAC	3.81%
PHH Mortgage	3.51%
Quicken Loans	2.03%
Flagstar Bancorp	1.80%
MetLife	1.60%

2019 Market Share	
Quicken	6.67%
United Shore	4.18%
Wells Fargo	2.86%
JPM Chase	2.07%
Fairway Indept.	1.81%
Loan Depot	1.80%
Caliber Homes	1.68%
Bank of America	1.65%
Freedom Mort.	1.36%
US Bank	1.16%

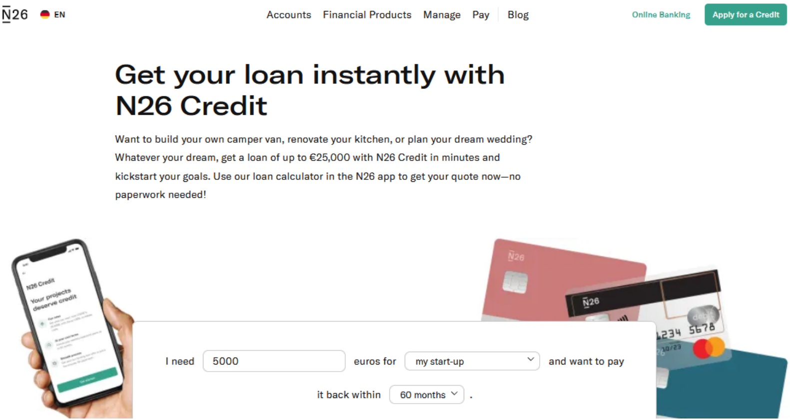
- Top 10 Mortgage Origination (Consumer Financial Protection Bureau) in 2019

The Rise of FinTech Lending

Name	Loans	Dollars	Average Loan Size	Mkt Share - Total Loans	Mkt Share - Total $
Rocket Mortgage, LLC	464,36	$127,577,235,000	$274,736	**5.60%**	5.00%
UNITED SHORE FINANCIAL SERVICES, LLC	348,42	$127,513,645,000	$365,982	**4.20%**	5.00%
Loandepot.Com, LLC	156,13	$52,531,740,000	$336,470	**1.90%**	2.10%
Wells Fargo Bank, National Association	142,26	$65,746,410,000	$462,163	**1.70%**	2.60%
Fairway Independent Mortgage Corporation	127,97	$40,808,695,000	$318,905	**1.50%**	1.60%

- Top 5 Mortgage Origination (Consumer Financial Protection Bureau) in 2022

The Rise of FinTech Lending



- In Europe?

The Role of Technology in Mortgage Lending

Fuster et al., 2019

	(1)	(2)	(3)	(4)	(5)
FinTech	-7.93***	-9.44***	-8.33***	-9.24***	-7.46***
	(0.52)	(0.61)	(0.43)	(0.48)	(0.45)
ln(loan amt)		4.47***		4.90***	6.10***
ln(income)		-0.56***		-1.00***	-0.45***
FHA		0.61***		0.23**	-0.40***
VA		1.67***		1.49***	1.87***
Jumbo		3.14***		5.28***	5.94***
Census tr. $\times$ Month FEs?	No	No	Yes	Yes	Yes
Loan controls?	No	Yes	No	Yes	Yes
R2	0.00	0.02	0.23	0.24	0.34
Observations	19159345	19159345	18551855	18551855	7185042
Sample	All	All	All	All	Nonbanks

The dependent variable is mortgage processing time: the time from loan application to closing. Other controls include indicators for gender and race of the borrower, and dummies for occupancy, presence of co-applicant, and pre-approval. Robust standard errors in parentheses (clustered by lender-month). ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively.

- FinTech lenders process applications 20% faster

The Role of Technology in Mortgage Lending

Fuster et al., 2019

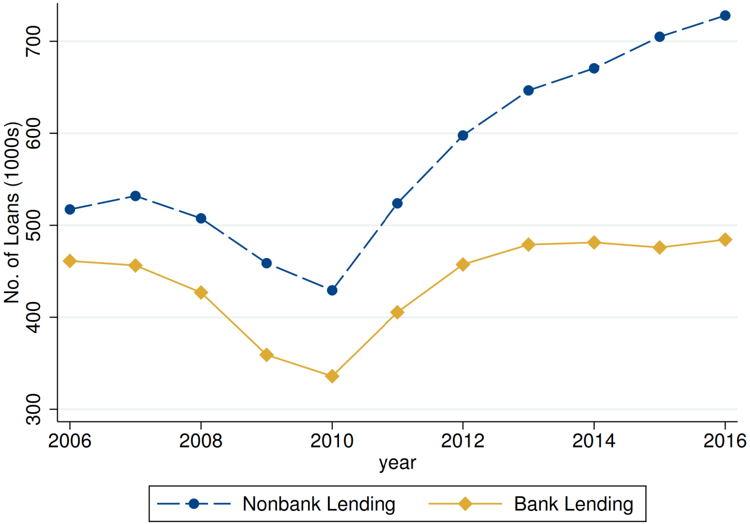
	(1)	(2)	(3)	(4)	(5)
FinTech	-1.29***	-0.97***	-0.93***	-1.51***	-0.79***
	(0.33)	(0.30)	(0.27)	(0.46)	(0.16)
Avg. P(default)	3.65	3.65	3.65	4.00	2.73
Loan Sample	All	All	All	Purch.	Refi
Purpose FE	No	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	No	No	No
MonthXState FE	No	No	Yes	Yes	Yes
Loan Controls	No	No	Yes	Yes	Yes
Observations	4097569	4097568	4097544	2966644	1130881

Standard errors clustered by issuer. Sample includes FHA 30-year FRMs originated 2013-2017.

- FinTech lenders process applications 20% faster
- With no increase in default rate (dependant variable)

Also in Small Business Lending!

Gopal and Schnabl, 2022



Also in Small Business Lending!

Gopal and Schnabl, 2022

$$\Delta Y_c = \beta \text{Bank Share}_{06,c} + \gamma X_c + FE_s + \varepsilon_c$$

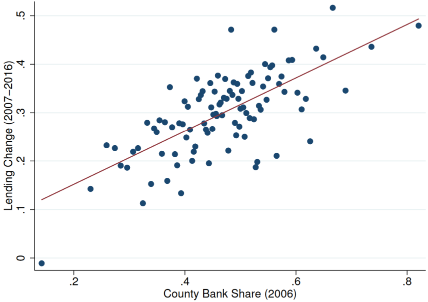
- ΔY_c = county lending growth or market share, 2007-2016
- Bank $\text{Share}_{06,c}$ = Market share of banks in county c in 2006
- X_c = county-level controls

Also in Small Business Lending!

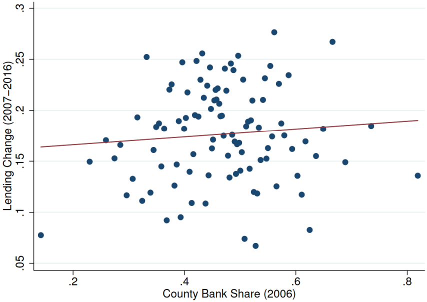
Gopal and Schnabl, 2022

$$\Delta Y_c = \beta \text{Bank Share}_{06,c} + \gamma X_c + FE_s + \varepsilon_c$$

Nonbank lending Growth (07-16)



Total lending growth (07-16)



What is the mechanism?

Machine Learning in FinTech Lending

Berg at al., 2020

- The growth of the internet leaves a trace of simple, easily accessible information about almost every user worldwide $\rightarrow$ **digital footprint**
- FinTech lenders use people's digital footprint to evaluate the creditworthiness of applicants who have no official credit score
- Banks' information advantage has eroded $\rightarrow$ really ?

Research questions:

i Does digital footprint helps augment information traditionally considered to be important for consumer default prediction?

ii Can it be used for the prediction of consumer payment behavior and defaults?

Machine Learning in FinTech Lending

Berg at al., 2020

- Unique dataset covering ≈ 250,000 purchases from an E-commerce company located in Germany
- Contains 10 digital footprint variables that are easily accessible for any firm operating in the digital sphere
  - Device type, the operating system, and the e-mail provider
  - Proxy for income, character, and reputation
- In addition to these digital footprint variables, the dataset also contains a credit score from a private credit bureau.

Machine Learning in FinTech Lending

Berg at al., 2020

- Difference in default rates between customers using iOS (Apple) and Android $\approx$ difference in default rates between a median and the 80th pct. of the credit score
- Bertrand and Kamenica (2017) show that owning an iOS device is :
  - among best predictors for being in the top quartile of the income distribution
  - also for character (status-seeking users are likely to buy an iOS device)
- Recent research documents that firms being named after their owners have a superior performance.
- So-called "eponymous effect" is mainly driven via a reputation channel (Belenzon, Chatterji, and Daley 2017).

Machine Learning in FinTech Lending

Berg at al., 2020

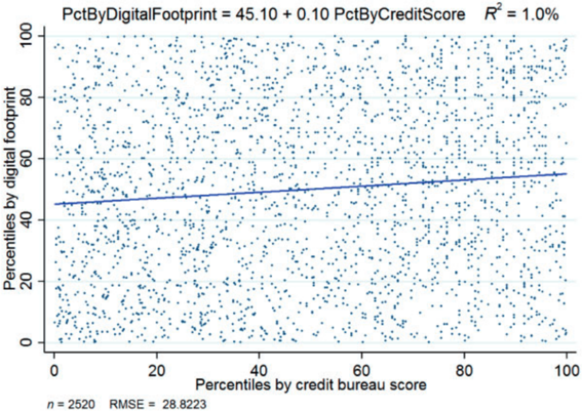


Figure 4
Correlation between digital footprint and credit bureau score (scorable customers)

– Digital footprints complement rather than substitute for credit bureau information

Machine Learning in FinTech Lending

Berg at al., 2020

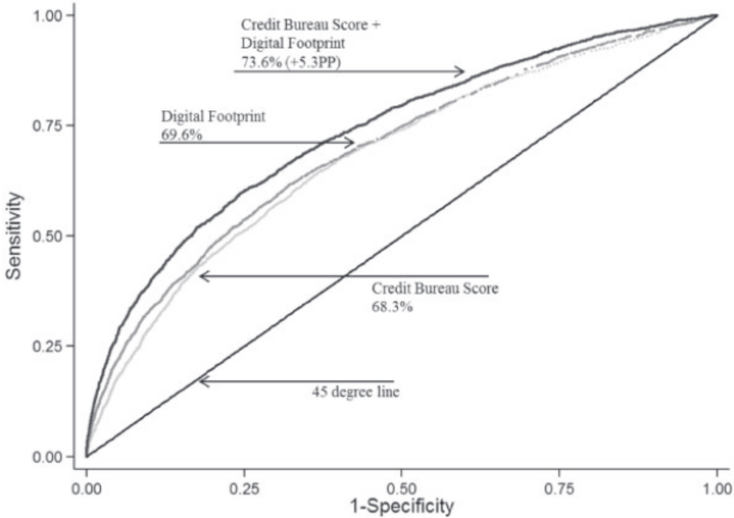
- Similar to Gambacorta et al., 2019 , use Logistic regression (Logit) to predict default rate

$$\mathbb{P}\left(\text{Default}_{i,t}\right) = \phi\left(\alpha\,\mathbf{X}_{i,t} + \beta\,\mathbf{Z}_{i,t} + \mu_r + \mu_t + \epsilon_{i,t}\right)$$

- **X** is (i) the credit score, (ii) digital footprint and (iii) both. **Z** are controls for Age, Gender, Item category, Loan. Month and region FE. amount

Machine Learning in FinTech Lending

Berg at al., 2020



Machine Learning in FinTech Lending

Berg at al., 2020

Table 5
Out-of-sample estimates

	(1) Baseline (in-sample)	(2) Out-of-sample	(3) Out-of-sample/out-of-time
AUC credit bureau score	0.683	0.681	0.691
N	254,819	254,819	74,543
AUC digital footprint	0.696	0.688	0.692
N	254,819	254,819	74,543
AUC credit bureau score + Digital footprint	0.736	0.728	0.739
N	254,819	254,819	74,543
AUC credit bureau score + Digital footprint, fixed effects	0.762	0.734	0.730
N	254,613	254,613	74,543

- Digital footprint alone is a **better predictor** of default than credit scores

Machine Learning in FinTech Lending

Berg at al., 2020

Table 8
Marginal AUC for digital footprint variables and combinations of digital footprint variables

A. Individual digital footprint variables (dependent variable: default (0/1))

Variable	Stand-alone AUC (%)	Marginal AUC (PP)
Computer & operating system	59.03	+1.71***
E-mail host	59.78	+2.44***
E-mail Host: paid versus nonpaid dummy	53.80	+0.98***
E-mail Host: Variation within nonpaid e-mail hosts	57.82	+1.79***
Channel	54.95	+0.70***
Checkout time	53.56	+0.63***
Do not track setting	50.40	+0.14*
Name in e-mail	54.61	+0.30**
Number in e-mail	54.15	+0.19**
Is lowercase	54.91	+1.15***
E-mail error	53.08	+1.78***

- OS and e-mail host work super well, as well as e-mail error and lowercase

Machine Learning in FinTech Lending

Berg at al., 2020

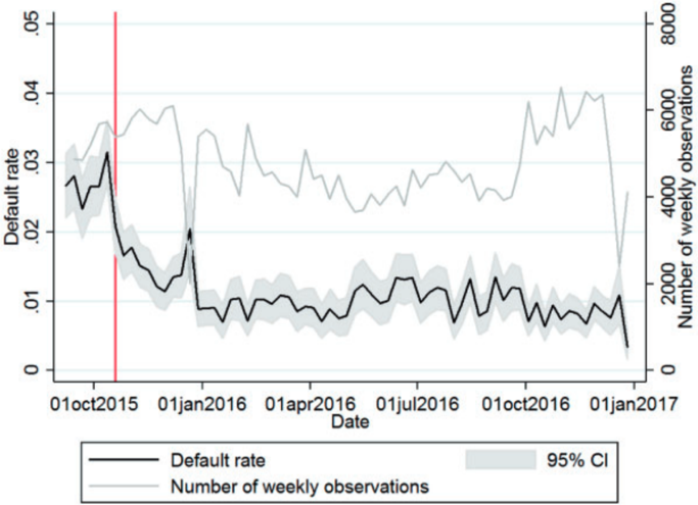


Figure 5
Default rates around the introduction of the digital footprint

Machine Learning in FinTech Lending

Berg at al., 2020

- Consumers with no credit score have comparable digital footprint-induced scores
- Digital footprints may help to overcome information asymmetries between lenders and borrowers when standard credit bureau information is not available.

Machine Learning in FinTech Lending

Berg at al., 2020

- Consumers with no credit score have comparable digital footprint-induced scores
- Digital footprints may help to overcome information asymmetries between lenders and borrowers when standard credit bureau information is not available.

→ Could extend access to credit for the unbanked. Remember that the lack of access to financial services affects around 2 billion working-age adults worldwide and is seen as one of the main drivers of inequality

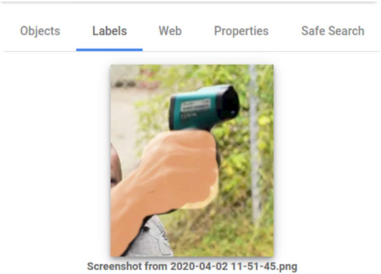
- Jagtiani et al (2019) find that US FinTech lenders tend to supply more mortgages to consumers with weaker credit scores than do banks ; they also have greater market shares in areas with lower credit scores and higher mortgage denial rates.

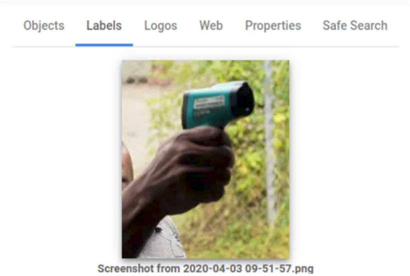
NB: results are subject to the **Lucas (1976) critique**

NB: results are subject to the **Lucas (1976) critique**, with customers potentially changing their online behavior if digital footprints are widely used

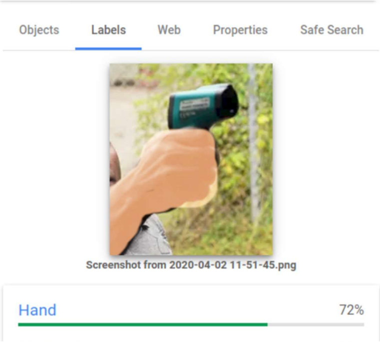
So, all good?

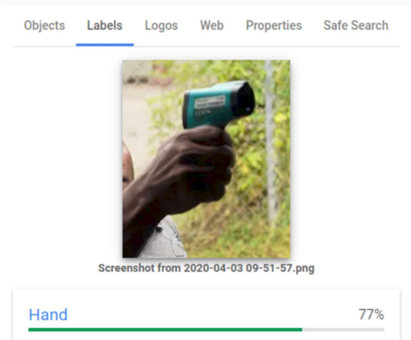
Artificial stupidity: Google Vision Cloud AI



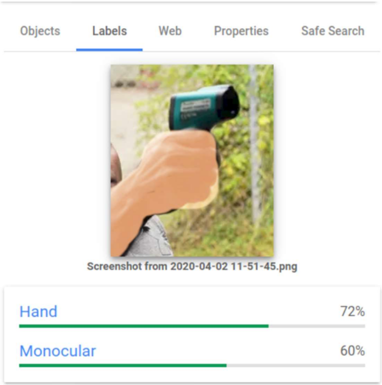


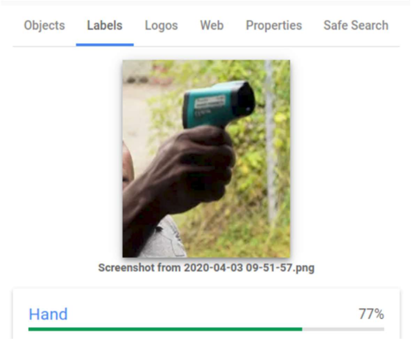
Artificial stupidity: Google Vision Cloud AI



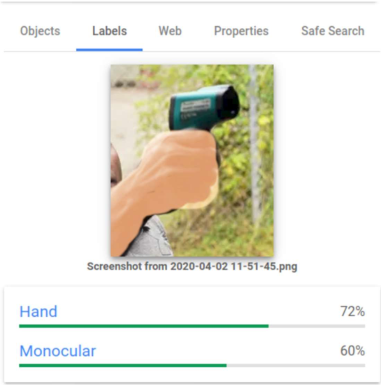


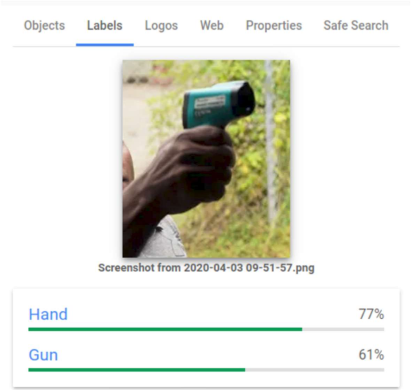
Artificial stupidity: Google Vision Cloud AI





Artificial stupidity: Google Vision Cloud AI






Fairness?