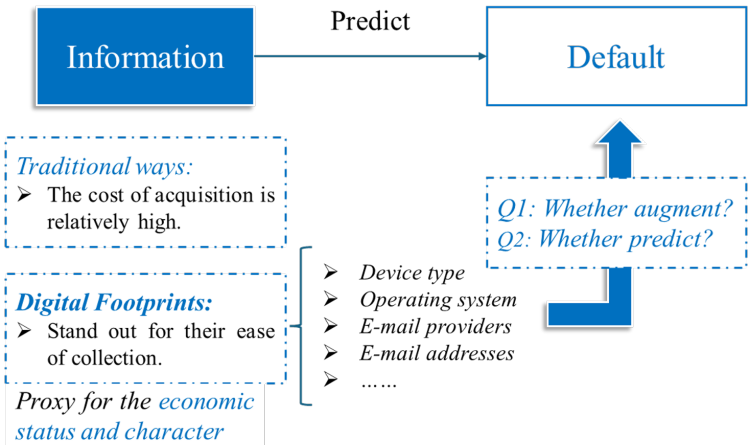
On the Rise of FinTechs: Credit Scoring Using Digital Footprints

The Review of Financial Studies
Tobias Berg et al.
Reporter: Yanrui Zhou

March 14, 2025

Outline

Framework


Information
Predict
Default
Traditional ways:
The cost of acquisition is relatively high.
Digital Footprints:
Stand out for their ease of collection.
Proxy for the economic status and character
Device type
Operating system
E-mail providers
E-mail addresses
……
Q1: Whether augment?
Q2: Whether predict?

Motivation

- Judging the creditworthiness(default rate) of customers is important.
  - Traditional methods of judging creditworthiness are based on specific information.
  - Digital footprints are easily accessible but are ignored in this area.
- This paper tries to combine digital footprints and creditworthiness judgment.

Research Questions

Q1: Whether the digital footprint helps augment information traditionally considered to be important for default prediction?

Q2: Can it be used for the prediction of consumer payment behavior and defaults?

Contributions

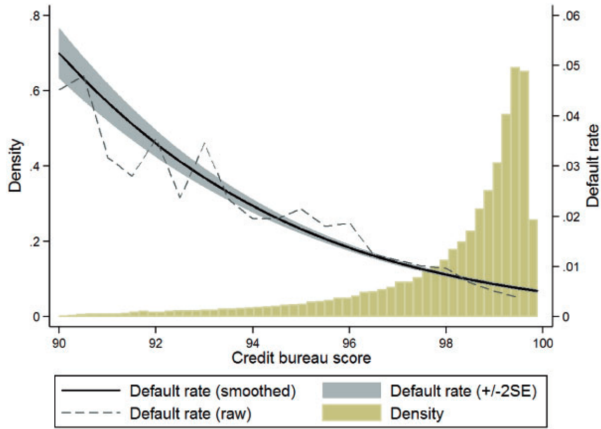
- The Literature on information for lending.
  - Previous Studies: focus on relationship-specific and nontraditional data(P2P, network).
  - Extension: use digital footprints which stand out for their ease of collection.
- The Literature on digital footprints.
  - Previous studies: digital footprints can reflect users' characters.
  - Extension: use footprints to judge default risk.

Data and Setup

- Access data about 270,399 purchases from an E-commerce company selling furniture in Germany between October 2015 and December 2016.
  1. Goods are shipped first and invoiced for later. The customer has 14 days to pay.
  2. If the customer does not pay on time, three reminders are sent out.
  3. A customer who does not pay after three reminders is in default.
- The company uses a digital footprint and information from private credit bureaus to decide whether customers have a sufficient creditworthiness.
  1. Credit bureau score draws on credit history from banks and payment behavior data.

Data and Setup

- Credit bureau score distribution and default rates:



Digital Footprint

- The digital footprint comprises easily accessible pieces of information known to proxy for the economic status of a person:
  1. Device type (desktop, tablet, mobile)
  2. Operating system (e.g., Windows, iOS, Android)
  3. E-mail providers in Germany (e.g., Gmx, Web, T-Online, Gmail, Yahoo, or Hotmail)
  4. Visit channel (paid ads clicks, direct visit, price comparison site + visit time)
  5. E-mail addresses

1. Univariate results

- Credit bureau score, digital footprint variables, and default rates.

Variable	Value	Observations	Proportion (%)	Default rate (%)	*t*-test against baseline
Credit bureau score (by quintile)	All	254,819	100	0.94	
	Q1 - lowest	50,980	20	2.12	Baseline
	Q2	50,949	20	1.02***	(−14.17)
	Q3	50,991	20	0.68***	(−19.51)
	Q4	51,181	20	0.47***	(−23.37)
	Q5 - highest	50,718	20	0.39***	(−24.89)
Device	All	254,819	100	0.94	
	Desktop	145,879	57	0.74	Baseline
	Tablet	45,575	18	0.91***	(3.62)
	Mobile	26,808	11	2.14***	(21.84)
	Do-not-track setting	36,557	14	0.88***	(2.90)
Operating system	All	254,819	100	0.94	
	Windows	124,605	49	0.74	Baseline
	iOS	41,478	16	1.07***	(6.35)
	Android	29,089	11	1.79***	(16.64)
	Macintosh	21,163	8	0.69	(−0.79)
	Other	1,927	1	1.09*	(1.74)
	Do-not-track setting	36,557	14	0.88***	(2.66)

2. Multivariate results: Digital footprint and default

- We use a *Logistic Regression* and report the AUC for every specification.
- X: Credit bureau score and Digital footprint variables
- y: Default dummy

AUC	*Bureau Score*	**0.683**	*Digital Footprint*	**0.696**	*Bureau Score* + *Digital Footprint*	**0.736**	*Bureau Score* + *Digital Footprint* + *Other Controls*	**0.762**
(SE)		**(0.006)**		**(0.006)**		**(0.005)**		**(0.005)**
Difference to AUC=50%		**0.183*****|	**0.196*****		**0.236*****		**0.262*****
Difference AUC to (1)				**0.013***		**0.053*****		**0.080*****

- Even simple, easily accessible variables from the digital footprint are as useful.

1. Economic mechanism

- Decompose the overall informational of the digital footprint into individual variables.
- The marginal AUC of variable X is defined as the AUC of the full model minus the AUC of the model using all variables except X.

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***

2. Default rates at the E-commerce firm

- Analyze the default rates around the introduction of the digital footprint.

Table 10
Development of default rates around the introduction of the digital footprint (multivariate results)

Dependent variable Method	(1) Default (0/1) Difference post vs. pre	(2) Default (0/1) Difference post vs. pre, add categories	(3) Default (0/1) Add time trend, controls and FEs	(4) Default (0/1) Add subcategories
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks
Post	−0.014*** (−9.12)			
Post x ScoreAndDFAdded		−0.014*** (−8.55)	−0.014*** (−5.88)	−0.015*** (−6.13)
Post x DFAdded		−0.013*** (−3.85)	−0.012*** (−3.04)	
Post x “DFAdded / High score”				−0.001 (−0.19)
Post x “DFAdded / Medium score”				0.003 (0.65)
Post x “DFAdded / Low score”				−0.026** (−2.51)
Post x “DFAdded / Unscorable”				−0.052*** (−2.72)

2. Default rates at the E-commerce firm

- Analyze the access to credit around the introduction of the digital footprint.

Table 11
Development of access to credit around the introduction of the digital footprint (multivariate results)

	(1)	(2)	(3)	(4)
Dependent variable	Invoice offered (0/1)	Invoice offered (0/1)	Invoice offered (0/1)	Invoice offered (0/1)
Method	Difference post vs. pre	Difference post vs. pre, add categories	Add time trend, controls and FEs	Add subcategories
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks
Post	0.002			
	(0.50)			
Post x ScoreAndDFAdded		−0.066***	−0.054***	−0.046***
		(−25.18)	(−10.31)	(−10.48)
Post x DFAdded		0.011*	0.017**	
		(1.85)	(2.63)	
Post x "DFAdded / High score"				0.030***
				(3.78)
Post x "DFAdded / Medium score"				0.044***
				(4.37)
Post x "DFAdded / Low score"				−0.022**
				(−2.58)
Post x "DFAdded / Unscorable"				0.009
				(1.57)

Idea

- Change the case (lending, time period, platform ,country)
- Lucas critique applies:
  - whether people will change their behavior after the introduction of digital footprint.
- Study other behaviors or footprints of the account.