On the Rise of FinTechs –
Credit Scoring using Digital Footprints

Tobias Berg, Frankfurt School of Finance & Management
Valentin Burg, Humboldt University Berlin
Ana Gombović, Frankfurt School of Finance & Management
Manju Puri, Duke University, FDIC and NBER

September 2019

Motivation

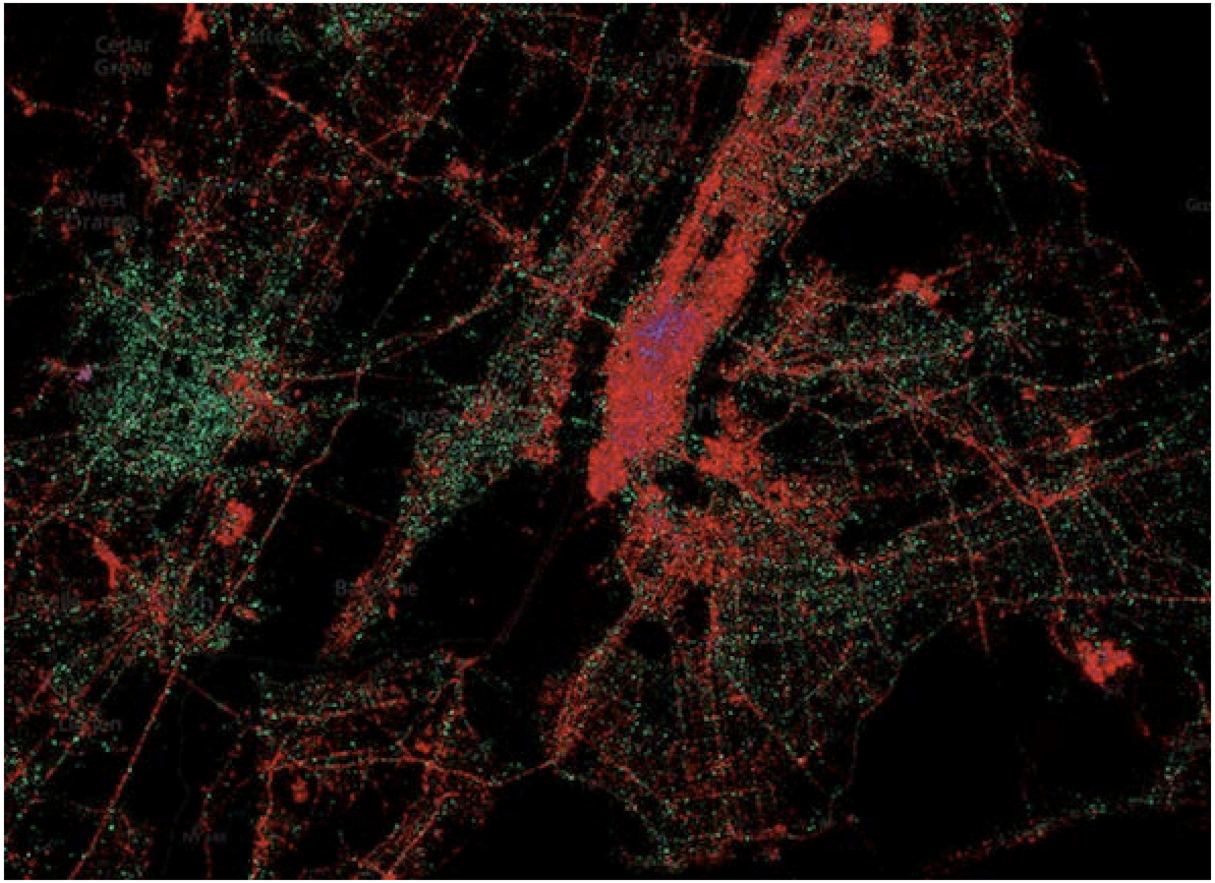
- Digital footprint: Trace of simple, easily accessible information about almost every individual worlwide

- One key reason for existence of financial intermediaries: Superior ability to access and process information for screening borrowers

- This paper: Informativeness of digital footprint for credit scoring

- Wide implications
  - Financial intermediaries‘ business models
  - Access to credit for unbanked
  - Behavior of consumers, firms, and regulators in the digital sphere

Motivation: New York – Use of operating systems



Red = iOS, Green = Android, Purple = Blackberry

Information about customers' operating system available to every website without any effort

Source: Gnip, MapBox, Eric Fischer, Data 2011-2013

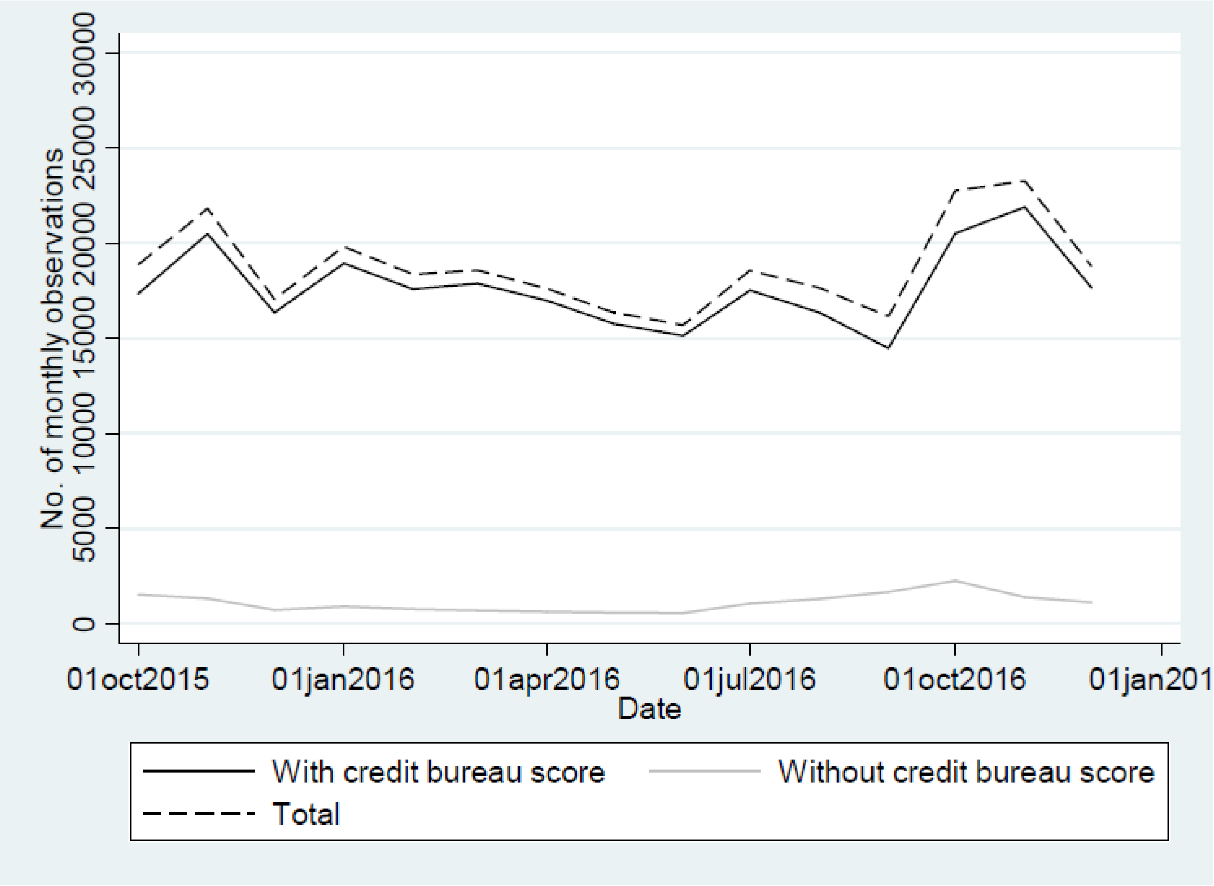
Dataset: Overview

- Sample:
  - 270,399 purchases from E-commerce company in Germany (similar to Wayfair)
  - Goods shipped first and paid later (~short term consumer loan)
  - Period: Oct2015 – Dec2016
  - Mean purchase volume: EUR 320 (~USD 350)
  - Mean age: 45 years
  - Geographical distribution similar to German population
  - Contains credit bureau score(s)

- Default rate: 0.9% (~3% annualized)
  - Default rate on all German consumer loans in 2016: 2.4%

- Data set limited to purchases > €100 and predicted default rate < 10%.
  - Benefit: more comparable to typical credit card, bank loan or P2P data set
  - For comparison: Lending club with minimum loan amount of USD 1,000 and minimum FICO of 640 (~15% default rate)

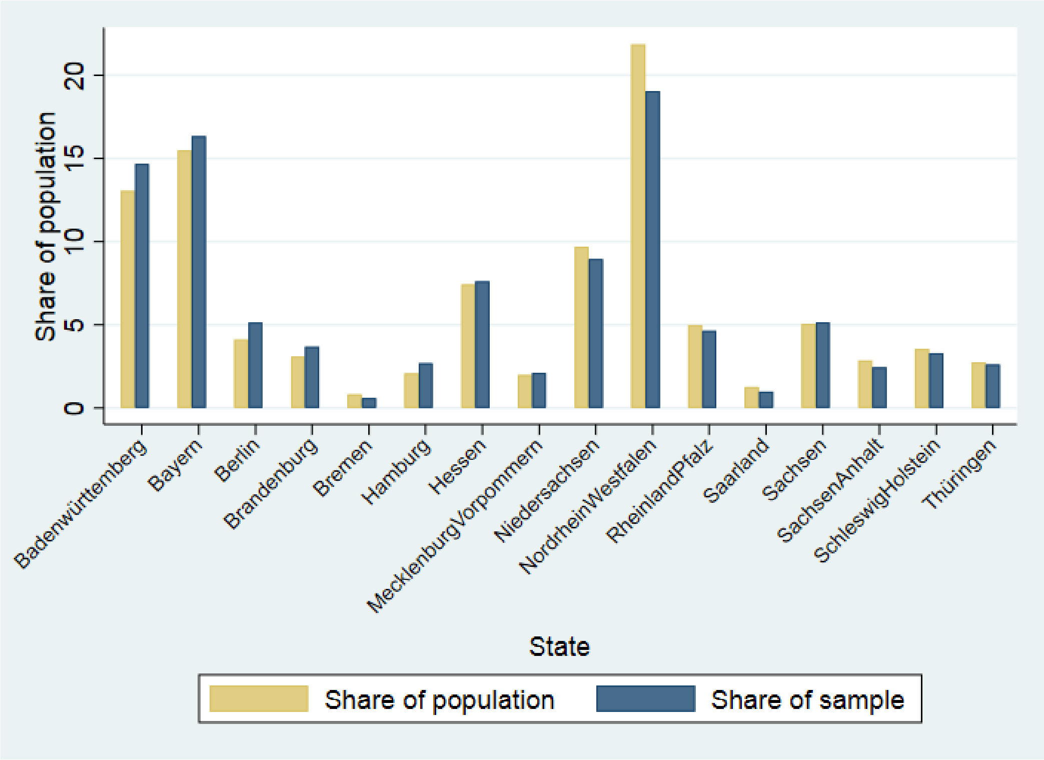
Distribution of observations over time



Roughly even distribution over time –
with slight increases in dark season (October/November)

Geographic distribution across states

This figure illustrates the share of customers by states in our sample compared to the German population by states.



Is dataset comparable to other loan data sets?

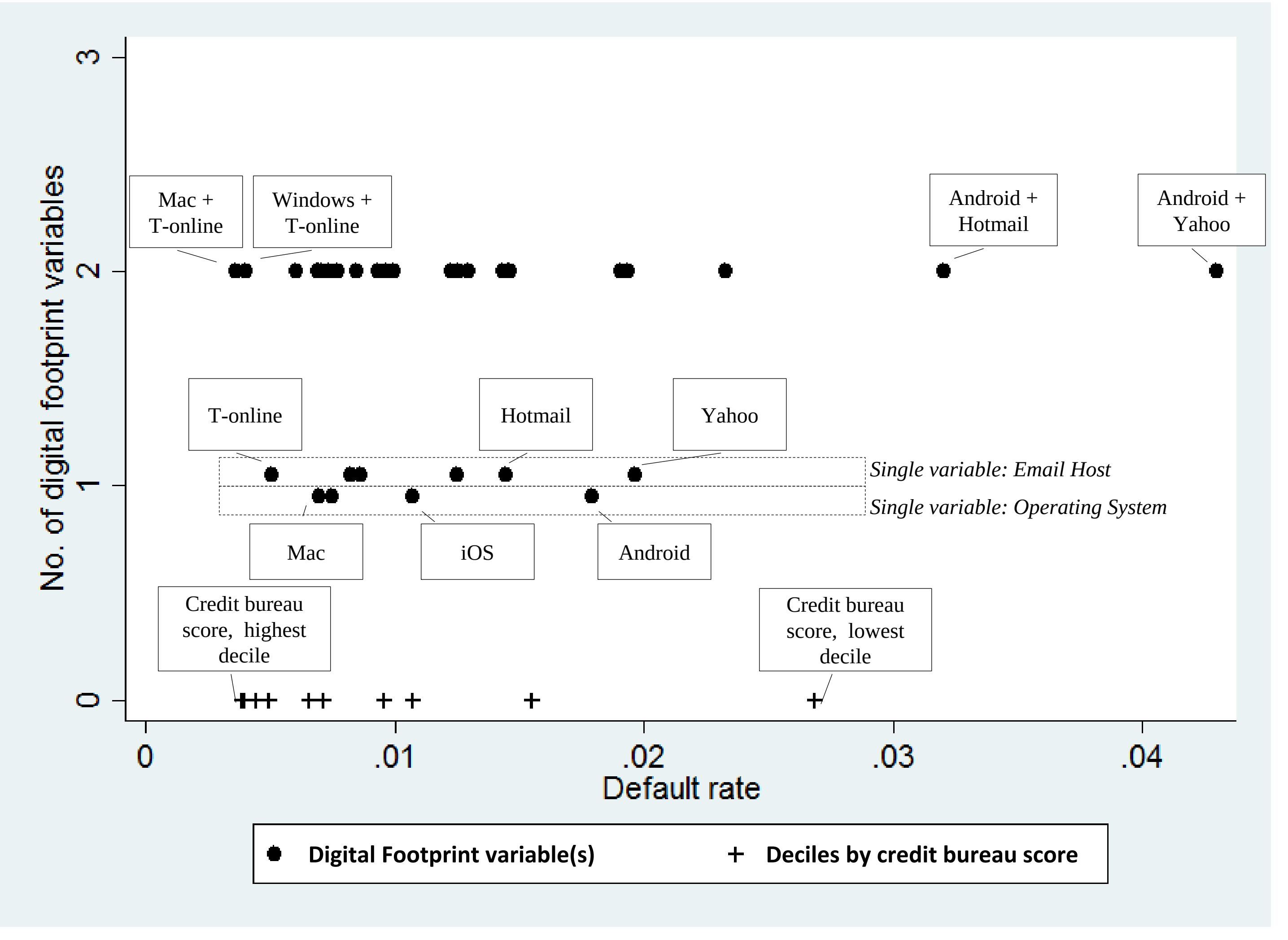
Study	Sample	Default rate	Time horizon	Default rate (annualized)
This study				
This study	270,399 purchases at a German E-Commerce company between October 2015 and December 2016	1.0%	~4 months	3.0%
Germany				
Berg, Puri, and Rocholl (2017)	100,000 consumer loans at a large German private bank, 2008-2010	2.5%	12 months	2.5%
Puri, Rocholl, and Steffen (2017)	1 million consumer loans at 296 German savings banks, 2004-2008	1.1%	12 months	1.1%
Schufa (2017) – study by the major credit bureau in Germany	17.4 million consumer loans covered by the main credit bureau in Germany in 2016	2.2%	12 months	2.2%
Schufa (2016)– study by the major credit bureau in Germany	17.3 million consumer loans covered by the main credit bureau in Germany in 2015	2.4%	12 months	2.4%
Deutsche Bank (2016)	All retail loans of Deutsche Bank (i.e., the largest German bank)	1.5% (Basel II PD estimate)	12 months	1.5%
Commerzbank (2016)	All retail loans of Commerzbank (i.e., the second largest German bank)	2.0% (Basel II PD estimate)	12 months	2.0%
United States				
Federal reserve	Charge-off rate on consumer loans, Q4/2016	2.09%	12 months (annualized quarterly data)	2.09%
Federal reserve	Charge-off rate on consumer loans, Q4/2015	1.76%	12 months (annualized quarterly data)	1.76%
Hertzberg, Liberman, and Paravisini (2016)	12,091 36-months loans from Lending Club issued between December 2012 and February 2013	9.2%	~26 months	4.2%
Lending Club (own analysis)	375,803 36-month loans from Lending Club issued between October 2015 and December 2016	5.11%	12 months	5.11%
Iyer, Khwaja, Luttmer, and Shue (2016)	17,212 36-months loans from Prosper.com issued between February 2007 and October 2008	30.6%	36 months	10.2%
Puri, Hildebrandt, and Rocholl (2017)	12,183 loans from Prosper.com between February 2007- April 2008	10.8%-18.6%	per 1,000 days	3.9%-6.8%

- Similar default rates compared to other German lending data sets
- Similar default rates compared to U.S. lending data sets
- Exception: P2P-lending studies using data from 2007/2008 with significantly higher default rates
- Data is also representative in terms of the age structure and geographic distribution in Germany

Digital footprint – 10 easily accessible variables

Variable	Description	Information content
Device Type	Main examples: Desktop, Tablet, Mobile.	**Income**
Operating System	Main examples: Windows, iOS, Android.	e.g. Bertrand and Kamenica (2018): iOS best predictor for being in Top-Quartile by income
Email Provider	Main examples: Gmail, Yahoo, T-Online.	
Channel	Channel through which customer has arrived at homepage of the firm. Main examples: paid click vs organic search; affiliate such as price comparison site; direct entering of URL	**Character** e.g. Rook (1987) and Wells et al. (2011): personality traits and impulse shopping
Check-Out Time	Time of day of purchase (morning, afternoon, evening, night)	
Do not track setting	Customer does not allow tracking of device and operating system information, and channel.	
Email Error	Email address contains an error in the first trial (Note: Clients can only order if they register with a correct email address).	
Name in Email	First or last name of customer is part of email address.	**Reputation**
Number in Email	Email address contains number.	e.g. Belenzon, Chatterji, and Daley (2017) and Stern and Guzman (2016): Eponymous Entrepreneurs Effect
Is Lower Case	First name, last name, street, or city are written in lower case.	

Bivariate results



Measure of association: Cramer's V

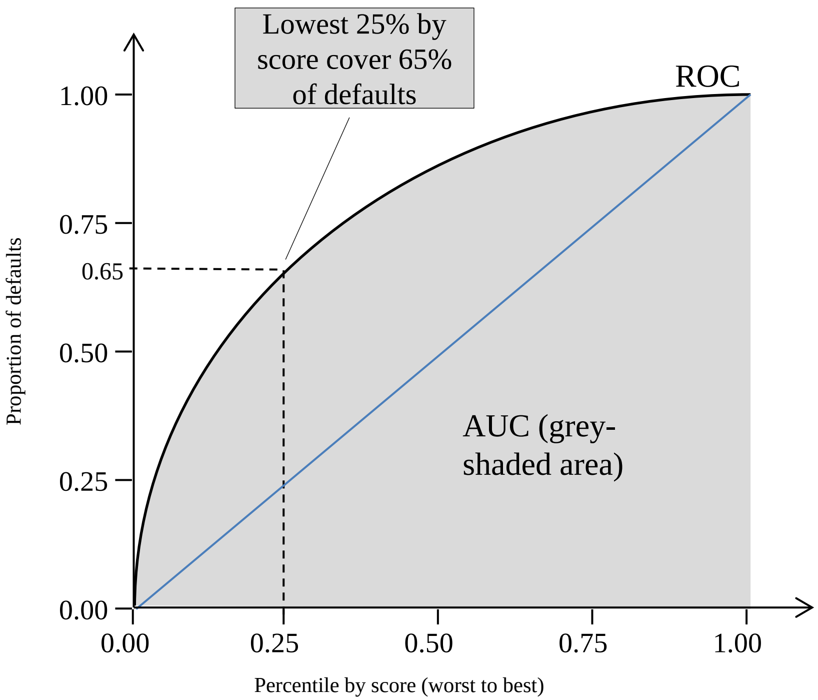
	Credit bureau score	Device Type	Operating System	Email Host	Channel	Check-Out Time	Name in Email	Number in Email	Is Lower Case	Email Error	Age	Order amount	Item category	Month
Main variables														
Credit bureau score [a]	1.00***	0.07***	0.05***	0.07***	0.03***	0.03***	0.01***	0.07***	0.02***	0.01	0.20***	0.01***	0.05***	0.01***
Device Type		1.00***	0.71*** [b]	0.07***	0.06*** [b]	0.04***	0.05***	0.06***	0.07***	0.01***	0.12***	0.03***	0.05***	0.06***
Operating System			1.00***	0.08***	0.06*** [b]	0.04***	0.06***	0.08***	0.06***	0.01***	0.10***	0.02***	0.04***	0.03***
Email Host				1.00***	0.03***	0.03***	0.08***	0.18***	0.04***	0.06***	0.16***	0.02***	0.02***	0.01***
Channel					1.00***	0.02***	0.01***	0.02***	0.04***	0.02***	0.09***	0.04***	0.06***	0.13***
Check-Out Time [a]						1.00***	0.01***	0.01***	0.01***	0.01*	0.06***	0.01***	0.03***	0.02***
Name in Email							1.00***	0.22***	0.01***	0.02***	0.04***	0.01	0.03***	0.01
Number in Email								1.00***	0.02***	0.00**	0.06***	0.01***	0.04***	0.01***
Is Lower Case									1.00***	0.03***	0.03***	0.02***	0.02***	0.02***
Email Error										1.00***	0.03***	0.01**	0.01***	0.01*
Control variables														
Age [a]											1.00***	0.05***	0.11***	0.03***
Order amount [a]												1.00***	0.27***	0.02***
Item category													1.00***	0.11***
Month														1.00***

[a] Transformed into quintiles.
[b] We exclude customers with a do-not-track setting, as the setting simultaneously applies to device, operating system, and channel information.

- Digital footprint variables not highly correlated with credit bureau score
- Correlations between other digital footprint variables in general low
- Device Type / Operating System highly correlated (for example: most desktops run on Windows) → we use most frequent combinations in multivariate regressions below

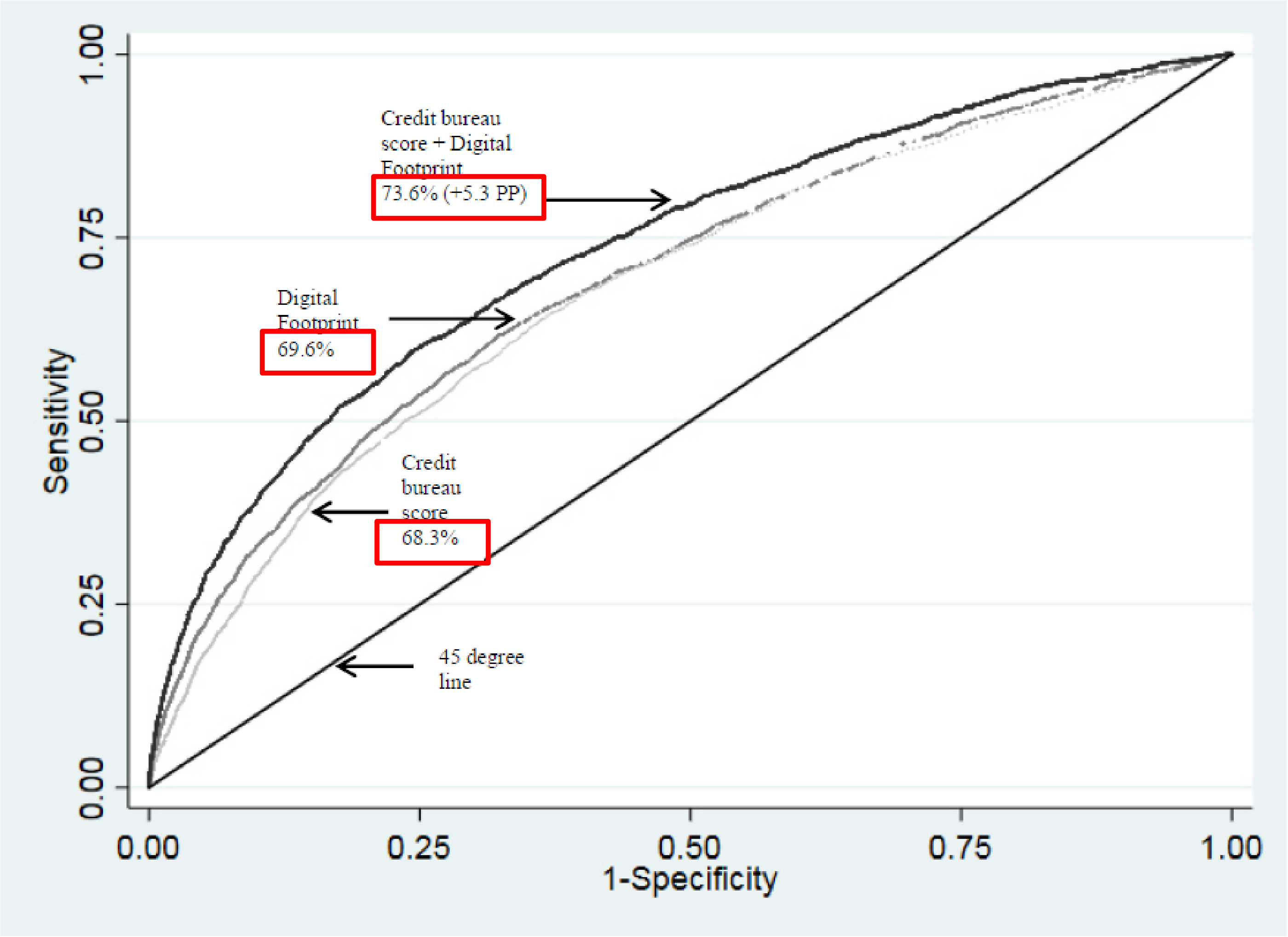
Judging discriminatory power: AUC

- Method: logistic regression with default dummy as the dependent variable
- Formal analysis of discriminatory power: Receiver Operating Characteristics (ROC) and Area-under-the-Curve (AUC)



- Range: 50% (random prediction) to ~ 100% (perfect prediction)
- Closely related to GINI: GINI = 2·AUC – 1
- Interpretation: Probability of correctly identifying good case if faced with random (good, bad)-pair
- Iyer, Khwaja, Luttmer, Shue (2016): 60% desirable in information-scarce environments, 70% in information-rich environments
- See also Vallee and Zeng (2018) and Fuster, Plosser, Schnabl, and Vickery (2018)

Area-under-Curve: Credit bureau score versus digital footprint



Area-under-Curve: Comparison to other studies

Study	Sample	AUC using credit score
Area Under the Curve (AUC) using credit bureau scores only		
This study	270,399 purchases at a German E-Commerce company in 2015/2016	68.3%
Berg, Puri, and Rocholl (2017)[#]	100,000 consumer loans at a large German private bank, 2008-2010	66.6%
Puri, Rocholl, and Steffen (2017)[#]	1 million consumer loans at 296 German savings banks, 2004-2008	66.5%
Iyer, Khwaja, Luttmer, and Shue (2016)	17,212 36-months loans from Prosper.com issued between February 2007 and October 2008	62.5%
Lending Club (own analysis)	375,803 36-month loans from Lending Club issued between October 2015 and December 2016[1]	59.8%
AUC and changes in the Area Under the Curve using other variables in addition to the credit score		
		AUC Change
This study	Digital footprint versus credit bureau score only	+ 5.3PP
Berg, Puri, and Rocholl (2017)[#]	Bank internal rating (which includes credit bureau score) versus credit bureau score only	+8.8PP
Puri, Rocholl, and Steffen (2017)[#]	Bank internal rating (which includes credit bureau score) versus credit bureau score only	+11.9PP
Iyer, Khwaja, Luttmer, and Shue (2016)	Interest rates versus credit score only	+5.7PP
Iyer, Khwaja, Luttmer, and Shue (2016)	All available financial and coded information (including credit score) versus credit score only	+8.9PP
Lending Club (own analysis)	Lending Club loan grade (which includes credit score) versus credit score only	+11.9PP

Multivariate regression (logistic)

	(1)		(2)		(3)		(4)	
	Credit bureau score		Digital footprint		Credit bureau score & digital footprint		Credit bureau score & digital footprint, further controls	
VARIABLES	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
Credit bureau score	-0.17***	(-7.89)			-0.15***	(-6.67)	-0.14***	(-5.91)
Computer & Operating system								
Desktop/Windows			Baseline		Baseline		Baseline	
Desktop/Macintosh			-0.07	(-0.53)	-0.13	(-1.03)	-0.19	(-1.51)
Tablet/Android			0.29***	(3.19)	0.29***	(3.06)	0.33***	(3.45)
Tablet/iOS			0.08	(1.05)	0.08	(0.97)	0.07	(0.91)
Mobile/Android			1.05***	(17.26)	0.95***	(15.34)	1.01***	(16.18)
Mobile/iOS			0.72***	(9.07)	0.57***	(6.73)	0.61***	(7.26)
Email Host [a]								
Gmx (partly paid)			Baseline		Baseline		Baseline	
Web (partly paid)			-0.00	(-0.01)	-0.02	(-0.22)	-0.01	(-0.11)
T-Online (affluent customers)			-0.40***	(-3.89)	-0.35***	(-3.34)	-0.27**	(-2.47)
Gmail (free)			0.34***	(3.79)	0.28***	(3.08)	0.27***	(2.82)
Yahoo (free, older service)			0.75***	(9.19)	0.72***	(8.98)	0.69***	(8.26)
Hotmail (free, older service)			0.35***	(3.70)	0.28***	(2.73)	0.25**	(2.38)
Channel								
Paid			Baseline		Baseline		Baseline	
Affiliate			-0.49***	(-5.33)	-0.54***	(-5.56)	-0.61***	(-6.32)
Direct			-0.27***	(-4.24)	-0.28***	(-4.43)	-0.26***	(-4.29)
Organic			-0.15*	(-1.79)	-0.15*	(-1.73)	-0.15*	(-1.82)
Check-Out Time								
Evening (6pm-midnight)			Baseline		Baseline		Baseline	
Morning (6am-noon)			0.28***	(4.52)	0.28***	(4.62)	0.29***	(4.74)
Afternoon (noon-6pm)			0.08	(1.42)	0.08	(1.47)	0.10*	(1.87)
Night (midnight-6am)			0.80***	(7.74)	0.75***	(7.11)	0.73***	(6.77)
Do-not-track setting			-0.02	(-0.25)	-0.07	(-0.90)	-0.09	(-1.22)
Name In Email			-0.28***	(-5.67)	-0.29***	(-5.69)	-0.29***	(-5.61)
Number In Email			0.26***	(4.50)	0.23***	(3.92)	0.23***	(3.86)
Is Lower Case			0.76***	(13.04)	0.74***	(13.16)	0.75***	(13.24)
Email Error			1.66***	(20.01)	1.67***	(20.37)	1.70***	(20.34)
Constant	12.43***	(5.77)	-4.92***	(-62.84)	9.97***	(4.49)	9.09***	(4.07)
Control for age, gender, item category, loan amount, month and region fixed effects	No		No		No		Yes	
Observations	254,808		254,808		254,808		254,592	
Pseudo R^2	**0.0244**		**0.0525**		**0.0718**		**0.0924**	
AUC	**0.683**		**0.696**		**0.736**		**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*****		**0.079*****	

(1) Credit bureau score with clear discriminatory ability

(2) All components of digital footprint exhibit discriminatory ability. Economic effects are significant. Example: Mobile/Android with exp(1.05)=2.86 times higher odds ratio of defaulting than Desktop/Windows.

(3) Coefficient estimates barely change. Suggests that digital footprint complements rather than substitutes for credit bureau score.

(4) Digital footprint not a simple proxy for region, date, or age

Contribution of individual variables to AUC

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

Variable	Standalone AUC	Marginal AUC
Computer & Operating system	59.03%	+1.71PP***
Email Host	59.78%	+2.44PP***
Email Host: paid versus non-paid dummy	53.80%	+0.98PP***
Email Host: Variation within non-paid email hosts	57.82%	+1.79PP***
Channel	54.95%	+0.70PP***
Check-Out Time	53.56%	+0.63PP***
Do not track setting	50.40%	+0.00PP
Name In Email	54.61%	+0.30PP**
Number In Email	54.15%	+0.19PP**
Is Lower Case	54.91%	+1.15PP***
Email Error	53.08%	+1.79PP***

Panel B: Combinations of digital footprint variables (dependent variable: default (0/1))

Variables	Standalone AUC	Marginal AUC
Potential proxy for income		
Potential proxy for income, financially costly to change (Computer & Operating system, Email host: paid vs. non-paid dummy)	61.03%	+2.20PP
Unlikely to be a proxy for income, not financially costly to change (Non-paid email host, Channel, Check-out time, Do not track setting, Name in Email, Number in Email, Is Lower Case, Email Error)	67.35%	+8.52PP
Impact on everyday behavior		
Requires one-time action only (Computer & Operating system, Email host, Do not track setting, Name in Email, Number in Email)	64.92%	+7.25PP
Requires thinking about how to behave during every individual purchase (Channel, Check-out time, Is Lower Case, Email Error)	62.30%	+4.63PP

- No single variable dominates
- All variables apart from "do not track" with significant marginal AUCs

- Non-income proxies more important than (potential) income proxies

- Mix between one-time actions and actions during current purchase process

External validity: Idea

- Evidence so far: Predictive power of digital footprint for short-term loans for products purchased online

- Now: Test whether digital footprint with predictive power for traditional loan products as well.

- Unfortunately, no data on other loans available. Idea: Does the digital footprint predict future changes in the credit bureau score?

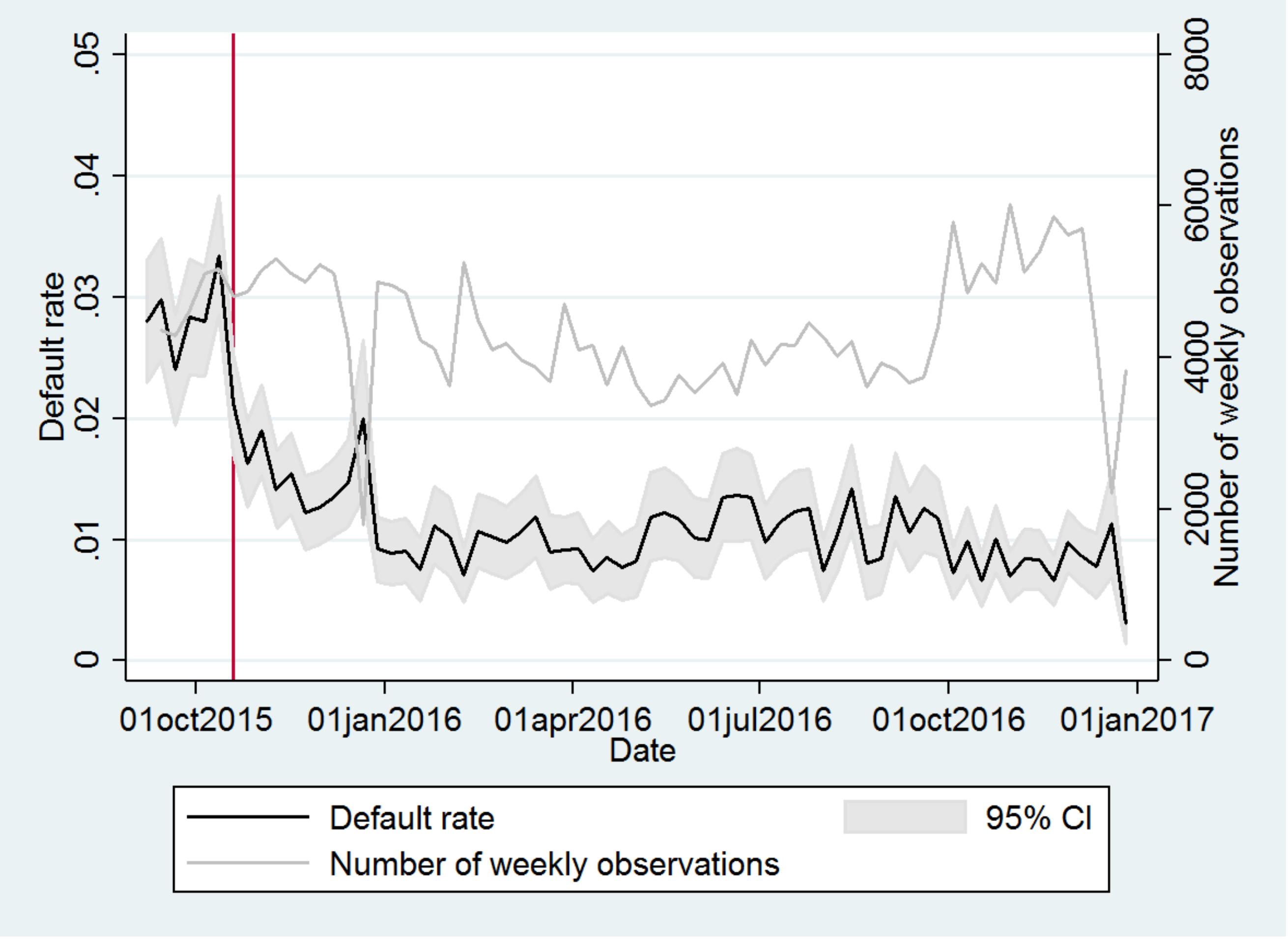
$$\Delta(CreditScore_{t+1}, CreditScore_t) = \beta_0 + \beta_1 \Delta(DF_t, CreditScore_t) + X + \varepsilon \qquad (1)$$

External validity: Digital footprint predicts future changes in credit bureau scores

Dependent variable	(1) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(2) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(3) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(4) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(5) Δ (CreditScore$_{t+1}$, CreditScore$_t$)
Δ (DigitalFootprint$_t$, CreditScore$_t$)	-74.56*** (-11.71)	-28.14*** (-4.56)	-29.74*** (-4.95)		-34.24*** (-4.23)
Q1 (-100% to -0.49%)				0.39** (2.38)	
Q2 (-0.49% to -0.25%)				0.15* (1.74)	
Q3 (-0.25% to -0.05%)				baseline	
Q4 (-0.05% to +0.35%)				0.08 (0.92)	
Q5 (+0.35% to +100%)				-0.39*** (-3.05)	
DigitalFootprint-Better-Than-CreditScore (0/1)					0.36** (2.45)
DigitalFootprint-Better-Than-CreditScore (0/1) x LowCreditScore					0.68** (2.05)
Q2					-0.02 (-0.11)
Q3					omitted
Q4					-0.19 (-1.06)
HighCreditScore					-0.02 (-0.06)
CreditScore$_t$		-0.43*** (-13.81)	-0.42*** (-13.66)	-0.42*** (-10.27)	FE for each credit score quintile
Constant	0.37*** (8.64)	42.31*** (13.84)	absorbed	absorbed	absorbed
Month & region fixed effects	No	No	Yes	Yes	Yes
Observations	17,645	17,645	17,645	17,645	17,645
Adj. R^2	2.74%	6.95%	7.95%	7.92%	7.13%

- Good digital footprint predicts improvement in credit bureau score (even after controlling for mean reversion)
- Good digital footprint predicts improvement in credit bureau score in particular for lower credit bureau scores

Economic impact of using a better scoring model



October 19, 2015 = Introduction of digital footprint and extension of bureau score

Digital footprint helps most for low-score and unscorable custom.

Linear model

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Default (0/1)	Default (0/1)	Default (0/1)	Default (0/1)	Default (0/1)	Default (0/1)
Method	Difference Post vs. Pre	Difference Post vs. Pre, add categories	add time trend, controls and FEs	add subcategories	Narrower window around Oct19-2015	Placebo test, 1-year later
Sample	+/- 6 weeks	+/- 6 weeks	+/- 6 weeks	+/- 6 weeks	+/- 4 weeks	+/- 4 weeks
Post	-0.014***					
	(-9.12)					
Post x ScoreAndDFAdded		-0.014***	-0.014***	-0.015***	-0.015***	0.001
		(-8.55)	(-5.88)	(-6.14)	(-4.30)	(0.30)
Post x DFAdded		-0.013***	-0.012***			
		(-3.85)	(-3.04)			
Post x "DFAdded / High score"				-0.001	0.000	0.002
				(-0.19)	(0.00)	(0.78)
Post x "DFAdded / Medium score"				0.003	0.003	0.004
				(0.65)	(0.47)	(1.06)
Post x "DFAdded / Low score"				-0.026**	-0.021*	-0.014
				(-2.50)	(-1.71)	(-1.48)
Post x "DFAdded / Unscorable"				-0.052***	-0.059***	0.007
				(-2.72)	(-2.66)	(0.43)
Time trend	No	No	0.000	0.001	0.001	-0.002
			(0.29)	(0.63)	(0.15)	(-0.81)
Category FE (=variables from interaction terms as non-interacted variables)	No	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	44,703	44,703	44,703	44,703	30,322	28,905
Adj. R^2	0.002	0.003	0.012	0.021	0.020	0.012

Implication 1: Information advantage of financial intermediaries

- One key reason for the existence of financial intermediaries: Superior ability to access and process information relevant for screening and monitoring of borrowers

- This paper: Digital footprint with valuable information for predicting defaults.
  - Likely proxy for some of the current relationship-specific information that banks have
  - Reduces gap between FinTechs and traditional financial intermediaries

- Implication: Informational advantage of banks threatened by digital footprint

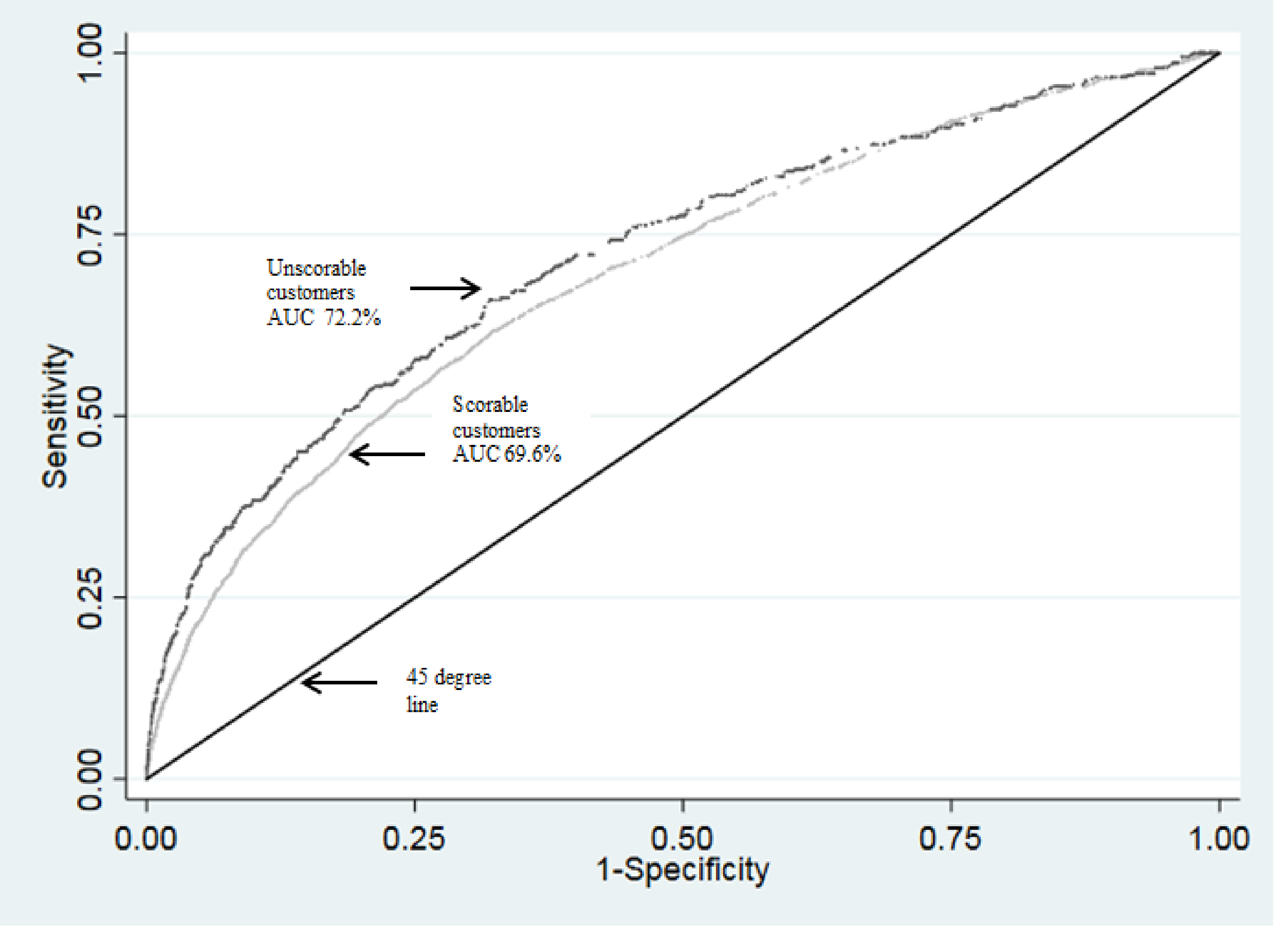
Implication 2: Access to credit for unbanked

- Two billion working-age adults lack access to financial services

- High expectations in digital footprints
  - World Bank: “Can digital footprints lead to Greater Financial Inclusion?”
  - Harvard Business Review: Fintech Companies Could Give Billions of People More Banking Options
  - Prior evidence on availability of credit and credit scores (Japelli and Pagano, 1993; Brown, Japelli, and Pagano, 2009; Djankov, McLiesh, and Shleifer, 2009; Beck, Demirguc-Kunt, and Honohan, 2009)

- Our paper: Digital footprint help to alleviate credit constraints for unscorables
  - ~6% of our sample: no credit bureau score (but: existence of customer confirmed and customer not in private bankruptcy)
  - Discriminatory power for unscorable customers is similar
  - Digital footprint helps to access credit for this sample
  - Subject to external validity concerns

Unscorable vs. scorable customers: AUC comparison



Implication 3: Behavior of consumers, firms, and regulators in digital sphere

- Lucas critique: Change in consumers behavior if digital footprint is used by intermediaries
  - Some variables costly to manipulate
  - Others require change in consumer habits

- If Lucas critique applies
  - Risk of costly signaling equilibrium (Spence 1973): expensive suit vs. expensive phone
  - If people change their behavior as a response to digital footprints being used, then people change their behavior (=impact on everyday life)

- Beyond consumer behavior
  - Firms: Response by firms associated with low-creditworthiness products
  - Statistical discrimination / fair lending acts: Proxy for prohibited variables such as race or gender → likely to be more important than for other alternative data sources
  - Lobbying: Incumbant banks might lobby regulators to intervene

Robustness tests: Overview

Out-of-sample tests
- Nx2-fold cross validation, N=100
- Results are not driven by over-fitting in-sample

Default definition
- Similar results if we focus on ultimate payment behavior (after effort by collection agency)
- Digital footprint predicts loss given default better than credit bureau score
- Digital footprint predicts both fraud (~10% of defaults) and non-fraud defaults

Sample splits
- Similar performance for large versus small orders
- Similar performance for male versus female customers
- Coefficient stability over time

Further tests
- Clustering on various dimensions (2-digit zip code, 3-digit zip codes, age, week)
- Control for type of purchased item

Difference analysis
- Pre-event trend: No trend
- Placebo test for all 52 weeks outside of event window: event window with largest effect and largest t-stat
- Default rate development consumer loans in Germany: no trend during our sample period
- Histogram of order amounts: No manipulation of order amounts
- Access to credit instead of default rate as dependent variable: Access to credit increases slightly when DF added

Robustness tests: 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	0.680	0.691
N	254,808	254,808	90,198
AUC Digital Footprint	0.696	0.688	0.692
N	254,808	254,808	90,198
AUC credit bureau score + Digital Footprint	0.736	0.728	0.738
N	254,808	254,808	90,198
AUC credit bureau score + Digital Footprint, fixed effects	0.762	0.734	0.732
N	254,592	254,592	90,198

Robustness tests (scorable customers): detailed results

Panel A: Default definition	(1) Baseline (Default = Transfer to collection agency)	(2) Default = Writedown	(3) Loss given default (R^2 reported)	
AUC credit bureau score	0.6826	0.6918	0.0126	
AUC Digital footprint	0.6960	0.7232	0.0650	
AUC credit bureau score + digital footprint	0.7360	0.7564	0.0715	
N	254,808	254,808	2,384	
Panel B: Sample splits	(1) Small orders < EUR 218.92	(2) Large orders ≥ EUR 218.92	(3) Female	(4) Male
AUC credit bureau score	0.6878	0.6784	0.6893	0.6696
AUC Digital footprint	0.7126	0.6910	0.6997	0.6999
AUC credit bureau score + digital footprint	0.7497	0.7306	0.7448	0.7245
N	127,404	127,404	168,366	86,442

Robustness test: Fraud

	(1) In-sample	(2) Out-of-sample
Panel 1: Scorable customers		
AUC Credit Bureau Score		
Baseline (as in paper)	0.683	0.680
Exclude fraud	0.680	0.680
Fraud as dependent variable	0.702	0.682
N	254,808	254,808
AUC Digital Footprint		
Baseline (as in paper)	0.696	0.688
Exclude fraud	0.691	0.681
Fraud as dependent variable	0.786	0.728
N	254,808	254,808
AUC Credit Bureau Score + Digital Footprint		
Baseline (as in paper)	0.736	0.728
Exclude fraud	0.730	0.720
Fraud as dependent variable	0.804	0.748
N	254,808	254,808
Panel 2: Unscorable customers		
AUC Digital Footprint		
Baseline (as in paper)	0.722	0.683
Exclude fraud	0.718	0.668
Fraud as dependent variable	0.837	0.710
N	15,591	15,591

Conclusion

- Is digital footprint useful for predicting payment behavior?
  - Simple, easily accessible variables with similar predictive power as credit bureau score
  - Complement rather than substitute to credit bureau score
  - Works equally well for unscorable customers

- Potentially wide implications
  - Financial intermediaries' business model: Digital footprint helps to overcome information asymmetries between lenders and borrowers
  - Access to credit for the unbanked
  - Behavior of consumers, firms, and regulators in the digital sphere