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

Berlin, March 2019

We live in a Tech world

Top 5 U.S. firms by MarketCap

1980	2000	2018[1]
IBM	GE	Microsoft
AT&T	Exxon	
Exxon	Pfizer	amazon
STANDARD	cisco	Google
Shell	citigroup	f

1 Excluding Berkshire Hathaway (#5 based on market capitalization as of December 31, 2018)
Source: CRSP, Own analysis

Do we live in a FinTech world?

Banks (U.S.)		FinTechs (U.S.)		FinTechs (ex U.S.)	
JPMorgan Chase & Co.	325 bn	PayPal	99 bn	蚂蚁金服 ANT FINANCIAL	150 bn*
Bank of America	242 bn.	Quicken Loans	N.A.	wirecard	19 bn
VISA	232 bn.	Square	17 bn.	adyen wherever people pay	15 bn
WELLS FARGO	217bn	stripe	20 bn*	Klarna	5 bn*
mastercard	193bn	LendingClub	1 bn		
		ondeck	0.5 bn		

Source: CRSP, Own analysis. as of December 31, 2018 (or closest data possible), all numbers in USD
* Unlisted firm, valuation based on last funding round

Ant Financial vs. German Banking Sector

Valuation: ~150bn USD	x4	Deutsche Bank (2018): ~€ 20 bn
Alipay (Payments): 520mn customers	x10	Savings Banks GER (2017): ~54mn cards
Ant Fortune (Asset Mgr): USD 211bn AuM	x30%	DWS (2017): € 700 bn AuM
MY Bank (SME loans): 17bn loan volume H1/2017	x40%	Savings Banks GER (2016): € 80 bn
AntCreditPay (cons loans): 100mn active users	x13	Germany (2016): 7,7 Mio. New loans
Insurance: 392mn active users	x5	Allianz (worldwide): ~85 mn
Sesame Credit (scoring): 257mn active users	x4	Schufa: ~70 mn

Source: AntFinancial, various annual reports, Schufa, CapGemini: World FinTech Report 2018
Remark: Multiplicators in same currency over the same time horizon

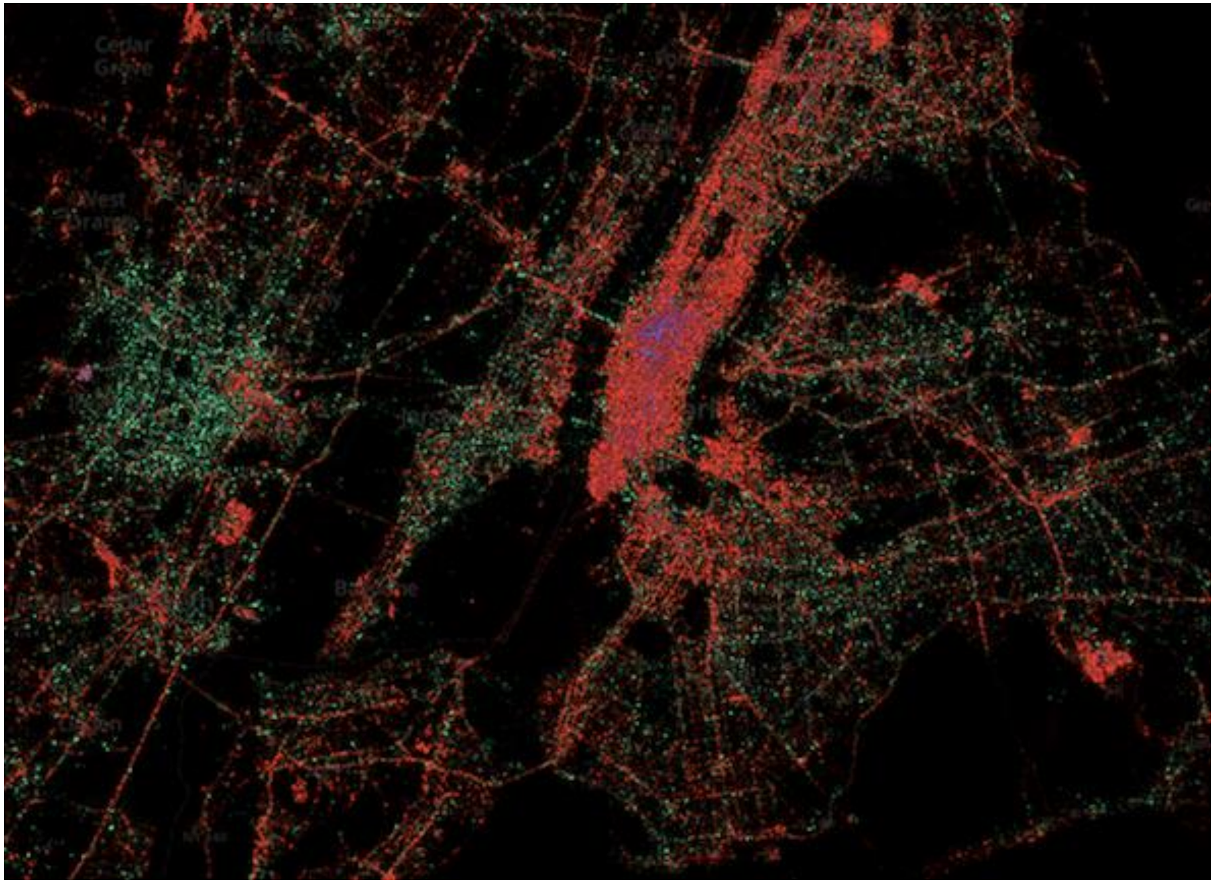
FinTech: Important topics

- Why/when is FinTech successful?
  - Regulatory arbitrage (Buchak et al., JFE 2018)
  - Operational efficiency (Fuster et al., RFS 2019)
  - UX
  - Information: Focus of this paper

- Why is it important?
  - Regulator: Financial stability
  - Bank perspective: loss of relationship advantage and increase in competition
  - Borrower perspective: Access to credit, who gains who looooses?
  - Inequality and fair lending acts: Do disadvantaged groups suffer more?
  - Development economics: Access to credit hindered by available data

- Q: Do digital footprints help FinTechs make better credit decisions?

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

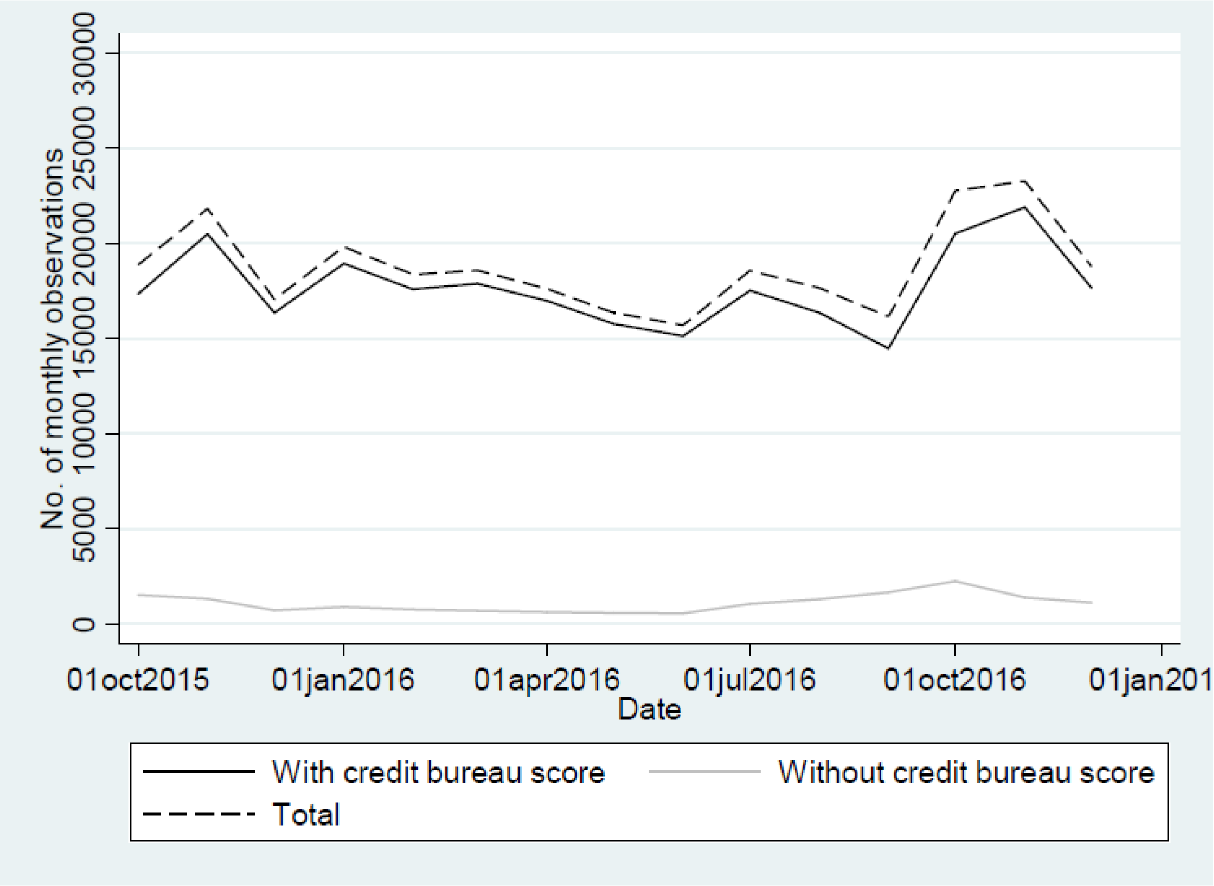
Dataset: Overview

- 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



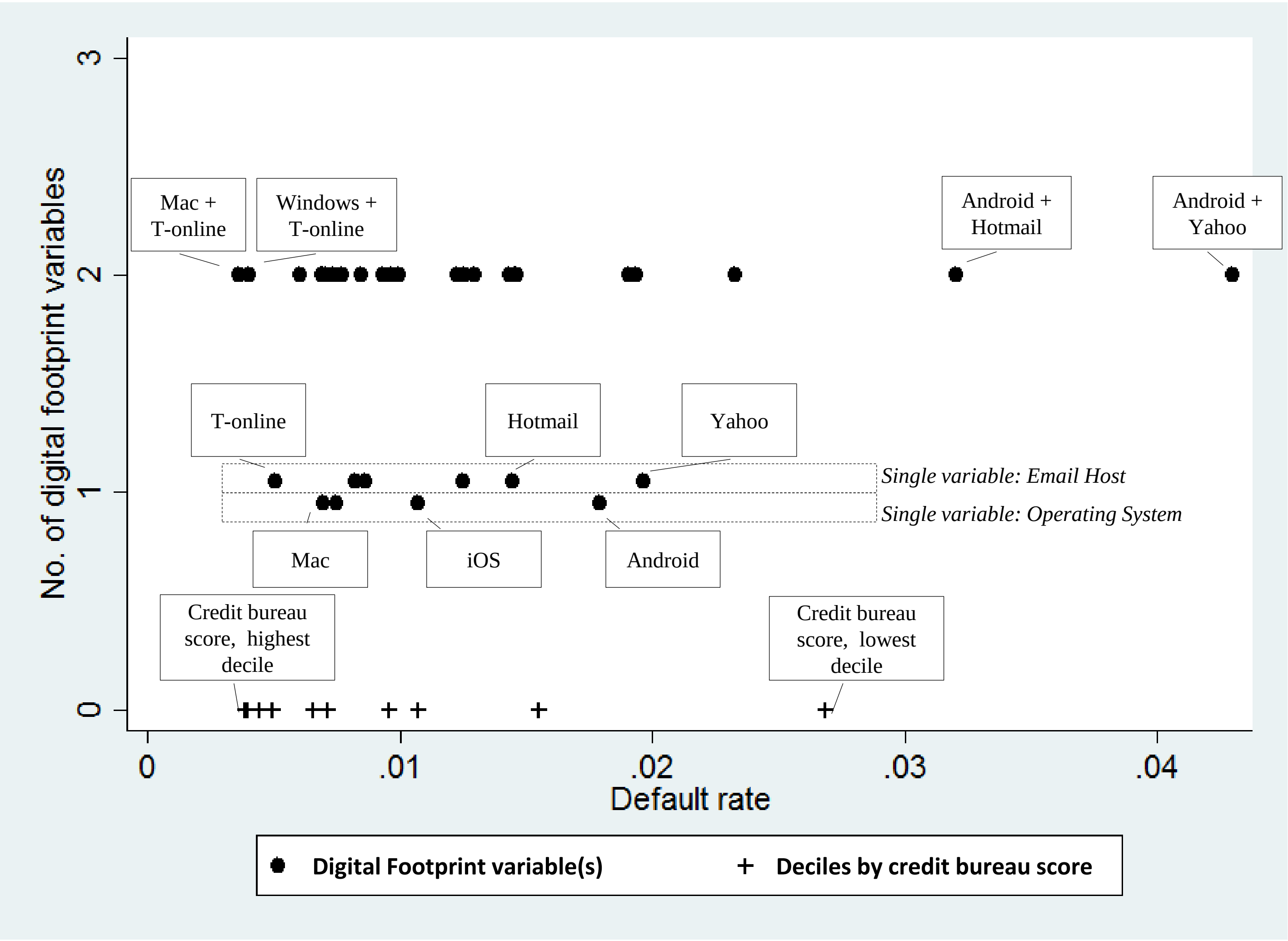
Roughly even distribution over time –
with slight increases in dark season (October/November)

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.	

Why these variables? Available for billions of people worldwide

Bivariate results



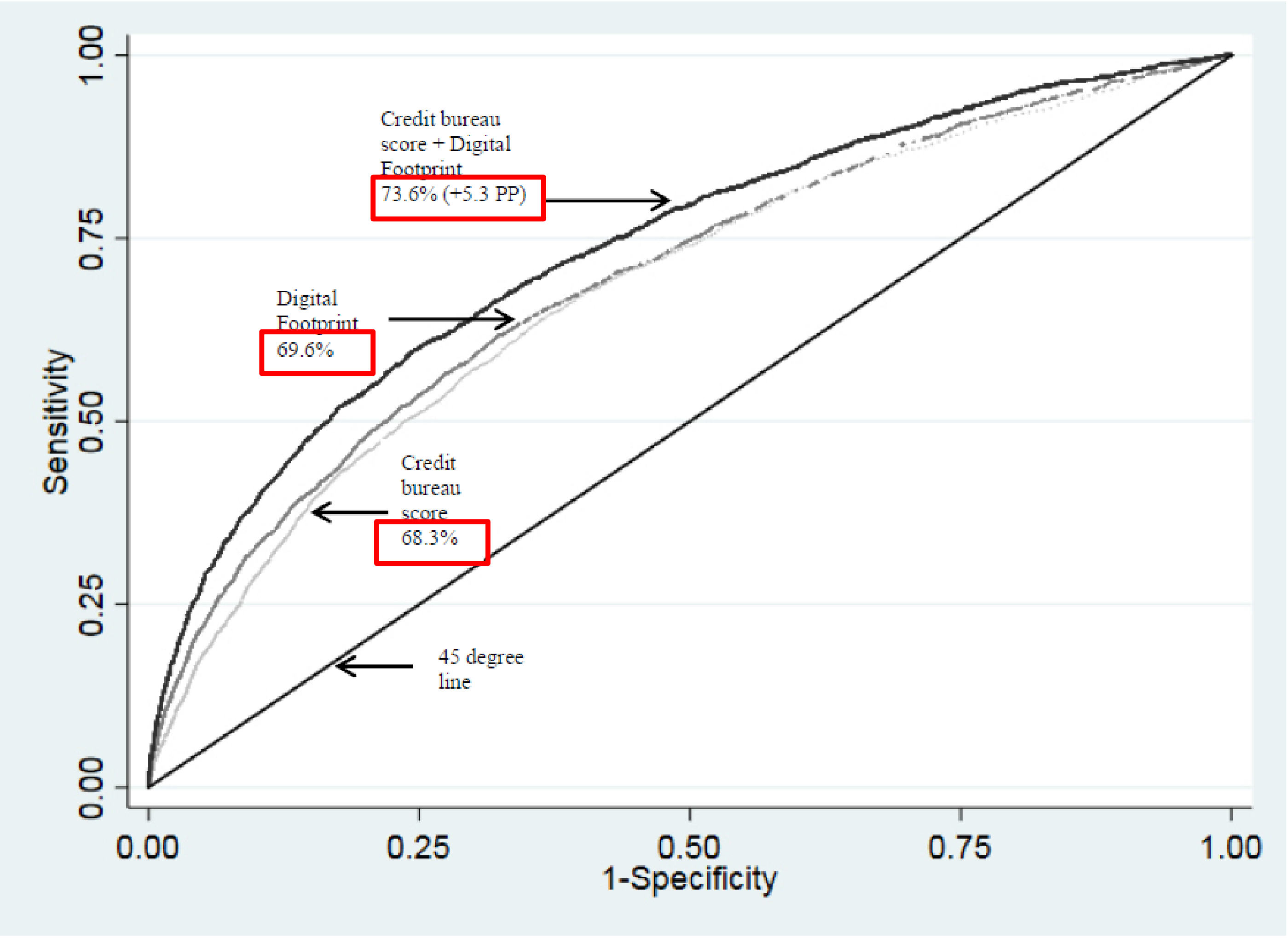
Outline of analysis

1. Digital footprint for default prediction
   - Area under the curve (AUC)
   - Which variables are important?
   - Comparison to bank-internal models

2. Impact on default rates and access to credit
   - Economic impact for firm and customers

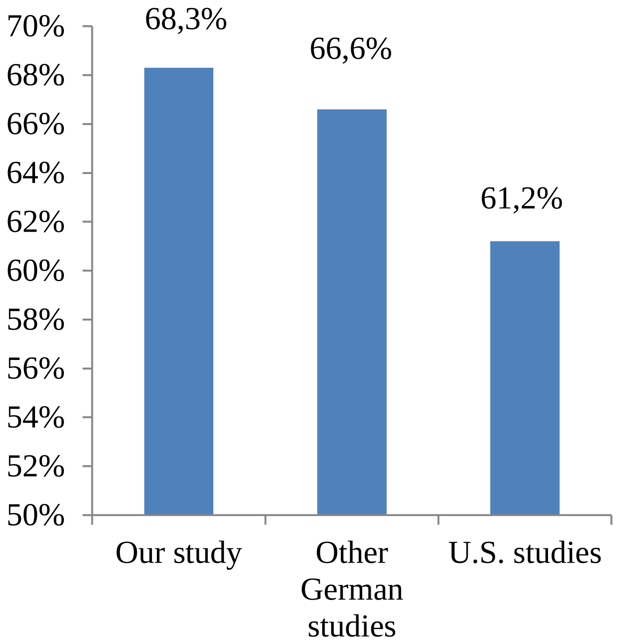
3. Discussion

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

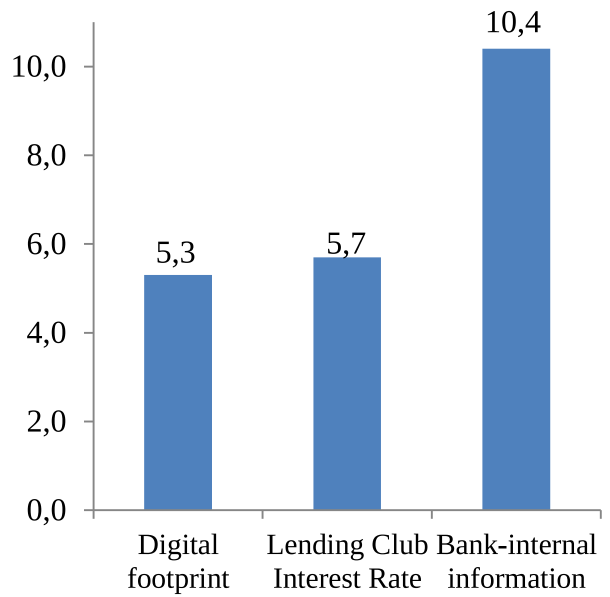


1) Area-under-the-Curve: Comparison

Discriminatory power
of credit bureau scores



Additional discriminatory power (in PP AUC) beyond credit bureau score



E-commerce sample not too special: German credit bureau score works well in our sample

Digital footprint brings us roughly halfway towards bank-internal information

Note: For details see Table A.2 in our paper. Note: Improvement of 10.4 PP using bank-internal information comes from a different data set, see Table A.2 for details.

1) Contribution of individual variables to AUC

Panel A: Individual digital footprint variables

Variable	Standalone AUC	Marginal AUC
Computer & Operating system	59.03%	+1.71PP***
Email Host	59.78%	+2.44PP***
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

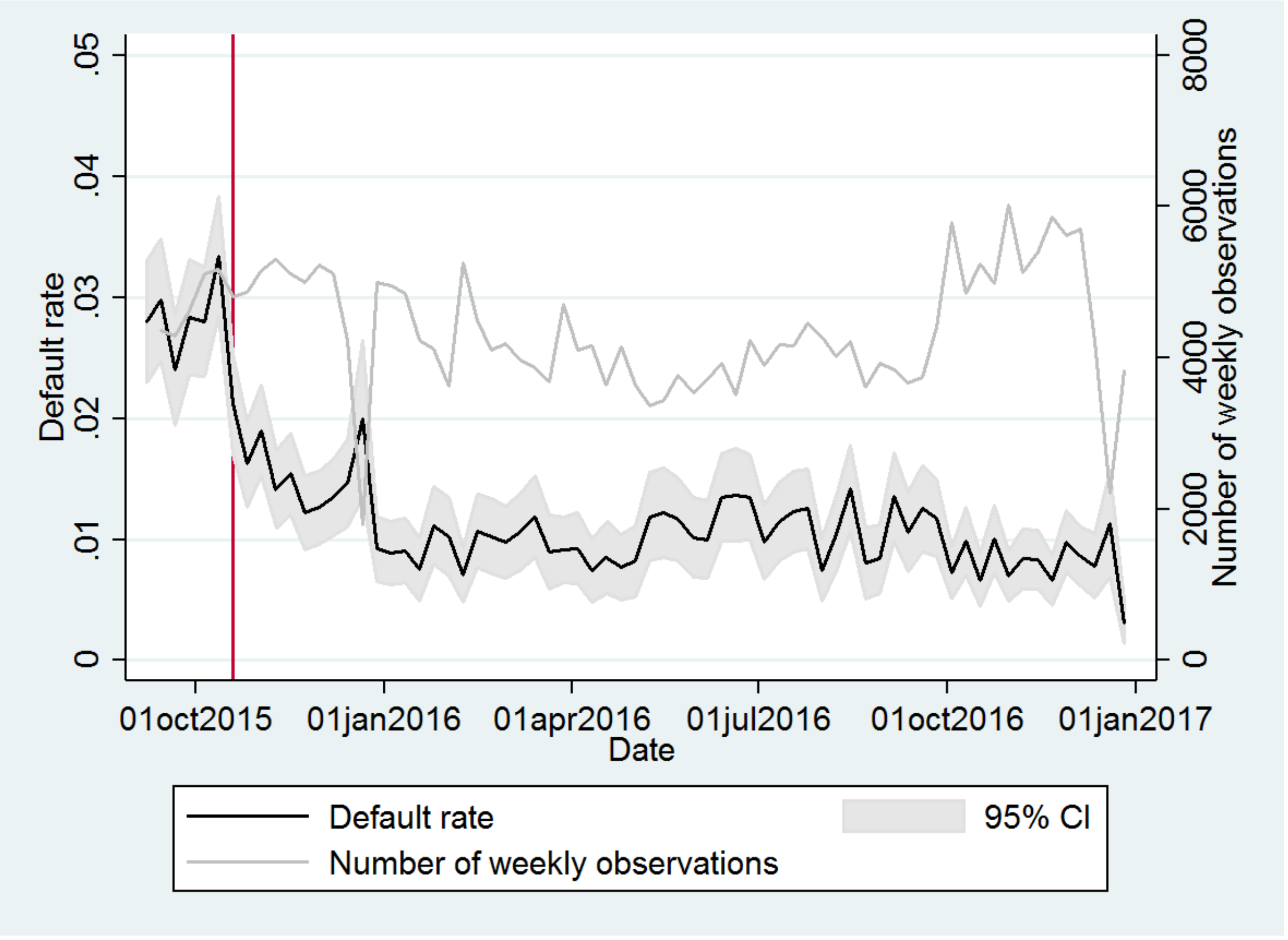
Variables	Standalone AUC	Marginal AUC
Proxy for income / costly to manipulate		
Potential proxy for income, financially costly to manipulate (Computer & Operating system, Email host: paid vs. non-paid dummy)	61.03%	+2.31PP
Unlikely to be a proxy for income, not financially costly to manipulate (Non-paid email host, Channel, Check-out time, Do not track setting, Name in Email, Number in Email, Is Lower Case, Email Error)	67.24%	+8.52PP
Impact on everyday behavior		
Requires one-time change only (Computer & Operating system, Email host, Do not track setting, Name in Email, Number in Email)	64.92%	+7.25PP
b) Requires thinking about how to behave during every individual buying process (Channel, Check-out time, Is Lower Case, Email Error)	62.30%	+4.63PP
Ease of manipulation		
Easy: financially cheap and requires one-time change only (Non-paid email host, Do not track setting, Name in Email, Number in Email)	60.88%	+2.27PP
Hard: financially costly or requires thinking about how to behave during every individual buying process (Computer & Operating system, Email host: paid vs. non-paid dummy, Channel, Check-out time, Is Lower Case, Email Error)	67.28%	+8.67PP

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

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

- Most important variables need effort to manipulate (financially or time-consuming)

2) Economic impact of better scoring model



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

2) Higher impact for low-score and unscorables

		Default rate			Invoice offered			Credit bureau score		
	N	Pre	Post	Δ	Pre	Post	Δ	Pre	Post	Δ
Panel A: Categories										
Sample 1: ScoreAndDFAdded	33,896	2.54%	1.19%	-1.36%***	96.65%	90.05%	-6.60%***	n.a.	98.26	n.a.
Sample 2: DFAdded	10,807	3.62%	2.33%	-1.29%***	39.00%	40.11%	1.11%***	97.82	97.84	0.02
Panel B: Sub-Categories of Sample 2 ("DFAdded")										
DFAdded / High score	3,614	0.84%	0.88%	0.04%	90.00%	90.94%	0.95%	99.42	99.42	0.00
DFAdded / Medium score	4,015	1.82%	2.15%	0.33%	85.31%	87.72%	2.41%***	98.17	98.16	0.00
DFAdded / Low score	2,096	6.30%	3.74%	-2.56%***	31.63%	27.64%	-3.98%***	94.46	94.41	-0.04
DFAdded / Unscorable	1,082	11.65%	6.44%	-5.22%***	10.14%	9.59%	-0.54%	n.a.	n.a.	n.a.

2) Further results

- Statistical tests (see paper for details)
  - Logistic regression
  - OOS (Nx2-fold cross validation), OOS-OOT
  - Subperiods
  - Machine learning (Random Forest): no significant improvement (WIP)
  - Sample split, default definition, loss given default

- Digital footprint predicts change in bureau scores
  - Q1 of digital footprint: bureau score increases by 0.39 over ~18m
  - Q5 of digital footprint: bureau score decreases by 0.39 over ~18m

- Access to credit
  - Allows access to credit for "unscorables" (basic information available, but no credit score available)
  - No effect for people with high credit scores (>25th percentile)

3) 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 (but bank-internal information still seems superior)

3) Implication 2: Access to credit for unbanked

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

- High expectations in digital footprints:
  - Digital footprints are special: ubiquitous, even in countries with few reliable records
  - World Bank: “Can digital footprints lead to Greater Financial Inclusion?”
  - Harvard Business Review: Fintech Companies Could Give Billions of People More Banking Options

- 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 in our sample (subject to ext. validity concerns)

3) 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

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