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
Motivation

• Digital footprint: Trace of simple, easily accessible information about almost every individual worldwide

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

• This paper: How much information does simply surfing the internet and registering on websites leave behind? How well does it predict default rates?

• 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
Digital footprint – 10 easily accessible variables

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

Graph showing the relationship between digital footprint variables and default rates, with Credit bureau score, highest decile and lowest decile highlighted. The graph includes categories such as Mac + T-online, Windows + T-online, Android + Hotmail, and Android + Yahoo. Single variable categories are Email Host and Operating System.
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: $\text{GINI} = 2 \cdot \text{AUC} - 1$
- 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
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.

- Idea: Does the digital footprint predict future changes in the credit bureau score? Answer is yes.
Economic impact of introducing digital footprint: Default rates go down

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

Pre-October 19:
• No digital footprint
• Credit bureau score for >€1100 and “unknowns” (“unknowns = customer not known to basic credit bureau”)

Post-October 19:
• Digital footprint for every observation
• Credit bureau score for every observation
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?”
  - 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
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
  - Lucas critique: consumers react to use of digital footprint. Implication: considerable impact on everyday’s life

- Beyond consumer behavior
  - Firms: Response by firms associated with low-creditworthiness products
  - Regulators: May intervene in case of violation of fair lending acts, incumbent banks might lobby regulators to intervene
Conclusion

• Is digital footprint useful for predicting payment behavior?
  – Simple, easily accessible variables with 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