Fintech and Credit Scoring for the Millennial Generation

Authors: Sumit Agarwal (NUS), Shashwat Alok (ISB), Pulak Ghosh (IIM-Bangalore), and Sudip Gupta (Fordham)

ABFER-2019

Research Question

- Does digital data available from individuals’ mobile phones predict loan outcomes?
    1. Likelihood of loan approval
    2. Loan Default
- Use data from a large fintech lending firm in India (CASHe)

Motivation

- 2 billion people around the world still lack bank accounts (IMF 2017)
  - Even those with bank account lack credit access
  - Primarily due to the lack of credit history/credit score
- Not just a developing country phenomena: 45 million Americans do not have a credit score (Consumer Financial Protection Bureau, 2015)
  - Could be 'good borrowers' if their `credit worthiness' could be evaluated by alternate means.
- In India, about 850 million individuals have never taken credit
  - Out of these, Transunion estimates - 220 million are credit eligible
  - But , 150 million such credit eligible customers lack a credit history and score

Motivation

- While credit history is absent, access to internet through mobile phones has been growing
  - **98.7 per cent mobile phone adoption in developing countries**
  - **Internet and Social media use has increased exponentially**
- An overlap between potentially creditworthy individuals who lack credit access but have an online presence
  - Large traces of data
- Quest for an alternate credit Score
  - Zhima/Seasame credit from Ant financial
    - Social media interactions and purchase transactions
  - Fintech lending platforms that rely on digital footprint data for loan decisions have been mushrooming

Big Picture Question?

- Can digital presence data be used to come up with an alternate credit score for unbanked customers?
- Why should we care?
  - A huge untapped market!
  - Lenders can potentially expand credit to the traditionally underserved
  - Policy implications for governments as they seek to expand credit access
- What do we know?
  - Berg, Burg, Gombovic and Puri (2019)
  - Iyer, Khwaja, Luttmer and Shue (2015)
  - D'Acunto, Rauter, Scheuch, and Weber (2019)

Our setting: Data

- Data: proprietary anonymized data on about 1,99,000 loan applicants from a mobile-only Fintech lending platform (CASHe) operating in India since 2016.
    - loans size: Rs 10,000 ($142) to 200,000 ($2846)
    - loan duration: minimum of 15 days to a maximum of 180 days.
    - 180,000 active users with 75% repeat borrowers
- All loan applications between October 2016 to January 2018
- The loan process: a customer has to download the app, enter all the requisite details, documentation and submit.
- Key variables:
    - Loan level: Loan size, interest rate, duration and purpose
    - Customer level: Age, salary, Job designation, Education level, Location
    - Customer's digital footprint:
        - Mode of login (Linkedin vs Facebook),
        - Types of apps installed: E-commerce (Ex Amazon, Flipkart etc) , travel (Airbnb, Tripadvisor etc) , dating (Tinder etc), social media (Whastapp, Facebook, Twitter etc), financial apps (Banking and Stock trading)
        - Type of smartphone: IOS vs Android
        - Other information: Number of calls, sms, contacts, and social media connections

Our setting: Specific Empirical Questions

- Whether and how the loan characteristics, the customer characteristics, and the digital footprints captured from mobile phones relate to
  1. loan approval decision
  2. Loan defaults
  3. Interaction between loan purpose and digital footprint for default prediction

Not a paper on causality!

Key findings

1. Digital footprint data has a significant correlation with the loan approval probability.
   - Those who log in through Facebook (Linkedin) are two (four) times more likely to get approved relative to those who log in by other means.
   - the number of contacts, the number of applications (apps from now) installed, and Mobile Loan app dummy is positively associated with loan approval.
2. Digital footprint variables have comparable (to credit score) and incremental power in predicting defaults
   - as compared to the credit score, the digital footprint variables taken together are able to explain 2 percentage points higher variation in defaults.
   - AUC of credit score is 58% and of digital footprint variables alone is 55%
   - Importantly, the AUC of a model that includes digital footprints, customer, and loan characteristics is 76%, 18 percentage points higher than the AUC of the model using only the credit bureau score and comparable to a model that uses includes credit score, customer, and loan characteristics
3. The discriminatory ability of the digital footprint variables varies based on the loan purpose
   - Customers who log in via Facebook are 20%, 26%, and 32% more likely to default when they take loans for making a purchase, meeting the EMI on an existing loan, or repayment of an existing loan respectively but less likely to default when they take a loan for medical needs.

Summary Statistics

	Approved (1)	Not Approved (2)	Difference (3)	Default (4)	Not Default (5)	Difference (6)
Loan Amount	23078.4	15254.4	7824.054 ***	30376.7	21154.2	9222.50 ***
Loanpurpose travel	0.09	0.046	0.044***	0.074	0.096	-0.022
Loanpurpose EMI	0.09	0.156	-0.07	0.076	0.1	-0.024
Loanpurpose purchase	0.15	0.144	0.006	0.138	0.154	-0.016
Loanpurpose Loanrepayment	0.07	0.1	-0.03	0.079	0.071	0.008
Loanpurpose Other	0.39	0.397	-0.007	0.382	0.393	-0.011
Log Interest Rate	1.4	NA		1.786	1.293	0.493
Age	31.3791	29.3428	2.03629 ***	31.583	31.32	2.03629 ***
Salary	37505.2	30213.4	7291.79 ***	38392.4	37271.2	1121.18 ***
CIBIL (>0, N=109 & 53k)	640.954	559.234	81.7195 ***	616.053	647.407	-31.3538 ***
Facebook Status	0.2889	0.2811	0.0078 ***	0.2878	0.28925	-0.00145 ***
Linkedin Status	0.027	0.01337	0.01363 ***	0.02407	0.02784	-0.00377 ***
Referral	0.1488	0.0457	0.1031 ***	0.1258	0.15492	-0.02912 ***
Sales App	0.2	0.164	0.036	0.182	0.199	-0.017
Dating App	0.03	0.023	0.007	0.028	0.029	-0.001
Finsavy app	0.85	0.003	0.847***	0.801	0.865	-0.064
Socialconnect app	0.9	0.003	0.897***	0.883	0.908	-0.025
Travel app	0.69	0.048	0.642***	0.637	0.71	-0.073
Mloan app	0.51	0.001	0.509***	0.47	0.526	-0.056
Referrer	0.223	0.0253	0.1977 ***	0.15693	0.24067	-0.083736 ***
# of SMS	1535.96	821.37	714.586 ***	1419.29	1566.52	-147.238 ***
# of Apps	44.046	37.09	6.956 ***	40.5931	44.9505	-4.35747 ***
# of Contacts	814.44	682.25	132.19 ***	803.085	817.43	-14.3453 ***
# of Connections	224.72	306.16	-81.44 ***	232.413	222.857	9.5556 ***
# of Calls	1953.8	1558.04	395.759 ***	1738.25	2010.28	-272.029 ***
IOS	0.11	0.061	0.049	0.105	0.108	-0.003
Prop of Repeat	0.6769	0.4475	0.2294 ***	0.4967	0.72449	-0.227792 ***
Education						
<High School	0.32	0.12	0.2***	0.13	0.12	0.01
High School	0.53	0.65	-0.12***	0.65	0.65	0
College	0.15	0.23	-0.08***	0.22	0.23	0.01
Job Designation						
Worker	0.42	0.42	0	0.37	0.44	0.07
Supervisor	0.23	0.25	-0.02	0.25	0.23	0.02
Manager	0.34	0.33	0.01	0.38	0.33	0.05
N	1,11,956	87,127		23,417	88,539	

Summary Statistics: Loan and Financial Variables

- Out of the 199,000 loan applications in our sample, 111,956 were approved while 87,127 were denied.
- Default rate in our sample is quite high at approximately 20%
  - 3% otherwise in India for retail loans
- Average credit score is 641 (sub-prime borrowers) and 5% of borrowers don't have a credit score
  - caters to higher risk customers
- Average loan size is Rs 23,078 ($328)
- Average annualized interest rate is 25%
- Average age is 31 (Younger customers)
- Average salary is Rs 37,505 ($534) per month or $6408 per annum
  - roughly 3 times the median per capita income of $2,134 in 2018.
  - Relatively well-off customers
- Loan purpose: 9% travel, 9% for EMI, 15 % for purchasing a good, 7% for the purpose of repaying a loan principal, and 21% for medical expenditure.

Summary Statistics: Digital Footprint Variables

- Log in
  - 29% of customers logged in using Facebook
  - 2.7% used Linkedin.
  - Rest by other means
- 85% customers have a mobile banking, mutual fund apps or stock trading app
  - Suggests potentially fintech savvy customers
- 51% of customers have installed another mobile-loan application
- 11% of the customers own an apple phone

Univariate Analysis: Loan approval

	Approved (1)	Not Approved (2)	Difference (3)
Loan Amount	23078.4	15254.4	7824.054 ***
Loanpurpose travel	0.09	0.046	0.044***
Loanpurpose EMI	0.09	0.156	-0.07
Loanpurpose purchase	0.15	0.144	0.006
Loanpurpose Loanrepayment	0.07	0.1	-0.03
Loanpurpose Other	0.39	0.397	-0.007
Log Interest Rate	1.4	NA	
Age	31.3791	29.3428	2.03629 ***
Salary	37505.2	30213.4	7291.79 ***
CIBIL (>0, N=109 & 53k)	640.954	559.234	81.7195 ***
Facebook Status	0.2889	0.2811	0.0078 ***
Linkedin Status	0.027	0.01337	0.01363 ***
Referral	0.1488	0.0457	0.1031 ***
Sales App	0.2	0.164	0.036
Dating App	0.03	0.023	0.007
Finsavy app	0.85	0.003	0.847***
Socialconnect app	0.9	0.003	0.897***
Travel app	0.69	0.048	0.642***
Mloan app	0.51	0.001	0.509***
Referrer	0.223	0.0253	0.1977 ***
# of SMS	1535.96	821.37	714.586 ***
# of Apps	44.046	37.09	6.956 ***
# of Contacts	814.44	682.25	132.19 ***
# of Connections	224.72	306.16	-81.44 ***
# of Calls	1953.8	1558.04	395.759 ***
IOS	0.11	0.061	0.049
Prop of Repeat	0.6769	0.4475	0.2294 ***

Univariate Analysis: Loan approval

	Approved (1)	Not Approved (2)	Difference (3)
Loan Amount	23078.4	15254.4	7824.054 ***
Loanpurpose travel	0.09	0.046	0.044***
Loanpurpose EMI	0.09	0.156	-0.07
Loanpurpose purchase	0.15	0.144	0.006
Loanpurpose Loanrepayment	0.07	0.1	-0.03
Loanpurpose Other	0.39	0.397	-0.007
Log Interest Rate	1.4	NA	
Age	31.3791	29.3428	2.03629 ***
Salary	37505.2	30213.4	7291.79 ***
CIBIL (>0, N=109 & 53k)	640.954	559.234	81.7195 ***
Facebook Status	0.2889	0.2811	0.0078 ***
Linkedin Status	0.027	0.01337	0.01363 ***
Referral	0.1488	0.0457	0.1031 ***
Sales App	0.2	0.164	0.036
Dating App	0.03	0.023	0.007
Finsavy app	0.85	0.003	0.847***
Socialconnect app	0.9	0.003	0.897***
Travel app	0.69	0.048	0.642***
Mloan app	0.51	0.001	0.509***
Referrer	0.223	0.0253	0.1977 ***
# of SMS	1535.96	821.37	714.586 ***
# of Apps	44.046	37.09	6.956 ***
# of Contacts	814.44	682.25	132.19 ***
# of Connections	224.72	306.16	-81.44 ***
# of Calls	1953.8	1558.04	395.759 ***
IOS	0.11	0.061	0.049
Prop of Repeat	0.6769	0.4475	0.2294 ***

Univariate Analysis: Loan Defaults

	Default (4)	Not Default (5)	Difference (6)
Loan Amount	30376.7	21154.2	9222.50 ***
Loanpurpose travel	0.074	0.096	-0.022
Loanpurpose EMI	0.076	0.1	-0.024
Loanpurpose purchase	0.138	0.154	-0.016
Loanpurpose Loanrepayment	0.079	0.071	0.008
Loanpurpose Other	0.382	0.393	-0.011
Log Interest Rate	1.786	1.293	0.493
Age	31.583	31.32	2.03629 ***
Salary	38392.4	37271.2	1121.18 ***
CIBIL (>0, N=109 & 53k)	616.053	647.407	-31.3538 ***
Facebook Status	0.2878	0.28925	-0.00145 ***
Linkedin Status	0.02407	0.02784	-0.00377 ***
Referral	0.1258	0.15492	-0.02912 ***
Sales App	0.182	0.199	-0.017
Dating App	0.028	0.029	-0.001
Finsavy app	0.801	0.865	-0.064
Socialconnect app	0.883	0.908	-0.025
Travel app	0.637	0.71	-0.073
Mloan app	0.47	0.526	-0.056
Referrer	0.15693	0.24067	-0.083736 ***
# of SMS	1419.29	1566.52	-147.238 ***
# of Apps	40.5931	44.9505	-4.35747 ***
# of Contacts	803.085	817.43	-14.3453 ***
# of Connections	232.413	222.857	9.5556 ***
# of Calls	1738.25	2010.28	-272.029 ***
IOS	0.105	0.108	-0.003
Prop of Repeat	0.4967	0.72449	-0.227792 ***

Univariate Analysis: Loan Defaults

	Default (4)	Not Default (5)	Difference (6)
Loan Amount	30376.7	21154.2	9222.50 ***
Loanpurpose travel	0.074	0.096	-0.022
Loanpurpose EMI	0.076	0.1	-0.024
Loanpurpose purchase	0.138	0.154	-0.016
Loanpurpose Loanrepayment	0.079	0.071	0.008
Loanpurpose Other	0.382	0.393	-0.011
Log Interest Rate	1.786	1.293	0.493
Age	31.583	31.32	2.03629 ***
Salary	38392.4	37271.2	1121.18 ***
CIBIL (>0, N=109 & 53k)	616.053	647.407	-31.3538 ***
Facebook Status	0.2878	0.28925	-0.00145 ***
Linkedin Status	0.02407	0.02784	-0.00377 ***
Referral	0.1258	0.15492	-0.02912 ***
Sales App	0.182	0.199	-0.017
Dating App	0.028	0.029	-0.001
Finsavy app	0.801	0.865	-0.064
Socialconnect app	0.883	0.908	-0.025
Travel app	0.637	0.71	-0.073
Mloan app	0.47	0.526	-0.056
Referrer	0.15693	0.24067	-0.083736 ***
# of SMS	1419.29	1566.52	-147.238 ***
# of Apps	40.5931	44.9505	-4.35747 ***
# of Contacts	803.085	817.43	-14.3453 ***
# of Connections	232.413	222.857	9.5556 ***
# of Calls	1738.25	2010.28	-272.029 ***
IOS	0.105	0.108	-0.003
Prop of Repeat	0.4967	0.72449	-0.227792 ***

Multivariate Analysis

$$\text{Loan Outcome}_{ilt} = \beta_0 + \sum_{j=1}^{M} \beta_j \text{Loan Characteristics}_{lt} + \sum_{j=1}^{N} \beta_j \text{Customer financials}_{it} + \sum_{j=1}^{O} \beta_j \text{Customer digital footprint}_{it} + \varepsilon_{ilt}$$

i refers to an individual, l refers to loan and t refers to the time of loan application

Loan outcome:

1. Approved is a dummy variable which takes the value one for loans that were approved and zero otherwise
2. Default which identifies loans in default

3. Loan purpose identifies one of the six different kinds of loan, Medical, Travel, EMI, Loan repayment, purchase and other purpose
4. Loan duration a dummy variable that identifies loans of 15, 60, 90, 120, and 180 days duration

We focus on loan approvals and defaults for now!

Dependent Variable: Loan Approval

VARIABLES	Logit coeff (1)	Logit coeff (2)	Logit coeff (3)	Logit coeff (4)	Logit coeff (5)	Logit coeff (6)
CIBIL	0.00276***	0.00268***				0.00207***
	(0.000)	(0.000)				(0.000)
Log of Salary		0.450***			0.172	0.137
		(0.018)			(0.145)	(0.142)
Log Loan Amount		0.161***			0.644***	0.662***
		(0.010)			(0.082)	(0.082)
Log Age		-0.243***			3.047***	2.794***
		(0.034)			(0.525)	(0.517)
Education Dummy		0.251***			-0.547***	-0.534***
		(0.014)			(0.136)	(0.134)
Travel.purpose cashe		0.326***			-0.835***	-0.915***
		(0.026)			(0.255)	(0.257)
EMI.purpose cashe		-0.477***			-0.683***	-0.723***
		(0.020)			(0.241)	(0.243)
Purchase.purpose cashe		-0.207***			0.238	0.176
		(0.019)			(0.284)	(0.286)
Loanrepayment.purpose cashe		-0.484***			-1.382***	-1.436***
		(0.023)			(0.231)	(0.235)
Other.purpose cashe		-0.192***			-0.261	-0.309
		(0.016)			(0.200)	(0.202)
Log no of SMS			-0.0246	-0.026	-0.0154	-0.0121
			(0.031)	(0.031)	(0.031)	(0.031)
Log No of Contacts			0.115*	0.113*	0.0435	0.0309
			(0.064)	(0.063)	(0.069)	(0.068)
Log no of Apps			0.177*	0.174*	0.175	0.149
			(0.104)	(0.105)	(0.109)	(0.106)
Log Calllog			-0.344***	-0.348***	-0.343***	-0.335***
			(0.055)	(0.055)	(0.056)	(0.053)
Dating App			0.376	0.4	0.631	0.623
			(0.455)	(0.455)	(0.459)	(0.459)
Finsavy app			-0.556*	-0.589*	-0.324	-0.365
			(0.321)	(0.324)	(0.331)	(0.332)
Socialconnect app			-	-	-	-
Travel app			0.196	0.201	0.249	0.293*
			(0.152)	(0.153)	(0.156)	(0.154)
Mloan app			0.415***	0.415***	0.384***	0.388***
			(0.127)	(0.127)	(0.128)	(0.128)
Facebook status			0.665***	0.669***	0.642***	0.680***
			(0.164)	(0.165)	(0.167)	(0.169)
Linkedin status			1.407**	1.393**	1.158	1.272*
			(0.709)	(0.709)	(0.715)	(0.715)
IOS Dummy				-1.106***	-0.975***	-0.959***
				(0.343)	(0.348)	(0.353)
Constant	-0.918***	-6.064***	7.079***	7.192***	-10.29***	-10.37***
	(0.022)	(0.156)	(0.570)	(0.571)	(1.700)	(1.709)
State Fixed Effects	N	N	N	N	N	Y
Observations	166,231	166,220	97,457	97,457	97,450	97,441
Pseudo R-squared	0.0393	0.0649	0.0258	0.0279	0.0838	0.0964

Key Takeaways

- loan applicants with higher credit score have a higher likelihood of getting approved
  - Credit score explains about 4% of the variation in the likelihood of loan approval.
- Customers that earn more and are more educated have a higher chance of approval.
- Younger customers are more likely to be approved.
- Travel and education loans have the highest likelihood of approval
- Medical loans have a higher likelihood of approval as compared to loans taken for the purpose of EMI, purchase, repayment and uncategorized loans.

Dependent Variable: Loan Approval

VARIABLES	Logit coeff (1)	Logit coeff (2)	Logit coeff (3)	Logit coeff (4)	Logit coeff (5)	Logit coeff (6)
CIBIL	0.00276***	0.00268***				0.00207***
	(0.000)	(0.000)				(0.000)
Log of Salary		0.450***			0.172	0.137
		(0.018)			(0.145)	(0.142)
Log Loan Amount		0.161***			0.644***	0.662***
		(0.010)			(0.082)	(0.082)
Log Age		-0.243***			3.047***	2.794***
		(0.034)			(0.525)	(0.517)
Education Dummy		0.251***			-0.547***	-0.534***
		(0.014)			(0.136)	(0.134)
Travel.purpose cashe		0.326***			-0.835***	-0.915***
		(0.026)			(0.255)	(0.257)
EMI.purpose cashe		-0.477***			-0.683***	-0.723***
		(0.020)			(0.241)	(0.243)
Purchase.purpose cashe		-0.207***			0.238	0.176
		(0.019)			(0.284)	(0.286)
Loanrepayment.purpose cashe		-0.484***			-1.382***	-1.436***
		(0.023)			(0.231)	(0.235)
Other.purpose cashe		-0.192***			-0.261	-0.309
		(0.016)			(0.200)	(0.202)
Log no of SMS			-0.0246	-0.026	-0.0154	-0.0121
			(0.031)	(0.031)	(0.031)	(0.031)
Log No of Contacts			0.115*	0.113*	0.0435	0.0309
			(0.064)	(0.063)	(0.069)	(0.068)
Log no of Apps			0.177*	0.174*	0.175	0.149
			(0.104)	(0.105)	(0.109)	(0.106)
Log Calllog			-0.344***	-0.348***	-0.343***	-0.335***
			(0.055)	(0.055)	(0.056)	(0.053)
Dating App			0.376	0.4	0.631	0.623
			(0.455)	(0.455)	(0.459)	(0.459)
Finsavy app			-0.556*	-0.589*	-0.324	-0.365
			(0.321)	(0.324)	(0.331)	(0.332)
Socialconnect app			-	-	-	-
Travel app			0.196	0.201	0.249	0.293*
			(0.152)	(0.153)	(0.156)	(0.154)
Mloan app			0.415***	0.415***	0.384***	0.388***
			(0.127)	(0.127)	(0.128)	(0.128)
Facebook status			0.665***	0.669***	0.642***	0.680***
			(0.164)	(0.165)	(0.167)	(0.169)
Linkedin status			1.407**	1.393**	1.158	1.272*
			(0.709)	(0.709)	(0.715)	(0.715)
IOS Dummy				-1.106***	-0.975***	-0.959***
				(0.343)	(0.348)	(0.353)
Constant	-0.918***	-6.064***	7.079***	7.192***	-10.29***	-10.37***
	(0.022)	(0.156)	(0.570)	(0.571)	(1.700)	(1.709)
State Fixed Effects	N	N	N	N	N	Y
Observations	166,231	166,220	97,457	97,457	97,450	97,441
Pseudo R-squared	0.0393	0.0649	0.0258	0.0279	0.0838	0.0964

Key Takeaways

- Customers who log in through either Facebook or Linkedin have a significantly higher chance of approval.
  - Those who login through Facebook (Linkedin) are two (four) times more likely to get approved relative to those who login by other means.
- The number of contacts, the number of apps installed, and Mloan app dummy are positively associated with loan approval.
- Surprisingly:
  - Customers with Financial apps are less likely to be approved.
  - Customers with an IOS device have a lower likelihood of approval.
    - prior studies highlight that owning an IOS device is a strong predictor of higher earnings (Bertrand and Kamenica (2018)).
    - And defaults (Berg, Burg, Gombovic and Puri (2019)

Credit Score vs Digital Footprint

- Focus on a model that includes all loan characteristics, customer characteristics, and digital footprint but excludes credit bureau score.
- Our objective here is two folds.
  1. Whether our results on digital footprint continue to hold once we control for other loan level and customer level characteristics.
  2. we want to examine if observable loan, customer, and digital foot print characteristics can explain a higher fraction of the variation in loan approval decisions as compared to a model that includes just the CIBIL score along with customer and loan characteristics.

  - Can digital footprint substitute for the credit score?

Dependent Variable: Loan Approval

VARIABLES	Logit coeff (1)	Logit coeff (2)	Logit coeff (3)	Logit coeff (4)	Logit coeff (5)	Logit coeff (6)
CIBIL	0.00276***	0.00268***				0.00207***
	(0.000)	(0.000)				(0.000)
Log of Salary		0.450***			0.172	0.137
		(0.018)			(0.145)	(0.142)
Log Loan Amount		0.161***			0.644***	0.662***
		(0.010)			(0.082)	(0.082)
Log Age		-0.243***			3.047***	2.794***
		(0.034)			(0.525)	(0.517)
Education Dummy		0.251***			-0.547***	-0.534***
		(0.014)			(0.136)	(0.134)
Travel.purpose cashe		0.326***			-0.835***	-0.915***
		(0.026)			(0.255)	(0.257)
EMI.purpose cashe		-0.477***			-0.683***	-0.723***
		(0.020)			(0.241)	(0.243)
Purchase.purpose cashe		-0.207***			0.238	0.176
		(0.019)			(0.284)	(0.286)
Loanrepayment.purpose cashe		-0.484***			-1.382***	-1.436***
		(0.023)			(0.231)	(0.235)
Other.purpose cashe		-0.192***			-0.261	-0.309
		(0.016)			(0.200)	(0.202)
Log no of SMS			-0.0246	-0.026	-0.0154	-0.0121
			(0.031)	(0.031)	(0.031)	(0.031)
Log No of Contacts			0.115*	0.113*	0.0435	0.0309
			(0.064)	(0.063)	(0.069)	(0.068)
Log no of Apps			0.177*	0.174*	0.175	0.149
			(0.104)	(0.105)	(0.109)	(0.106)
Log Calllog			-0.344***	-0.348***	-0.343***	-0.335***
			(0.055)	(0.055)	(0.056)	(0.053)
Dating App			0.376	0.4	0.631	0.623
			(0.455)	(0.455)	(0.459)	(0.459)
Finsavy app			-0.556*	-0.589*	-0.324	-0.365
			(0.321)	(0.324)	(0.331)	(0.332)
Socialconnect app			-	-	-	-
Travel app			0.196	0.201	0.249	0.293*
			(0.152)	(0.153)	(0.156)	(0.154)
Mloan app			0.415***	0.415***	0.384***	0.388***
			(0.127)	(0.127)	(0.128)	(0.128)
Facebook status			0.665***	0.669***	0.642***	0.680***
			(0.164)	(0.165)	(0.167)	(0.169)
Linkedin status			1.407**	1.393**	1.158	1.272*
			(0.709)	(0.709)	(0.715)	(0.715)
IOS Dummy				-1.106***	-0.975***	-0.959***
				(0.343)	(0.348)	(0.353)
Constant	-0.918***	-6.064***	7.079***	7.192***	-10.29***	-10.37***
	(0.022)	(0.156)	(0.570)	(0.571)	(1.700)	(1.709)
State Fixed Effects	N	N	N	N	N	Y
Observations	166,231	166,220	97,457	97,457	97,450	97,441
Pseudo R-squared	0.0393	0.0649	0.0258	0.0279	0.0838	0.0964

Key Takeaways

- Customer's salary, number of contacts, number of apps installed, and Linkedin login dummy is no longer statistically significant.
  - Suggests that digital footprint variables proxy for some customer and loan level characteristics
- Digital footprint model explain 8% of the variation in loan approvals.
  - 2% higher than the model with credit score
- Overall, digital footprint variables have significant explanatory power for loan approval decisions over and above the credit bureau score.

Dependent Variable: Loan defaults

VARIABLES	Logit coeff (1)	Logit coeff (2)	Logit coeff (3)	Logit coeff (4)	Logit coeff (5)	Logit coeff (6)
CIBIL	-0.00138***	-0.00156***				-0.00128***
	(0.000)	(0.000)				(0.000)
Log of Salary		-0.960***			-0.845***	-0.826***
		(0.035)			(0.039)	(0.038)
Log Loan Amount		0.634***			0.601***	0.594***
		(0.033)			(0.037)	(0.036)
Log Age		-0.075			-0.413***	-0.355***
		(0.051)			(0.056)	(0.056)
Education Dummy		-0.126***			-0.150***	-0.146***
		(0.019)			(0.021)	(0.021)
Travel.purpose cashe		-0.394***			-0.335***	-0.313***
		(0.033)			(0.037)	(0.037)
EMI.purpose cashe		-0.320***			-0.299***	-0.288***
		(0.032)			(0.035)	(0.035)
Purchase.purpose cashe		-0.383***			-0.377***	-0.359***
		(0.027)			(0.029)	(0.029)
Loanrepayment.purpose cashe		-0.274***			-0.249***	-0.230***
		(0.033)			(0.036)	(0.036)
Other.purpose cashe		-0.264***			-0.277***	-0.258***
		(0.021)			(0.023)	(0.023)
Log no of SMS			0.0167***	0.0163***	0.00384	0.00289
			(0.004)	(0.004)	(0.005)	(0.005)
Log No of Contacts			-0.0196**	-0.0202**	-0.0428***	-0.0389***
			(0.009)	(0.009)	(0.010)	(0.010)
Log no of Apps			-0.0975***	-0.101***	-0.101***	-0.0887***
			(0.014)	(0.014)	(0.016)	(0.016)
Log Calllog			-0.0300***	-0.0313***	-0.0253***	-0.0263***
			(0.007)	(0.007)	(0.007)	(0.007)
Dating App			0.139***	0.147***	0.165***	0.168***
			(0.045)	(0.045)	(0.049)	(0.049)
Finsavy app			-0.487***	-0.495***	-0.382***	-0.374***
			(0.031)	(0.031)	(0.034)	(0.034)
Socialconnect app			0.120*	0.105	0.0792	0.0725
			(0.068)	(0.068)	(0.074)	(0.074)
Travel app			-0.210***	-0.209***	-0.349***	-0.358***
			(0.019)	(0.019)	(0.021)	(0.022)
Mloan app			-0.0938***	-0.0934***	-0.129***	-0.133***
			(0.017)	(0.017)	(0.018)	(0.018)
Facebook status			0.00318	0.00472	0.0509***	0.0443**
			(0.018)	(0.018)	(0.019)	(0.019)
Linkedin status			0.0125	0.00748	0.0413	0.0307
			(0.049)	(0.049)	(0.054)	(0.054)
IOS Dummy				-0.725***	-0.816***	-0.823***
				(0.085)	(0.090)	(0.090)
Constant	-0.440***	2.935***	-0.264***	-0.209**	3.494***	3.918***
	(0.030)	(0.258)	(0.091)	(0.092)	(0.301)	(0.302)
State Fixed Effects	N	N	N	N	N	N
Observations	113,245	113,235	98,391	98,391	98,381	98,372
Pseudo R-squared	0.00674	0.146	0.00889	0.00978	0.15	0.154
AUC	0.5826	0.7606	0.554	0.5573	0.7634	0.7674

Key Takeaways

- AUC of the model using only the credit score for predicting defaults is 58%
  - is lower than 62% reported by Iyer, Khwaja, Luttmer and Shue (2015) based on a sample of loans from a US based peer to peer lending platform, “Propser.com” and 68.3% reported by Berg, Burg, Gombovic and Puri (2019) based on a sample of purchases from a German e-retailer.
  - Suggests that Discriminatory ability of credit score maybe lower in emerging markets
  - The marginal value of digital footprint variables (additional information) is likely to be higher in such environments
- Default likelihood is lower for all categories of loans (Travel, EMI, Purchase, Repayment, and other) relative to loans taken for medical needs.
  - Consistent with the idea that health shocks are correlated with financial distress (Kalda (2019)).
- Salary, and education are negatively related to defaults.

Dependent Variable: Loan defaults

VARIABLES	Logit coeff (1)	Logit coeff (2)	Logit coeff (3)	Logit coeff (4)	Logit coeff (5)	Logit coeff (6)
CIBIL	-0.00138***	-0.00156***				-0.00128***
	(0.000)	(0.000)				(0.000)
Log of Salary		-0.960***			-0.845***	-0.826***
		(0.035)			(0.039)	(0.038)
Log Loan Amount		0.634***			0.601***	0.594***
		(0.033)			(0.037)	(0.036)
Log Age		-0.075			-0.413***	-0.355***
		(0.051)			(0.056)	(0.056)
Education Dummy		-0.126***			-0.150***	-0.146***
		(0.019)			(0.021)	(0.021)
Travel.purpose cashe		-0.394***			-0.335***	-0.313***
		(0.033)			(0.037)	(0.037)
EMI.purpose cashe		-0.320***			-0.299***	-0.288***
		(0.032)			(0.035)	(0.035)
Purchase.purpose cashe		-0.383***			-0.377***	-0.359***
		(0.027)			(0.029)	(0.029)
Loanrepayment.purpose cashe		-0.274***			-0.249***	-0.230***
		(0.033)			(0.036)	(0.036)
Other.purpose cashe		-0.264***			-0.277***	-0.258***
		(0.021)			(0.023)	(0.023)
Log no of SMS			0.0167***	0.0163***	0.00384	0.00289
			(0.004)	(0.004)	(0.005)	(0.005)
Log No of Contacts			-0.0196**	-0.0202**	-0.0428***	-0.0389***
			(0.009)	(0.009)	(0.010)	(0.010)
Log no of Apps			-0.0975***	-0.101***	-0.101***	-0.0887***
			(0.014)	(0.014)	(0.016)	(0.016)
Log Calllog			-0.0300***	-0.0313***	-0.0253***	-0.0263***
			(0.007)	(0.007)	(0.007)	(0.007)
Dating App			0.139***	0.147***	0.165***	0.168***
			(0.045)	(0.045)	(0.049)	(0.049)
Finsavy app			-0.487***	-0.495***	-0.382***	-0.374***
			(0.031)	(0.031)	(0.034)	(0.034)
Socialconnect app			0.120*	0.105	0.0792	0.0725
			(0.068)	(0.068)	(0.074)	(0.074)
Travel app			-0.210***	-0.209***	-0.349***	-0.358***
			(0.019)	(0.019)	(0.021)	(0.022)
Mloan app			-0.0938***	-0.0934***	-0.129***	-0.133***
			(0.017)	(0.017)	(0.018)	(0.018)
Facebook status			0.00318	0.00472	0.0509***	0.0443**
			(0.018)	(0.018)	(0.019)	(0.019)
Linkedin status			0.0125	0.00748	0.0413	0.0307
			(0.049)	(0.049)	(0.054)	(0.054)
IOS Dummy				-0.725***	-0.816***	-0.823***
				(0.085)	(0.090)	(0.090)
Constant	-0.440***	2.935***	-0.264***	-0.209**	3.494***	3.918***
	(0.030)	(0.258)	(0.091)	(0.092)	(0.301)	(0.302)
State Fixed Effects	N	N	N	N	N	N
Observations	113,245	113,235	98,391	98,391	98,381	98,372
Pseudo R-squared	0.00674	0.146	0.00889	0.00978	0.15	0.154
AUC	0.5826	0.7606	0.554	0.5573	0.7634	0.7674

Key Takeaways

- AUC of this specification is 55% and lies in the confidence interval of the AUC estimate using just the credit bureau score.
- Explains about 2% additional variation in loan defaults as compared to just the credit bureau score.
- Individuals without a financial app are about one and a half times more likely to default relative to those that have such an app installed.
  - may be correlated with the financial sophistication of a customer.
- Customers with some other mobile loan application (Mloan dummy) are about 9% less likely to default.
- Those with a dating app (any other social network app) are 15% (13%) more likely to default.
- Customers with a travel app are about 19% less likely to default.
- those with an Android phone are twice as likely to default as those with an Apple phone.

- Caveat: We can't pin down the underlying causal mechanism
  - However, these results indicate that the nature of apps installed on the phone have significant discriminatory power in default prediction.

Dependent Variable: Loan defaults

VARIABLES	Logit coeff (1)	Logit coeff (2)	Logit coeff (3)	Logit coeff (4)	Logit coeff (5)	Logit coeff (6)
CIBIL	-0.00138***	-0.00156***				-0.00128***
	(0.000)	(0.000)				(0.000)
Log of Salary		-0.960***			-0.845***	-0.826***
		(0.035)			(0.039)	(0.038)
Log Loan Amount		0.634***			0.601***	0.594***
		(0.033)			(0.037)	(0.036)
Log Age		-0.075			-0.413***	-0.355***
		(0.051)			(0.056)	(0.056)
Education Dummy		-0.126***			-0.150***	-0.146***
		(0.019)			(0.021)	(0.021)
Travel.purpose cashe		-0.394***			-0.335***	-0.313***
		(0.033)			(0.037)	(0.037)
EMI.purpose cashe		-0.320***			-0.299***	-0.288***
		(0.032)			(0.035)	(0.035)
Purchase.purpose cashe		-0.383***			-0.377***	-0.359***
		(0.027)			(0.029)	(0.029)
Loanrepayment.purpose cashe		-0.274***			-0.249***	-0.230***
		(0.033)			(0.036)	(0.036)
Other.purpose cashe		-0.264***			-0.277***	-0.258***
		(0.021)			(0.023)	(0.023)
Log no of SMS			0.0167***	0.0163***	0.00384	0.00289
			(0.004)	(0.004)	(0.005)	(0.005)
Log No of Contacts			-0.0196**	-0.0202**	-0.0428***	-0.0389***
			(0.009)	(0.009)	(0.010)	(0.010)
Log no of Apps			-0.0975***	-0.101***	-0.101***	-0.0887***
			(0.014)	(0.014)	(0.016)	(0.016)
Log Calllog			-0.0300***	-0.0313***	-0.0253***	-0.0263***
			(0.007)	(0.007)	(0.007)	(0.007)
Dating App			0.139***	0.147***	0.165***	0.168***
			(0.045)	(0.045)	(0.049)	(0.049)
Finsavy app			-0.487***	-0.495***	-0.382***	-0.374***
			(0.031)	(0.031)	(0.034)	(0.034)
Socialconnect app			0.120*	0.105	0.0792	0.0725
			(0.068)	(0.068)	(0.074)	(0.074)
Travel app			-0.210***	-0.209***	-0.349***	-0.358***
			(0.019)	(0.019)	(0.021)	(0.022)
Mloan app			-0.0938***	-0.0934***	-0.129***	-0.133***
			(0.017)	(0.017)	(0.018)	(0.018)
Facebook status			0.00318	0.00472	0.0509***	0.0443**
			(0.018)	(0.018)	(0.019)	(0.019)
Linkedin status			0.0125	0.00748	0.0413	0.0307
			(0.049)	(0.049)	(0.054)	(0.054)
IOS Dummy				-0.725***	-0.816***	-0.823***
				(0.085)	(0.090)	(0.090)
Constant	-0.440***	2.935***	-0.264***	-0.209**	3.494***	3.918***
	(0.030)	(0.258)	(0.091)	(0.092)	(0.301)	(0.302)
State Fixed Effects	N	N	N	N	N	N
Observations	113,245	113,235	98,391	98,391	98,381	98,372
Pseudo R-squared	0.00674	0.146	0.00889	0.00978	0.15	0.154
AUC	0.5826	0.7606	0.554	0.5573	0.7634	0.7674

Key Takeaways

- the coefficient estimate of IOS dummy remains statistically significant even after controlling for the customer's monthly salary.
  - implies that owing an Apple device captures an unobservable aspect of individuals which is not fully absorbed by earnings.
- Customers who log in through Facebook are more likely
- The AUC of this specification is 76%
  - 18 percentage points higher than the AUC of the model using only the credit bureau score
  - equal to the model which includes CIBIL score combined with customer and loan characteristics.
- Overall, digital footprint variables can be used to predict the likelihood of default and can perform at least as well as the credit score.

Defaults, Digital Footprint, and Loan Purpose

- Interact loan purpose with digital footprint variables
  - examine whether digital footprint variables have greater discriminatory power in predicting defaults depending on the purpose of the loan.
  - Facebook login may capture the propensity of a consumer to engage in conspicuous consumption (Immorlica et al. (2017))
  - Default rates - higher for loans taken for the purpose of purchase by such customers?
- Base loan category in these tests is Medical loans
  - default rates are measured relative to the default rates for medical loans.

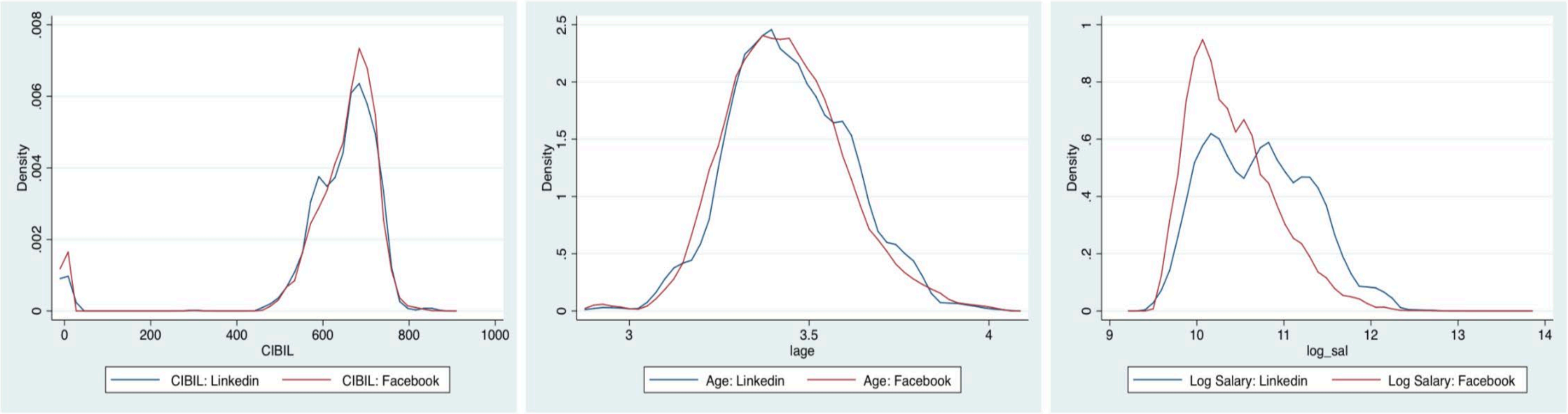
Defaults, Digital Footprint, and Loan Purpose

	(0.262)
Travel.purpose_cashe x Facebook status	-0.0304
	(0.083)
EMI.purpose_cashe x Facebook status	0.233***
	(0.079)
Purchase_purpose_cashe x Facebook status	0.186***
	(0.066)
LoanPayment.purpose_cashe x Facebook status	0.275***
	(0.079)
Other_purpose_cashe x Facebook status	-0.00117
	(0.052)
Travel.purpose_cashe x Linkedin status	-0.109
	(0.220)
EMI.purpose_cashe x Linkedin status	-0.106
	(0.204)
Purchase.purpose_cashe x Linkedin status	-0.536***
	(0.179)
LoanRepayment.purpose_cashe x Linkedin status	-0.339
	(0.217)
Other.purpose_cashe x Linkedin status	-0.291**
	(0.148)

Key Takeaways

- Customers who log in through Facebook are 20%, 26%, and 32% more likely to default when they take Purchase loans, EMI loans, and Repayment loans respectively.
- Customers who log in through Linkedin are 42% less likely to default when they take purchase loans.
- Are Facebook customers are on average of low creditworthiness?
  - plot the kernel density distribution of of CIBIL score, Salary, and Age for customers that login through Facebook and Linkedin.

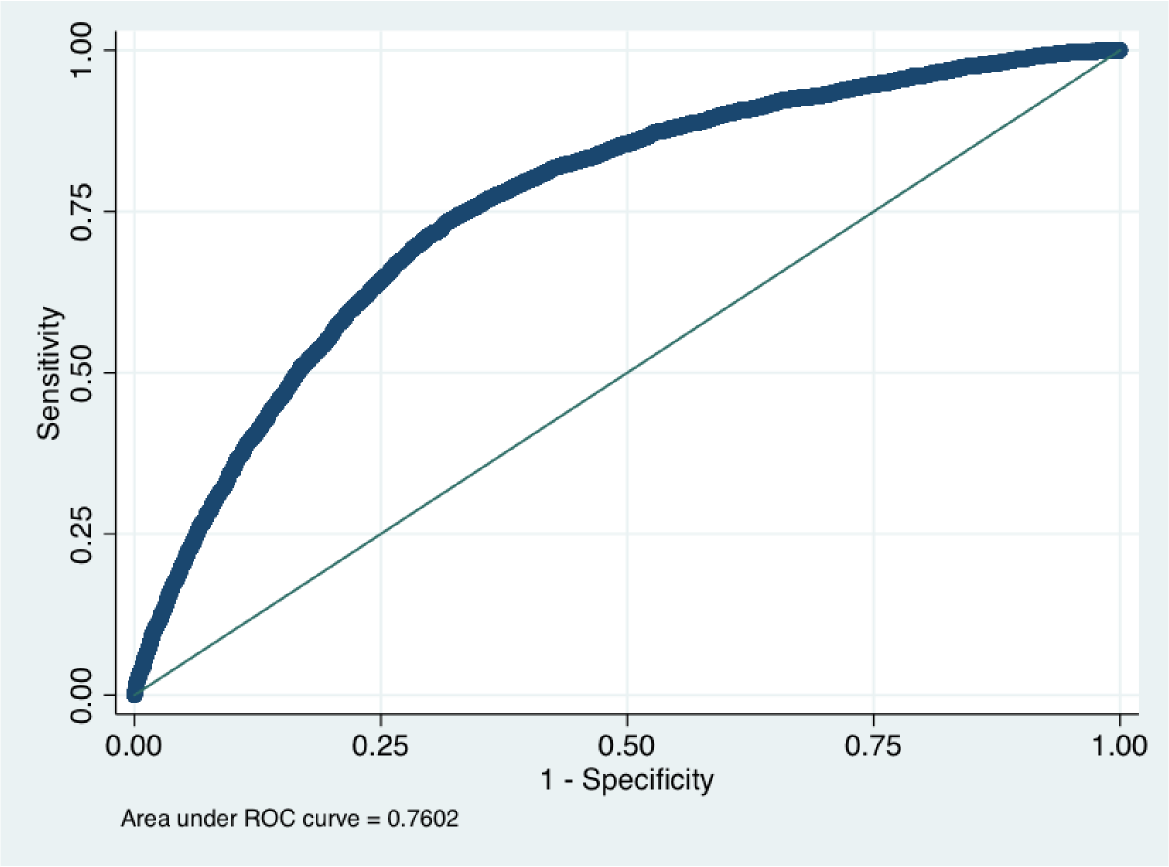
Defaults, Digital Footprint, and Loan Purpose



(1.1) CIBIL (1.2) Age (1.3) Log(Salary)

Figure 1: **Kernel Density Plots: Facebook vs Linkedin Login**
This figure plots the kernel density distribution of CIBIL score, Salary, and Age for customers that login through Facebook and Linkedin.

Out of Sample Test



Marginal Contribution

- Our work complements and builds on Berg, Burg, Gombovic and Puri (RFS forthcoming)
  - data covering approximately 250,000 purchases from an E-Commerce company located in Germany
  - digital footprint complements rather than substitutes for credit bureau information
  - informative even for customers who do not have credit bureau scores
- Our paper is similar in spirit to their work
  - While related, there are important differences
  - our paper further builds on and complements their findings

Marginal Contribution

1. Our data is from a stereotypical fintech lender operating in a developing country and covers all kinds of loans and not just those for e-commerce purchases.
   - allows us to extrapolate the importance of digital footprints in measuring creditworthiness for loans taken for different purposes and not just an e-commerce purchase.
2. Our data capture very different aspects of the digital footprint from the mobile phones of customers.
   - the large majority of customers in their sample access the digital world through desktop
   - important given that globally, about 50% of the users access the Internet through mobile phones, and 5% through tablets.
   - particularly true in a developing country setting. For instance, 80% of the Internet access time in India is through mobiles.
   - even in developed countries like the UK, USA, and Germany, the fraction of users that access the Internet primarily through mobile phones is increasing.
   - findings are potentially generalizable to other developing countries and the millennial generation.

Marginal Contribution

3. Because we have data on the salary, education, and job of the customers we can disentangle whether digital footprint simply proxies for these characteristics or provides incremental information.

   - For, instance we find that owning an IOS device has predictive power even after controlling for earnings.
   - Collecting additional data on overall savings and investments portfolios of customers

4. Our data allows us study a richer set of loan outcomes which includes the likelihood of loan approval .

- Working on loan duration, interest rate and loan purpose

Marginal Contribution

- Document that digital footprints can allow lenders to estimate the likelihood of default based on the end use of the loans.
- So, two customers with otherwise same credit scores and earnings may have a differing propensity to default for different kinds of loans.
  - For instance, compared to other customers, borrowers who log in via Facebook are less likely to default when they borrow to meet medical expenses but more likely to default when they borrow to make a purchase.
- Implication: The same customer can have different creditworthiness (and consequently credit score) conditional on the purpose of the loan and digital footprints

Ongoing Work

- Dynamic Call logs data
- Dynamic app installation data
- Customers’ overall financial portfolio
- Determinants of loan duration and purpose
- Extending the data to current time period

Conclusion

- We document statistically and economically significant role of individuals' digital footprint variables in the loan approval process.
  - In absence of sufficient credit history and credit scores for millennial customers to judge their credit worthiness, the fintech lender uses individuals digital footprint as an alternative credit screening process.
- Individual's digital footprint and preferred social media mode for log in have significant predictive power in predicting default.
  - Importantly, these variables have incremental predictive power over and above the CIBIL credit score.
- The discriminatory power of digital footprint variables varies conditional on the loan purpose
- Overall, our study has implications for expanding access to credit to those who don't have a credit history but who leave a large trace of unstructured information on their mobile phones that can be used to predict loan outcomes.