Fintech and Credit Scoring for the Millennial Generation

Sumit Agarwal [*]
Shashwat Alok[†]
Pulak Ghosh[‡]
Sudip Gupta[§]

March 2019
(Preliminary Draft)
Please do not cite or circulate

Abstract

Using a unique and proprietary loan-level data from a large Fintech firm in India, we analyze whether unstructured data regarding a consumer's digital mobile footprint such as borrower's preferred social media platform, the type of mobile phone applications, number of applications on the phone, type of operating software used by a loan applicant etc, can act as a substitute for traditional credit bureau scores. We find that the digital mobile footprint of an individual performs at least as good as the credit score in predicting loan approvals and defaults. Moreover borrower heterogeneity coupled with their preferred social media platform points to the role of conspicuous consumption in loan performance. 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.

[*]Email: ushakri@yahoo.com. National University of Singapore.
[†]Email: shashwat_alok@isb.edu. Indian School of Business.
[‡]Email: pulak.ghosh@iimb.ac.in. Indian Institute of Management, Bangalore.
[§]Email: sgupta24@fordham.edu. Fordham University.

1 Introduction

A recent survey in the US showed that almost half of the millennials in the US feels that their credit score is holding them back [1] . Younger people suffer from shorter credit history and hence often denied credit by traditional financial institutions or charged very high interest rates for loans to prohibit them access to credit. This in turn exacerbates the evaluation of their credit worthiness problem as they cannot build a longer credit history to have a decent credit score. Many such individuals may actually be 'good borrowers' if their 'credit worthiness' could be evaluated by alternate means.

The problem of lack of credit history for the millennials is a word-wide phenomenon. For example, according to a recent industry report, 156 million Indians who comprise the 'urban mass' and urban middle' section representing an annual income of USD 3000 and above have the potential of mass adoption of consumer credit. Of this the 'urban mass' constituting approximately 129 millions have been mostly deprived of credit, due to a lack of credit history.[2] This led to the quest for alternative data for credit scoring for the millennials.

While there are millions across the country who have never obtained a bank loan, they are Internet users who shop online, have a good social media presence, have a stable residential status and also have been using their mobile phones actively. These traces of unstructured data ("digital footprint", see **?**) that individuals leave through their online behavior and mobile phone usage can potentially be used to predict their loan behavior. Consistent with this idea, a plethora of fintech firms have mushroomed all around the world, that aim to service such customers by leveraging unstructured data and big data analytics to predict their default behavior. However, thus far there is limited evidence on whether or not "digital footprint" of an individual can substitute for traditional credit

[1] MarketWatch News Article [Accessed on 14th March, 2019].The survey looked into the credit knowledge and experience of 2,000 Americans ages 18 to 34, and found that many young adults are suffering the consequences of bad credit. In fact, 24 percent of those surveyed said they never learned how to build good credit in the first place, and 15 percent reported that their level of debt is unmanageable, with 1 in 5 admitting that they don't have control over their finances.

[2] Financial Expressed News Article [Accessed on 14th March, 2019]

bureau scores. The aim of this paper is to further the early work in this area.

Towards, this end, we use data from one of the largest Fintech lending firm in India to examine the discriminatory ability of digital footprint variables in predicting loan outcomes. Specifically, we want to understand whether and how the digital footprint are associated with loan level outcomes such as the likelihood of loan approval, the purpose of the loan, the duration of the loan, and the likelihood of default. More importantly, we want to understand whether these variables can be used to predict the likelihood of default for a borrower without any credit history and consequently a credit bureau score. Our goal is not to pin down the causal channels through which a customer's digital footprint may effect her creditworthiness, rather further along the lines of work of **?**, to analyze the association between the digital footprint and credit worthiness of individuals with lack of credit history.

We obtain the universe of loan applications made to one of the largest fintech lender named CASHe in India, between the period of October 2016 to January 2018. Unlike prior studies, we also have access to loan applications that were eventually denied allowing is to examine the determinants of loan approval. Out of the 199,000 loan applications in our sample, 111,956 were approved while 87,127 were denied. CASHe is a mobile-only lending platform targeted towards meeting the short-term credit needs of the salaried millennial. It grants loans ranging from a minimum of ₹10,000 to a maximum of ₹200,000 for 15, 30, 90, 120, and a maximum loan duration of 180 days.

To apply for a loan, an individual needs to submit regulation mandated identification and address documents, along with bank statement, salary slip. The potential borrower authorizes CASHe to use its digital footprint variables for evaluation of his credit worthiness and research. They also provide CASHe data on their CIBIL−Transunion credit score (if available), education, and job designation. Importantly for our study, CASHe also collects digital information from the individuals' mobile phone such as the mode of login (for example, Facebook and Linkedin), the various applications installed, number of calls, number of contacts on phone, number of social connections, and the kind of mobile operating system such as IOS and Android. We have anonymized access to detailed data on the kind of mobile applications that an individual uses that we club into 6 broad categories: *Sales apps* which includes applications for e-commerce such as Amazon, Flipkart, Snapdeal among others, *Social Network apps* such as Whatsapp, Twitter, Messenger services, *Financial Apps* such as Mobile banking and stock trading applications, *Travel apps* such as Airbnb, Tripadvisor, and MakeMyTrip, *Mloan app* which includes other mobile-based lending platforms (**Such as?**), and *Dating apps* such as Tinder. This kind of digital information on the number of social connections or kind of applications that a customer uses can potentially proxy for hard to quantify and unobservable aspects of individual behavior that is unavailable to traditional banks.

We begin by analyzing whether and how the loan characteristics, the customer characteristics, and the digital footprints relate to loan approval decisions. As one would expect, we find that a loan applicant with a higher credit score, salary, and education is more likely to get approved. Importantly, we find that the digital footprint of an individual is significantly related to her likelihood of approval. Interestingly, we find that the preferred social media platform to log in and apply for CASHe loan has a significant impact on the loan approval probability. Specifically, 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.

Focusing on other digital footprint variables, we find that the number of contacts, the number of applications (apps from now) installed, and Mloan app dummy is positively associated with loan approval. Similarly, installing a financial application is likely to be correlated with an individual's financial literacy. The discriminatory ability of digital footprint variables is robust to controlling for the credit bureau scores, customer's earnings, age, education, location as well as the duration and purpose of the loan. This suggests that digital footprint variables provide incremental information that is important for predicting loan outcomes beyond what is captured in the credit score.

Next, we examine the ability of digital footprint variables in predicting defaults. Here, we rely on both the economic and statistical significance of individual explanatory variables as well as Area Under the Curve (AUC) - an easy and commonly used measure of the predictive power of credit scores (**?**, **?**). We first note that the AUC of the model using only the credit score for predicting defaults is 58%. The AUC of credit score in our sample, while significantly different from chance (AUC of 50%) is lower than 62% reported by **?** based on a sample of loans from peer to peer lending platform, "Propser.com" and 68.3% reported by **?** based on a sample of purchases from a German e-retailer. This suggests that the discriminatory ability of the credit score in predicting defaults is likely to vary across geographies and intermediaries. To the extent that digital footprint variables complement the information content of credit score, the marginal value of such information is higher in contexts where the credit score itself has lower discriminatory power. Thus, fintech firms that rely on the digital footprint for screening borrowers maybe even more important to expand credit access in countries with weak information environment and lower levels of financial inclusion.

The AUC of a model that relies exclusively on digital footprint to predict defaults at 55% is comparable to and lies in the confidence interval of the AUC of the model using only the credit score. Further, as compared to the credit score, the digital footprint variables taken together are able to explain 2 percentage points higher variation in defaults. Focusing on the individual variables, our results suggest that digital footprint variables may be capturing hard to quantify aspects of individuals' behavior which has implications for the likelihood of default. For instance, customers without a financial application installed on a phone are about one and a half times more likely to default relative to those who have such an application installed. Similarly, customers with some other mobile loan application (Mloan dummy) are about 9% less likely to default. This is consistent with the idea that installing financial applications may proxy for the financial sophistication of a customer. In contrast, those with a dating application (any other social network app) are 15% (13%) more likely to default. Interestingly, customers with a travel application are about 19% less likely to default.

Consistent, with the evidence in **?**, we find that owning an Apple device is significantly and negatively associated with the likelihood of default. Specifically, those with an IOS phone are half as likely to default as compared to those with an Android phone. These results hold after controlling for customer's salary, age, and education. In this respect, our finding complements the evidence reported in **?**. Given that they do not have information on earnings or education of the customer, they are unable to disentangle whether owning an Apple phone simply proxies for potentially quantifiable financial characteristics of an individual or some unobservable aspect of individuals' behavior which matters for default prediction. This is important because if digital footprint simply proxies for easily measurable financial characteristics, then fintech lending firms should directly collect data on those characteristics rather than trying to infer it from the digital footprint variables. Indeed such digital information holds more promise if it captures some soft information which would be otherwise difficult to measure using financial characteristics. In such a case, first, digital footprints can be used to improve traditional credit scoring models.

Second, we can use digital information to build credit scoring models for and make loans to individuals without credit or financial history, thereby expanding credit access. Our results suggest that owning an Apple device captures an unobservable aspect of individuals which is not fully absorbed by earnings, education, or credit score. 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 equal to the model which includes the credit score combined with customer and loan characteristics. In other words, a predictive model which includes loan characteristics, customer characteristics, and digital footprint performs as well in predicting defaults as a model which includes credit bureau score, loan characteristics, and customer characteristics. Overall, these findings suggest that digital footprint variables complement credit bureau score and observable customer characteristics.

Finally, a unique aspect of our paper is that we can examine whether the discriminatory power of digital footprint variables varies based on the purpose for which a loan is taken. For instance, we find a significant predictive power of the individuals' preferred mode of log in (Facebook vs. Linkedin) in predicting default. If Facebook and more broadly social network users have a higher propensity to engage in costly status-seeking conspicuous consumption (**?**) beyond one's financial means, then we should expect the default rates to be higher amongst such customer especially when they take a loan for making a purchase. Consistent with this conjecture, we find that as compared to customers who do not log in through Facebook or Linkedin, 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. Interestingly, in contrast to what we find for Facebook login, customers who log in through login through Linkedin are 42% less likely to default when they take loans for making a purchase.

Our results suggest that there is significant variation in the discriminatory power of digital footprint variables in predicting defaults depending on the purpose for which a loan is taken. More specifically, the default likelihood and consequently the creditworthiness of a customer estimated using digital footprints can vary depending on the end use of the loan. For instance, on average, as compared to an individual who did not log in through Facebook or Linkedin, a borrower who logged in via Facebook is less likely to default when they borrow to meet medical expenses but more likely to default when they borrow to make a purchase. This finding to the best of our knowledge is novel to the literature.

Overall, our study documents that digital footprint variables have significant discriminatory power in both loan approvals and default prediction. Importantly, with the use of big data, fintech lenders can potentially build credit scores and can expand access to credit to even customers with little or no credit history that are underserved by the traditional banks. Consistent with this conjecture, the average individual in our sample is a sub-prime borrower with a credit score of 641.[3] Moreover, an economically significant 5% of borrowers in our sample do not have a credit score. This is in contrast to the USA, where fintech lenders primarily cater to borrowers who already have access to credit via traditional banks (**?**, **?**).

[3]The credit scores and associated risk tiers in India are: 801–900 (*Prime plus*), 751-800 (*Prime*), 651–750 ((*Near prime*)), and 300–650 (*Subprime*)

The paper closest to our study is **?**. Using data covering approximately 250,000 purchases from an E-Commerce company located in Germany, **?** document that the digital footprint complements rather than substitutes for credit bureau information, and is informative even for customers who do not have credit bureau scores. While related, our paper further builds on and complements their findings. First, 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.

Second, the large majority of customers in their sample access the digital world through desktop, while our data capture very different aspects of the digital footprint from the mobile phones of customers. This is important given that globally, about 50% of the users access the Internet through mobile phones, and 5% through tablets. This is particularly true in a developing country setting. For instance, 80% of the Internet access time in India is through mobiles.[4] Moreover, even in developed countries like the UK, USA, and Germany, the fraction of users that access the Internet primarily through mobile phones is increasing. Thus, given the mobile-based digital footprints and the developing country setting, our findings are potentially generalizable to other developing countries and the millennial generation.

Third, 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. Fourth, given the nature of our data, we study a richer set of loan outcomes which includes the likelihood of approval. This allows us to document whether and how lenders use digital footprints in their loan approval decisions. Moreover, our setting 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.

Finally, we document that digital footprints can allow lenders to estimate the likeli-

[4] **Add similar stats for other countries.**

hood 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 access 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. The implication is that with the use of big data on digital footprint, the same customer can have different creditworthiness (and consequently credit score) conditional on the purpose of the loan.

2 Data and Summary Statistics

We obtain proprietary data on about 1,99,000 loan applicants from a mobile-only Fintech lending platform named CASHe operating in India since 2016. CASHe aims to provide short-term credit to young salaried professionals by using their mobile, digital footprints, and social behavior to determine their creditworthiness even when a credit history may not be available. CASHe provides loans of amount ranging from a minimum of ₹10,000 ($142) to ₹200,000 ($2846).[5] The loan duration ranges from a minimum of 15 days to a maximum of 180 days. Currently, they have 180,000 active customers with 75% repeat users. A total of ₹6500 million ($92 million) worth of loans have been disbursed since its inception in 2016. To get a loan, a customer has to download the CASHe app, enters all the requisite details, documentation and submit. The application download allows CASHe to gather additional information, on the mode of login, the various apps installed, number of calls and SMSs, number of contacts on phone, number of social connections, and the kind of mobile operating system such as IOS and Android. We obtained data from CASHe for all loans granted during the period October 2016 to January 2018.

[5] Based on the nominal exchange rate of $1=₹70.28 as of March 2019.

2.1 Summary Statistics: Loan and Financial Variables

Table 4 reports the summary statistics. Out of the 199,000 loan applications in our sample, 111,956 were approved while 87,127 were denied. The default rate in our sample is quite high at approximately 20%[6]. This is significantly higher than the delinquency rate of 3% for retail loans given out by all banks across India.[7] This suggests that CASHe caters to higher risk customers. The average loan size is ₹23,078 ($328) age of a customer is 31 consistent with the idea that CASHe's target segment is a young salaried customer.[8] The average credit score is 640 and is obtained from TransUnion CIBIL. The average interest rate charged on a loan is 25% (log value of 1.4). On average, a customer earns ₹37,505 ($534) per month or $6408 per annum. Thus, the income of a customer is our sample is roughly 3 times the median per capita income of $2,134 in 2018. Thus, CASHe caters to relatively to higher income customers. CASHe also records the purpose for which loan is taken which can be of the following: Medical, Travel, EMI, Purchases, Loan Repayment, Others. Amongst the sample of approved loans, 9% were taken for the purpose of travel, 9% for EMI, 15 % for purchasing a good, about 7% for the purpose of repaying a loan and the rest 21% for medical expenditure.

2.2 Summary Statistics: Digital Footprint Variables

In addition to the credit bureau score, and other customer level variables, CASHe also captures digital footprint data on the various kinds of mobile applications installed on the user's phone: such as Facebook, Linkedin, financial apps, dating apps, e-commerce apps, and travel apps. CASHe also collects data on other variables that may capture the social behavior and status of the customer such as the number of calls, the number of SMSs, the number of contacts on the phone, the number of social media connections, and the kind of mobile operating system such as IOS and Android. Facebook (Linkedin) dummy variables identify customers that logged in to CASHe using Facebook (Linkedin).

[6] $\frac{\text{23417 defaults}}{\text{111,956 approvals}}$

[7] The default rate for all retail loans disbursed by banks was obtained from RBI bulletin.

[8] The average loan amount of $328 is comparable to the average purchase amount of $350 in ?.

About 29% of customers logged in to CASHe using Facebook, while 2.7% used Linkedin. On average, 85% of the customers have a banking or stock trading app. About 51% of customers have installed another mobile-loan application suggesting that they look for loans on other platforms as well, while 11% of the customers own an apple phone (ios dummy).

3 Results

3.1 Univariate Analysis

In columns 1-3 of Table 4, we compare the customer and loan characteristics of loans that were approved and those that were denied. Surprisingly, the average size of the loan demanded is about 51% higher for loan applications that were approved.[9] Consistent with conventional wisdom, we also find that customers with a higher salary, credit score, and older customers have a higher likelihood of approval. Focusing on the digital footprint variables, we find that, approved customers are more likely to log in through either Facebook or Linkedin. Approved customers are also more likely to have installed a financial app (Banking apps, Mutual Fund apps, and stock tracking apps), social networking app (Facebook, Twitter, Whatsapp, and other chat apps). Whether or not the customer installs a dating app or an e-commerce application (such as Amazon and Flipkart captured in the *Sales* dummy) does not seem to be associated with the likelihood of loan approval. Customers that have either been referred by others (*Referral* dummy) and those who have referred others (*Referrer* dummy) are also more likely to be approved. On average, approved customers have a higher number of apps, send and receive a greater number of SMSs and calls, have a higher number of contacts but fewer connections on a social platform. Approved customers are also 5% more likely to own an Iphone (*IOS opsys* dummy). However, the difference between the fraction of customers who own an Iphone across the sample of approved and not approved customers is statistically indistinguishable from

[9] $\frac{(23078.4.-15254.4)*100}{21154.2}$

zero.

In columns 1-3 of Table 4, we analyze the customer and loan characteristics that can potentially predict the likelihood of default. Customers who default on average borrow 44% more than those who don't.[10]. While customers who default on average are charged a higher interest rate, the difference in the interest rate between loans in default and loans that are current is not statistically significant. Surprisingly, customers who default one average are older and have a greater salary as compared to customers that have not defaulted. Not surprisingly, customers who default have lower credit scores.

Focusing on the digital footprint variables, we find that customers who default are less likely to have logged in through either Facebook or Linkedin. This suggests that the mode of login has predictive power for the likelihood of default. We don't find that the kind of apps installed (*sales, dating, finsavvy, social conect, travel, and Mloan*) has any discriminatory power for predicting defaults. We do find that digital footprint variables that capture various aspects of social behavior have a bearing on the likelihood of default. For instance, customers those were referred by others, and those who refer others are less likely to default. This is consistent with the marketing and economics literature that finds that customers or employees acquired through referrals have a stronger sense of commitment and attachment to the firm (**?**, **?**). Using data on referred customers of a German bank, **?** find that such customers have a higher retention rate and are more valuable in both the long and short term. Along similar lines, **?** find that referred workers yield substantially higher profits per worker than non-referred workers. To the extent that the likelihood of referring or being referred is associated with the strength of an individual's social connections, our finding suggests that social ties may have positive spillover effects on the customer's attitude towards default. Consistent with this idea that customers who do not default send and receive a greater number of SMSs and calls have a higher number of contacts but fewer connections on a social platform. These variables again potentially capture the strength of social ties of a customer. The number of apps also seem to have a discriminatory ability to predict defaults as defaulting customers have

[10] $\frac{(30376.7-21154.2)*100}{21154.2}$

fewer apps. In our univariate analysis, we do not find any significant difference in the likelihood of owning an Iphone between across the set of defaulter and non-defaulters.

3.2 Multivariate Analysis

We now move on to the discussion of our multivariate analysis. Formally, we run a logit or multinomial logit regressions of loan outcome measures on loan and customer characteristics:

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

Where i identifies a unique customer, l identifies a unique loan, and t refers to a year-month. The *Loan outcome* is one of the following: *Approved* is a dummy variable which takes the value one for loans that were approved and zero otherwise, *Loan purpose* identifies one of the six different kinds of loan, Medical, Travel, EMI, Loan repayment, purchase and other purpose, *Loan duration* a dummy variable that identifies loans of 15, 60, 90, 120, and 180 days duration, and *Default* which identifies loans in default. Loan Characteristics refer to loan size, and loan purpose. Customer financial refers to customer age, salary, education, and job designation. Customer digital footprint refers to all the variables summarized and discussed in the previous section.

3.2.1 Loan Approvals

We begin our multivariate analysis by examining the determinants of loan application approval. The dependent variable in these tests is a dummy variable which takes the value one for loans that were approved and zero otherwise. Table 2 reports the results of our analysis. Column (1) reports the results using only the credit bureau score (*CIBIL*) as the explanatory variable. Not surprisingly, loan applicants with higher credit score have a higher likelihood of getting approved. The R^2 of the regressions is 0.039 implying that credit scores explain about 4% of the variation in the likelihood of loan approval.

In column (2) we repeat these tests after including other loan and customer characteristics. We find that customers that earn more and more educated have a higher chance of approval. We also include loan purpose dummies in these tests where a medical loan is the base category. We find that medical loans have a higher likelihood of approval as compared to loans taken for the purpose of EMI, purchase, repayment and uncategorized loans. Interestingly, travel loans have the highest likelihood of approval and younger customers are more likely to be approved.

In column (3), we report the results for digital footprint variables. Since IOS dummy has significant predictive power for loan outcomes (see ?), to make sure that our results are not just driven by IOS variable, we don't include it in column (3). Consistent with our univariate analysis, we find that customers who log in through either Facebook or Linkedin have a significantly higher chance of approval. The odds ratio indicates that those who login through Facebook (Linkedin) are two (four) times more likely to get approved relative to those who login by other means.

Focusing on other digital foot print variables, we find that 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. These results continue to hold when we include IOS dummy in column (4). We find that customers with an IOS device have a lower likelihood of approval. This is surprising given that prior studies highlight that owning an IOS device is a strong predictor of higher earnings (?). One possibility, is that owning an IOS device may also proxy for status-seeking individuals or those with a propensity for conspicuous consumption. Over all results indicate that digital footprint variables have significant explanatory power for the likelihood loan approval.

Column (5), includes all loan characteristics, customer characteristics, and digital footprint but excludes credit bureau score. Our objective here is two folds. First, we want to understand whether our results on digital footprint continue to hold once we control for other loan level and customer level characteristics. For instance, some of the variables such as owing an IOS device may simply be a proxy for the income of the customer and thus may not have any independent explanatory power over the customer's salary. Second, 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 just the CIBIL score. We find that customer's salary, number of contacts, number of apps installed, and Linkedin login dummy is no longer statistically significant. However, focusing on the R^2, we find that taken together these variables explain 8% (twice that of CIBIL score) of the variation in loan approvals.

Finally, in column (6) we also include CIBIL score and state fixed effects. The results remain qualitative similar. The R^2 increases to 9.6% suggesting that loan characteristics, customer characteristics, and digital footprint, have some complementary information beyond what is captured in the CIBIL score. Summarizing, the key takeaway from this section for the purpose of our study is that digital footprint variables have significant explanatory power for loan approval decisions even in the absence of a credit bureau score.

3.2.2 Defaults

In this section, we focus on analyzing the relationship between digital foot print variables, loan, and customer characteristics and default. The dependent variable in these tests is a dummy variable which takes the value one for loans that are delinquent. Table **??** reports the results from these tests. Column (1) reports the results using only the credit bureau score (*CIBIL*) as the explanatory variable. Not surprisingly, a higher credit bureau score is associated with a significantly lower likelihood of default. The AUC of the CIBIL score in our sample is 58%. The AUC of credit score in our sample, while significantly different from chance (AUC of 50%) is lower than 62% reported by **?** based on a sample of loans from peer to peer lending platform, "Propser.com" and 68.3% reported by **?** based on a sample of purchases from a German e-retailer. This suggests that the discriminatory ability of the credit score in predicting defaults is likely to vary across countries and type of financial intermediary. Focusing on pseudo-R^2, we find that the credit score explains about 7% of the variation in loan defaults.

In column (2), we include other customer and loan level characteristics excluding digital footprint variables. The increase in AUC and R^2 suggesting that these characteristics have incremental information for predicting default beyond what is captured by the CIBIL score alone. Focusing on individual explanatory variables, we find that salary, and education is negatively related to defaults. Interestingly, we also find that default likelihood is lower for all categories of loans (Travel, EMI, Purchase, Repayment, and other) relative to loans taken for medical needs. This is consistent with the idea that health shocks are correlated with financial distress (**?**). Thus, the likelihood of default is higher for customers taking loans to meet medical expenditure as compared to loans taken for leisure/consumption purposes.

In column (3), we report the results for digital footprint variables. Since IOS dummy has significant predictive power for loan outcomes (see **?**), to make sure that our results are not just driven by IOS variable, we don't include it in column (3). The AUC of this specification is 55% and lies in the confidence interval of the AUC estimate using just the credit bureau score. The pseudo-R^2 is about 9%, implying that digital footprint variables explain about 2% additional variation in loan defaults as compared to just the credit bureau score.

Focusing on the individual variables, we find that digital footprint variables may proxy for hard to quantify aspect of individual behavior which has implications for the likelihood of default. We find that individuals that have a financial app installed on their phone have a significantly lower likelihood of default. The odds ratio of *Finsavvy* dummy is 0.61 implying that 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. Similarly, customers with some other mobile loan application (*Mloan* dummy) are about 9% less likely to default.[11] This suggests that *Finsavvy* dummy, *Mloan* dummy may be correlated with the financial sophistication of a customer. In contrast, those with a dating app (any other social network app) are 15% (13%) more likely to default.[12]. Interestingly, customers with a travel app are about 19% less likely to default. As mentioned before, it is difficult to pin down the channel through which these variables may be affecting the likelihood of default. However, to the extent that the objective in a credit scoring exercise is to increase the precision of predicting default, these results indicate that the nature of apps installed on the phone have significant discriminatory power in default prediction.

In column (4), we also include the IOS dummy. The statistical and economic significance of other digital footprint variables remains qualitatively similar. In line with the evidence in **?**, we find that borrowers with IOS operating system (Apple) are significantly less likely to default relative to the Android operating system. The odds ratio of *IOS* dummy is 0.484 implying that those with an android phone are twice as likely to default as those with an Apple phone.

As before, in column (5), includes all loan characteristics, customer characteristics, and digital footprint but excludes credit bureau score. We find that the coefficients of the digital footprint variables change somewhat suggesting that these variables are correlated with some loan and customer characteristics. Interestingly, the coefficient estimate of *IOS* dummy remains statistically significant even after controlling for the customer's monthly salary. Our study complements **?** who conjecture that discriminatory ability of owning an apple device is presumably driven by its correlation with earnings. Specifically, our finding implies that owing an Apple device captures an unobservable aspect of individuals which is not fully absorbed by earnings.

Interestingly, we also find that, customers who log in through Facebook are more likely to default once we control for loan and customer characteristics. Importantly, the AUC of this specification is 76% , 18 percentage points higher than the AUC of the model using

[11]The odds ratio of *Mloan* dummy is 0.911.

[12]The odds ratio of *Dating* dummy is 1.15 and that of *Socialconnect* dummy is 1.13

only the credit bureau score and equal to the model which includes CIBIL score combined with customer and loan characteristics. This suggests that digital footprint variables not only complements credit bureau score but also that a predictive model which includes loan characteristics, customer characteristics, and digital footprint performs as well in predicting defaults as a model which includes credit bureau score, loan characteristics, and customer characteristics.

Finally, in column (6) we also include CIBIL score and state fixed effects for robustness. The results remain qualitative similar.

Overall, we document that digital footprint variables can be used to predict the likelihood of default and can perform at least as well as the credit score. Our findings have implications for expanding credit access to those without a credit history and consequently a credit score so long as we can capture enough aspects of their digital footprint.

3.2.3 Defaults, Digital Footprint, and Loan Purpose

Table 4 reports the results from our tests examining the impact of the loan purpose on the probability of default. Note that the base loan category in these tests is *Medical loans*, so the default rates are measured relative to the default rates for medical loans. In these tests, we interact loan purpose with digital footprint variables to examine whether digital footprint variables have greater discriminatory power in predicting defaults depending on the purpose of the loan. For instance, if Facebook login also captures the propensity of a consumer to engage in conspicuous consumption (**?**), then we should expect the default rates to be higher for loans taken for the purpose of purchase (*Purchase loans*) if the customer taking the loan logged in through Facebook. Consistent with this conjecture, we find that the likelihood of default is significantly higher for customers who log in to CASHe through Facebook when they take loans for the purpose of purchase, EMI payments (*EMI* loans), or loan repayment (*Repayment* loans). Specifically, as compared to customers who do not log in through Facebook or Linkedin, 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.

In contrast, logging in through Linkedin has incremental discriminatory power only for purchase and uncategorized loans (*Other loans*). Interestingly, in contrast to what we find for Facebook login, customers who log in through login through Linkedin are significantly less likely to default when they take a *Purchase loan.* Specifically, as compared to customers who do not log in through Facebook or Linkedin, customers who log in through Linkedin are 42% less likely to default when they take purchase loans. Linkedin login customers are also 25% less likely to default when they borrow for uncategorized purpose.

One possibility for difference in default rates between customers that login through Facebook and Linkedin could be driven by self-selection if say customers that login through Facebook are on average of low creditworthiness. To examine this possibility in figure 1 we plot the kernel density distribution of of CIBIL score, Salary, and Age for customers that login through Facebook and Linkedin. We note from figure 1.1 that the distribution of CIBIL score is similar for both types of customers suggesting that the difference in default rates is unlikely to be driven by creditworthiness. Focusing on figures 1.2 and 1.3, we find that again the age profile of customers is also similar, while the salary of Linkedin customers is slightly higher. We conclude that the propensity of login through Facebook and Linkedin captures an unobservable aspect of individual behavior that is correlated with default but not absorbed by either credit score, salary, or age.

In summary, the key takeaway from this section is that with the use of big data on digital footprints, the credit score/default prediction should be a function of the loan purpose as well. So lenders should use digital footprint data and base their loan decisions conditional on loan purpose. In other words, the default likelihood and consequently the credit score for a customer can vary depending on the mode of login and the purpose of the loan.

4 Conclusion

In this paper we have used an unique and proprietary dataset to analyze the impact of digital footprint and preferred social media of individual borrowers in predicting loan outcome. Our dataset come from a leading fintech lending company in India. We find that the digital footprint and social media preference for log in has significantly more predictive power than traditional credit score used by banks.

We found a number of interesting results. First, 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. This is consistent with the wide use of social media based credit scoring recently adopted by fintech companies wordwide.

We also find that individual's digital media footprint and preferred social media mode for log in have significant predictive power in predicting default. Specifically, Individuals who log in via friendship network like Facebook has significantly higher probability of default compared to individuals who prefer to log-in via professional network like Linkedin. More importantly, these variables have significantly higher predictive power than traditional credit rating variables like Transunion based CIBIL credit score. Exploiting borrower heterogeneity, their mobile phone app usage and purpose of the loan information, we find supporting evidence in favor of our conjecture that conspicuous consumption motive for borrowing is the main driver of this result.

Overall, our paper underscores the importance of individuals' digital footprint, everyday behavior and social media preference in predicting consumer loan approval and default prediction. These have wider policy implications as we design new modes of financial intermediation, services and their regulations in the era of 'big data'.

Table 1: **Summary Statistics**

This table reports summary statistics on the customer and loan characteristics. Columns 1-3 compares these characteristics for loan applications that were approved and those that were denied. Columns 4-6 compares these characteristics for approved and disbursed loans that were in default and those that were not in default. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

	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	

Table 2: **Approval of Loans**

This table reports the estimates from our logit regressions examining the determinants of loan approval. The dependent variable, *Approved* takes the value one for loan applications that were approved and zero for those that were denied. The specification in column (1) only includes the credit bureau score (CIBIL) as the explanatory variable. Column (2) also includes other loan and customer characteristics excluding digital footprint. Column (3) includes only digital footprint variables excluding IOS dummy. Column (4) includes only digital footprint variables along with IOS dummy. Column (5) includes all loan and customer characteristics and digital footprint variables but not the CIBIL score. Column (6) includes all variables including the CIBIL score and state fixed effects. Standard errors are clustered at the state level. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

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

Table 3: **Loan Defaults**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, and customer characteristics and likelihood of default. The dependent variable, *Default* takes the value one for loans that are delinquent and zero otherwise. The specification in column (1) only includes the credit bureau score (CIBIL) as the explanatory variable. Column (2) also includes other loan and customer characteristics excluding digital footprint. Column (3) includes only digital footprint variables excluding IOS dummy. Column (4) includes only digital footprint variables along with IOS dummy. Column (5) includes all loan and customer characteristics and digital footprint variables but not the CIBIL score. Column (6) includes all variables including the CIBIL score and state fixed effects. Standard errors are clustered at the state level. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

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

Table 4: **Loan Purpose, Digital Footprint, and default**
This table reports the estimates from our logit regressions examining the relationship between digital footprint variables loan, and customer characteristics and likelihood of default. The dependent variable, *Default* takes the value one for loans that are delinquent and zero otherwise. In these tests, we interact loan purpose with CIBIL score and digital footprint variables to examine whether the discriminatory ability of these variables varies with the purpose of the loan. We include digital footprint variables loan, and customer characteristics, and state fixed effects in this test. Standard errors are clustered at the state level. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Logit coeff (1)
CIBIL	-0.00143*** (0.000)
Facebook status	-0.013 (0.041)
Linkedin status	0.309*** (0.118)
Dating App	0.0147 (0.117)
Finsavy app	-0.591*** (0.067)
Socialconnect app	0.0347 (0.149)
Travel app	-0.332*** (0.043)
Mloan app	-0.210*** (0.038)
Travel.purpose cashe	-1.243*** (0.385)
EMI.purpose cashe	-0.565* (0.341)
Purchase.purpose cashe	0.00317 (0.303)
Loanrepayment.purpose cashe	0.893** (0.390)
Other.purpose cashe	-0.25 (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)
Travel_purpose_cashe x Dating App	-0.076 (0.182)
EMI_purpose_cashe x Dating App	0.0708 (0.186)
Purchase_purpose_cashe x Dating App	0.0655 (0.177)
LoanRepayment.purpose_cashe x Dating App	-0.14 (0.218)
Other.purpose_cashe x Dating App	0.266* (0.143)
Travel_purpose_cashe x Finsavy app	0.673*** (0.166)
EMI_purpose_cashe x Finsavy app	0.370** (0.160)
Purchase_purpose_cashe x Finsavy app	0.165 (0.118)

LoanRepayment.purpose_cashe x Finsavy app	0.383**
	(0.162)
Other.purpose_cashe x Finsavy app	0.292***
	(0.085)
Travel_purpose_cashe x Socialconnect app	0.494
	(0.375)
EMI_purpose_cashe x Socialconnect app	0.0563
	(0.330)
Purchase_purpose_cashe x Socialconnect app	-0.493*
	(0.262)
LoanRepayment.purpose_cashe x Socialconnect app	-0.779**
	(0.343)
Other.purpose_cashe x Socialconnect app	-0.0184
	(0.183)
EMI_purpose_cashe x Travel app	0.0759
	(0.082)
Purchase_purpose_cashe x Travel app	0.0421
	(0.068)
LoanRepayment.purpose_cashe x Travel app	0.237***
	(0.088)
Other.purpose_cashe x Travel app	-0.00576
	(0.053)
Travel_purpose_cashe x Mloan app	-0.0203
	(0.076)
EMI_purpose_cashe x Mloan app	0.221***
	(0.074)
Purchase_purpose_cashe x Mloan app	0.065
	(0.061)
LoanRepayment.purpose_cashe x Mloan app	0.139*
	(0.075)
Other.purpose_cashe x Mloan app	0.183***
	(0.048)
Travel.purpose cashe x CIBIL	-0.000135
	(0.000)
EMI.purpose cashe x CIBIL	-0.000306
	(0.000)
Purchase.purpose cashe x CIBIL	0.00033
	(0.000)
Loanrepayment.purpose cashe x CIBIL	-0.000912***
	(0.000)
Other.purpose cashe x CIBIL	0.000315**
	(0.000)
IOS Dummy	-0.765***
	(0.089)
Constant	4.201***
	(0.352)
Financial Variables	Y
Digital Variables	Y
Observations	98,577
Pseudo R-squared	0.159
AUC	0.7717

Table 5: **Default Heterogeneity by Credit Score**

The following table looks at the borrower heterogeneity based on credit rating on the impact of default with demonetization exercise in India (announced Nov 8th 2016) as a natural experiment to establish the causal link of borrower type to default probability. Demonetization was treated as a shock to the individual's budget constraint. The variable demon takes value 1 if it is after the demonetization period and 0 otherwise. The triple interaction term (demon x loan purpose x Facebook status) is aimed to capture the relative (causal) effect of conspicuous consumption on borrowers with Facebook status. Borrowers with Facebook association with very high (>700 CIBIL) and very low (<650 CIBIL) credit ratings are significantly more likely to default when they borrow to repayment of (previous) loans. Also borrowers with sales app are more likely to default for low credit rating individuals and borrowers with Dating_App app are more likely to default for high credit rating individuals. Borrowers with financial savy apps (stocks, banking, payments) are less likely to default for high credit rating individuals. Control variables include (salary, credit rating, loan amount, interest rate and state fixed effects etc.) Interestingly borrowers with IOS operating system (Apple) are significantly less likely to default relative to the Samsung operating system. Heteroskedastic robust standard errors are clustered at the state.

VARIABLES	Default Regressions: Low CreditRating (1)	Default Regressions: Medium CreditRating (2)	Default Regressions: High CreditRating (3)
Facebook status	-0.195*	0.112	-0.0282
	(0.102)	(0.107)	(0.126)
Linkedin status	-0.0118	-0.0307	0.995**
	(0.141)	(0.255)	(0.499)
Travel.purpose cashe	-0.408***	0.942	-0.0399
	(0.108)	(0.743)	(0.521)
EMI.purpose cashe	-0.460***	1.949	-0.117
	(0.105)	(1.396)	(0.966)
Purchase.purpose cashe	-0.596***	2.872	0.292
	(0.099)	(2.085)	(1.436)
Loanrepayment.purpose cashe	-0.427***	4.003	0.433
	(0.124)	(2.761)	(1.939)
Other.purpose cashe	-0.545***	5.161	0.93
	(0.119)	(3.464)	(2.396)
Travel.purpose_cashe x Facebook status	0.0403	-0.242	-0.0328
	(0.149)	(0.196)	(0.240)
EMI.purpose_cashe x Facebook status	0.217	0.139	0.286
	(0.202)	(0.240)	(0.260)
Purchase_purpose_cashe x Facebook status	0.208*	0.314*	-0.0852
	(0.126)	(0.167)	(0.239)
LoanPayment.purpose_cashe x Facebook status	0.380**	0.234	0.325
	(0.174)	(0.211)	(0.253)
Other_purpose_cashe x Facebook status	0.0996	-0.0734	0.0217
	(0.110)	(0.144)	(0.164)
Travel.purpose_cashe x Linkedin status	-0.231	0.145	-0.763
	(0.425)	(0.449)	(0.841)
EMI.purpose_cashe x Linkedin status	0.000917	0.366	-1.127
	(0.382)	(0.479)	(0.767)
Purchase.purpose_cashe x Linkedin status	-0.381	-0.133	-0.872
	(0.256)	(0.543)	(0.706)
LoanRepayment.purpose_cashe x Linkedin status	0.0787	0.162	-1.469**
	(0.361)	(0.509)	(0.600)
Other.purpose_cashe x Linkedin status	0.12	-0.291	-1.099**
	(0.219)	(0.274)	(0.453)
CIBIL	-0.00116***	-0.00283	0.00374
	(0.000)	(0.004)	(0.003)
IOS Dummy	-0.895***	-0.707***	-1.002**
	(0.235)	(0.177)	(0.399)
Constant	4.389***	6.539***	1.859
	(0.561)	(2.343)	(1.824)
Financial Variables	Y	Y	Y
Digital Variables	Y	Y	Y
State Fixed Effects	Y	Y	Y
Observations	39,591	33,944	26,976
Pseudo R-squared	0.165	0.156	0.159

Table 6: **Default Heterogeneity by Age**

The following table looks at the borrower heterogeneity based on age on the impact of default with demonetization exercise in India (announced Nov 8th 2016) as a natural experiment to establish the causal link of borrower type to default probability. Demonetization was treated as a shock to the individual's budget constraint. The variable demon takes value 1 if it is after the demonetization period and 0 otherwise. The triple interaction term (demon x loan purpose x Facebook status) is aimed to capture the relative (causal) effect of conspicuous consumption on borrowers with Facebook status. Older borrowers (>33.8 years of age) with Facebook association are more likely to default when they borrow for loan repayment or purchase, while borrowers with LinkedIn association are significantly less likely to default for the same purpose. Also borrowers with sales app are more likely to default for older individuals (>34 years). Borrowers with financial savy apps (stocks, banking, payments) are less likely to default for all age groups. Control variables include (salary, credit rating, loan amount, interest rate and state fixed effects etc.) Interestingly borrowers with IOS operating system (Apple) are significantly less likely to default relative to the Samsung operating system. Heteroskedastic robust standard errors are clustered at the state.

VARIABLES Age Heterogeneity	Default Regressions: Age 1Q (1)	Default Regressions: Age 3Q (2)
Facebook status	0.376***	-0.220*
	(0.140)	(0.128)
Linkedin status	0.713***	-0.0977
	(0.248)	(0.221)
Travel.purpose cashe	-0.114	-0.506***
	(0.148)	(0.141)
EMI.purpose cashe	-0.609***	-0.628***
	(0.193)	(0.140)
Purchase.purpose cashe	-0.499**	-0.460**
	(0.199)	(0.202)
Loanrepayment.purpose cashe	-0.506**	-0.309
	(0.233)	(0.260)
Other.purpose cashe	-0.534**	-0.2
	(0.224)	(0.271)
Travel.purpose_cashe x Facebook status	-0.789**	0.362
	(0.327)	(0.241)
EMI.purpose_cashe x Facebook status	-0.569*	0.753**
	(0.317)	(0.321)
Purchase_purpose_cashe x Facebook status	-0.315	0.295
	(0.274)	(0.184)
LoanPayment.purpose_cashe x Facebook status	-0.155	0.328*
	(0.300)	(0.197)
Other_purpose_cashe x Facebook status	-0.409**	0.0465
	(0.198)	(0.168)
Travel.purpose_cashe x Linkedin status	-1.666***	0.515
	(0.548)	(0.648)
EMI.purpose_cashe x Linkedin status	0.0231	0.667
	(0.572)	(0.752)
Purchase.purpose_cashe x Linkedin status	-0.990***	0.249
	(0.296)	(0.433)
LoanRepayment.purpose_cashe x Linkedin status	-0.501	-0.252
	(1.851)	(0.446)
Other.purpose_cashe x Linkedin status	-0.932**	0.105
	(0.363)	(0.269)
CIBIL	-0.00180***	-0.00128***
	(0.000)	(0.000)
IOS Dummy	-1.168***	-0.728**
	(0.251)	(0.299)
Constant	1.05	0.935
	(2.600)	(1.148)
Financial Variables	Y	Y
Digital Variables	Y	Y
Observations	16,546	27,983
Pseudo R-squared	0.148	0.174

Table 7: **Default Heterogeneity by Salary**

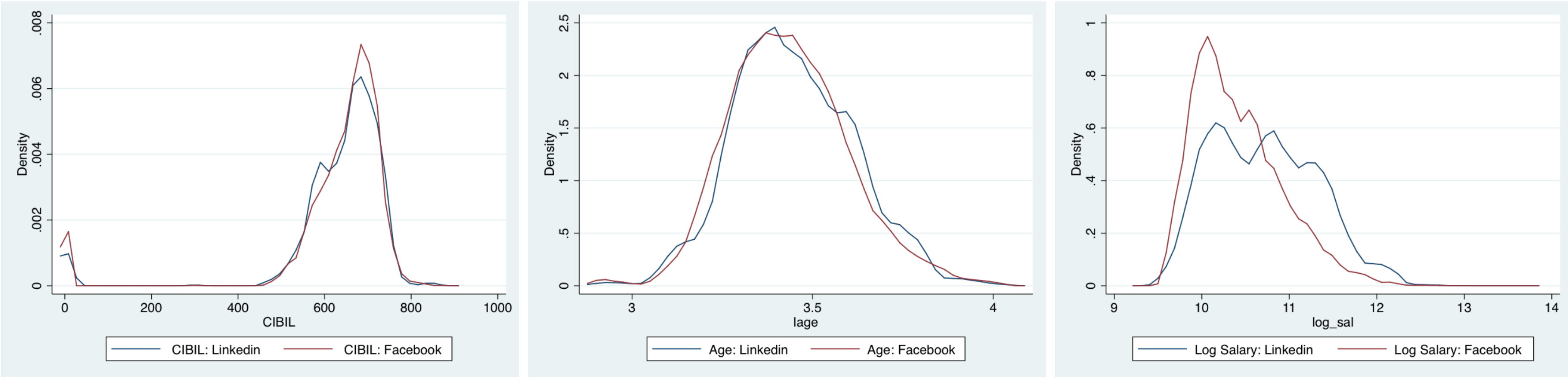
The following table looks at the borrower heterogeneity based on salary levels on the impact of default with demonetization exercise in India (announced Nov 8th 2016) as a natural experiment to establish the causal link of borrower type to default probability. Demonetization was treated as a shock to the individual's budget constraint. The variable demon takes value 1 if it is after the demonetization period and 0 otherwise. The triple interaction term (demon x loan purpose x Facebook status) is aimed to capture the relative (causal) effect of conspicuous consumption on borrowers with Facebook status. Both borrowers with high and low salary with Facebook association are more likely to default when they take a loan for conspicuous consumption, however borrowers with lower salary are significantly more likely to default if they have Facebook association than that of higher salary.Also borrowers with sales and Dating_App apps are more likely to default for low salary individuals (<20k per month). Borrowers with financial savy apps (stocks, banking, payments) are less likely to default for both salary groups. Control variables include (salary, credit rating, loan amount, interest rate and state fixed effects etc.) Interestingly borrowers with IOS operating system (Apple) are significantly less likely to default relative to the Samsung operating system. Heteroskedastic robust standard errors are clustered at the state.

VARIABLES Salary Heterogeneity	Default Regressions: Salary 1Q (1)	Default Regressions: Salary 3Q (2)
Facebook status	-0.0672	-0.183
	(0.145)	(0.115)
Linkedin status	-0.406	0.104
	(0.540)	(0.204)
Travel.purpose cashe	-0.245	-0.432***
	(0.177)	(0.134)
EMI.purpose cashe	-0.331*	-0.467***
	(0.174)	(0.154)
Purchase.purpose cashe	-0.617***	-0.352
	(0.196)	(0.230)
Loanrepayment.purpose cashe	-0.667***	-0.26
	(0.206)	(0.299)
Other.purpose cashe	-0.554**	-0.274
	(0.242)	(0.299)
Travel.purpose_cashe x Facebook status	-0.191	0.134
	(0.289)	(0.179)
EMI.purpose_cashe x Facebook status	-0.00178	0.765***
	(0.300)	(0.276)
Purchase_purpose_cashe x Facebook status	0.712***	0.25
	(0.275)	(0.185)
LoanPayment.purpose_cashe x Facebook status	0.640**	0.400**
	(0.272)	(0.193)
Other_purpose_cashe x Facebook status	0.0851	0.310*
	(0.205)	(0.166)
Travel.purpose_cashe x Linkedin status		0.00685
		(0.498)
EMI.purpose_cashe x Linkedin status	0.377	0.249
	(1.131)	(0.382)
Purchase.purpose_cashe x Linkedin status	0.225	-0.570*
	(0.720)	(0.312)
LoanRepayment.purpose_cashe x Linkedin status		-0.0162
		(0.340)
Other.purpose_cashe x Linkedin status	0.174	0.00166
	(0.615)	(0.260)
CIBIL	-0.00161***	-0.00140***
	(0.000)	(0.000)
IOS Dummy	-1.407***	-0.445***
	(0.349)	(0.138)
Constant	7.048**	1.704*
	(3.535)	(1.010)
Financial Variables	Y	Y
Digital Variables	Y	Y
Pseudo R-squared	0.0878	0.185

Table 8: **Default Heterogeneity by Designation**

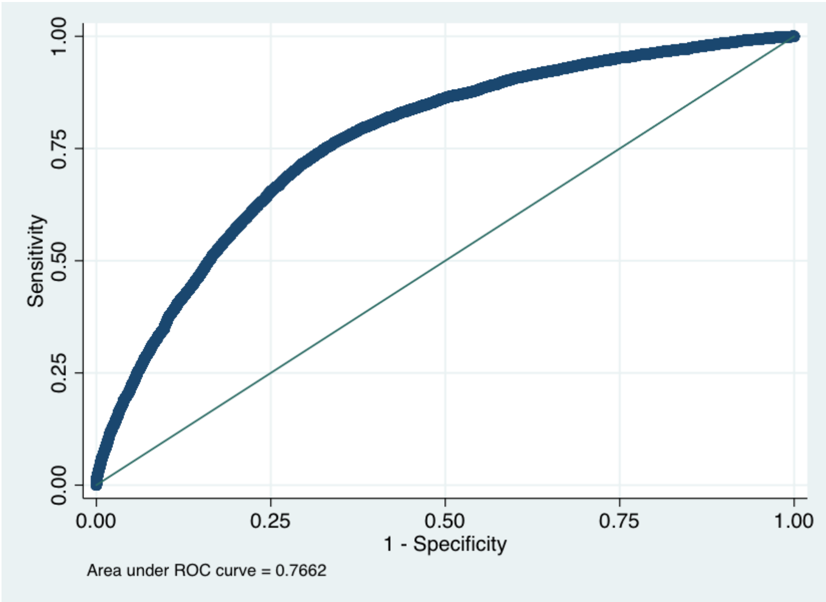
The following table looks at the borrower heterogeneity based on job descriptions (workers, supervisors, managers) on the impact of default with demonetization exercise in India (announced Nov 8th 2016) as a natural experiment to establish the causal link of borrower type to default probability. Demonetization was treated as a shock to the individual's budget constraint. The variable demon takes value 1 if it is after the demonetization period and 0 otherwise. The triple interaction term (demon x loan purpose x Facebook status) is aimed to capture the relative (causal) effect of conspicuous consumption on borrowers with Facebook status.Borrower with Facebook association with higher job descriptions (Managers, Supervisors) are more likely to default for conspicuous consumption related loans.Also borrowers with sales and Dating_App apps are more likely to default for workers. Borrowers with financial savy apps (stocks, banking, payments) are less likely to default for all job groups. Control variables include (salary, credit rating, loan amount, interest rate and state fixed effects etc.) Interestingly borrowers with IOS operating system (Apple) are significantly less likely to default relative to the Samsung operating system. Heteroskedastic robust standard errors are clustered at the state.

VARIABLES Designation Heterogeneity	Default Regressions: Worker (1)	Default Regressions: Supervisor (2)	Default Regressions: Manager (3)
Facebook status	-0.0489	0.0436	-0.0943
	(0.095)	(0.119)	(0.098)
Linkedin status	0.416**	0.585	-0.126
	(0.168)	(0.477)	(0.183)
Travel.purpose cashe	-0.430***	-0.0407	-0.339**
	(0.096)	(0.141)	(0.147)
EMI.purpose cashe	-0.586***	-0.407**	-0.172
	(0.096)	(0.175)	(0.150)
Purchase.purpose cashe	-0.736***	-0.419**	-0.262*
	(0.148)	(0.181)	(0.157)
Loanrepayment.purpose cashe	-0.565***	-0.234	-0.264
	(0.176)	(0.214)	(0.226)
Other.purpose cashe	-0.648***	-0.18	-0.219
	(0.162)	(0.236)	(0.220)
Travel.purpose_cashe x Facebook status	0.163	-0.452	-0.109
	(0.200)	(0.277)	(0.231)
EMI.purpose_cashe x Facebook status	0.0633	0.285	0.522*
	(0.159)	(0.268)	(0.286)
Purchase_purpose_cashe x Facebook status	0.259*	0.169	0.188
	(0.145)	(0.226)	(0.163)
LoanPayment.purpose_cashe x Facebook status	0.214	0.339**	0.402*
	(0.210)	(0.164)	(0.217)
Other_purpose_cashe x Facebook status	0.0272	0.0619	-0.00862
	(0.119)	(0.156)	(0.154)
Travel.purpose_cashe x Linkedin status	-0.989**	-0.915	0.429
	(0.489)	(0.864)	(0.507)
EMI.purpose_cashe x Linkedin status	-0.838*	0.616	0.145
	(0.446)	(0.741)	(0.318)
Purchase.purpose_cashe x Linkedin status	-0.875**	-0.92	0.21
	(0.362)	(0.595)	(0.338)
LoanRepayment.purpose_cashe x Linkedin status	-0.284	-0.689	0.011
	(0.451)	(0.549)	(0.322)
Other.purpose_cashe x Linkedin status	-0.577**	-0.752	0.27
	(0.227)	(0.650)	(0.257)
CIBIL	-0.00178***	-0.00135***	-0.00134***
	(0.000)	(0.000)	(0.000)
IOS Dummy	-0.752**	-1.010***	-0.826***
	(0.302)	(0.297)	(0.147)
Constant	5.227***	6.638***	3.648***
	(0.749)	(1.048)	(0.687)
Financial Variables	Y	Y	Y
Digital Variables	Y	Y	Y
Pseudo R-squared	0.162	0.179	0.154

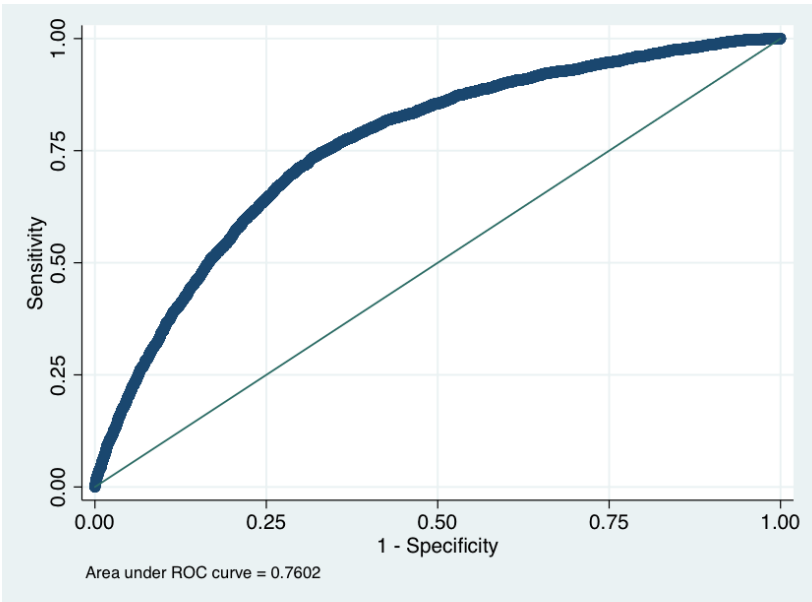


(1.1) CIBIL (1.2) Age (1.3) Log(Salary)

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.



(1.1)



(1.2)

Figure 2: **Out of Sample Tests**
Sudip: Can you add the description.

Internet Appendix A

Table A1: **Loan Duration - Panel A**

This table reports the estimates from our multionomial logit regressions examining the determinants of loan duration. The dependent variable, *Loan Duration* takes the value one for loans of 15-days duration (Columns 1-6 of panel A) and zero for 180-day loans. The specification in column (1) only includes the credit bureau score (CIBIL) as the explanatory variable. Column (2) also includes other loan and customer characteristics excluding digital footprint. Column (3) includes only digital footprint variables excluding IOS dummy. Column (4) includes only digital footprint variables along with IOS dummy. Column (5) includes all loan and customer characteristics and digital footprint variables but not the CIBIL score. Column (6) includes all variables including the CIBIL score and state fixed effects. Columns 7-12 are arranged similarly and the dependent variable, *Loan Duration* takes the value one for loans of 30-days duration and zero for 180-day loans. Standard errors are clustered at the state level. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Logit coeff 15Days (1)	Logit coeff 15Days (2)	Logit coeff 15Days (3)	Logit coeff 15Days (4)	Logit coeff 15Days (5)	Logit coeff 15Days (6)	Logit coeff 30Days (7)	Logit coeff 30Days (8)	Logit coeff 30Days (9)	Logit coeff 30Days (10)	Logit coeff 30Days (11)	Logit coeff 30Days (12)
CIBIL	0.00572***	0.00147***				0.00185***	0.00743***	0.00157***				0.00175***
	(0.000)	(0.000)				(0.000)	(0.000)	(0.000)				(0.000)
Log of Salary		14.72***			14.95***	14.90***		13.00***			13.10***	13.05***
		(0.303)			(0.355)	(0.359)		(0.301)			(0.352)	(0.356)
Log Loan Amount		-19.00***			-19.73***	-19.73***		-17.21***			-17.95***	-17.95***
		(0.319)			(0.375)	(0.378)		(0.318)			(0.373)	(0.377)
Log Age		11.11***			10.09***	9.947***		11.87***			10.72***	10.58***
		(0.298)			(0.314)	(0.313)		(0.294)			(0.310)	(0.308)
Education Dummy		-0.656***			-0.458***	-0.468***		-0.791***			-0.565***	-0.574***
		(0.126)			(0.138)	(0.138)		(0.124)			(0.136)	(0.136)
Travel.purpose cashe		0.546**			0.915***	0.875***		0.487**			0.857***	0.818***
		(0.220)			(0.248)	(0.249)		(0.217)			(0.246)	(0.247)
EMI.purpose cashe		0.561***			0.610***	0.581***		0.570***			0.636***	0.609***
		(0.201)			(0.213)	(0.213)		(0.199)			(0.211)	(0.211)
Purchase.purpose cashe		0.336**			0.619***	0.593***		0.348**			0.561***	0.536***
		(0.171)			(0.187)	(0.188)		(0.169)			(0.185)	(0.185)
Loanrepayment.purpose cashe		-0.0105			0.384*	0.337		-0.0593			0.29	0.244
		(0.199)			(0.223)	(0.223)		(0.195)			(0.219)	(0.219)
Other.purpose cashe		0.662***			0.853***	0.811***		0.640***			0.827***	0.787***
		(0.144)			(0.155)	(0.156)		(0.143)			(0.153)	(0.154)
Facebook status			0.103	0.104	0.0998	0.0948			-0.372***	-0.370***	-0.346***	-0.352***
			(0.114)	(0.114)	(0.128)	(0.128)			(0.113)	(0.113)	(0.126)	(0.126)
Linkedin status			-0.308	-0.309	-0.437	-0.418			-1.349***	-1.351***	-1.134***	-1.116***
			(0.226)	(0.226)	(0.278)	(0.278)			(0.225)	(0.225)	(0.274)	(0.274)
Log no of SMS			-0.0880***	-0.0876***	-0.0016	0.00135			-0.0891***	-0.0886***	0.00749	0.0104
			(0.032)	(0.032)	(0.034)	(0.034)			(0.032)	(0.032)	(0.034)	(0.034)
Log No of Contacts			0.242***	0.243***	0.333***	0.324***			0.260***	0.261***	0.392***	0.384***
			(0.043)	(0.043)	(0.058)	(0.058)			(0.042)	(0.042)	(0.057)	(0.057)
Log no of Apps			0.188**	0.186**	0.386***	0.363***			0.0686	0.067	0.379***	0.357***
			(0.083)	(0.083)	(0.103)	(0.104)			(0.082)	(0.082)	(0.102)	(0.103)
Log Calllog			0.218***	0.219***	0.399***	0.399***			0.245***	0.246***	0.394***	0.394***
			(0.038)	(0.038)	(0.052)	(0.052)			(0.037)	(0.037)	(0.051)	(0.052)
Dating App			0.389	0.392	0.674*	0.643*			0.276	0.279	0.683*	0.651*
			(0.343)	(0.344)	(0.373)	(0.373)			(0.342)	(0.342)	(0.370)	(0.369)
Finsavy app			0.268	0.276	0.148	0.144			0.368**	0.376**	0.322	0.318
			(0.178)	(0.177)	(0.237)	(0.239)			(0.176)	(0.175)	(0.234)	(0.236)
Socialconnect app			1.457***	1.443***	4.078***	4.099***			2.970***	2.955***	5.626***	5.654***
			(0.190)	(0.190)	(0.292)	(0.293)			(0.190)	(0.190)	(0.304)	(0.305)
Travel app			-0.815***	-0.814***	-0.879***	-0.893***			-1.046***	-1.045***	-0.926***	-0.942***
			(0.153)	(0.153)	(0.190)	(0.192)			(0.152)	(0.152)	(0.188)	(0.190)
Mloan app			-0.588***	-0.588***	-0.711***	-0.702***			-0.499***	-0.498***	-0.710***	-0.701***
			(0.116)	(0.116)	(0.131)	(0.132)			(0.115)	(0.115)	(0.130)	(0.131)
IOS Dummy				-0.218	0.0702	0.0777				-0.24	0.178	0.186
				-0.374	-0.424	-0.424				-0.371	-0.418	-0.418
Pseudo Rsq	0.0018	0.475	0.0098	0.0099	0.4734	0.4751	0.0018	0.475	0.0098	0.0099	0.4734	0.4751

Observations	113,013	113,006	98,260	98,260	98,253	98,244	113,013	113,006	98,260	98,260	98,253	98,244

Table A1: **Loan Duration - Panel B**

This table reports the estimates from our multionomial logit regressions examining the determinants of loan duration. The dependent variable, *Loan Duration* takes the value one for loans of 60-days duration (Columns 1-6 of panel B) and zero for 180-day loans. The specification in column (1) only includes the credit bureau score (CIBIL) as the explanatory variable. Column (2) also includes other loan and customer characteristics excluding digital footprint. Column (3) includes only digital footprint variables excluding IOS dummy. Column (4) includes only digital footprint variables along with IOS dummy. Column (5) includes all loan and customer characteristics and digital footprint variables but not the CIBIL score. Column (6) includes all variables including the CIBIL score and state fixed effects. Columns 7-12 are arranged similarly and the dependent variable, *Loan Duration* takes the value one for loans of 90-days duration and zero for 180-day loans. Standard errors are clustered at the state level. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Logit coeff 60Days (1)	Logit coeff 60Days (2)	Logit coeff 60Days (3)	Logit coeff 60Days (4)	Logit coeff 60Days (5)	Logit coeff 60Days (6)	Logit coeff 90Days (7)	Logit coeff 90Days (8)	Logit coeff 90Days (9)	Logit coeff 90Days (10)	Logit coeff 90Days (11)	Logit coeff 90Days (12)
CIBIL	0.00465***	0.00292***				0.00292***	0.00668***	0.000429**				0.000714***
	(0.000)	(0.000)				(0.000)	(0.000)	(0.000)				(0.000)
Log of Salary		11.57***			11.72***	11.63***		3.488***			3.501***	3.498***
		(0.301)			(0.352)	(0.356)		(0.271)			(0.317)	(0.321)
Log Loan Amount		-15.84***			-16.53***	-16.52***		-4.549***			-4.705***	-4.716***
		(0.317)			(0.372)	(0.376)		(0.279)			(0.327)	(0.332)
Log Age		11.64***			10.46***	10.25***		4.415***			4.002***	3.936***
		(0.298)			(0.315)	(0.314)		(0.267)			(0.277)	(0.276)
Education Dummy		-0.893***			-0.708***	-0.716***		-0.128			-0.00237	-0.00988
		(0.126)			(0.139)	(0.138)		(0.116)			(0.127)	(0.126)
Travel.purpose cashe		0.264			0.658***	0.599**		0.166			0.348	0.335
		(0.221)			(0.249)	(0.250)		(0.205)			(0.230)	(0.231)
EMI.purpose cashe		0.28			0.398*	0.355*		-0.297			-0.255	-0.262
		(0.202)			(0.214)	(0.214)		(0.188)			(0.195)	(0.195)
Purchase.purpose cashe		0.153			0.359*	0.314*		0.0438			0.176	0.178
		(0.172)			(0.188)	(0.189)		(0.159)			(0.172)	(0.173)
Loanrepayment.purpose cashe		-0.0542			0.323	0.264		-0.133			0.111	0.0917
		(0.200)			(0.223)	(0.223)		(0.180)			(0.201)	(0.201)
Other.purpose cashe		0.515***			0.700***	0.644***		0.295**			0.385***	0.368**
		(0.145)			(0.156)	(0.156)		(0.135)			(0.144)	(0.144)
Facebook status			-0.435***	-0.432***	-0.379***	-0.376***			-0.274**	-0.273**	-0.329***	-0.339***
			(0.116)	(0.116)	(0.128)	(0.128)			(0.114)	(0.114)	(0.118)	(0.118)
Linkedin status			-1.702***	-1.705***	-1.365***	-1.331***			-0.709***	-0.708***	-0.624***	-0.632***
			(0.245)	(0.245)	(0.287)	(0.287)			(0.225)	(0.225)	(0.232)	(0.230)
Log no of SMS			-0.0962***	-0.0958***	0.0012	0.00522			-0.0590*	-0.0584*	-0.0246	-0.0244
			(0.032)	(0.032)	(0.034)	(0.034)			(0.032)	(0.032)	(0.032)	(0.032)
Log No of Contacts			0.195***	0.196***	0.395***	0.382***			0.345***	0.346***	0.145***	0.142***
			(0.043)	(0.043)	(0.058)	(0.058)			(0.042)	(0.042)	(0.052)	(0.052)
Log no of Apps			0.00646	0.00438	0.397***	0.363***			0.0773	0.0758	0.0668	0.0578
			(0.083)	(0.084)	(0.103)	(0.104)			(0.082)	(0.082)	(0.094)	(0.094)
Log Calllog			0.219***	0.220***	0.384***	0.383***			0.197***	0.198***	0.161***	0.161***
			(0.038)	(0.038)	(0.052)	(0.052)			(0.037)	(0.037)	(0.049)	(0.049)
Dating App			0.264	0.27	0.702*	0.679*			0.386	0.388	0.720**	0.708**
			(0.349)	(0.349)	(0.374)	(0.373)			(0.342)	(0.343)	(0.349)	(0.349)
Finsavy app			0.333*	0.339*	0.318	0.3			-0.189	-0.179	-0.0896	-0.095
			(0.182)	(0.181)	(0.238)	(0.241)			(0.176)	(0.175)	(0.217)	(0.220)
Socialconnect app			2.061***	2.048***	5.750***	5.692***			2.093***	2.074***	1.905***	1.919***
			(0.196)	(0.197)	(0.360)	(0.353)			(0.189)	(0.189)	(0.228)	(0.231)
Travel app			-1.226***	-1.225***	-1.088***	-1.088***			-0.544***	-0.543***	-0.448**	-0.470**
			(0.154)	(0.154)	(0.190)	(0.192)			(0.152)	(0.152)	(0.182)	(0.184)
Mloan app			-0.328***	-0.328***	-0.565***	-0.552***			-0.549***	-0.548***	-0.545***	-0.541***
			(0.118)	(0.118)	(0.132)	(0.132)			(0.116)	(0.116)	(0.123)	(0.124)
IOS Dummy				-0.493	-0.000354	0.00648				-0.0708	-0.246	-0.246
				-0.385	-0.427	-0.427				-0.371	-0.373	-0.373
Pseudo Rsq	0.0018	0.475	0.0098	0.0099	0.4734	0.4751	0.0018	0.475	0.0098	0.0099	0.4734	0.4751

Observations	113,013	113,006	98,260	98,260	98,253	98,244	113,013	113,006	98,260	98,260	98,253	98,244

Table A2: **Loan Purpose**

This table reports the estimates from our multionomial logit regressions examining the determinants of loan purpose. The dependent variable, *Loan Purpose* identifies Medical Loans in columns 1-2, Travel loans in columns 3-4, EMI loans in columns 5-6, Purchase loans in columns 7-8, and Repayment loans in columns 9-10. The base category is the uncategorized loans (*Other loans*) in all columns. Standard errors are clustered at the state level. (***), (**), (*) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Logit coeff Medical (1)	Logit coeff Medical (2)	Logit coeff Travel (3)	Logit coeff Travel (4)	Logit coeff EMI (5)	Logit coeff EMI (6)	Logit coeff Purchases (7)	Logit coeff Purchases (8)	Logit coeff Loan Repayment (9)	Logit coeff Loan Repayment (10)
Facebook status		0.0489**		0.163***		-0.0311		0.0336		0.0143
		(0.020)		(0.027)		(0.026)		(0.022)		(0.029)
Linkedin status		-0.162***		0.348***		0.383***		0.429***		0.217***
		(0.061)		(0.068)		(0.064)		(0.055)		(0.075)
Log no of SMS		0.00501		0.00201		0.0262***		0.000342		0.000786
		(0.005)		(0.007)		(0.006)		(0.005)		(0.007)
Log No of Contacts		0.0430***		-0.0413***		-0.133***		-0.0277***		-0.0613***
		(0.010)		(0.013)		(0.012)		(0.010)		(0.013)
Log no of Apps		-0.127***		-0.228***		0.0923***		0.114***		0.0397*
		(0.016)		(0.022)		(0.021)		(0.018)		(0.022)
Log Calllog		-0.00268		-0.0794***		-0.0566***		-0.0678***		-0.101***
		(0.007)		(0.010)		(0.009)		(0.008)		(0.011)
Dating App		0.0457		0.532***		0.566***		0.0262		0.130*
		(0.054)		(0.059)		(0.058)		(0.058)		(0.073)
Finsavy app		-0.152***		-0.453***		0.416***		0.226***		0.175***
		(0.037)		(0.052)		(0.056)		(0.046)		(0.057)
Socialconnect app		-0.527***		-3.145***		-1.233***		-1.083***		-1.212***
		(0.071)		(0.106)		(0.083)		(0.076)		(0.086)
Travel app		0.0970***		3.930***		0.103***		0.0556**		0.245***
		(0.021)		(0.064)		(0.028)		(0.024)		(0.032)
Mloan app		0.0794***		-0.0368		0.0236		0.0451**		0.121***
		(0.019)		(0.025)		(0.024)		(0.021)		(0.027)
CIBIL	-0.00112***		-0.00243***		-0.00194***		-0.00143***		-0.00239***	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
IOS Dummy		0.124		0.265**		0.311***		0.437***		0.522***
		(0.080)		(0.103)		(0.099)		(0.080)		(0.099)
Pseudo Rsq	0.0004		0.0004		0.0004		0.0004		0.0004	
Observations	166,628	98,816	166,628	98,816	166,628	98,816	166,628	98,816	166,628	98,816