Fintech and Credit Scoring for the Millennials

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

October 2019

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 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 outperforms the credit score in predicting loan approvals and defaults. Importantly, including measures of borrower's "deep digital footprints" based on call logs significantly improves default prediction. Our study has implications for expanding access to credit to those who do not 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 feel that their credit score is holding them back.[1] [2] Younger people suffer from shorter credit history and hence are often denied credit by traditional financial institutions or are charged prohibitively high interest rates, which limits their access to credit. This, in turn, exacerbates the evaluation of their creditworthiness by limiting their ability to build a good credit history. Many such individuals may actually be 'good borrowers' if their 'creditworthiness' could be evaluated using alternate data. The problem of lack of credit history for the millennials is a world-wide phenomenon and especially true for developing countries. For example, according to a recent industry report, 156 million Indians who comprise the 'urban mass' representing an annual income of USD 3000 and above have the potential of mass adoption of consumer credit. Of this 'urban mass,' approximately 129 million have been mostly deprived of credit due to a lack of credit history.[3] This led to the quest for alternative data for credit scoring for the millennials.

While millions across the country 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 Berg et al. (2018)) 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 mobile footprint" of an individual can substitute for traditional credit bureau scores. This paper aims to further the early work in this area.

Towards this end, we use data from one of the largest Fintech lending firms in India to examine the discriminatory ability of digital mobile footprint variables in predicting loan outcomes. Specif-

[1] Wall Street Journal Blog [Accessed on 17th October, 2019].According to Wall Street Journal and Transunion; Around 53 million consumers are not scoreable due to lack of information at the three major credit bureaus, and this population is heavily skewed towards those under 35..

[2] MarketWatch News Article [Accessed on 14th March, 2019].The survey looked into the credit 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.

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

ically, we want to understand whether and how the digital footprint is associated with loan level outcomes such as the likelihood of loan approval 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 affect her creditworthiness, rather further along the lines of work of Berg et al. (2018), to analyze the association between the digital footprint and credit worthiness of individuals.

We obtain the universe of loan applications made to one of the largest fintech lender in India, between the period of February 2016 to November 2018. Unlike prior studies, we also have access to loan applications that were eventually denied allowing us to examine the determinants of loan approval. Out of about 417,000 loan applications in our sample, about 272,000 were approved while rest were denied. The mobile-only lending platform of the fintech lender 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 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 statements, salary slip. The potential borrower authorizes the lender to use its digital footprint variables for the evaluation of her creditworthiness and research. They also provide the fintech lender data on their CIBIL−Transunion credit score (if available), education, and job designation. Importantly for our study, The lender 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 access to detailed anonymized 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, and *Dating apps* such as Tinder. In addition, we have detailed information on call logs of individuals. 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 that larger is the digital mobile footprint of an individual, the higher is her likelihood of loan approval. Specifically, we find that the number of contacts, the number of applications (apps from now) installed, the number of calls made or received, and the presence of financial and mobile loan apps are positively associated with the loan approval. 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 (Iyer et al. (2015), Berg et al. (2018)). We first note that the AUC of the model using only the credit score for predicting defaults is 58.6%. The AUC of credit score in our sample, while significantly different from flipping a coin (AUC of 50%) is lower than 62% reported by Iyer et al. (2015) based on a sample of loans from peer to peer lending platform, "Propser.com", and 68.3% reported by Berg et al. (2018) based on a sample of purchases from a German e-retailer, and comparable to the AUC of 59.8% using U.S. credit scores from Lending Club reported in Berg et al. (2018).

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 environments and lower levels of financial inclusion.

The AUC of a model that relies exclusively on the digital footprint to predict defaults at 60.4% is approximately 2% more than the AUC of the model using only the credit score. 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 their phones are about one and a half times more likely to default relative to those who have such an application installed. 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 25% (33%) more likely to default. Interestingly, customers who log in to the application via Linked or Facebook are 27% and 9% more likely to default respectively relative to those who log-in via other means.

Consistent with the evidence in Berg et al. (2018), 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 Berg et al. (2018). Given that they do not have information on earnings or education of the customer, they are unable to disentangle whether owning an Apple phone only 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 only proxies for easily measurable financial or customer 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 or hard information that would be otherwise difficult to measure or verify. In such a case, first, digital footprints can be used to improve traditional credit scoring models.

Our results suggest that digital mobile footprint captures an unobservable aspect of individuals which is not fully absorbed by earnings, education, or credit score. Importantly, the AUC of this specification is 74% , 15 percentage points higher than the AUC of the model using only the credit bureau score and two percentage points higher than the model, which includes CIBIL score combined with customer and loan characteristics. In other words, a predictive model that includes loan characteristics, customer characteristics, and digital footprint performs better 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 the credit bureau score and observable customer characteristics.

Further, 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. To strengthen the evidence in favor of this thesis, we examine the predictive ability of digital mobile footprint in predicting defaults for the set of customers without a credit score or history. The AUC of the digital mobile footprint model for this sample is 58% and comparable to the predictive performance of the credit bureau score in the primary sample for customers with a credit bureau score.

Our analysis of default prediction thus far is based on crude measures of digital footprint such as the nature of apps installed, the number of apps installed, the number of calls, etc., to predict defaults. We now seek to understand whether we can use "deeper digital footprint" of customers from their call logs to improve upon the default prediction. Using various proxies based on the frequency and duration of daily incoming, outgoing, and missed calls that attempt to capture the breadth and strength of an individual's social capital, we find that these measures are strongly correlated with the likelihood of default.[4] Specifically, we find that defaulters are more likely to have their call concentrated over a smaller number of individuals. Consistent with this, defaulters seem to have stronger ties with individuals in their contact list as measured by the average number of calls and duration of calls per person. Delinquent customers have a smaller duration of incoming calls but have a higher duration of outgoing calls, which along with their frequency of missed calls, suggests that defaulters are less likely to respond to calls initiated by others.

Most importantly, the AUC of a model that includes call log measures along with other digital mobile footprint variables is 66%, an 8% improvement over the model with credit score alone. This is better than the 5.7 percentage points AUC improvement reported in Iyer et al. (2015) who compare the AUC using the Experian credit score to the AUC in a setting where, in addition to the credit score, lenders have access to a large set of borrower financial information as well and comparable to the improvement in the AUC by +8.8 percentage points reported by Berg et al. (2017) in a consumer loan sample of a large German bank in a setting where, in addition to the credit score, lenders have access to account data, as well as socio-demographic data and income information.

We also have access of the detailed credit reports for a subset of the borrowers in our sample.

[4]The underlying idea behind these tests builds on prior work which documents that call log patterns can be used to infer an individual's social capital (Singh and Ghosh (2017)), which is an important predictor of loan defaults (Karlan (2005)).

The credit report has various 'deep' financial information like the borrower's spending and income patterns, number of transactions, other borrowing information etc. over last three months. These are typical reports that any financial institution use during the loan approval process besides borrowers credit scores. These credit reports are accessed by the fintech lender during the loan application process. We find that for the subset of the borrowers for whom we have access to this credit report, the 'deep' digital information has more predictive power of borrower's credit risk than 'deep' financial information.

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, if installing financial applications captures the financial literacy of the individual and propensity of a consumer to engage in strategic defaults, then we should expect the default rates to be higher for loans taken for repaying an existing loan or meeting the EMI of another loan, if the customer taking the loan has installed financial applications. Consistent with this conjecture, we find that the likelihood of default is significantly higher for such customers when they take loans for EMI payments (*EMI* loans), or loan repayment (*Repayment* loans). Specifically, as compared to customers who do not have financial apps installed on their phones, those who do are 34%, and 56% more likely to default when they take EMI loans and Repayment loans, respectively. Along similar lines, customers who have installed another loan application app, are also more likely to default when they undertake a loan for EMI or loan repayment. These results suggest that there is a 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.

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.[5] Moreover, an economically significant 5% of borrowers in our sample do not have a credit

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

score. This is in contrast to the USA, where fintech lenders primarily cater to borrowers who already have access to credit via traditional banks (Buchak et al. (2018), Tang (2018)).

The paper closest to our study is Berg et al. (2018). Using data covering approximately 250,000 purchases from an E-Commerce company located in Germany, Berg et al. (2018) 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. Moreover, even in developed countries like the UK, the 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 only 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.

Fourth, we find that the default prediction can be improved significantly by using proxies that capture deeper aspects ("deep digital footprint") of an individual's digital presence. Finally, we document that digital footprints can allow lenders to estimate the likelihood of default based on

prime)), and 300–650 (*Subprime*)

the end-use of the loans. The implication is that with the use of big data on the 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 417,578 loan applicants from a mobile-only Fintech lending platform operating in India since 2016. The lender 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. The fintech lender provides loans of amount ranging from a minimum of ₹10,000 ($141) to ₹200,000 ($2816).[6] The loan duration ranges from a minimum of 15 days to a maximum of 180 days. Currently, they have 180,000 active customers, with about 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 lending app, submit all the requisite details and documentations. The borrower also gives permission to the lender to gather additional information on the mode of login, the various apps installed, the number of calls and SMSs, number of contacts on the phone, number of social connections, and the kind of mobile operating systems such as IOS and Android. We obtained data from the lending firm for all loans granted from February 2016 to November 2018.

2.1 Summary Statistics: Loan and Financial Variables

Table 1 reports the summary statistics. Out of the 417,578 loan applications in our sample, 272,931 were approved, while 144,647 were denied. The default rate in our full sample is quite high at approximately 13.5%[7]. However, the higher default rate is primarily driven by loans given out during the early period of it's operations. The default rates for loans advanced during 2018 is 3%, comparable to the delinquency rate of 3% for retail loans given out by all banks across India.[8] This suggests that the lender now caters to customers that are comparable in terms of default risk to the average retail borrower in India. The average loan size is ₹22,174 ($312) age of a customer is

[6] Based on the nominal exchange rate of $1=₹71 as of October 2019.

[7] $\frac{\text{32,555 defaults}}{\text{240,376 approvals}}$

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

32 consistent with the idea that lending firm target segment is a young salaried customer.[9] The average credit score is 634 and is obtained from TransUnion CIBIL. The average interest rate charged on loan is 25% (log value of 1.4). On average, a customer earns ₹37,524 ($527) per month or $6324 per annum. Thus, the income of a customer is our sample is roughly three times the median per capita income of $2,134 in 2018. Thus, the lender caters to relatively higher-income customers. The application process 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, 8% were taken for the purpose of travel, 9% for EMI, 13 % for purchasing a good, about 8% for the purpose of repaying a loan, 22% for medical expenditure, and rest is uncategorized.

2.2 Summary Statistics: Digital Footprint Variables

In addition to the credit bureau score, and other customer level variables, the lender 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. The app 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 systems such as IOS and Android. Facebook (Linkedin) dummy variables identify customers that logged in to the app using Facebook (Linkedin). About 27% of customers logged in to the app using Facebook, while 2.1% used Linkedin. On average, 68% of the customers have a banking or stock trading app. About 42% of customers have installed another mobile-loan application suggesting that they look for loans on other platforms as well, while 12% of the customers own an apple phone (ios dummy).

[9]The average loan amount of $312 based on the exchange rate of $1=₹71 is comparable to the average purchase amount of $350 in Berg et al. (2018).

3 Results

3.1 Univariate Analysis

In columns 1-3 of Table 1, 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 29% higher for loan applications that were approved.[10] 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 (less) likely to log in through Linkedin or Google (Facebook). Approved customers are also significantly 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* dummy).

In columns 1-3 of Table 1, we analyze the customer and loan characteristics that can potentially predict the likelihood of default. Customers who default on average borrow 71% more than those who don't.[11]. Customers who default on average are charged a higher interest rate ex-ante, consistent with such customers being riskier. Surprisingly, customers who default one average are slightly 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 more 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. Further, delinquent customers are less likely to have installed a financial app but more likely to have installed a social network/travel app. We

10 $\frac{(22174.26-17182.04)*100}{17182.04}$

11 $\frac{(35228.33-20509.83)*100}{20509.83}$

also find that other digital footprint variables that capture various aspects of social behavior have a bearing on the likelihood of default. For instance, customers who 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 (Schmitt et al. (2011), Burks et al. (2015)). Using data on referred customers of a German bank, Schmitt et al. (2011) find that such customers have a higher retention rate and are more valuable in both the long and short term. Along similar lines, Burks et al. (2015) 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 the social ties of a customer. The number of apps also seem to have a discriminatory ability to predict defaults as defaulting customers have fewer apps. Finally, owning an Apple phone is negatively associated with the likelihood of default.

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:

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

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, 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 for the full sample. Not surprisingly, loan applicants with higher credit scores have a higher likelihood of getting approved. The R^2 of the regressions is 0.009, implying that credit scores explain only about 0.9% of the variation in the likelihood of loan approval. In column 2, we repeat the analysis for the subsample of loan applicants with non-missing values of all digital mobile footprint variables and customer characteristics. For this sub-sample, we do not find that credit score is associated with loan approval suggesting that the Fintech lender relies primarily on other parameters for loan approval.

In column (3), we repeat these tests after including other loan and customer characteristics. We find that customers that earn more are older, and need smaller loans, 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 loan purpose is not associated with the likelihood of approval.

In column (4), we report the results for digital mobile footprint variables. Since the IOS dummy has significant predictive power for loan outcomes (see Berg et al. (2018)), to make sure that our results are not just driven by the IOS variable, we do not include it in column (4). We find that the number of contacts, the number of apps installed, the presence of financial and mobile loan apps (Finsavvy and Mloan dummy variables) are positively associated with the loan approval. These results continue to hold when we include the IOS dummy in column (5). We find that customers with an IOS device have a 44% higher likelihood of approval compared to those without an IOS device. This is consistent with prior studies, which highlight that owning an IOS device is a strong predictor of higher earnings (Bertrand and Kamenica (2018)). Overall, these results indicate that digital footprint variables have significant explanatory power for the likelihood of loan approval. The AUC of the model with digital mobile footprint variables at 54.4 is significantly higher than the AUC of 50.8 for the model with credit score alone in column (2). The results remain robust to including credit score in column (6).

Column (7), 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 footprint 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, finsavvy, mloan, and IOS dummies continue to be statistically significant. Further, the AUC of the model with loan, customer, and digital footprint characteristics is 8% more than that of the model with CIBIL score alone. This suggests that loan characteristics, customer characteristics, and digital footprint, have some complementary information beyond what is captured in the CIBIL score.

Finally, in columns (8) and (9), we also include CIBIL score and state fixed effects. The results remain qualitatively similar. 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.[12]

3.2.2 Defaults

In this section, we focus on analyzing the relationship between digital footprint variables, loans, and customer characteristics and default. The dependent variable in these tests is a dummy variable which takes the value one for delinquent loans. Table 3 reports the results from these tests. Column (1) reports the results using only the credit bureau score (*CIBIL*) as the explanatory variable for the sample of approved loans. Not surprisingly, a higher credit bureau score is associated with a significantly lower likelihood of default. In column (2), we repeat the analysis for the subsample of loan applicants with non-missing values of all digital mobile footprint variables and customer

[12] In Table A1 of Appendix A, we repeat these tests with the subsample of customers without a credit score. The AUC of a model using digital footprint variables to predict loan approval is 71.2%, again supporting our claim that the fintech lender relies heavily on digital variables in the absence of a credit bureau score.

characteristics. For this sub-sample, we don't find that credit score is associated with loan approval suggesting that the Fintech lender relies primarily on other parameters for loan approval. The AUC of the CIBIL score in this sample is 59%. The AUC of credit score in our sample, while significantly different from chance (AUC of 50%) is lower than 62% reported by Iyer et al. (2015) based on a sample of loans from peer to peer lending platform, "Propser.com" and 68.3% reported by Berg et al. (2018) based on a sample of purchases from a German e-retailer but comparable to the AUC of 59.8% using U.S. credit scores from Lending Club reported in Berg et al. (2018). This suggests that the discriminatory ability of the credit score in predicting defaults is likely to vary across countries and types of financial intermediary.

In column (3), 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, age, and education are 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 (Kalda (2019)). 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 (4), we report the results for digital footprint variables. Since the IOS dummy has significant predictive power for loan outcomes (see Berg et al. (2018)), to make sure that our results are not just driven by the IOS variable, we do not include it in column (4). The AUC of this specification is 60.4% and approximately 2% more than the AUC estimate using just the credit bureau score.

Focusing on the individual variables, we find that digital footprint variables may proxy for hard to quantify aspects of individual behavior, which has implications for the likelihood of default. We find that individuals that have a financial app installed on their phones have a significantly lower likelihood of default. The odds ratio of *Finsavvy* dummy is 0.71, 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. This suggests that *Finsavvy* dummy may be correlated with the financial sophistication of a customer. In contrast, those with a dating app (any other social network app) are 25% (33%) more likely to default.[13]. Interestingly, customers with a travel app are about 3% more likely to default than those without. Finally, those who log in to the application via Linked or Facebook are 27% and 9% more likely to default respectively relative to those who via other means. 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 has significant discriminatory power in default prediction.

In column (5), 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 Berg et al. (2018), 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 0.495, implying that those with an android phone are twice as likely to default as those with an Apple phone. In column (6), we include both the CIBIL score and mobile digital footprint variables together. We note that the AUC of this model is 60.8% and 2.2 percentage points higher than that of a model using only the credit bureau score.

As in Table 2, column (7) includes all loan characteristics, customer characteristics, and digital mobile footprint variables but excludes the credit bureau score. We find that the coefficients of the digital mobile footprint variables largely remain unchanged, suggesting that these variables have incremental predictive power over loan and customer characteristics. More specifically, the digital mobile footprint seems to be capturing unobservable aspects of customer behavior, which is not absorbed by education, age, salary, or job designation of the customer. Interestingly, the coefficient estimate of *IOS* dummy remains statistically significant even after controlling for the customer's monthly salary. In this respect, our study complements Berg et al. (2018) 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 that is not fully absorbed by earnings.

Interestingly, we also find that customers who log in through Facebook are more likely to default even once we control for loan and customer characteristics. Importantly, the AUC of this specification is 74% , 15 percentage points higher than the AUC of the model using only the

[13]The odds ratio of *Dating* dummy is 1.25 and that of *Socialconnect* dummy is 1.33

credit bureau score and two percentage points higher than 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 better in predicting defaults as compared to a model which includes credit bureau score, loan characteristics, and customer characteristics.

Finally, in columns (8) and (9), we also include CIBIL score and state fixed effects for robustness. The results remain qualitatively similar.

One concern with our evidence so far could be that it is driven by a subsample of customers in our sample. For instance, digital footprint variables may have predictive power only for customers with a high credit score or salary. This would limit the promise of using digital mobile footprints to score customers without a credit bureau score/history. To further strengthen the evidence regarding the discriminatory ability of digital mobile footprint variables in predicting defaults, in tables A2-A5 of Appendix A, we repeat our baseline models on subsamples based on credit score, age, salary, and job designation terciles. We find that the digital variables retain their discriminatory abilities across such subsamples.

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. To further strengthen this thesis, in the next section, we focus on predicting defaults using digital footprints for borrowers without a credit score..

3.2.3 Predicting defaults for customers without credit score

While our analysis so far suggests that the digital mobile footprint has incremental explanatory power for predicting defaults, customers who lack credit history and credit score may be very different from the set of customers with a credit bureau score. To examine if these results are generalizable to the set of unscorable customers, in Table 4, we focus on the set of customers without a credit score and examine whether and how does the digital mobile footprint perform in default prediction for this subsample. In column (1), we only include customer and loan characteristics and find that these have significant discriminatory power. The AUC of the model is 78.5%. In column (2), we include digital footprint variables and find that the AUC of the model is 58% and comparable to the predictive performance of the credit bureau score in Table 3. Importantly, in column (3), we include customer characteristics, loan characteristics, and digital mobile footprint variables together to examine if digital footprint variables have incremental explanatory power over customer and loan characteristics. As compared to column (1), including digital variables improves the AUC by 2.4% which is considered to be a significant improvement.][14] Summarizing, these findings suggest that digital mobile footprints can indeed be used to score customers without a credit history and conventional credit score.

3.2.4 Defaults, digital footprint, and loan purpose

Table 5 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 purposes with digital footprint variables to examine whether digital footprint variables have greater discriminatory power in predicting defaults conditional on the purpose of the loan. For instance, if installing financial applications captures the financial literacy of the individual and propensity of a consumer to engage in strategic defaults, then we should expect the default rates to be higher for loans taken for repaying an existing loan or meeting the EMI of another loan, if the customer taking the loan has installed financial applications. Consistent with this conjecture, we find that the likelihood of default is significantly higher for such customers when they take loans for EMI payments (*EMI* loans), or loan repayment (*Repayment* loans). Specifically, as compared to customers who do not have financial apps installed on their phones, those who do are 34%, and 56% more likely to default when they take EMI loans, and Repayment loans, respectively. Along similar lines, customers who have installed another loan application app, are also more likely to default when they undertake a loan for EMI or loan repayment.

One possibility for difference in default rates of customers that install financial apps could be driven by a selection bias if, for instance, customers that install financial apps are on average of low creditworthiness. To examine this possibility in figure 1, we plot the kernel density distribution of

[14] See for instance, Iyer et al. (2015) and Berg et al. (2018)

CIBIL score, Salary, and Age for customers that install financial apps and those who do not. We note from Figure 1, subfigure (a) 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 subfigures (b)and (c), we find that again, the age profile of customers is also similar. Similarly, from subfigures (d), (e), and (f), we note that there is no observable difference between the customers who have installed another mobile application and those who have not. We conclude that the propensity to install a financial app 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 the 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.

3.2.5 Predicting defaults using deep digital footprints from call logs

Thus far, we have relied on rudimentary and crude measures of digital footprint such as the nature of apps installed, the number of apps installed, the number of calls, etc; to predict defaults. We now seek to understand whether we can use "deeper digital footprint" of customers to improve upon the default prediction. For instance, if the presence of a financial app on a customer's phone can predict defaults, it would not be unreasonable to conjecture that the duration of time spent across different kinds of apps, time spent on social media, nature and time of online searches etc; could have incremental explanatory power for default prediction. Unfortunately, we do not have detailed information regarding the customer's usage of different installed applications. We do however, have detailed call logs for a large subsample of borrowers in the data. Prior literature highlights that call log patterns can be used to infer an individual's social capital (Singh and Ghosh (2017)), which is known to be an important predictor of loan defaults (Karlan (2005)).

Following prior literature, we create two kinds of proxies using call logs that attempt to capture the breadth and strength of an individual's social capital. We proxy for breadth using total frequency and duration of daily incoming, outgoing, and missed calls. Singh and Ghosh (2017) find that the frequency of missed calls and duration of incoming vsȯutgoing calls is also related to reciprocity– the propensity of an individual to respond to and engage in calls associated by others. We proxy for the strength of an individual's social connections using the average number and duration of calls per person. The underlying idea is that an individual is likely to make a greater number of calls or longer duration calls to people with whom they have stronger ties. Finally, we create a Herfindal index, which captures whether the calls of an individual are concentrated over a few connections or spread across multiple contacts. These measures are constructed both ex-ante based on the call logs information available prior to loan approval, and ex-post based on the call logs information available in the first 15 days after loan approval.

Table A6 of Appendix A provides the details of how we construct these measures, and panel A of Table 6 reports the univariate summary statistics. Focusing on the total and the average number of missed calls per person, we see that defaulters, on average, are less likely to accept calls initiated by others. Defaulters are also more likely to have their calls concentrated over a smaller number of individuals, as evidenced by the HHI index for all measures of incoming/outgoing calls. Consistent with this, defaulters seem to have stronger ties with individuals in their contact list as measured by the average number of calls and duration of calls per person. Delinquent customers have a smaller duration of incoming calls but have a higher duration of outgoing calls suggesting that defaulters, which along with their frequency of missed calls, suggest that defaulters are less likely to respond to calls initiated by others. These patterns are consistent across ex-ante and ex-post call logs based measures.

In Table 7, we again use our baseline multivariate logit model to examine whether the measures based on call logs predict defaults. Given that the various call based measures are correlated with each other, it is important to note that our goal is not to understand the direction of causality but rather to understand whether a model that includes these variables does a good job of predicting loan defaults. We start by analyzing the predictive ability of the credit bureau score for the subsample of customers for whom call details are available in column (1). The AUC of the credit score at 58.3% is comparable to what we observed in the full sample in Table 3. In column (2), we include only deep digital footprints based on call logs. The AUC of this model is remarkably high and 6% more than the model with credit score alone. In columns 3 and 4, we include call log measures along with other digital mobile footprint variables and credit score respectively and find that the AUC goes up to 66%, an 8% improvement over the model with credit score alone. This is better than the 5.7 percentage points AUC improvement reported in Iyer et al. (2015) who compare the AUC using the Experian credit score to the AUC in a setting where, in addition to the credit score, lenders have access to a large set of borrower financial information as well and comparable to the improvement in the AUC by +8.8 percentage points reported by Berg et al. (2017) in a consumer loan sample of a large German bank in a setting where, in addition to the credit score, lenders have access to account data, as well as socio-demographic data and income information.

We next examine whether digital mobile footprints taken together have incremental explanatory power over and above a model that includes credit score, loan, and customer characteristics. In column (5) of Table 7, we include loan and customer characteristics. The AUC of this model is 71.3%. In column (6), we include digital footprint variables along with customer and loan characteristics and find that the AUC of this model outperforms the model in column 5 by about 5.7%. Finally, in column (7), we include credit score, loan, and customer characteristics, and digital mobile footprint variables together. We find that including credit score does not improve the AUC significantly over a model with loan characteristics, customer characteristics, and digital footprint variables.[15]

Finally, table 8 reports the relative performance of 'deep' financial vs 'deep' digital information for a subset of the borrowers in predicting defaults. As mentioned earlier, the 'deep' financial information like spending in last three months, other borrowing, number of transactions in the bank account etc. are found in the credit reports of the borrower accessed during the loan application process. Column (4) reports the performance of the borrower's 'deep' digital information in predicting defaults. It has an AUC of 54%. Column (1) reports the performance of digital footprint variables which has a significantly higher AUC of about 60% in predicting default.

4 Conclusion

In this paper, we have used a unique and proprietary dataset to analyze the impact of the digital mobile footprint of individual borrowers in predicting loan outcomes. Our dataset comes from

[15] In Table A7 of the appendix, we also include ex-post measures based on call logs information during the first 15 days after the loan was granted and obtain similar results. In Table A8 of the appendix, we repeat these tests with the subsample of customers without a credit bureau score and obtain qualitatively similar results. We do not report these in the main tables as the sample of customers without a credit score for whom call logs information is also available is small.

a leading fintech lending company in India. We find that the digital footprint and social media preference for login has significantly more predictive power than traditional credit score used by banks.

We find a number of interesting results. First, we document a statistically and economically significant role of individuals' digital footprint variables in the loan approval process. In the absence of sufficient credit history and credit scores for millennial customers to judge their creditworthiness, 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 worldwide.

We also find that a simple predictive model in which an individual's both crude digital media mobile footprint and deeper digital footprint based on call logs significantly outperforms a model with a credit score in predicting defaults. 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 regulations in the era of 'big data.'

References

Berg, T., Burg, V., Gombović, A., and Puri, M. (2018). On the rise of fintechs–credit scoring using digital footprints. *Working paper.*

Berg, T., Puri, M., and Rocholl, J. (2017). Loan officer incentives, internal ratings and default rates. *Working paper.*

Bertrand, M. and Kamenica, E. (2018). Coming apart? cultural distances in the united states over time. *Working paper.*

Buchak, G., Matvos, G., Piskorski, T., and Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483.

Burks, S. V., Cowgill, B., Hoffman, M., and Housman, M. (2015). The value of hiring through employee referrals. *The Quarterly Journal of Economics*, 130(2):805–839.

Immorlica, N., Kranton, R., Manea, M., and Stoddard, G. (2017). Social status in networks. *American Economic Journal: Microeconomics*, 9(1):1–30.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2015). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577.

Kalda, A. (2019). Peer financial distress and individual leverage. *Working paper.*

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*, 95(5):1688–1699.

Schmitt, P., Skiera, B., and Van den Bulte, C. (2011). Referral programs and customer value. *Journal of marketing*, 75(1):46–59.

Singh, V. K. and Ghosh, I. (2017). Inferring individual social capital automatically via phone logs. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):95.

Tang, H. (2018). Peer-to-peer lenders versus banks: substitutes or complements? *Review of Financial Studies.*

Table 1: **Panel A: Summary Statistics of Customer and loan characteristics**

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	22174.26	17182.04	-4992.22***	35228.33	20509.83	-14718.49***
Log Interest Rate	1.445	0.892	-0.552***	1.857	1.393	-0.463***
Loanpurpose Medical	0.214	0.095	-0.118***	0.247	0.209	-0.037***
Loanpurpose Travel	0.082	0.024	-0.057***	0.075	0.082	0.007***
Loanpurpose EMI	0.087	0.081	-0.005***	0.073	0.088	0.014***
Loanpurpose purchase	0.133	0.068	-0.065***	0.129	0.133	0.004***
Loanpurpose Loanrepayment	0.081	0.047	-0.034***	0.082	0.081	-0.0006
Loanpurpose Other	0.405	0.232	-0.172***	0.395	0.405	0.010***
Age	31.89	29.45	-2.44***	32.00	31.88	-0.117***
Salary	37524.53	30346.39	-7178.13***	39262.32	37342.43	-1919.89***
CIBIL (>0, N=219k & 16k)	634.40	470.82	-163.58***	602.04	639.10	37.06***
Facebook Status	0.267	0.296	0.029***	0.274	0.263	-0.011***
Linkedin Status	0.021	0.015	-0.006***	0.023	0.021	-0.002**
Googleplus_status	1.712	1.690	-0.021***	1.700	1.714	0.013***
Referral	0.116	0.039	-0.077***	0.115	0.118	0.002*
Sales App	0.195	0.198	0.003	0.188	0.196	0.007***
Dating App	0.029	0.028	-0.001	0.029	0.029	0.0006
Finsavy app	0.679	0.034	-0.645***	0.677	0.862	-0.019***
Socialconnect app	0.714	0.036	-0.677***	0.760	0.708	-0.051
Travel app	0.567	0.048	-0.518***	0.576	0.566	-0.010***
Mloan app	0.423	0.020	-0.403***	0.423	0.423	-0.0002
Referrer	0.234	0.034	-0.200***	0.167	0.243	0.075
# of SMS	2481.71	1109.00	-1372.71***	1949.25	2548.19	598.94***
# of Apps	54.53	41.26	-13.27***	47.07	55.47	8.40***
# of Contacts	844.84	683.81	-161.02***	827.64	847.03	19.38***
# of Connections	525.89	452.39	-73.50	413.23	539.15	125.92***
# of Calls	3136.50	2071.97	-1064.53***	2394.96	3229.05	834.08***
IOS	0.119	0.066	-0.053***	0.112	0.120	0.007***
Education						
<High School	0.112	0.310	0.197***	0.119	0.112	-0.007***
High School	0.645	0.546	-0.090***	0.647	0.645	-0.002
College	0.241	0.144	-0.097***	0.233	0.243	0.010***
Job Designation						
Worker	0.354	0.410	0.056 ***	0.347	0.354	0.007***
Supervisor	0.248	0.254	0.005***	0.245	0.248	0.003
Manager	0.398	0.335	-0.063***	0.407	0.397	-0.010***
N	272,931	144,647		32,555	240,376	

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 (Log of CIBIL) as the explanatory variable with observations from the full sample. Column (2) includes the credit bureau score (Log of CIBIL) with observations from only the subsample. Column (3) also includes other loan and customer characteristics excluding digital footprint. Column (4) includes only digital footprint variables excluding IOS dummy. Column (5) includes only digital footprint variables along with IOS dummy. Column (6) includes only digital footprint variables and CIBIL score and IOS Dummy. Column (7) includes all loan and customer characteristics and digital footprint variables but not the CIBIL score. Column (8) includes all variables including the CIBIL score. Column (9) 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	Odds Ratio (1)	Odds Ratio (2)	Odds Ratio (3)	Odds Ratio (4)	Odds Ratio (5)	Odds Ratio (6)	Odds Ratio (7)	Odds Ratio (8)	Odds Ratio (9)
Log of cibil	1.190***	1.016	1.024*			1.014		1.020	1.017
	(0.000)	(0.229)	(0.083)			(0.297)		(0.146)	(0.228)
Log of Salary			1.171***				1.119***	1.114**	1.097**
			(0.000)				(0.009)	(0.013)	(0.035)
Log Loan Amount			0.733***				0.736***	0.735***	0.732***
			(0.000)				(0.000)	(0.000)	(0.000)
Log Age			1.491***				1.758***	1.705***	1.717***
			(0.000)				(0.000)	(0.000)	(0.000)
High School Dummy			1.071				1.071	1.058	1.074
			(0.153)				(0.146)	(0.239)	(0.143)
College Dummy			1.065				1.071	1.059	1.073
			(0.252)				(0.197)	(0.294)	(0.206)
Supervisor			0.875***				0.881***	0.873***	0.873***
			(0.001)				(0.001)	(0.001)	(0.001)
Manager			0.903***				0.914**	0.902***	0.910**
			(0.005)				(0.012)	(0.005)	(0.012)
Travel.purpose cashe			0.990				0.972	0.959	0.966
			(0.873)				(0.650)	(0.512)	(0.593)
EMI.purpose cashe			0.910				0.903*	0.897*	0.901*
			(0.115)				(0.085)	(0.071)	(0.088)
purchase.purpose cashe			0.982				0.973	0.966	0.972
			(0.726)				(0.587)	(0.511)	(0.585)
Loanrepayment.purpose cashe			0.994				0.974	0.976	0.979
			(0.920)				(0.658)	(0.690)	(0.732)
Other purpose.purpose cashe			0.952				0.949	0.951	0.962
			(0.219)				(0.185)	(0.204)	(0.344)
Log no of SMS				0.989	0.990	0.987	0.996	0.993	0.992
				(0.175)	(0.204)	(0.116)	(0.585)	(0.400)	(0.301)
Log No of Contacts				0.990	0.989	0.987	0.996	0.997	1.003
				(0.562)	(0.516)	(0.469)	(0.828)	(0.878)	(0.887)
Log no of Apps				1.172***	1.178***	1.183***	1.190***	1.194***	1.187***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log Callog				1.034***	1.036***	1.034***	1.031***	1.029**	1.029**
				(0.004)	(0.002)	(0.004)	(0.008)	(0.013)	(0.015)
Dating App				0.924	0.921	0.946	0.952	0.977	0.995
				(0.351)	(0.333)	(0.527)	(0.563)	(0.790)	(0.957)
Finsavy App				1.203***	1.205***	1.218***	1.177***	1.187***	1.183***
				(0.002)	(0.002)	(0.001)	(0.006)	(0.005)	(0.007)
Socialconnect App				0.916	0.954	0.946	0.902	0.888	0.837*
				(0.340)	(0.610)	(0.559)	(0.270)	(0.221)	(0.080)
Travel App				1.005	1.001	0.986	1.057	1.043	1.029
				(0.883)	(0.969)	(0.698)	(0.134)	(0.257)	(0.462)
Mloan App				1.076**	1.076**	1.080**	1.081**	1.085**	1.094***
				(0.018)	(0.019)	(0.015)	(0.013)	(0.010)	(0.005)
Facebook status				1.020	1.018	1.017	1.017	1.016	1.024
				(0.555)	(0.584)	(0.624)	(0.612)	(0.633)	(0.482)
Linkedin status				0.963	0.968	0.951	1.018	1.006	1.032
				(0.689)	(0.735)	(0.603)	(0.850)	(0.948)	(0.749)
IOS Dummy					1.439***	1.427***	1.507***	1.493***	1.459***
					(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Constant	4.531***	35.951***	35.708***	17.301***	16.055***	15.264***	13.642***	15.011***	15.406***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
State Fixed Effects	N	N	N	N	N	N	N	N	Y
Observations	235,765	189,055	189,055	194,093	194,093	189,055	194,093	189,055	185,162
Pseudo R2	0.00881	3.21e-05	0.00509	0.00225	0.00256	0.00260	0.00781	0.00798	0.00921
AUC	0.581	0.508	0.567	0.541	0.544	0.544	0.577	0.578	0.584

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 (Log of CIBIL) as the explanatory variable with observations from the full sample. Column (2) includes the credit bureau score (Log of CIBIL) with observations from only the subsample. Column (3) also includes other loan and customer characteristics excluding digital footprint. Column (4) includes only digital footprint variables excluding IOS dummy. Column (5) includes only digital footprint variables along with IOS dummy. Column (6) includes only digital footprint variables and CIBIL score and IOS Dummy. Column (7) includes all loan and customer characteristics and digital footprint variables but not the CIBIL score. Column (8) includes all variables including the CIBIL score. Column (9) 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	Odds Ratio (1)	Odds Ratio (2)	Odds Ratio (3)	Odds Ratio (4)	Odds Ratio (5)	Odds Ratio (6)	Odds Ratio (7)	Odds Ratio (8)	Odds Ratio (9)
Log of cibil	0.877***	0.900***	0.872***			0.906***		0.882***	0.885***
	(0.000)	(0.000)	(0.000)			(0.000)		(0.000)	(0.000)
Log of Salary			0.240***				0.265***	0.265***	0.260***
			(0.000)				(0.000)	(0.000)	(0.000)
Log Loan Amount			4.136***				4.193***	4.272***	4.302***
			(0.000)				(0.000)	(0.000)	(0.000)
Log Age			0.545***				0.333***	0.365***	0.363***
			(0.000)				(0.000)	(0.000)	(0.000)
High School Dummy			0.822***				0.855***	0.852***	0.848***
			(0.000)				(0.000)	(0.000)	(0.000)
College Dummy			0.724***				0.745***	0.742***	0.740***
			(0.000)				(0.000)	(0.000)	(0.000)
Supervisor Dummy			0.994				0.979	1.003	1.020
			(0.749)				(0.281)	(0.874)	(0.321)
Manager Dummy			0.932***				0.951***	0.964**	0.992
			(0.000)				(0.007)	(0.048)	(0.682)
Travel.purpose cashe			0.759***				0.808***	0.807***	0.802***
			(0.000)				(0.000)	(0.000)	(0.000)
EMI.purpose cashe			0.838***				0.842***	0.863***	0.869***
			(0.000)				(0.000)	(0.000)	(0.000)
purchase.purpose cashe			0.816***				0.845***	0.843***	0.846***
			(0.000)				(0.000)	(0.000)	(0.000)
Loanrepayment.purpose cashe			0.806***				0.817***	0.836***	0.840***
			(0.000)				(0.000)	(0.000)	(0.000)
Other purpose.purpose cashe			0.861***				0.869***	0.864***	0.851***
			(0.000)				(0.000)	(0.000)	(0.000)
Log no of SMS				0.969***	0.968***	0.972***	0.953***	0.957***	0.958***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log No of Contacts				0.964***	0.966***	0.976***	0.965***	0.971***	0.968***
				(0.000)	(0.000)	(0.007)	(0.000)	(0.002)	(0.001)
Log no of Apps				0.659***	0.653***	0.656***	0.632***	0.635***	0.636***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log Callog				0.917***	0.913***	0.915***	0.921***	0.922***	0.925***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dating App				1.246***	1.252***	1.218***	1.209***	1.187***	1.161***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Finsavy App				0.709***	0.706***	0.745***	0.753***	0.802***	0.816***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Socialconnect App				1.331***	1.233***	1.288***	1.358***	1.457***	1.661***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Travel App				1.034*	1.041**	1.038**	0.915***	0.911***	0.907***
				(0.052)	(0.019)	(0.031)	(0.000)	(0.000)	(0.000)
Mloan App				1.002	1.002	1.003	0.976	0.978	0.977
				(0.908)	(0.892)	(0.822)	(0.115)	(0.162)	(0.153)
Facebook status				1.091***	1.095***	1.099***	1.094***	1.096***	1.086***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Linkedin status				1.270***	1.256***	1.251***	1.196***	1.169***	1.172***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
IOS Dummy					0.495***	0.511***	0.427***	0.444***	0.458***
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.321***	0.256***	8.816***	1.760***	2.031***	2.979***	109.509***	112.540***	142.081***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
State Fixed Effects	N	N	N	N	N	N	N	N	Y
Observations	219,219	184,423	184,423	189,295	189,295	184,423	189,295	184,423	180,701
Pseudo R-squared	0.00417	0.00238	0.0903	0.0207	0.0222	0.0225	0.111	0.113	0.115
AUC	0.601	0.586	0.723	0.604	0.607	0.608	0.742	0.744	0.746

Table 4: **Predicting loan defaults (subsample without credit score)**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, and customer characteristics and likelihood of default using the sample of observations with no credit bureau score available. The dependent variable, Default takes the value one for loans that are delinquent and zero otherwise. The specification in column (1) includes customer and loan characteristics. Column (2) includes the digital footprint variables for the same sample. Column (3) includes loan and customer characteristics with digital footprint variables. Standard errors are clustered at the state level. (∗∗∗), (∗∗), (∗) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Odds Ratio (1)	Odds Ratio (2)	Odds Ratio (3)
Log of Salary	0.067***		0.068***
	(0.000)		(0.000)
Log Loan Amount	10.639***		12.061***
	(0.000)		(0.000)
Log Age	0.504***		0.410***
	(0.000)		(0.000)
High School Dummy	0.799***		0.846***
	(0.000)		(0.004)
College Dummy	0.655***		0.690***
	(0.000)		(0.000)
Supervisor	0.816***		0.870***
	(0.000)		(0.004)
Manager	0.883***		0.977
	(0.003)		(0.610)
Travel.purpose cashe	0.883		0.713***
	(0.113)		(0.007)
EMI.purpose cashe	0.812***		0.897
	(0.006)		(0.160)
purchase.purpose cashe	0.875**		0.907
	(0.032)		(0.133)
Loanrepayment.purpose cashe	0.608***		0.641***
	(0.000)		(0.000)
Other purpose.purpose cashe	0.963		1.015
	(0.402)		(0.756)
Log no of SMS		0.973***	0.953***
		(0.001)	(0.000)
Log No of Contacts		0.979	0.968
		(0.298)	(0.156)
Log no of Apps		0.875***	0.766***
		(0.000)	(0.000)
Log Callog		0.934***	0.942***
		(0.000)	(0.000)
Finsavy App		0.260***	0.265***
		(0.000)	(0.000)
Socialconnect App		8.625***	12.019***
		(0.000)	(0.000)
Travel App		0.927	1.211*
		(0.177)	(0.076)
Mloan App		0.957	0.852*
		(0.582)	(0.092)
Facebook status		0.919**	1.019
		(0.039)	(0.679)
Linkedin status		1.103	1.165
		(0.413)	(0.257)
IOS Dummy		0.846	0.616***
		(0.203)	(0.001)
Constant	181.810***	0.336***	482.318***
	(0.000)	(0.000)	(0.000)
Observations	47,152	45,473	45,425
Pseudo R2	0.0205	0.0237	0.0242
AUC	0.785	0.578	0.809

Table 5: **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	Odds Ratio (1)
Log of cibil	0.872*** (0.010)
Facebook status	1.098*** (0.019)
Linkedin status	1.165*** (0.057)
Log of Salary	0.264*** (0.007)
Log Loan Amount	4.268*** (0.070)
Log Age	0.357*** (0.018)
High School Dummy	0.845*** (0.020)
College Dummy	0.735*** (0.020)
Travel.purpose cashe	0.618* (0.167)
EMI.purpose cashe	0.754 (0.180)
purchase.purpose cashe	0.590** (0.121)
Loanrepayment.purpose cashe	0.403*** (0.098)
Other purpose.purpose cashe	1.241 (0.179)
Log no of SMS	0.957*** (0.004)
Log No of Contacts	0.969*** (0.009)
Log no of Apps	0.635*** (0.008)
Log Callog	0.923*** (0.005)
Dating App	1.245** (0.118)
Finsavy App	0.694*** (0.038)
Socialconnect App	2.084*** (0.217)
Travel App	0.880*** (0.032)
Mloan App	0.862*** (0.028)
IOS Dummy	0.451*** (0.025)
Travel.purpose cashe x Dating App	0.756* (0.123)
EMI.purpose cashe x Dating App	1.038 (0.181)
purchase.purpose cashe x Dating App	0.838 (0.127)
Loanrepayment.purpose cashe x Dating App	0.873 (0.159)
Other purpose.purpose cashe x Dating App	1.048 (0.125)
Travel.purpose cashe x Finsavy App	1.276* (0.167)
EMI.purpose cashe x Finsavy App	1.346** (0.182)
purchase.purpose cashe x Finsavy App	0.998 (0.100)
Loanrepayment.purpose cashe x Finsavy App	1.563*** (0.212)
Other purpose.purpose cashe x Finsavy App	1.232*** (0.088)
Travel.purpose cashe x Socialconnect App	1.157 (0.283)
EMI.purpose cashe x Socialconnect App	0.931 (0.217)
purchase.purpose cashe x Socialconnect App	1.120 (0.212)
Loanrepayment.purpose cashe x Socialconnect App	0.935 (0.197)
Other purpose.purpose cashe x Socialconnect App	0.412*** (0.052)

Travel.purpose cashe x Travel App	1.000
	(0.000)
EMI.purpose cashe x Travel App	1.018
	(0.075)
purchase.purpose cashe x Travel App	1.117*
	(0.068)
Loanrepayment.purpose cashe x Travel App	1.002
	(0.073)
Other purpose.purpose cashe x Travel App	1.055
	(0.048)
Travel.purpose cashe x Mloan App	1.012
	(0.068)
EMI.purpose cashe x Mloan App	1.196***
	(0.078)
purchase.purpose cashe x Mloan App	1.134**
	(0.062)
Loanrepayment.purpose cashe x Mloan App	1.282***
	(0.083)
Other purpose.purpose cashe x Mloan App	1.205***
	(0.049)
Travel.purpose cashe x Log of CIBIL	0.989
	(0.024)
EMI.purpose cashe x Log of CIBIL	0.972
	(0.021)
purchase.purpose cashe x Log of CIBIL	1.018
	(0.020)
Loanrepayment.purpose cashe x Log of CIBIL	1.042
	(0.026)
Other purpose.purpose cashe x Log of CIBIL	1.028*
	(0.015)
Constant	119.510***
	(31.291)
Financial Variables	Y
Digital Variables	Y
Observations	184,423
Pseudo R-squared	0.114
AUC	0.745

Table 6: **Summary Statistics of Call Logs and Financial Transactions**
This table reports summary statistics on call log variables. Columns 1-3 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.

Panel A: Call logs Metrics

Call Log Metric	Default (1)	Not Default (2)	Difference (3)
First 15 days: Per day Per person Avg No. of Incoming calls	1.53	1.49	-0.045***
First 15 days: Per day Per person Avg No. of Outgoing calls	2.14	2.03	-0.118***
First 15 days: Per day Per person Avg No. of Missed calls	1.57	1.48	-0.095***
First 15 days: Per day Per person Avg Duration of Incoming calls	156.14	157.55	1.40
First 15 days: Per day Per person Avg Duration of Outgoing calls	154.19	154.80	0.617
First 15 days: Per day No. of persons called	15.00	13.60	-1.40***
Past days: Per day Per person Avg No. of Incoming calls	1.58	1.52	-.051***
Past days: Per day Per person Avg No. of Outgoing calls	2.22	2.10	-.124***
Past days: Per day Per person Avg No. of Missed calls	1.61	1.51	-.099***
Past days: Per day Per person Avg Duration of Incoming calls	167.05	560.02	392.96***
Past days: Per day Per person Avg Duration of Outgoing calls	194.89	167.28	-27.61***
Past days: Per day No. of persons called	15.69	14.32	-1.37***
First 15 days: Per day Total No. of Incoming calls	10.97	9.73	-1.23***
First 15 days: Per day Total No. of Outgoing calls	20.76	17.53	-3.23***
First 15 days: Per day Total No. of Missed calls	7.44	5.80	-1.63***
First 15 days: Per day Total Duration of Incoming calls	1023.86	943.72	-80.13***
First 15 days: Per day Total Duration of Outgoing calls	1346.78	1205.45	-141.32***
Past days: Per day Total No. of Incoming calls	11.61	10.44	-1.172***
Past days: Per day Total No. of Outgoing calls	22.45	19.13	-3.31***
Past days: Per day Total No. of Missed calls	7.84	6.16	-1.67***
Past days: Per day Total Duration of Incoming calls	1113.62	1415.59	301.97 ***
Past days: Per day Total Duration of Outgoing calls	1561.83	1360.44	-201.38***
First 15 days: HHI of No. of Incoming calls	1049.70	890.56	-159.14***
First 15 days: HHI of No. of Outgoing calls	965.42	835.32	-130.09***
First 15 days: HHI of Total Duration of Incoming calls	1766.61	1597.18	-169.42***
First 15 days: HHI of Total Duration of Outgoing calls	1805.54	1681.39	-124.14***
First 15 days:HHI of No. of Missed calls	1430.33	1265.01	-165.31***
Past days: HHI of No. of Incoming calls	202.09	123.67	-78.41***
Past days: HHI of No. of Outgoing calls	201.19	128.40	-72.79***
Past days: HHI of Total Duration of Incoming calls	467.19	307.44	-159.75***
Past days: HHI of Total Duration of Outgoing calls	499.17	347.80	-151.36***
Past days:HHI of No. of Missed calls	291.76	176.06	-115.70***
N	17,095	89,052	
Panel B: Financial Transaction Metrics			
Debits to credits ratio	0.699	0.707	-0.007
# of Transactions	169.09	159.65	-9.44***
Expenditure to Income ratio	101.51	101.75	-0.321
Avg 2 Month Appreciation in Balance	411.90	-855.62	-1267.22
N	1,189	15,299	

Table 7: **Loan Defaults and Call Logs with ex-ante measures**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, customer characteristics and call logs 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 (Log of CIBIL) as the explanatory variable. Column (2) includes only call log variables. Column (3) includes call logs with digital footprint variables. Column (4) includes call logs, digital footprints and credit bureau score. Column (5) includes CIBIL score, loan characteristics and customer characteristics. Column (6) includes all variables of loan and customer characteristics, digital footprints and call logs excluding CIBIL score. Column (7) includes all variables including the CIBIL score. Standard errors are clustered at the state level. (∗ ∗ ∗), (∗∗), (∗) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Odds Ratio (1)	Odds Ratio (2)	Odds Ratio (3)	Odds Ratio (4)	Odds Ratio (5)	Odds Ratio (6)	Odds Ratio (7)
Log of cibil	0.889***			0.911***	0.872***		0.895***
	(0.000)			(0.000)	(0.000)		(0.000)
Past days: Per day Per person Avg No. of Incoming calls		1.080***	1.089***	1.082***		1.103***	1.099***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Past days: Per day Per person Avg No. of Outgoing calls		1.017	1.015	1.012		1.001	0.995
		(0.334)	(0.283)	(0.359)		(0.963)	(0.705)
Past days: Per day Per person Avg No. of Missed calls		0.950***	0.947***	0.945***		0.970	0.967
		(0.000)	(0.000)	(0.000)		(0.218)	(0.112)
Past days: Per day Per person Avg Duration of Incoming calls		0.000***	0.000***	0.000***		0.000***	0.000***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Past days: Per day Per person Avg Duration of Outgoing calls		0.916	0.942	0.962		0.897	0.961
		(0.733)	(0.718)	(0.755)		(0.592)	(0.717)
Past days: Per day No. of persons called		0.828***	0.888***	0.894***		0.923***	0.928***
		(0.000)	(0.000)	(0.000)		(0.005)	(0.008)
Past days: Per day Total Duration of Incoming calls		0.020	0.826	0.606		0.002*	0.002*
		(0.328)	(0.958)	(0.889)		(0.091)	(0.076)
Past days: Per day Total No. of Incoming calls		0.990	0.983	0.990		0.982	0.989
		(0.688)	(0.455)	(0.665)		(0.455)	(0.634)
Past days: Per day Total No. of Outgoing calls		1.368***	1.282***	1.256***		1.329***	1.309***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Past days: Per day Total Duration of Outgoing calls		0.776**	0.825**	0.858**		0.765***	0.774***
		(0.029)	(0.021)	(0.030)		(0.006)	(0.000)
Past days: Per day Total No. of Missed calls		1.380***	1.353***	1.353***		1.326***	1.327***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Past days: HHI of No. of Incoming calls		0.820***	0.825***	0.824***		0.827***	0.827***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Past days: HHI of No. of Outgoing calls		1.013	1.014	1.021		0.990	0.995
		(0.783)	(0.767)	(0.658)		(0.848)	(0.925)
Past days: HHI of Total Duration of Incoming calls		1.317***	1.297***	1.299***		1.306***	1.308***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Past days: HHI of Total Duration of Outgoing calls		1.230**	1.185*	1.164*		1.223**	1.201*
		(0.017)	(0.056)	(0.100)		(0.047)	(0.081)
Past days: HHI of No. of Missed calls		1.093***	1.078***	1.076***		1.087***	1.085***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Constant	0.312***	0.117***	1.288**	2.073***	24.753***	27.480***	35.856***
	(0.000)	(0.000)	(0.016)	(0.000)	(0.000)	(0.000)	(0.000)
Customer Characteristics	N	N	N	N	Y	Y	Y
Financial Variables	N	N	N	N	Y	Y	Y
Digital Variables	N	N	Y	Y	N	Y	Y
Observations	144,103	147,223	147,223	144,103	144,103	147,223	144,103
Pseudo R-squared	0.00279	0.0361	0.0529	0.0521	0.0836	0.135	0.135
AUC	0.584	0.644	0.662	0.661	0.715	0.762	0.762

Table 8: **Loan Defaults and Financial Transactions**

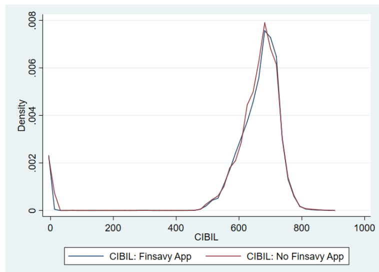
This table reports the estimates from our logit regressions examining the relationship between Financial transactions, digital footprint variables, loan-customer characteristics, call logs and likelihood of default. The dependent variable, Default takes the value one for loans that are delinquent and zero otherwise. The specification in column (1) includes variables correponding to only Call logs and Digital Footprints. Column (2) includes Call logs, Digital footprints and credit bureau score. Column (3) includes the credit bureau score (Log of CIBIL) with call logs, Digital footprints, customer and loan characteristics. Column (4) includes only the Financial Transactions. Column (5) includes Financial Transactions with call logs and digital footprints. Column (6) includes Financial transactions, call logs, digital footprints with credit bureau score. Column (7) includes Financial transactions, call logs, digital footprint variables, credit bureau score along with customer and loan characteristics. Standard errors are clustered at the state level. $(***)$, $(**)$, $(*)$ denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Odds Ratio (1)	Odds Ratio (2)	Odds Ratio (3)	Odds Ratio (4)	Odds Ratio (5)	Odds Ratio (6)	Odds Ratio (7)
Debit to credit ratio				1.026	1.021	1.024	1.040
				(0.571)	(0.671)	(0.614)	(0.488)
# of Transactions				1.141***	1.128***	1.129***	1.009
				(0.000)	(0.002)	(0.002)	(0.846)
Log Expenditure to Income ratio				1.083*	1.073	1.074	1.055
				(0.099)	(0.182)	(0.177)	(0.338)
Avg 2 Month Appreciation in Balance				1.025	1.027	1.025	1.012
				(0.659)	(0.623)	(0.654)	(0.699)
Log of cibil		1.150**	1.036			1.151**	1.041
		(0.010)	(0.571)			(0.010)	(0.520)
Log of Salary			0.045***				0.045***
			(0.000)				(0.000)
Log Loan Amount			39.857***				40.609***
			(0.000)				(0.000)
Log Age			0.196***				0.191***
			(0.000)				(0.000)
High School Dummy			1.095				1.128
			(0.569)				(0.465)
College Dummy			1.144				1.170
			(0.423)				(0.367)
Supervisor Dummy			1.177				1.174
			(0.174)				(0.184)
Manager Dummy			1.079				1.067
			(0.480)				(0.549)
Travel.purpose cashe			1.077				1.107
			(0.661)				(0.551)
EMI.purpose cashe			1.041				1.102
			(0.852)				(0.653)
purchase.purpose cashe			1.119				1.115
			(0.439)				(0.462)
Loanrepayment.purpose cashe			1.026				1.040
			(0.870)				(0.805)
Other purpose.purpose cashe			1.235*				1.205*
			(0.050)				(0.092)
Log no of SMS	0.998	1.001	0.974		0.992	0.995	0.972
	(0.903)	(0.963)	(0.247)		(0.664)	(0.794)	(0.222)
Log No of Contacts	1.003	0.996	1.006		1.001	0.993	1.013
	(0.931)	(0.914)	(0.903)		(0.987)	(0.856)	(0.807)
Log no of Apps	0.931	0.902	0.739***		0.934	0.904	0.744***
	(0.262)	(0.118)	(0.000)		(0.290)	(0.128)	(0.000)
Log Callog	0.946**	0.955*	0.987		0.945**	0.954*	0.995
	(0.048)	(0.099)	(0.676)		(0.043)	(0.090)	(0.869)
Dating App	1.456**	1.493**	1.900***		1.450**	1.486**	1.996***
	(0.029)	(0.020)	(0.001)		(0.031)	(0.022)	(0.000)
Finsavy App	0.976	1.004	0.956		0.998	1.027	0.996
	(0.891)	(0.980)	(0.842)		(0.990)	(0.886)	(0.987)
Socialconnect App	1.047	1.051	1.327		1.189	1.197	1.436

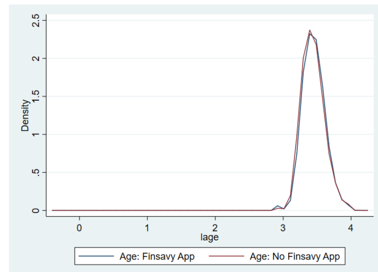
	(0.907)	(0.899)	(0.442)		(0.671)	(0.661)	(0.343)
Travel App	1.301**	1.305**	1.164		1.224*	1.228*	1.090
	(0.014)	(0.013)	(0.235)		(0.062)	(0.059)	(0.500)
Mloan App	1.448***	1.473***	1.512***		1.417***	1.443***	1.525***
	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Facebook status	1.104	1.113	1.073		1.096	1.104	1.056
	(0.205)	(0.175)	(0.425)		(0.248)	(0.215)	(0.552)
Linkedin status	0.736	0.735	0.698		0.714	0.711	0.683
	(0.219)	(0.217)	(0.193)		(0.189)	(0.183)	(0.194)
IOS Dummy	1.629***	1.555***	1.201		1.558***	1.483**	1.157
	(0.004)	(0.010)	(0.383)		(0.010)	(0.024)	(0.488)
Past days: Per day Per person Avg No. of Incoming calls	0.869	0.856	0.954		0.863	0.849	0.912
	(0.162)	(0.122)	(0.688)		(0.147)	(0.109)	(0.422)
Past days: Per day Per person Avg No. of Outgoing calls	1.039	1.056	1.076		1.065	1.087	1.262**
	(0.634)	(0.490)	(0.379)		(0.477)	(0.345)	(0.018)
Past days: Per day Per person Avg No. of Missed calls	0.948	0.944	1.090		0.962	0.958	1.154
	(0.536)	(0.502)	(0.423)		(0.660)	(0.624)	(0.175)
Past days: Per day Per person Avg Duration of Incoming calls	0.049*	0.044**	0.000		0.053**	0.047**	0.000
	(0.062)	(0.037)	(0.191)		(0.034)	(0.021)	(0.383)
Past days: Per day Per person Avg Duration of Outgoing calls	0.384	0.372	0.232		0.406	0.395	0.221
	(0.250)	(0.245)	(0.112)		(0.281)	(0.278)	(0.102)
Past days: Per day No. of persons called	0.748*	0.762	0.864		0.807	0.828	1.079
	(0.087)	(0.109)	(0.416)		(0.234)	(0.297)	(0.695)
Past days: Log of Per day Total Duration of Incoming calls	1.080	1.093	1.129		1.061	1.073	1.060
	(0.286)	(0.215)	(0.334)		(0.421)	(0.342)	(0.638)
Past days: Per day Total No. of Incoming calls	1.072	1.086	0.929		1.062	1.078	0.933
	(0.547)	(0.473)	(0.637)		(0.605)	(0.519)	(0.656)
Past days: Per day Total No. of Outgoing calls	1.235	1.205	1.304		1.168	1.133	1.107
	(0.198)	(0.262)	(0.152)		(0.363)	(0.471)	(0.597)
Past days: Per day Total Duration of Outgoing calls	1.483	1.569	1.403		1.581	1.677	1.456
	(0.428)	(0.379)	(0.528)		(0.350)	(0.306)	(0.477)
Past days: Per day Total No. of Missed calls	1.238***	1.218***	1.135		1.212**	1.190**	1.063
	(0.004)	(0.009)	(0.205)		(0.010)	(0.024)	(0.546)
Past days:Log HHI of No. of Incoming calls	1.031	1.053	1.373***		1.067	1.092	1.450***
	(0.740)	(0.580)	(0.003)		(0.487)	(0.357)	(0.001)
Past days: HHI of No. of Outgoing calls	0.684	0.706	0.652		0.660	0.685	0.648
	(0.470)	(0.507)	(0.585)		(0.443)	(0.483)	(0.578)
Past days: HHI of Total Duration of Incoming calls	0.985	0.999	0.775		1.014	1.028	0.770
	(0.907)	(0.995)	(0.201)		(0.913)	(0.824)	(0.161)
Past days: HHI of Total Duration of Outgoing calls	2.064*	2.009*	3.112*		2.041*	1.983	3.058*
	(0.085)	(0.097)	(0.079)		(0.091)	(0.105)	(0.079)
Past days: HHI of No. of Missed calls	1.189	1.175	1.010		1.159	1.145	1.005
	(0.217)	(0.251)	(0.953)		(0.271)	(0.317)	(0.978)
Constant	0.097***	0.042***	0.061*	0.078***	0.097***	0.042***	0.049**
	(0.000)	(0.000)	(0.054)	(0.000)	(0.000)	(0.000)	(0.043)
Observations	11,991	11,866	11,855	12,316	11,566	11,443	11,432
Pseudo R2	0.0156	0.0171	0.310	0.00247	0.0174	0.0191	0.312
AUC	0.596	0.601	0.872	0.544	0.600	0.604	0.874

Figure 1: **Kernel density plots**

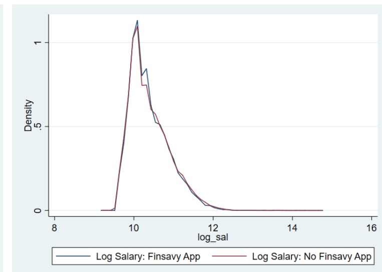
This figure plots the kernel density distribution of CIBIL score, Salary, and Age for customers with and without a financial application installed on their phones in subfigures (a), (b), and (c). Subfigures (d), (e), and (f) plot the kernel density distribution of CIBIL score, Salary, and Age for customers with and without a financial application installed on their phones



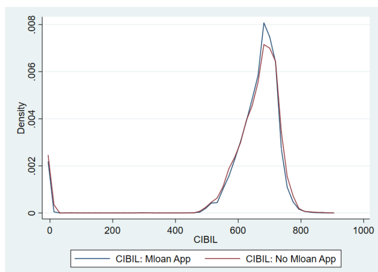
(a) CIBIL



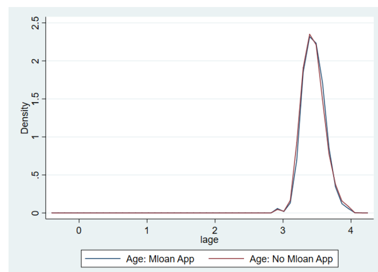
(b) Age



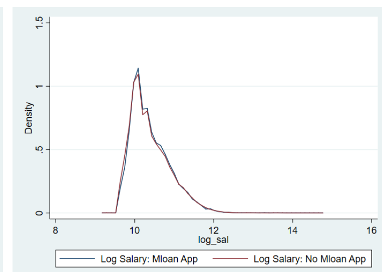
(c) Log(Salary)



(d) CIBIL



(e) Age



(f) Log(Salary)

Appendix A

Table A1: **Approval of Loans**

This table reports the estimates from our logit regressions examining the determinants of loan approval using the sample of observations with no credit bureau score available. 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) includes customer and loan characteristics. Column (2) includes the digital footprint variables for the same sample. Column (3) includes loan and customer characteristics with digital footprint variables. Standard errors are clustered at the state level. ($***$), ($**$), ($*$) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Odds Ratio	Odds Ratio	Odds Ratio
	(1)	(2)	(3)
Log of Salary	2.578***		2.344***
	(0.000)		(0.000)
Log Loan Amount	0.740***		0.709***
	(0.000)		(0.000)
Log Age	4.627***		6.636***
	(0.000)		(0.000)
High School Dummy	3.425***		3.051***
	(0.000)		(0.000)
College Dummy	4.419***		3.887***
	(0.000)		(0.000)
Supervisor	1.013		1.030
	(0.465)		(0.127)
Manager	0.980		0.946***
	(0.238)		(0.002)
Travel.purpose cashe	1.479***		1.415*
	(0.000)		(0.056)
EMI.purpose cashe	0.598***		0.648***
	(0.000)		(0.000)
purchase.purpose cashe	0.841***		0.809***
	(0.000)		(0.000)
Loanrepayment.purpose cashe	0.812***		0.813***
	(0.000)		(0.000)
Other purpose.purpose cashe	0.866***		0.879***
	(0.000)		(0.000)
Log no of SMS		1.148***	1.156***
		(0.000)	(0.000)
Log No of Contacts		1.379***	1.224***
		(0.000)	(0.000)

Log no of Apps		1.604***	1.580***
		(0.000)	(0.000)
Log Callog		1.017***	1.025***
		(0.003)	(0.000)
Finsavy App		1.237	1.413
		(0.282)	(0.122)
Socialconnect App		24.158***	30.966***
		(0.000)	(0.000)
Travel App		1.724***	0.958
		(0.000)	(0.809)
Mloan App		1.050	1.062
		(0.762)	(0.712)
Facebook status		0.650***	0.660***
		(0.000)	(0.000)
Linkedin status		0.970	0.825***
		(0.600)	(0.003)
IOS Dummy		4.231***	3.827***
		(0.000)	(0.000)
Constant	0.000***	0.007***	0.000***
	(0.000)	(0.000)	(0.000)
Observations	98,293	97,531	97,480
Pseudo R2	0.107	0.113	0.194
AUC	0.719	0.712	0.783

Table A2: **Default Heterogeneity by Credit Score**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, and customer characteristics and likelihood of default for customers in different terciles of the credit score distribution. The dependent variable, Default takes the value one for loans that are delinquent and zero otherwise. The specification includes all variables including the credit score, loan characteristics, customer characteristics, and digital mobile footprint variables. Standard errors are clustered at the state level. $(***)$, $(**)$, $(*)$ denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Default Regressions: Low CreditRating	Default Regressions: Medium CreditRating	Default Regressions: High CreditRating
	(1)	(2)	(3)
Log of Salary	0.287***	0.259***	0.220***
	(0.015)	(0.020)	(0.015)
Travel.purpose cashe	0.785***	0.906	0.781**
	(0.068)	(0.077)	(0.083)
EMI.purpose cashe	0.916	0.888*	0.876
	(0.081)	(0.059)	(0.099)
purchase.purpose cashe	0.795***	0.972	0.971
	(0.059)	(0.060)	(0.078)
Loanrepayment.purpose cashe	0.912	0.748***	0.819*
	(0.064)	(0.057)	(0.095)
Other purpose.purpose cashe	0.845***	0.910*	0.906
	(0.051)	(0.046)	(0.063)
Facebook status	1.007	1.231***	1.001
	(0.048)	(0.051)	(0.056)
Linkedin status	1.176*	1.111	1.003
	(0.103)	(0.130)	(0.158)
Log of cibil	0.938***	0.006***	3.177
	(0.011)	(0.005)	(2.263)
Log no of SMS	0.957***	0.957***	0.957***
	(0.010)	(0.009)	(0.010)
Log Loan Amount	4.156***	4.089***	4.687***
	(0.155)	(0.262)	(0.236)
Log Age	0.351***	0.417***	0.304***
	(0.067)	(0.056)	(0.052)
Log No of Contacts	0.972	0.978	0.954*
	(0.021)	(0.021)	(0.026)
Log no of Apps	0.670***	0.619***	0.637***
	(0.018)	(0.019)	(0.022)
Log Callog	0.916***	0.914***	0.950***
	(0.014)	(0.013)	(0.017)

Dating App	1.080	1.438***	1.298***
	(0.093)	(0.176)	(0.130)
Finsavy App	0.826***	0.717***	0.765***
	(0.054)	(0.082)	(0.076)
Socialconnect App	1.724***	2.011***	1.565***
	(0.165)	(0.290)	(0.189)
Travel App	0.918*	0.889*	0.883*
	(0.041)	(0.053)	(0.056)
Mloan App	0.959	0.946	1.118*
	(0.037)	(0.044)	(0.065)
IOS Dummy	0.377***	0.548***	0.504***
	(0.043)	(0.064)	(0.063)
High School Dummy	0.906	0.811***	0.770***
	(0.067)	(0.063)	(0.059)
College Dummy	0.827*	0.725***	0.643***
	(0.081)	(0.058)	(0.055)
Constant	44.626***	2.310e+16***	0.084
	(25.366)	-1.30E+17	(0.392)
Financial Variables	Y	Y	Y
Digital Variables	Y	Y	Y
Observations	62,276	66,377	52,176
Pseudo R-squared	0.112	0.108	0.114
AUC	0.738	0.743	0.749

Table A3: **Default Heterogeneity by Age**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, and customer characteristics and likelihood of default for customers in different terciles of the age distribution. The dependent variable, Default takes the value one for loans that are delinquent and zero otherwise. The specification includes all variables including the credit score, loan characteristics, customer characteristics, and digital mobile footprint variables. Standard errors are clustered at the state level. $(***)$, $(**)$, $(*)$ denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Default Regressions: Age 1Q	Default Regressions: Age 3Q
	(1)	(2)
Log of Salary	0.239***	0.293***
	(0.062)	(0.021)
Travel.purpose cashe	0.858	0.669***
	(0.138)	(0.058)
EMI.purpose cashe	0.840	0.935
	(0.094)	(0.100)
purchase.purpose cashe	0.925	0.796**
	(0.098)	(0.071)
Loanrepayment.purpose cashe	0.869	0.706***
	(0.114)	(0.060)
Other purpose.purpose cashe	0.841***	0.784***
	(0.056)	(0.065)
Facebook status	1.021	1.036
	(0.082)	(0.045)
Linkedin status	0.881	1.198
	(0.155)	(0.134)
Log of cibil	0.915***	0.848***
	(0.017)	(0.025)
Log no of SMS	0.945***	0.963***
	(0.014)	(0.010)
Log Loan Amount	4.585***	3.838***
	(0.378)	(0.199)
Log Age	0.772	1.456
	(0.305)	(0.418)
Log No of Contacts	1.037	0.962
	(0.033)	(0.024)
Log no of Apps	0.607***	0.658***
	(0.042)	(0.023)
Log Callog	0.955**	0.932***
	(0.019)	(0.014)

Dating App	1.053	1.307
	(0.089)	(0.236)
Finsavy App	0.832	0.714***
	(0.135)	(0.055)
Socialconnect App	1.944***	1.620***
	(0.396)	(0.220)
Travel App	0.861	0.921
	(0.079)	(0.046)
Mloan App	1.054	0.956
	(0.070)	(0.040)
IOS Dummy	0.243***	0.701**
	(0.075)	(0.114)
High School Dummy	0.894	0.811***
	(0.071)	(0.046)
College Dummy	0.782**	0.701***
	(0.092)	(0.047)
Constant	6.244	0.856
	(14.895)	(0.728)
Financial Variables	Y	Y
Digital Variables	Y	Y
Observations	20,853	60,847
Pseudo R-squared	0.123	0.106
AUC	0.747	0.738

Table A4: **Default Heterogeneity by Salary**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, and customer characteristics and likelihood of default for customers in different terciles of the salary distribution. The dependent variable, Default takes the value one for loans that are delinquent and zero otherwise. The specification includes all variables including the credit score, loan characteristics, customer characteristics, and digital mobile footprint variables. Standard errors are clustered at the state level. $(***)$, $(**)$, $(*)$ denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Default Regressions: Salary 1Q	Default Regressions: Salary 3Q
	(1)	(2)
Log of Salary	0.194***	0.332***
	(0.071)	(0.023)
Travel.purpose cashe	0.701***	0.758***
	(0.086)	(0.053)
EMI.purpose cashe	0.928	0.836*
	(0.117)	(0.079)
purchase.purpose cashe	1.000	0.809***
	(0.112)	(0.051)
Loanrepayment.purpose cashe	0.803	0.735***
	(0.118)	(0.060)
Other purpose.purpose cashe	0.889	0.873**
	(0.081)	(0.052)
Facebook status	1.129*	1.144***
	(0.075)	(0.046)
Linkedin status	0.831	1.129
	(0.221)	(0.115)
Log of cibil	0.893***	0.854***
	(0.016)	(0.024)
Log no of SMS	0.959***	0.975**
	(0.012)	(0.010)
Log Loan Amount	2.973***	4.235***
	(0.263)	(0.243)
Log Age	0.425***	0.324***
	(0.094)	(0.056)
Log No of Contacts	0.990	0.996
	(0.039)	(0.025)
Log no of Apps	0.635***	0.632***
	(0.033)	(0.023)
Log Callog	0.910***	0.941***
	(0.027)	(0.014)

Dating App	1.322	1.257***
	(0.491)	(0.085)
Finsavy App	0.761**	0.979
	(0.104)	(0.098)
Socialconnect App	1.905***	1.460***
	(0.415)	(0.171)
Travel App	0.942	0.898
	(0.066)	(0.070)
Mloan App	0.908	1.046
	(0.056)	(0.039)
IOS Dummy	0.183***	0.654***
	(0.048)	(0.055)
High School Dummy	0.834**	0.849**
	(0.069)	(0.059)
College Dummy	0.767**	0.746***
	(0.081)	(0.055)
Constant	32,408.998***	10.268***
	(117,792.210)	(5.971)
Financial Variables	Y	Y
Digital Variables	Y	Y
Observations	26,744	51,227
Pseudo R-squared	0.0652	0.109
AUC	0.6901	0.7404

Table A5: **Default Heterogeneity by Designation**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, and customer characteristics and likelihood of default for customers in three different employment category: workers, supervisors, and managers. The dependent variable, Default takes the value one for loans that are delinquent and zero otherwise. The specification includes all variables including the credit score, loan characteristics, customer characteristics, and digital mobile footprint variables. Standard errors are clustered at the state level. $(***)$, $(**)$, $(*)$ denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Default Regressions: Worker	Default Regressions: Supervisor	Default Regressions: Manager
	(1)	(2)	(3)
Log of Salary	0.254***	0.198***	0.299***
	(0.016)	(0.018)	(0.016)
Travel.purpose cashe	0.873*	0.696***	0.816**
	(0.065)	(0.089)	(0.067)
EMI.purpose cashe	0.918	0.839	0.887
	(0.056)	(0.091)	(0.099)
purchase.purpose cashe	0.960	0.727***	0.898
	(0.064)	(0.075)	(0.064)
Loanrepayment.purpose cashe	0.870*	0.685***	0.872**
	(0.066)	(0.080)	(0.061)
Other purpose.purpose cashe	0.901**	0.739***	0.928
	(0.042)	(0.070)	(0.051)
Facebook status	1.100**	1.087	1.095**
	(0.052)	(0.060)	(0.042)
Linkedin status	1.137	1.172	1.156
	(0.176)	(0.180)	(0.103)
Log of cibil	0.892***	0.883***	0.876***
	(0.021)	(0.019)	(0.015)
Log no of SMS	0.954***	0.951***	0.970***
	(0.007)	(0.013)	(0.011)
Log Loan Amount	4.199***	4.719***	4.148***
	(0.180)	(0.289)	(0.181)
Log Age	0.428***	0.458***	0.268***
	(0.081)	(0.127)	(0.030)
Log No of Contacts	0.991	0.941**	0.971
	(0.020)	(0.029)	(0.022)
Log no of Apps	0.652***	0.679***	0.595***
	(0.019)	(0.026)	(0.020)
Log Callog	0.931***	0.906***	0.929***

	(0.013)	(0.019)	(0.014)
Dating App	1.208**	1.498***	1.179
	(0.097)	(0.208)	(0.160)
Finsavy App	0.726***	0.710**	0.892
	(0.067)	(0.094)	(0.069)
Socialconnect App	1.867***	1.741***	1.533***
	(0.218)	(0.276)	(0.206)
Travel App	0.859***	0.919	0.997
	(0.040)	(0.057)	(0.050)
Mloan App	0.986	0.947	1.025
	(0.042)	(0.035)	(0.048)
IOS Dummy	0.365***	0.403***	0.617***
	(0.056)	(0.120)	(0.058)
High School Dummy	0.784***	0.840**	1.021
	(0.058)	(0.063)	(0.087)
College Dummy	0.667***	0.737***	0.894
	(0.058)	(0.064)	(0.085)
Constant	77.568***	462.753***	89.627***
	(58.078)	(435.993)	(44.164)
Financial Variables	Y	Y	Y
Digital Variables	Y	Y	Y
Observations	64,879	44,457	71,493
Pseudo R-squared	0.112	0.119	0.111
AUC	0.744	0.749	0.743

Table A6: Average and Total Call Log Metrics

The below metrics have been calculated for every customer in the Cashe Database. We divide the above metrics into ex-ante (all days before start date of loan) and ex-post (first 15 days after start date of loan) call logs. $C_{i,j}$ is the total number of calls made to the person i on j^{th} day. k is the number of contacts in the customer's contact list and n_j is the number of persons contacted on the j^{th} day.

Metric	Formula
First 15 day Average: Per day Per person	
First 15 days: Per day Per person Avg No. of Incoming calls	$\frac{\sum_{j=1}^{j=15} \frac{\sum_{i=1}^{n_j} C_{i,j}}{n_j}}{15}$
First 15 days: Per day Per person Avg No. of Outgoing calls	
First 15 days: Per day Per person Avg No. of Missed calls	
First 15 days: Per day Per person Avg Duration of Incoming calls	
First 15 days: Per day Per person Avg Duration of Outgoing calls	
Past History Average: Per day Per person	
Past days: Per day Per person Avg No. of Incoming calls	$\frac{\sum_{\forall j \leq 0} \frac{\sum_{i=1}^{n_j} C_{i,j}}{n_j}}{\sum_{\forall j \leq 0} 1}$
Past days: Per day Per person Avg No. of Outgoing calls	
Past days: Per day Per person Avg No. of Missed calls	
Past days: Per day Per person Avg Duration of Incoming calls	
Past days: Per day Per person Avg Duration of Outgoing calls	
First 15 day Average: Per day	
First 15 days: Per day No. of persons called	$\frac{\sum_{j=1}^{j=15} n_j}{15}$
First 15 days: Per day Total Duration of Incoming calls	$\frac{\sum_{j=1}^{j=15} \sum_{i=1}^{n_j} C_{i,j}}{15}$
First 15 days: Per day Total No. of Incoming calls	
First 15 days: Per day Total No. of Outgoing calls	
First 15 days: Per day Total Duration of Outgoing calls	
First 15 days: Per day Total No. of Missed calls	
Past History Average: Per day	
Past days: Per day No. of persons called	$\frac{\sum_{\forall j \leq 0} n_j}{\sum_{\forall j \leq 0} 1}$
Past days: Per day Total Duration of Incoming calls	$\frac{\sum_{\forall j \leq 0} \sum_{i=1}^{n_j} C_{i,j}}{\sum_{\forall j \leq 0} 1}$
Past days: Per day Total No. of Incoming calls	
Past days: Per day Total No. of Outgoing calls	
Past days: Per day Total Duration of Outgoing calls	
Past days: Per day Total No. of Missed calls	

Table A6: Herfindahl-Hirschman Call Log Index

The below metrics have been calculated for every customer in the Cashe Database. We divide the above metrics into ex-ante (all days before start date of loan) and ex-post (first 15 days after start date of loan) call logs. $C_{i,j}$ is the total number of calls made to the person i on j^{th} day. k is the number of contacts in the customer's contact list and n_j is the number of persons contacted on the j^{th} day.

Metric	Formula
First 15 days: Herfindahl-Hirschman Index	
First 15 days: HHI of No. of Incoming calls	$\sum_{i=1}^{i=k}\left[\frac{\frac{\sum_{j=1}^{j=15} C_{i,j}}{15}}{\sum_{i=1}^{i=k}\frac{\sum_{j=1}^{j=15} C_{i,j}}{15}}\times 100\right]^2$
First 15 days: HHI of No. of Outgoing calls	
First 15 days: HHI of Total Duration of Incoming calls	
First 15 days: HHI of Total Duration of Outgoing calls	
First 15 days: HHI of No. of Missed calls	
Past History : Herfindahl-Hirschman Index	
Past days: HHI of No. of Incoming calls	$\sum_{i=1}^{i=k}\left[\frac{\frac{\sum_{\forall j\leq 0} C_{i,j}}{\sum_{\forall j\leq 0} 1}}{\sum_{i=1}^{i=k}\frac{\sum_{\forall j\leq 0} C_{i,j}}{\sum_{\forall j\leq 0} 1}}\times 100\right]^2$
Past days: HHI of No. of Outgoing calls	
Past days: HHI of Total Duration of Incoming calls	
Past days: HHI of Total Duration of Outgoing calls	
Past days: HHI of No. of Missed calls	

Table A7: **Loan Defaults and Call Logs with ex-post and ex-ante measures**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, customer characteristics and call logs 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 (Log of CIBIL) as the explanatory variable. Column (2) includes credit bureau score with customer and loan characteristics. Column (3) includes only call log variables. Column (4) includes call logs with credit bureau score. Column (5) includes call logs with customer and loan characteristics. Column (6) includes call logs and digital footprint variables. Column (7) includes all call variables, loan characteristics, customer characteristics and digital footprint variables but not the CIBIL score. Column (8) includes all variables including the CIBIL score. Standard errors are clustered at the state level. $(***)$, $(**)$, $(*)$ denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of cibil	0.900***	0.891***		0.915***				0.913***
	(0.000)	(0.000)		(0.000)				(0.000)
First 15 days: Per day Per person Avg No. of Incoming calls			1.031**	1.022	1.034**	1.030**	1.032**	1.025
			(0.040)	(0.147)	(0.034)	(0.048)	(0.044)	(0.125)
First 15 days: Per day Per person Avg No. of Outgoing calls			1.014	1.014	1.007	1.009	1.002	1.003
			(0.333)	(0.350)	(0.646)	(0.539)	(0.899)	(0.838)
First 15 days: Per day Per person Avg No. of Missed calls			0.981	0.978	0.994	0.978	0.991	0.989
			(0.137)	(0.101)	(0.656)	(0.102)	(0.541)	(0.447)
First 15 days: Per day Per person Avg Duration of Incoming calls			0.928***	0.937***	0.931***	0.927***	0.931***	0.938***
			(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.004)
First 15 days: Per day Per person Avg Duration of Outgoing calls			0.991	0.996	0.992	0.998	0.997	1.001
			(0.613)	(0.813)	(0.634)	(0.885)	(0.852)	(0.953)
First 15 days: Per day No. of persons called			0.998	0.997	0.978	1.009	0.989	0.987
			(0.945)	(0.918)	(0.476)	(0.768)	(0.720)	(0.685)
Past days: Per day Per person Avg No. of Incoming calls			1.050***	1.049***	1.053***	1.059***	1.060***	1.062***
			(0.003)	(0.004)	(0.003)	(0.000)	(0.000)	(0.000)
Past days: Per day Per person Avg No. of Outgoing calls			1.013	1.009	1.003	1.015	1.006	1.000
			(0.398)	(0.579)	(0.873)	(0.315)	(0.708)	(0.990)
Past days: Per day Per person Avg No. of Missed calls			0.961***	0.960***	0.973	0.958***	0.969**	0.969***
			(0.000)	(0.000)	(0.244)	(0.000)	(0.014)	(0.010)
Past days: Per day Per person Avg Duration of Incoming calls			0.000	0.000	0.000	0.000	0.000	0.000
			(0.229)	(0.210)	(0.376)	(0.113)	(0.272)	(0.246)
Past days: Per day Per person Avg Duration of Outgoing calls			1.025	1.012	1.023	1.010	1.006	0.995
			(0.803)	(0.896)	(0.806)	(0.880)	(0.935)	(0.944)
Past days: Per day No. of persons called			0.881***	0.889***	0.902***	0.921**	0.951	0.955
			(0.000)	(0.001)	(0.005)	(0.018)	(0.164)	(0.210)
First 15 days: Per day Total Duration of Incoming calls			1.039	1.029	1.037	1.048*	1.048*	1.038
			(0.121)	(0.261)	(0.144)	(0.057)	(0.063)	(0.136)
First 15 days: Per day Total No. of Incoming calls			0.976	0.990	0.974	0.971	0.971	0.980
			(0.385)	(0.721)	(0.362)	(0.285)	(0.298)	(0.481)

First 15 days: Per day Total No. of Outgoing calls			1.088***	1.087***	1.115***	1.073***	1.093***	1.092***
			(0.001)	(0.002)	(0.000)	(0.007)	(0.001)	(0.002)
First 15 days: Per day Total Duration of Outgoing calls			0.992	0.988	0.999	0.990	0.999	0.995
			(0.711)	(0.565)	(0.977)	(0.628)	(0.973)	(0.824)
First 15 days: Per day Total No. of Missed calls			1.077***	1.076***	1.074***	1.071***	1.067***	1.067***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past days: Per day Total Duration of Incoming calls			0.001	0.001	0.000**	0.003	0.000**	0.000**
			(0.114)	(0.112)	(0.014)	(0.212)	(0.028)	(0.025)
Past days: Per day Total No. of Incoming calls			1.003	1.006	1.016	1.007	1.018	1.023
			(0.925)	(0.843)	(0.612)	(0.833)	(0.570)	(0.486)
Past days: Per day Total No. of Outgoing calls			1.279***	1.248***	1.298***	1.219***	1.237***	1.215***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past days: Per day Total Duration of Outgoing calls			0.789***	0.847**	0.732***	0.849**	0.774***	0.824***
			(0.002)	(0.027)	(0.000)	(0.020)	(0.000)	(0.009)
Past days: Per day Total No. of Missed calls			1.291***	1.292***	1.264***	1.273***	1.245***	1.245***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First 15 days: HHI of No. of Incoming calls			1.003	0.995	1.019	1.006	1.029	1.020
			(0.917)	(0.840)	(0.460)	(0.825)	(0.275)	(0.467)
First 15 days: HHI of No. of Outgoing calls			0.993	0.991	0.976	0.994	0.979	0.977
			(0.765)	(0.704)	(0.343)	(0.816)	(0.400)	(0.378)
First 15 days: HHI of Total Duration of Incoming calls			1.019	1.028	0.994	1.004	0.972	0.981
			(0.422)	(0.248)	(0.801)	(0.877)	(0.257)	(0.440)
First 15 days: HHI of Total Duration of Outgoing calls			1.684***	1.672***	1.502***	1.644***	1.442***	1.431***
			(0.000)	(0.000)	(0.001)	(0.000)	(0.003)	(0.004)
First 15 days: HHI of No. of Missed calls			1.066***	1.066***	1.057***	1.054***	1.044***	1.045***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.004)
Past days: HHI of No. of Incoming calls			0.854***	0.853***	0.865***	0.853***	0.870***	0.871***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past days: HHI of No. of Outgoing calls			1.064	1.076	1.062	1.058	1.051	1.058
			(0.360)	(0.280)	(0.439)	(0.404)	(0.529)	(0.483)
Past days: HHI of Total Duration of Incoming calls			1.285***	1.286***	1.289***	1.275***	1.272***	1.273***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past days: HHI of Total Duration of Outgoing calls			1.123	1.102	1.127	1.100	1.101	1.086
			(0.347)	(0.443)	(0.415)	(0.461)	(0.531)	(0.602)
Past days: HHI of No. of Missed calls			1.046***	1.042***	1.048***	1.039***	1.040**	1.038**
			(0.003)	(0.007)	(0.003)	(0.009)	(0.013)	(0.020)
Constant	0.359***	42.540***	0.161***	0.277***	5.377***	1.405**	38.185***	43.234***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)
Customer Characteristics	N	Y	N	N	Y	N	Y	Y
Financial Variables	N	Y	N	N	Y	N	Y	Y
Digital Variables	N	N	N	N	N	Y	Y	Y

Observations	97,069	97,069	97,105	94,962	97,105	97,105	97,105	94,962
Pseudo R2	0.00239	0.0781	0.0391	0.0393	0.111	0.0529	0.126	0.126
AUC	0.586	0.707	0.646	0.648	0.739	0.662	0.750	0.751

Table A8: **Predicting Loan Defaults with call logs (Subsample without credit score)**

This table reports the estimates from our logit regressions examining the relationship between digital footprint variables, loan, and customer characteristics and likelihood of default using the sample of observations with no credit bureau score available. The dependent variable, Default takes the value one for loans that are delinquent and zero otherwise. The specification in column (1) includes loan and customer characteristics. Column (2) includes the digital footprint variables for the same sample. Column (3) includes digital footprint variables with call log variables. Column (4) includes loan and customer characteristics with digital footprint variables and call logs. Standard errors are clustered at the state level. ($***$), ($**$), ($*$) denote statistical significance at 1%, 5%, and 10% levels respectively.

VARIABLES	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio
	(1)	(2)	(3)	(4)
Log of Salary	0.325***			0.445***
	(0.000)			(0.000)
Log Loan Amount	4.204***			3.661***
	(0.000)			(0.000)
Log Age	0.356***			0.213***
	(0.002)			(0.000)
High School Dummy	0.665***			0.882
	(0.003)			(0.432)
College Dummy	0.512***			0.711*
	(0.000)			(0.075)
Supervisor	1.152			1.074
	(0.208)			(0.596)
Manager	1.107			1.155
	(0.373)			(0.285)
Travel.purpose cashe	0.707**			0.694*
	(0.049)			(0.081)
EMI.purpose cashe	0.543***			0.503***
	(0.003)			(0.004)
purchase.purpose cashe	0.781			0.732*
	(0.107)			(0.070)
Loanrepayment.purpose cashe	0.365***			0.430***
	(0.000)			(0.000)
Other purpose.purpose cashe	0.889			0.929
	(0.293)			(0.578)
Log no of SMS		0.915***	0.910***	0.926***
		(0.000)	(0.000)	(0.005)
Log No of Contacts		0.834***	0.826***	0.832***
		(0.002)	(0.002)	(0.006)
Log no of Apps		0.671***	0.650***	0.644***
		(0.000)	(0.000)	(0.000)
Log Callog		0.908**	0.971	1.018

	(0.023)	(0.522)	(0.723)
Finsavy App	0.397***	0.485***	0.467***
	(0.000)	(0.000)	(0.000)
Socialconnect App	0.388***	0.360***	0.426**
	(0.003)	(0.001)	(0.026)
Travel App	1.276**	1.369***	1.332**
	(0.024)	(0.007)	(0.031)
Mloan App	0.824**	0.822**	0.726***
	(0.031)	(0.039)	(0.001)
Facebook status	1.236**	1.312**	1.503***
	(0.048)	(0.016)	(0.001)
Linkedin status	1.062	1.224	1.081
	(0.842)	(0.535)	(0.800)
IOS Dummy	0.541	0.546	0.447*
	(0.184)	(0.132)	(0.063)
Past days: Per day Per person Avg No. of Incoming calls		1.150**	1.118
		(0.050)	(0.135)
Past days: Per day Per person Avg No. of Outgoing calls		1.062	1.089
		(0.347)	(0.235)
Past days: Per day Per person Avg No. of Missed calls		0.975	1.036
		(0.730)	(0.633)
Past days: Per day Per person Avg Duration of Incoming calls		0.000	0.000
		(0.216)	(0.353)
Past days: Per day Per person Avg Duration of Outgoing calls		0.544	0.279
		(0.445)	(0.202)
Past days: Per day No. of persons called		0.838	0.824
		(0.231)	(0.226)
Past days: Per day Total Duration of Incoming calls		324,000,000,000,000,000	104,800,000
		(0.204)	(0.547)
Past days: Per day Total No. of Incoming calls		0.904	1.011
		(0.505)	(0.945)
Past days: Per day Total No. of Outgoing calls		1.547***	1.579***
		(0.000)	(0.001)
Past days: Per day Total Duration of Outgoing calls		0.280**	0.316**
		(0.013)	(0.044)
Past days: Per day Total No. of Missed calls		1.249***	1.180**
		(0.000)	(0.017)
Past days: HHI of No. of Incoming calls		0.886	0.871
		(0.295)	(0.241)
Past days: HHI of No. of Outgoing calls		0.719*	0.756
		(0.077)	(0.143)

Past days: HHI of Total Duration of Incoming calls			1.191	1.187
			(0.127)	(0.174)
Past days: HHI of Total Duration of Outgoing calls			1.879***	1.683**
			(0.002)	(0.014)
Past days: HHI of No. of Missed calls			1.087	1.065
			(0.419)	(0.538)
Constant	1.873	62.533***	20.748***	61.398**
	(0.621)	(0.000)	(0.000)	(0.019)
Observations	3,252	3,155	3,133	3,132
Pseudo R-squared	0.109	0.0746	0.126	0.209
AUC	0.732	0.675	0.729	0.801

Variable Definitions:

SNo.	Variable type	Variable name	Variable definition
1	Financial Transactions	Debit to Credit ratio	Ratio of total debit to total credit in 3-month window before start of loan.
2		No. of Transactions	No of transactions in 3-month window before start of loan.
3		Log Expenditure to Income ratio	Log of ratio of Expense to Income for the 3-month window before start of loan.
4		Avg 2 Month Appreciation of Account Balance	Avg increase in account balance between account snapshots in the 2 months before start of loan. Data consists of snapshots spaced out at 10 day gaps. Variable captures the pace at which money is put into the account (Time2-Time1). For someone whose savings increase as the month progresses, the variable should be positive.
5	CIBIL	Log of cibil	Log of Credit bureau score
6	Customer Characteristics	Log of Salary	Log of customer's salary
7		Log Age	Log of customer's age.
8		High School Dummy	Dummy takes value 1 if customer's highest qualification is High School.
9		College Dummy	Dummy takes value 1 if customer's highest qualification is College.
10		Supervisor Dummy	Dummy takes value 1 if customer's designation falls in the supervisor category.
11		Manager Dummy	Dummy takes value 1 if customer's designation falls in the manager category.
12	Loan Characteristics	Log Loan Amount	Log of Loan Amount of the loan.
13		Travel.purpose cashe	Dummy takes value 1 if purpose of loan is travel.

14		EMI.purpose cashe	Dummy takes value 1 if purpose of loan is to pay EMI.
15		Loan repayment.purpose cashe	Dummy takes 1 if purpose of loan is to pay another loan.
16		Other purpose.purpose cashe	Dummy takes 1 if purpose of loan is other than travel, EMI, loan repayment and medical.
17	Digital variables	Log no of SMS	Log of Total No. of SMS.
18		Log no of Contacts	Log of No. of people in contact list.
19		Log no of Apps	Log of no. of applications in phone.
20		Log Callog	Log of Total No. of calls.
21		Dating App	Dummy takes 1 if customer has a dating app.
22		Finsavy App	Dummy takes 1 if customer has a financial services app (stocks, banking, payment and wallet).
23		Socialconnect App	Dummy takes 1 if customer has a social connect app (messaging app, video streaming app, music streaming app, social network app, dating app, video call app).
24		Travel App	Dummy takes 1 if customer has a Travel app.
25		Mloan App	Dummy takes 1 if customer has another loan app.
26		Facebook Status	Dummy takes 1 if customer logged into Cashe app using Facebook.
27		Linkedin Status	Dummy takes 1 if customer logged into Cashe app using Linkedin.
28		IOS Dummy	Dummy takes 1 if customer has an Apple phone.
29	Call Log variables	Per day Per person Avg No. of Incoming calls	No. of incoming calls received from a person on average in a day.

30		Per day Per person Avg No. of Outgoing calls	No. of outgoing calls made to a person on average in a day.
31		Per day Per person Avg No. of Missed calls	No. of missed calls received from a person on average in a day.
32		Per day Per person Avg Duration of Incoming calls	Duration of incoming calls with a person on average in a day.
33		Per day Per person Avg Duration of Outgoing calls	Duration of outgoing calls with a person on average in a day.
34		Per day No. of persons called	No. of persons called (includes incoming, outgoing and missed) in a day.
35		Log of Per day Total Duration of Incoming calls	Total Talk time of incoming calls in a day.
36		Per day Total No. of Incoming calls	No. of incoming calls in a day.
37		Per day Total No. of Outgoing calls	No. of outgoing calls in a day.
38		Per day Total Duration of Outgoing calls	Total Talk time of outgoing calls in a day.
39		Per day Total No. of Missed calls	No. of missed calls in a day.
40		HHI of No. of Incoming calls	Herfindahl-Hirschman index of incoming calls. To compute this measure, we first calculate the no. of calls received from a person for every day (for a customer). We then take average across all days to get the no. of calls received from the person per day. We then assign share of calls to every person and compute HHI for the customer.
41		HHI of No. of Outgoing calls	Herfindahl-Hirschman index of outgoing calls. To compute this measure, we first calculate the no. of calls made

			to a person for every day (for a customer). We then take average across all days to get the no. of calls made to the person per day. We then assign share of calls to every person and compute HHI for the customer.
42		HHI of Total Duration of Incoming calls	Herfindahl-Hirschman index of duration of incoming calls. To compute this measure, we first calculate the duration of calls received from a person for every day (for a customer). We then take average across all days to get duration of calls per day. We then assign share of durations to every person and compute HHI for the customer.
43		HHI of Total Duration of Outgoing calls	Herfindahl-Hirschman index of duration of outgoing calls. To compute this measure, we first calculate the duration of calls made to a person for every day (for a customer). We then take average across all days to get duration of calls per day. We then assign share of durations to every person and compute HHI for the customer.
44		HHI of No. of Missed calls	Herfindahl-Hirschman index of missed calls. To compute this measure, we first calculate the no. of missed calls received from a person for every day (for a customer). We then take average across all days to get the no. of missed calls received from the

			person per day. We then assign share of missed calls to every person and compute HHI for the customer.