Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment

Daniel Björkegren[a] and Darrell Grissen[b]

October 28, 2015

Abstract:

Many households in developing countries lack formal financial histories, making it difficult for banks to allocate capital, and for potential borrowers to obtain loans. However, many unbanked households have mobile phones, and even prepaid phones generate rich data about their behavior. This project shows that behavioral signatures in mobile phone data predict default with accuracy approaching that of credit scoring methods that rely on financial histories. The method is demonstrated using call records matched to loan outcomes for a sample of borrowers in a Caribbean country. Individuals in the highest quartile of risk by our measure are 6 times more likely to default than those in the lowest quartile.

Keywords: credit scoring, microfinance, mobile phones, big data

Acknowledgements:

Thanks to Entrepreneurial Finance Lab and partners for data, and Nathan Eagle, Seema Jayachandran, and Jeff Berens for helpful discussions. This work was supported by the Stanford Institute for Economic Policy Research through the Shultz Fellowship in Economic Policy.

[a]Brown University, Department of Economics. E-mail: danbjork@brown.edu, Web: http://dan.bjorkegren.com (corresponding author)

[b]Entrepreneurial Finance Lab. E-mail: darrell.grissen@eflglobal.com

1 Introduction

Many studies have found that small firms in developing countries have access to opportunities with high returns that, puzzlingly, remain untapped (De Mel et al., 2008; McKenzie and Woodruff, 2008; Banerjee and Duflo, 2014). One reason these opportunities may remain untapped is if potential lenders have difficulty identifying the investments that will be profitable.

Developing country banks that would lend to small firms face several particular challenges. In developed countries, banks have access to robust information on borrower reputation through credit bureaus, which aggregate information on an individual's historical management of credit. The credit bureau model has been copied in many developing countries (e.g., Luoto et al., 2007; de Janvry et al., 2010), but many remain sparse: many households in developing countries do not interact with formal institutions that generate the necessary data. As a result, lenders have very little formal information on potential borrowers.[1] This is particularly problematic, as banks who would lend to small or informal businesses may have little recourse if a borrower were to default. Even when banks can rely on institutions like police and courts, it is costly to follow up on small loans.

Traditional microfinance has presented one solution to the repayment problem, relying on community members to aid in monitoring and selection of loans. However, it is not clear that the investments currently selected by microfinance have led to transformative effects for borrowers (Banerjee et al., 2015; Karlan and Zinman, 2011; Banerjee et al., 2014).

This paper introduces a new method to identify profitable investments, using information on potential borrowers that is already being collected by mobile phone networks.

Although unbanked households lack the formal records needed for traditional credit scores, many have maintained a rich history of interaction with a formal institution over an extended period of time—their mobile phone activity, recorded by their operator. In 2011, there were 4.5 billion mobile phone accounts in developing countries (ITU, 2011). Even with prepaid plans, operator records can yield rich information about individual behavior and social networks. If indicators derived from this data are predictive of creditworthiness, they can help banks identify profitable opportunities. There are many straightforward indicators of behavior that are plausibly related to loan repayment. For example, a responsible borrower may keep their phone topped up to a minimum threshold so they have credit in case of emergency, whereas one prone to default may allow it to run out and depend on others to call them. Or, an individual whose calls to others are returned may have stronger social connections that allow them to better follow through on entrepreneurial opportunities.

[1] Some researchers have explored using this limited information to identify borrowers; for example, Schreiner (2004) finds simple characteristics can complement the judgment of loan officers.

This paper demonstrates that indicators of behavior derived from mobile phone transaction records are predictive of loan repayment using data from a Caribbean country. To gauge the predictive quality of the method, this paper combines bank data from loans that have been completed with borrowers' mobile phone records. It predicts who among these individuals ended up repaying their loan, based on how they used their mobile phones before taking a loan. We find that the predictive accuracy of the method approaches that of credit scoring methods using traditional data in more developed settings. Using a common metric of performance, the area under the receiver operating characteristic curve (AUC), our best model obtains 0.68 and comparison models using financial data in other settings obtain 0.67-0.79. Individuals in the highest quartile of risk by our measure are 6 times more likely to default than those in the lowest quartile. We obtain this performance despite observing very little behavior. Our sample is poor and uses phones infrequently: in our data we observe the median individual place 22 minutes of calls and send 1 SMS, spending a total of $2.85 over a period of 11 weeks.

2 Data

The organizational partner is EFL (Entrepreneurial Finance Lab), which works on alternative credit scoring methods in developing and emerging markets, with an emphasis on the unbanked.[2] EFL obtained linked phone and completed loan data for a sample of borrowers in a Caribbean country. First, anonymous records were obtained for a sample of borrowers who took out a small loan from a local microfinance institution. These records include basic demographics such as age and gender, the terms of the loan provided, and whether the loan was defaulted on (defined by 90 days of nonpayment). Many of these borrowers also have mobile phone accounts with a large operator in the country. These borrowers were matched to their prepaid phone accounts using encrypted (anonymous) individual identifiers. Mobile phone transaction records (CDR) were obtained for these matched accounts, including metadata for each call, SMS, top up, and data access, for the year of 2012. Fields include an identifier for the other party, time stamps, tower locations, durations, charges, and handset models used. They include no information on the content of any communication.

Because we aim to predict default based on the information available at the time a loan was granted, only mobile phone transactions that precede the loan date are included. We focus on the bank's small loan product. The data includes 3,131 loans granted between January and September 2012, of which 12.8% ended in default. Borrowers have a median

[2]From their website, "EFL Global develops credit scoring models for un-banked and thin-file consumers and MSMEs, using many types of alternative data such as psychometrics, mobile phones, social media, GIS, and traditional demographic and financial data. We work with lenders across Latin America, Africa and Asia." http://www.eflglobal.com

Table 1: Description of Individuals

	Mean	SD	Median
Borrowers			
Gender is female	71.8%	-	-
Age (years)	35.7	8.3	34.5
Mobile phone use preceding loan			
Weekly...			
Calls out, number	7.8	11.5	4.4
Calls out, minutes	4.0	6.3	2.1
SMS sent	0.8	2.7	0.1
Data use	18,182	257,703	0
Top Ups	\$0.49	\$0.82	\$0.25
Spend	\$0.53	\$0.89	\$0.27
Balance	\$0.26	\$2.83	\$0.05
Days of mobile phone data preceding loan	77	41	74
Loan			
Size	\$188	\$84	\$184
Term (days)	142	22	150
Default (90 day)	12.8%	-	-
N	3,131		

loan size of \$184 and term of 5 months.[3] Many of these loans were granted early in 2012, so due to the temporal overlap of the two data sets, we have relatively few days of phone data on which to estimate the model: 77 days for the mean borrower.

Descriptive statistics for the sample are presented in Table 1. Most borrowers are female (71.8%), with a mean age of 35.7. There is high variation in mobile phone usage. Mobile phone accounts are prepaid; subscribers keep low balances relative to how much they use the phone. The median subscriber has an average weekly balance of \$0.05 but tops up \$0.25 and spends \$0.27.

3 Method and Results

The goal is to predict the likelihood of default using behavioral features derived from mobile phone usage. The model is estimated using data on loans that have already been completed; these estimates are used to predict whether a loan would end in default based on the information available at the time the loan was granted. Because this sample of individuals

[3] All results reported in US dollars.

Table 2: Individual Features Correlated with Default

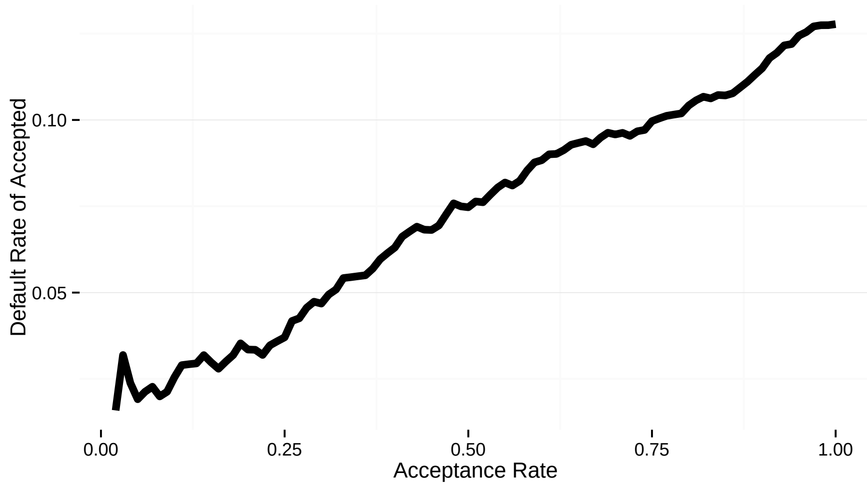
	Correlation with default	t-stat
Bank data		
Female	-0.060	-
Age	-0.055	-3.11
Loan size	0.019	1.07
Loan term	0.004	0.22
Features derived from phone usage		
Most predictive features:		
Variation in usage	-0.179	-10.20
Periodicity of usage	-0.156	-8.81
Mobility	-0.143	-8.05
N	3,131	

did obtain loans, risk is reported among those who were allocated loans based on the bank's decision process at the time; default risk among unbanked populations may differ, and can be explored in follow up work.

The loan data provides an indicator for whether a particular borrower defaulted on their loan (90 days past due). From the phone data we derive 6,841 features with variation. These include various quantifications of the intuitive features presented in the introduction as well as other measures. In total, features include measures of usage: intensity and distribution over space and time, top up and depletion patterns, mobility, the pattern of handset use, and strength and diversity of social network connections. The performance of these features will be compared against a benchmark using the characteristics that the bank recorded at the time of the loan: gender, age, loan size, and the loan term in days.

A first question is how individual features correlate with default. Table 2 presents the single variable correlation with default, for demographic features recorded by the bank and for the most correlated sets of features derived from phone usage. The demographic features have very low correlation with default (magnitudes between 0.004 and 0.06), which is reasonable: the bank should already be incorporating the features it observes into the decision of whether to extend credit. From the phone data many features measure similar concepts, for example variation in call durations, number of calls, and dollars spent are all measures of variation of usage. We report the categories with the highest magnitude of correlation with default, in the bottom panel of Table 2. Several categories of features have highly significant correlations with magnitudes between 0.12 and 0.18, including variation and periodicity of usage, and several measures of mobility. A popular intuition associates successful repayment with stability, but the correlations suggest that variation in usage and

Figure 1: Model Performance: Default Rate by Proportion of Borrowers Accepted



Model using phone indicators, chosen using stepwise OLS from a random set of starting variables.

mobility are associated with lower default risk. One reason this could be the case is if in the weeks prior to signing a loan, successful borrowers have variable behavior, for example an entrepreneur may be scouting out potential locations for a business or discussing plans with a variety of contacts.

While individual features are significantly correlated with default, the best prediction method will take into account multiple features at the same time. Because of the large number of potential predictors, including all of them in a simple method like ordinary least squares (OLS) would lead to overfitting. Instead, two types of models are estimated, including a specification of OLS chosen by a model selection procedure (stepwise search using the Bayesian Information Criterion) and random forests. To best measure how the method will perform out of sample, all estimates are computed using cross validation with 5 folds. The data is randomly divided into 5 folds, and the outcomes for each fold are predicted using a model estimated on the omitted folds.

As a first check of the method's performance, we consider how well the best model separates low and high risk borrowers. Figure 1 shows how the default rate varies with the fraction of borrowers accepted (where borrowers with lowest predicted default are accepted first). Individuals with the highest 25% of risk scores are 6 times more likely to default than those with the lowest 25%. The bank could have reduced defaults by 41% while still accepting 75% of these borrowers.

Table 3 presents more technical measures of the method's performance. First, the table reports the area under the receiver operating characteristic curve (AUC). The receiver operating characteristic curve (ROC) plots the true positive rate of a classifier against the false positive rate. A naïve classifier would generate an AUC of 0.5 and a perfect classifier would generate an AUC of 1.0. The table also reports the Kolmogorov–Smirnov statistic of the deviation of the ROC from that of a naïve classifier, and the H measure which has been proposed as an improvement over the AUC (Hand, 2009).

The benchmark models using only demographic and loan data perform very poorly, with AUCs in the range of 0.53-0.54, as shown in the first panel of Table 3. This low performance could be because few characteristics are available, or because the bank already used these characteristics when deciding to extend loans to this sample (in which case a benchmark may perform better in a general sample). Adding features derived from phone data improves AUCs to the range of 0.66-0.68. This performance approaches that of a sample of published AUC estimates from other studies that use traditional credit scoring methods in more developed settings, shown in the second panel of the table (ranging from 0.67-0.79).

The method achieves this performance despite using relatively sparse data: our sample of borrowers has very low incomes, and spends little on telephony. Each week the median borrower places only 2.1 minutes of outgoing calls and spends $0.27 on the phone network. Further, our panel of data is short: we observe each borrower's phone usage for 11 weeks on average.

To account for the possibility of unforeseen shocks, ideally the model would be tested not only on out of sample individuals, but also on out of sample time periods. Since the data spans only a short time this is left for future work.

4 Conclusion

This paper demonstrates a method to predict default among borrowers without formal financial histories, using behavioral patterns revealed by mobile phone usage. This method is promising even for poor borrowers whose mobile phone usage is extremely sparse.

The method quantifies rich aspects of behavior typically considered 'soft' information, making it legible to formal institutions (Berger and Udell, 2006). The results suggest that this information is useful even in addition to current screening methods that rely on 'soft' information gathered in person. It may represent a complement to these current methods, or it may represent a substitute that could be used independently. If it can be used independently, it could enable new models of lending.

The method could be implemented in three ways. First, it can be used to extend telecom-specific credit, within the telecoms that already possess the necessary data. In developing countries, many households find it difficult to save to purchase a handset or to maintain a consistent airtime balance. With a user's permission, a scoring model can be used to extend credit for a handset purchase or for airtime.[4] Second, it can be used to extend general forms of credit, by turning telecom data into a credit score that can be used by a bank, either through mobile banking platforms or an independent credit bureau. Third, it can be used to extend general forms of credit, by obtaining usage data independently. For example, lending could be provided through a smartphone app that asks for permission to view call history.

There remain open questions. Some indicators are 'gameable' in the sense that a subscriber may be able to manipulate their score if they knew the algorithm; it is preferable to use indicators that are less susceptible (for example, manipulating spending or travel can be costly). A related question is whether the best model is stable across different regions, which may have different customs of phone usage, or as time passes and new behaviors are adopted (such as increasing data usage).[5] Any implementation will also need to carefully consider individual privacy and legal questions.[6]

References

Baesens, B., T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen (2003): "Benchmarking state-of-the-art classification algorithms for credit scoring," *Journal of the Operational Research Society*, 54, 627–635.

Banerjee, A., E. Duflo, C. Kinnan, and R. Glennerster (2014): "The Miracle of Microfinance? Evidence from a Randomized Experiment," .

Banerjee, A., D. Karlan, and J. Zinman (2015): "Six Randomized Evaluations of Microcredit: Introduction and Further Steps," *American Economic Journal: Applied Economics*, 7, 1–21.

Banerjee, A. V. and E. Duflo (2014): "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program," *The Review of Economic Studies*, 81, 572–607.

Berger, A. N. and G. F. Udell (2006): "A more complete conceptual framework for SME finance," *Journal of Banking & Finance*, 30, 2945–2966.

[4]Many developing country operators already offer small airtime loans like this; a scoring model could improve their provision.

[5]Phone sharing is common in many developing countries, but this does not represent a problem for the method as long as the practice does not differ between estimation and implementation. In that case, the method will capture the behavior of phone owners as well as those they choose to lend to.

[6]While anonymized data can be used to estimate the scoring model, a lending decision would need to use a potential borrower's data. In an implementation, the use of this data can be opt-in.

Calabrese, R. and S. A. Osmetti (2013): "Modelling small and medium enterprise loan defaults as rare events: the generalized extreme value regression model," *Journal of Applied Statistics*, 40, 1172–1188.

de Janvry, A., C. McIntosh, and E. Sadoulet (2010): "The supply- and demand-side impacts of credit market information," *Journal of Development Economics*, 93, 173–188.

De Mel, S., D. McKenzie, and C. Woodruff (2008): "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 123, 1329–1372.

Hand, D. J. (2009): "Measuring classifier performance: a coherent alternative to the area under the ROC curve," *Machine Learning*, 77, 103–123.

ITU (2011): "World telecommunication/ICT indicators database," Tech. rep., International Telecommunication Union.

Karlan, D. and J. Zinman (2011): "Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation," *Science*, 332, 1278–1284.

Luoto, J., C. McIntosh, and B. Wydick (2007): "Credit Information Systems in Less Developed Countries: A Test with Microfinance in Guatemala," *Economic Development and Cultural Change*, 55, 313–334.

McKenzie, D. and C. Woodruff (2008): "Experimental Evidence on Returns to Capital and Access to Finance in Mexico," *The World Bank Economic Review*, 22, 457–482.

Schreiner, M. (2004): "Scoring Arrears at a Microlender in Bolivia," *Journal of Microfinance / ESR Review*, 6, 65–88.

Van Gool, J., W. Verbeke, P. Sercu, and B. Baesens (2012): "Credit scoring for microfinance: is it worth it?" *International Journal of Finance & Economics*, 17, 103–123.

Table 3: Model Performance

	AUC	KS	H-measure
Credit Scoring Models			
Demographics and loan characteristics			
Random Forest	0.536	0.061	0.015
OLS, stepwise BIC	0.531	0.085	0.009
Phone indicators, demographics, and loan characteristics			
Random Forest	0.660	0.249	0.084
OLS, stepwise BIC	0.678	0.279	0.103

Comparison to traditional credit scoring	Best AUC	Features
Baesens et al. (2003):		
UK	0.668-0.758	16-19 unspecified predictors, in 4 samples
Belgium/Netherlands/Luxembourg	0.776-0.791	33 unspecified predictors, in 2 samples
Van Gool et al. (2012):		
Bosnia microfinance	0.707	Demographics, earnings, capital, debt, loan characteristics
Calabrese and Osmetti (2013):		
Italian small and medium enterprises	0.723	Firm leverage, liquidity, profitability

For the top panel, statistics estimate out of sample performance using 5 fold cross validation. Demographics include age and gender; loan characteristics include size of loan and term of loan in days. AUC represents the area under the receiver operating characteristic curve. KS represents the Kolmogorov-Smirnoff statistic. Stepwise OLS is repeated several times, seeded with a random sample of the indicators; the mean along each dimension is shown. Baesens et al. (2003) also report results from publicly available Australian and German data sets, but it is not clear whether the outcomes are defaults so they have been omitted.