Title
Essays on the Economic Impacts of Mobile Phones in Haiti

Author
Barriga Cabanillas, Oscar Eduardo

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays on the Economic Impacts of Mobile Phones in Haiti

By

OSCAR EDUARDO BARRIGA-CABANILLAS
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in

AGRICULTURE AND RESOURCE ECONOMICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Travis J. Lybbert (Chair)

Stephen R. Boucher

Ashish Shennoy

Committee in Charge
2021

Contents

List of Figures

List of Tables

Acknowledgements

I can only express gratitude to my dissertation committee. Travis Lybbert, who is one of the clearest and sharpest people that I know. Thank you for always being happy to discuss new ideas and providing invaluable guidance and mentorship –I could not have hoped for a better mentor. Steve, who not only has an incomparable attention to detail but an amazing capacity to know what really matters, both as an academic and at a human level; thank you for always keeping the door of your office (and home) open. Ashish, who always finds a better method to improve how we answer a research question; thank you for pointing at me how to make better research. I have learned an extraordinary amount from all of you.

I will always treasure the memories of my Ph.D. years. Many people were with me throughout these years: Aleks Michuda (best officemate ever), Juan Sebastian Correa, Cristina Chiarella, Matthieu Stigler, Jesus Arellano, Armando Rangel, Francois Castonguay, Karen Ortiz, Tim Lee, Tina Kotsakou, Javier Parada, Gerardo Aragon, Irvin Rojas, John Blanchette, Felipe Aviles, Saloni Chopra, Laila Salimi, Felipe Pinilla, Laura van der Werf, Carlos Rondon, Natalia Garbiras, Camilo Salcedo, and Daniel Jaramillo. I could only hope everybody was able to find such a strong and giving network.

I could not explain any of my achievements in life without the love and support from my family. The new guard: Cynthia, with whom I have traveled through life; and Sofia, who in only three years has given me a lifetime of joy. As well as the old guard: My parents, who always pushed me to think big while having the certainty of their unconditional love and support; my brother, with whom no topic is off-limits; and my grandparents Oscar, whose stories made a child dream; Mary, the warmest human being; and Ceci, with whom I spend endless hours playing (you might not, but I remember all the riddles you taught me). This list would be incomplete without acknowledging my aunt and uncles: You filled my childhood with happiness and are present since my earliest memories. To all of you, I owe you everything.

Abstract

This dissertation adds to our understanding of how mobile phones play a role in improving development opportunities. They not only provide a channel to communicate with others, but generate troves of data about the economic lives of subscribers, including those with little "economic footprint". I contribute to the literature by examining three aspects of the potential of cell phones as development tools.

In chapter one, I study the most widespread form of digital credit: airtime loans. This service allows prepaid customers to borrow small airtime advances for a fee instead of purchasing recharges that must be paid upfront. Relying on rich administrative data from a mobile network operator in Haiti, I show access to loans increases total communication expenditure by 16% but with distinctly heterogeneous impacts. When airtime loans become available, poorer customers more than double their mobile communication spending, while access to loans leaves expenditure of the highest tercile unchanged. These differences exist despite relatively uniform patterns of loan usages between the poor and non-poor. I argue these differences are driven by distinct motivations for requesting airtime loans, with poorer customers using the loans to relax short-term liquidity constraints at critical communication times whereas non-poor customers primarily use these loans for convenience, as it gives them more discretion in when to visit airtime vendors.

In chapter two, I build on the evidence that demonstrates mobile phone metadata, in conjunction with machine learning algorithms, can be used to estimate the wealth of individual subscribers, and to target resources to poor segments of society. This paper uses survey data from an emergency cash transfer program in Haiti, in combination with mobile phone data from potential beneficiaries, to explore whether similar methods can be used for impact evaluation. A conventional regression discontinuity-based impact evaluation using survey data shows positive impacts of cash transfers on household food security and dietary diversity. However, machine learning predictions of food security derived from mobile phone data do not show statistically significant effects; nor do the predictions accurately differentiate beneficiaries from non-beneficiaries at baseline. Our analysis suggests that the poor performance is likely due to the homogeneity of the study population; when the same algorithms are applied to a more diverse Haitian population, performance improves markedly. We conclude with a discussion of the implications and limitations for predicting welfare outcomes using big data in poor countries.

In chapter three, I provide evidence on the determinants of the adoption of mobile money services. In contrast with previous research that centers on adoption after a service is launched, I study a mature platform that experiences a stagnating user base. I combine a survey with a randomized component with mobile money transaction logs to test if informational videos induce people to open an account and try new products. My results show that awareness of mobile money services is high and, even if having an account is free, many people use the service indirectly by asking others to make transactions for them. The intervention increased adoption by 5.4%. However, a large share of new users came from the group that declined the opportunity to watch the videos, indicating more than simply information drove their decision to adopt. I do not find the videos increased the usage of additional services by people with an account at the time of the survey. Taken together, my results show further growth of the mobile money platform requires increasing the number of services available to attract additional customers and incentivize the daily usage of mobile money for economic transactions.

Introduction

In the last decade, more than 700 million people around the world obtained their first cellphone (GSMA, 2019). Most of the new subscribers live in developing countries, with the lion's share of new users coming from the bottom of the income distribution. Their devices might differ in hardware capabilities, from fancy iPhones to simple feature phones with a flashlight. However, they share a common denominator: they are a powerful tool for delivering life-enhancing information and services. From a device to make and receive calls, mobile phones have become a springboard for the delivery of additional services, including entertainment, education, healthcare, and financial services, as well as a new tool for social research where troves of data can help better inform policies.

Mobile phones have the potential to play a positive role improving development opportunities, in particular in circumstances where asymmetric information and high transaction costs limit the functioning (or existence) of markets. Their ability connect people over long distances and poor infrastructure gives cellphones the capacity to solve asymmetric information problems. As a communication device, mobile phones provide a more economical and reliable alternative to other sources of information such as personal travel, local social networks, radio, and television (Aker and Blumenstock, 2015), with basic calling and texting increasing the flow of information, and with that, the ability of users to learn about labor opportunities and the best places to buy and sell goods (Jensen, 2007; Aker and Mbiti, 2010; Fafchamps, 1992).

Mobile phones are also platforms for delivering services. The impact of these "add-on" services is greater for sectors where there is a high cost of developing a purpose-built infrastructure. Mobile financial services (MFS) is, perhaps, the most representative sector built on top of the cellphone infrastructure, providing new payment, credit, and insurance alternatives. In developed countries, since traditional financial providers are well-established, MFS represents only a new delivery method. Still, it has spurred a whole new industry that competes for customers with affordability and convenience. In contrast, MFS have been able to bypass the poor infrastructure and high transaction costs that have inhibited the development of traditional financial providers in the developing world, providing for many low-income customers the first opportunity to reduce their dependence of a cash-only economy. Evidence shows that even these simple financial services offer the opportunity to reach a wider risk-sharing network during lean times, increasing the resilience of households to negative income shocks, and providing an active area of research to understand how the growing network of mobile money users can leverage new saving, insurance, and payment opportunities (Suri, 2017).

Repurposing the troves of data that cellphone usage generates offers a new window to the economic lives of individuals whose "economic footprint" is hard for private companies and the public sector to detect. Combining this information with machine learning algorithms has allowed the private sector to launch new services, some of which have achieved long sought development goals, such as expanding access to formal credit services to millions of people whose record of economic transactions was invisible to traditional credit scoring mechanisms. More importantly, small tweaks in the algorithms that allows private companies to target ads and create credits scores offer the public sector new opportunities to fight poverty with data. Current applications include the possibility to reduce the cost to create (and update) socioeconomic indicators and innovative methods to target social programs (Blumenstock, 2016b, 2018a; Blumenstock, Cadamuro, and On, 2015).

Through out three chapters, I answer specific questions about the role of cellphones in increasing access to credit services, improving the implementation of social programs, and the barriers to the growth of mobile money services. In particular, in chapter one I explore how prepaid cellphone subscribers react to the possibility to finance their cellphone expenditures, especially in the presence of high levels of liquidity constraints; chapter two studies the potential and limitations of cell phone data for the targeting and impact evaluation of a cash transfer program implemented by the World Food Program in Haiti; finally, chapter three studies why mobile money might fail to take hold and become a large payment ecosystem even in the absence of strong competition from traditional financial providers.

All of the chapters of this dissertation take place in Haiti. The poorest country in the Americas and a place where poor infrastructure, widespread market failures, and low state capacity limit its development opportunities. The material draws from fieldwork in the country, several focus groups, lots of informal chats with mobile phone owners, and *a lot* of data from the largest telecommunications operator. Throughout three chapters, I empirically investigate the role cellphones play in the Haitian economy, with the aim of better grasping how they can help improve the quality of life of the people on the island.

Understanding the impact of cellphone technology is a moving target: The growing number of new subscribers, shifting usage patterns, and the development of new services, sometimes with market-specific rules, creates an almost infinite amount of opportunities to learn about the positive and negative impacts of cellphone technology. Far from providing an exhaustive list, my research in Haiti provides a lookout into future lines of research as well as some answered questions.

First, data access remains the main constraint to analyzing how individuals, firms, and public entities react to the introduction of mobile phones and the development of new services. Future lines of research should focus on providing greater and more transparent access to raw data. The value of data is made clear in the context of mobile financial services. Data are at the center of the industry's capacity to provide services for people excluded from traditional financial providers. In contrast with centralized credit scoring systems, a customer's good financial behavior remains hidden from competitors, a situation where customers lack the tools to showcase their good payment history, and increases the costs of entry for new competitors as they need to bear the full costs of gathering data. How to properly regulate data sharing and compatibility across platforms is at the heart of the question in how to make the market of mobile financial services more competitive. Keeping information "private" implies granting exclusive usage rights to private companies: In the era of information, this is a new form of market power and a barrier for public entities and researchers from using this data for the public good. It is simply not defensible that the most advanced data and tools remain at the disposal of private companies and are cumbersome to access when families' lives and well-being are at stake.

A second element is that new services can not be studied in the abstract. Communities are not only spectators on their economic lives and actively engage in activities to deal with information problems and the lack of formal markets. Research on the unintended consequences of new ways to access information and financial services must consider not only how a person is affected directly but also how new services interact with preexisting formal and informal arrangements. Properly studying these interactions needs more than collecting data from a company's server and combine it with a limited sets of indicators from surveys: It requires a deep understanding of the relationships and strategies people (and communities) have developed, and their potential interactions (and disruption) with new technologies.

Finally, cellphone technology is not a silver bullet to solve all of the ills affecting the people living in the country. It has disrupted the development prospects in dramatic and (mostly) positive way, offering Haitians opportunities to communicate, access a wider risk sharing network, and receive social assistance from the government and non-governmental organizations. However, it can only solve very specific problems out of the plethora of development challenges the country faces. In fact, the sector's development is not immune to poor economic performance, political instability, and weak institutional incentives that constitute a barrier to other economic activities.

In what follows, I briefly summarize each chapter. Chapter one studies digital credit, a service that uses cellphone-derived indicators to overcome imperfections in the credit market that, traditionally, have excluded the working poor from formal credit sources. The original research proposal investigated a small line of credit for which approval depended exclusively on cellphone-derived indicators to provide credit scores, and funding disbursement and repayment relied on the network's provider mobile money platform. Despite the success of digital credit in Kenya, where more than 25% of the population has taken out at least one digital loan (Blumenstock, 2018a), the experience in Haiti reveals how difficult from a technical and regulatory perspective it is to develop and launch similar services. Instead, the final project ended up exploiting an often ignored but widely available form of digital credit: airtime loans, a service that provides prepaid customers small airtime advances, on average less than USD$0.50 for a 10% fee, that can only cover cellphone usage. A particularity of this product is that, in a country where 46% of adults lack access to formal financial service, and two of every three loans come from informal lenders (FinScope, 2018), airtime loans represent the first formal financial transaction of their lives for many poor customers.

I use the loan eligibility rule, that grants access to the service five weeks after a line is activated, to implement an event-study design. I find that even under flexible prepaid plans, high liquidity constraints limit the capacity of low-income subscribers to use their cellphones. Specifically, airtime loans allow poorer cellphone users to more than double their mobile communication spending, an increase above the additional loan fee costs, which is caused by the availability of credit driving new additional expenditure. Better-off subscribers present a similar demand for loans, but there is no observable change in their total spending. Analysis of the expenditure patterns suggest that distinct motivations for loan demand cause the differential impacts. Poorer customers appear to use loans to relax short-term liquidity constraints at critical communication times, with loan availability making them less sensitive about which calling opportunities to take. In contrast, non-poor customers primarily use these loans for convenience, as it gives them more discretion on when to visit airtime vendors.

My second chapter investigates the role of Call Detail Records (CDRs) on a social program's targeting and evaluation. My objectives are twofold. First, I investigate if cell phone data provides a mechanisms to target a World Food Programme unconditional cash transfer program. Second, I test if predicted individual-level food security outcomes using cellphone records are precise enough as to detect the changes detected by the program's formal impact evaluation. Passively generated data from cellphone usage offers an opportunity to identify vulnerable individuals and measure real-time changes in their economic well-being. The program's implementation provides a unique opportunity to evaluate the potential of CDR-based methods since all survey instruments and data protocols had access to the cellphone data in mind, allowing us to combine them.

Results are not encouraging. Despite providing a controlled setting to test the potential of CDR-based methods, none of the models replicate the targeting or the survey-based RD results. In a postmortem assessment, I discuss four reasons why CDR predictions failed. Taken together, these four factors show the excitement generated by CDRs as a data source in development economics should be tempered by some very real limitations. First, phone ownership is not universal and is biased towards the rich, creating a natural limitation to the application of CDR-based methods. Second, there is fundamental empirical tension between algorithmic predictions using machine learning methods, on the one hand, and causal identification, targeting and program evaluation on the other. This tension is multi-faceted, but based on the notion that effective outcome prediction requires statistical variation across a broad support of the underlying outcome distribution. This requirement runs counter to causal identification, which hinges on counterfactuals for comparable individuals or households. Third, the same pattern emerges in the context of targeting: The more effective targeting is at identifying a sub-population with relatively homogeneous needs or characteristics, the less effective outcome predictions will be among this sub-population. Finally, the encouraging results showing that CDR data could predict wealth levels appear, at least in our setting, not to replicate when predicting consumption variables, the main outcome variable for the study.

My third chapter studies the success (or lack thereof) of mobile money in Haiti. The success of mobile money services in Africa suggests that a dense network of agents and little competition from traditional financial providers are key components for the success and rapid adoption of mobile money, in particular among those with little access to brick-and-mortar institutions. Haiti experiences several of these conditions: A large and relatively dense network of mobile money agents, and one of the lowest densities of bank and microfinance institutions in the world. Nevertheless, after ten years in the market, and a relaunch of the service in 2015, the service has failed to reach the adoption levels seen in Sub-Saharan African countries, with a stagnant level of subscribers and transactions.

Why is mobile money in Haiti not replicating the success observed in other countries? My research provides the first formal test of the industry's working theory; specifically, that lack knowledge about how to use the service, including difficulties with the codes to navigate the menus on feature phones, is the main constraint for mobile money growth. I formally test if providing information about how to use the service can attract new users and increase the services used. The experimental design includes a short survey, where access to videos with information about how to join and use mobile money is randomized. Results from the survey component of the intervention show that account ownership is not a good measure of mobile money usage. Even if opening and maintaining an account is free, 32% of the people interviewed without an account still use the service by asking subscribers to do transactions for them. Going back to the information campaign, I find individuals in the treatment group are 5.4% more likely to open a mobile money account in the weeks following the intervention, with some evidence of repeated usage. Still, the results are not conclusive about how much of the impact comes from access to information, as a large percentage of people adopters in the treatment groups declined watching any video, suggesting that people in this group were able to learn by themselves about how to use the product. Additionally, even if participants with an account showed interest in watching videos for services they were not currently using, I detect no effect on additional service usage.

A broad view of the results reveals that knowledge gaps is only one of many constraints to the growth of the mobile money platform. The stagnant number of subscribers reflects a deeper problem with the current service offering, which only provide limited opportunities to use the service. Launching additional services in a market such as Haiti is challenging; the country has high poverty levels, political instability and is very vulnerable to shocks (both economic and natural).

With that said, experience from African countries showcases that the markets with the highest levels of mobile money users tend to be the those with the most diversified set of services, including receiving international remittances, savings, digital credit, and paying informal vendors.

Chapter 1

Liquidity or Convenience? Heterogeneous Impacts of Mobile Airtime Loans on Network Usage and Communication Expenditure

This chapter is co-authored with Travis J. Lybbert (Professor, University of California, Davis)

Credit market imperfections can decrease welfare by increasing vulnerability to shocks and destabilizing consumption. Meta data from individual cellphone users have enabled a proliferation of mobile financial services in markets where information asymmetries and high provision costs tend to deter formal financial institutions. As the first such financial product typically offered to new users, airtime loans provide prepaid customers small airtime advances for a fee as an alternative to recharges purchased from network agents. Relying on rich administrative data from a mobile network operator in Haiti, we study the impact of airtime loans on consumer cellphone expenditure and network usage. We find that access to loans increases total communication expenditure by 16% due to a crowding-in of additional network usage. This expenditure response to airtime loans is distinctly heterogeneous. Poorer customers in the lowest tercile of initial expenditure more than double their mobile communication spending when airtime loans become available, while access to loans leaves expenditure of the highest tercile unchanged. These differences in the expenditure impacts of airtime loans exist despite relatively uniform patterns of loan usages between the poor and non-poor. We find suggestive evidence that these differences are driven by distinct motivations for requesting airtime loans: Poorer customers appear to use loans to relax short-term liquidity constraints at critical communication times whereas non-poor customers primarily use these loans for convenience, as it gives them more discretion in when to visit airtime vendors. Despite systematic differences in cell phone usage by gender, we find no evidence of gender differentiated impacts of airtime loans.

1.1 Introduction

As cell phones have spread around the world and entered the lives of rich and poor alike, they have ushered in unprecedented financial inclusion opportunities. Mobile Financial Services (MFS) have enabled a flurry of innovation, with products that bring new ways of paying, saving, borrowing, and insuring to people and regions previously under-served by traditional financial institutions. These impressive and innovative gains have in many ways redefined economic development opportunities and even macroeconomic policy possibilities (Suri, 2017; Aron et al., 2017).

The proliferation of MFS offerings we observe today in many middle and low-income countries first hinged on the widespread availability of inexpensive hardware — especially feature phones — and simple prepaid plans that allowed customers to add airtime as needed in increments of nickels and dimes. Although these breakthroughs made cellular service accessible to almost everyone, including the poor, with this access came new financial dilemmas as these costs stretched the already limited resources of poor households. As cellphones became the newest necessity, managing this new asset and the expenses associated with this vital connection to one's network became an essential part of daily budgeting. For example, one fifth of Kenyan users reports to forgo other expenditures such as food, bus fares, or utility bills in order to keep their cellphones active, and studies across different countries show poor households spend between 10 and 25% of their disposable income on mobile phone usage (Agüero, de Silva, and Kang, 2011; The World Bank, 2012).[1] Nickles and dimes spent on prepaid cell service can add up fast for those living on a dollar or two a day, particularly since demand for this service is often as frequent as demand for food. Against this backdrop, managing one's prepaid balance against expected communication needs and the opportunity cost of short-term (often intra-day) liquidity becomes a non-trivial, pressing and ever-present financial

[1] The survey was conducted on a representative sample of cellphone customers, see The World Bank (2012)

imperative.

As cheap phones and prepaid plans enabled mobile network operators to reach customers that had rarely been reached by the formal sector before (much less, formal financial services), the stage was set for MFS to create entirely new financial inclusion opportunities by sidestepping the informational asymmetries and institutional limitations that continue to stymie the development of financial markets in the developing world. These traditional challenges are especially salient for those who lack formal financial histories, want small loans, and are often located in areas that are difficult to serve. Cellphone technology has two key features that make it ideal to alleviate these challenges. First, cellphone usage creates a personalized data-trail from which insights about credit worthiness can be extracted. For the unbanked poor with few collateralizable assets, such alternative credit score sources can provide a key point of departure for financial inclusion. Second, it allows for remote, and automated, processing of the transactions lowering administrative costs (Björkegren and Grissen, 2018; Bharadwaj, Jack, and Suri, 2019).[2] The potential of financial inclusion gains from these innovations became clear in the past decade as individuals with access to MFS were able to better manage and share risk, smooth consumption, and take advantage of productive opportunities (Bharadwaj, Jack, and Suri, 2019; Suri and Jack, 2016). The potential for future gains is similarly massive as familiarity grows among the billion individuals who currently have a cellphone but not a bank account (GSMA, 2014).

In this paper, we study the first rung in the financial inclusion ladder provided by MFS and evaluate its impact on communication expenditure. This MFS product provides customers airtime advances in the form of very small and very short-term loans. Each loan is, on average, less than

[2]The cellphone provider does not need to be directly involved in the provision of the services with many services working over multiple platforms. However, given their competitive edge, cellphone companies are the biggest players in the market.

USD$0.50. In a country where 46% of adults lack access to any formal financial service and two of every three loans come from informal lenders, friends, and family (FinScope, 2018), airtime loans represent for many poor customers the first formal financial transaction of their lives. Airtime loans are popular, with 40% of eligible customers using them every month. On average, eligible customers finance 30% of his cellphone expenditure with airtime loans, each loan incurring a 10% fee. As elsewhere, the popularity of these loans is easy to appreciate given that the flexibility of prepaid phone service comes at the cost of frequent recharges and the risk of running out of airtime at a critical moment (Jack and Smith, 2020).[3] The cost of missed calls, unsent SMS messages, and frequent visits to airtime vendors are obviously difficult to quantify, but clearly these costs fall disproportionately on liquidity-constrained customers who must weigh them against locking up their limited liquidity in the form of airtime balance. For such customers, airtime loans soften this daily dilemma and potentially alter how and how often they use the mobile network.

We use a unique dataset from the largest cellphone provider in Haiti that contains the full set of transactions in the network in 2019. We exploit the eligibility rule, which makes that customers can request their first loan five weeks after initial activation, to implement an event study to identify the impact of airtime loans on subsequent communication patterns and expenditure. We find that access to loans increases total expenditure by 16%, which represents a crowding-in of new communication expenditure well beyond the fees associated with the loan. This result parallels the finding from the U.S. that increasing credit limits on credit cards induces an immediate and significant increase in consumption (Gross and Souleles, 2002).[4]

There are three mechanisms through which airtime loans could conceivably increase communica-

[3]Cellphone technology is ideal for this billing scheme as there is permanent communication between the device and the service provider, and the cost of each transaction can be made explicit to the consumer. Aker and Mbiti (2010) identifies this flexibility as a key factor contributing to the rapid adoption of cellphones by less wealthy consumers. From the provider's perspective, prepayment has the advantage that it avoids costly enforcing of contracts.

[4]The author find that every $100 increase in the credit card limit increases spending between $10 to $14 dollars.

tion expenditure. First, airtime loans could relax binding liquidity constraints that force customers to reduce their cellphone activity when they are short on cash to prepay for mobile communications. Second, airtime loans could reduce the salience of marginal communication costs and thereby increase network activity. Finally, airtime loans could lower the transaction costs of physically recharging one's prepaid account through a local agent. These mechanisms are not mutually exclusive and may simultaneously apply in the case of a particular customer. Since they likely shape different customer motives differently, we explore heterogeneous impacts of airtime loans by income. Despite relative uniform patterns of loan usages between the poor and non-poor we find distinctly heterogeneous results. Poorer customers in the lowest tercile of initial expenditure more than double their mobile communication spending when airtime loans become available. Meanwhile, access to loans leaves expenditure of the highest tercile unchanged. Why would access to airtime loans lead to an increase in the expenditure of the poor while still being attractive to better-off customers? We find suggestive evidence that these differences are driven by distinct motivations for requesting airtime loans. Consistent with binding liquidity constraints, poorer customers who often survive day-to-day on razor-thin cash balances appear to use loans to relax liquidity constraints at times critical communication times. Considering that loans are paid relatively quickly and have high repayment rates, these results indicate the consumption levels of this group are highly sensitive to the timing of their income. This is a typical of results in the presence of credit markets imperfections where the tools available to individuals in terms of cash management, savings, and borrowing are not enough to isolate consumption for (daily) cash cycles (Gelman et al., 2014).

Non-poor customers, on the other hand, seem to use airtime loans much more out of convenience. By borrowing airtime, subscribers avoid looking for airtime vendors and can strategically shift their recharges towards times with lower transaction costs. In our setting, the loan facilitation fee of 10%

provides an upper bound for the perceived transaction costs of recharging at inconvenient times. This willingness to pay for convenience has been observed in other financial products where new technologies have been introduced. For example, Buchak et al. (2018) find that online lenders are able to charge a premium for the possibility to apply for a loan by computer with a user-friendly interface. This premium exists even when they offer a similar product as brick-and-mortar banks, and it is higher than what regulatory or financial costs could explain.

We contribute to the empirical literature on the effects of credit access in several ways. First, we investigate the most popular digital credit product and show that eligibility increases total cellphone expenditure, a result that we explain by credit access lifting pressing liquidity constrains. This result is important since digital credit products are quickly proliferating among the newly-connected poor. Comparing our results with other studies is difficult as the study of airtime loans has been overlooked in favor of products that provide larger uncollaterallized loans that can be converted into cash. However, adoption of these products has been slow due to high levels of risk and deployment costs. Bharadwaj, Jack, and Suri (2019), the only evidence available on the effects of extending credit access through digital financial services,[5] confirms that digital credit can reach consumers excluded by formal financial services. Although the digital credit product they study is quite different than the airtime loans we study here, .[6] the loans are often used to directly finance prepaid airtime. Similar to our finding, they observe a large demand for additional liquidity that digital credit seems to be ideally designed to supply. Their results are encouraging as they show digital credit increases the resilience of households to shocks, with credit access allowing households to finance unexpected expenditures that otherwise the household lacks the liquidity to pay.[7] Their

[5]There is an ongoing study of a digital credit product in Tanzania, but results are not yet available.

[6]The loans they study can be converted into cash and are, on average, ten times larger that the average airtime loan (USD$4.8 versus USD$0.50).

[7]Nearly 34% of the eligible population taking, at least one loan, within two years after eligibility.

identification strategy allows them to identify effects only for people with high levels of poverty and no significant access to other credit sources. In contrast, we test for differential impacts of airtime loans on poor and non-poor customers.

Second, we contribute to the understanding of how demand for digital loans responds to both liquidity needs and, for some, convenience. This convenience effect dominates for customers with relatively high income and raises concerns about how digital credit might contribute to overindebetdness. Evidence suggests that the expedience of fund delivery has a detrimental effect on the repayment probability (Bulando, Kuhn, and Prina, 2020), with consumer credit having positive effects only when borrowing responds to unexpected shocks. Otherwise, borrowers are likely to fall into overindebetdness, and lose future access to formal credit options (Carrell and Zinman, 2014; Skiba and Tobacman, 2019; Ausubel, 1991; Bond, Musto, and Yilmaz, 2009; Morse, 2011; Zinman, 2010; Karlan and Zinman, 2010).

Finally, we contribute to the emerging MFS literature by testing for gender-differentiated uses and impacts of airtime loans. Compared to much of the developing world, the gender gap in mobile phone ownership in Haiti is remarkably small: conditional on age, men and women own mobile phones at roughly the same rate. However, women tend to be poorer, more likely to work as self-employed in the informal sector (e.g., fruit or handicraft sales on street corners, cleaning or other services, etc.) and spend about 25% less on network communications. The combination of similar mobile phone ownership rates and systematic economic differences by gender would seem to suggest that airtime loans might have distinctly different impacts on women relative to men. This prior motivates our test for gender-differentiated impact. Somewhat surprisingly, we find that within income (proxy) terciles airtime loans have essentially the same impact on women as on men.

The paper continues as follows. Section 1.2 explains how airtime loans fit into the ecosystem of MFS, with an emphasis on what characteristics have contributed to their rapid widespread in the developing world. This section also describes the cellphone market in Haiti and the specific conditions or airtime loans in the country. Section 1.4 outlines out empirical strategy, and the data. Section 1.5 develops the econometric methodology. Section 1.6 analyzes the heterogeneity in people's responses to credit, focusing on the role of liquidity constraints and convenience. We extend this heterogeneity analysis to gender in Section 1.7 and offer concluding thoughts in Section 1.8.

1.2 Background

1.2.1 Mobile Financial Services

Since the early 2012, more than 50 billion dollars have been invested in the Financial Technology sector by adding digital options to existing financial products and by creating new services that appeal customers that the formal financial sector has struggled to serve (McKinsey, 2017). Financial technologies (FinTech) includes all the services that improve and automate the delivery of financial services. As part of the FinTech sector, Mobile Financial Services (MFS) includes products where mobile phones are an integral part of the user's experience. Depending on the service they provide, MFS can be divided in four categories: mobile money, insurance, savings, and credit.

Digital credit[8] allows subscribers to access short-term loans from a mobile device, with the whole application process processed remotely. It has several advantages over existing formal financial institutions that allow it to serve low income customers. It manages information asymmetries using non-traditional sources of data. By not relying on 'hard' data such as proof of income, employment or a formal credit histories, it can screen customers for which traditional services find

[8] Also called mobile credit and digital lending

difficult to asses their risk level. Additionally, the costs of services tends to be lower, in particular in remote areas, as digital products rely on the infrastructure already in place by cellphone companies. Furthermore, the digitalization of products can dramatically lower the wait-time of transactions that applying on physical locations and the manual sorting of applications entails.[9]

The risk profile of each digital credit product depends on how it combines alternative credit scoring, usage of mobile money to distribute funds, and the amount of collateral required. Consider three tiers of such products. The first tier, where airtime loans reside, involves the lowest level of risk exposure for providers. Products in this category leverage the cellphone's data-trail as a credit scoring system, with only a person's number as collateral. The loan can be used to make calls, send SMS or use the internet. The only consequence of defaulting a loan is that the customer loses the ownership of his cellphone number. Airtime loans present several advantages over other products that explain its rapid adoption in most cellphone markets. First, low risk makes them attractive for a Mobile Network Operator (MNO) with no previous experience with credit products (GSMA, 2014). Second, they can be launched as a stand alone product that does not require the development complementary services, in particular of a network of mobile money agents with enough liquidity to manage large withdraws; a factor that several MNOs have find challenging (Suri, 2017). Third, the product does not require a partnership with a financial institution, and tends not to be subject to regulatory approval. Although statistics are hard to come by, it seems that nearly every network operator that offers a prepaid service also offers at least one version of an airtime loan, with slight differences in the terms of the service and the size of the loans it makes available. While airtime credit products are not included GSMA's Mobile Money Deployment Tracker making difficult to obtain information of similar products around the world (GSMA, 2014), we have found at least

[9]Several products deviate from one or more characteristics but are still part of the ecosystem.

one MNO offering a similar product in every market we surveyed in Latin American, sub-Saharan Africa, and Asia. It seems safe to assume that these popular financial products exist in every country in the world unless explicitly prohibited by law. Due to their popularity with MNOs and customers alike, airtime loans provide a ubiquitous first rung in the financial inclusion ladder for billions of mobile phone users who have never before had access to formal financial services.

The second tier consists on digital credit products that use formal credit histories and rely on bank accounts to disburse funds. These products are provided by traditional lenders that use mobile applications as a way to reduce frictions during the loan application process. Evidence suggests that the expedience of fund delivery has a detrimental effect on the repayment probability Bulando, Kuhn, and Prina (2020).[10] Additionally, the easier application process allows lenders that use digital channels to charge a 'convenience' premium over their brick-and-mortar competitors that cannot be explained by differences in the cost of regulation or raising funds between different provider (Buchak et al., 2018).

The third tier contains products that similarly to tier one rely on cellphone metadata to screen customers, but with the additional characteristic that loans can be converted into cash.[11] The first product of this kind was launched in Kenyan in 2012, and during the first two years of its existence made over 20 million loans, many for sums of a few dollars, to 2.6 million borrowers (Cook and McKay, 2015).[12] This higher-end digital credit products have experience a slower pace of adoption for reasons that include higher risk, the need to develop an ecosystem of services to support the

[10] The authors exploits that loan are disbursed in batches to identify the how longer delays affect loan repayment, finding that one additional hour of delay causes a 0.4 percentage points increase in the repayment probability. Since the usage of the funds is not observed, they are constrained on the mechanisms behind this finding. The results is not driven by customers, who wait the longest, repaying the loan with the funds just provided; an option that is allowed. Loans are paid close to maturity independently of the waiting period.

[11] Some products use additional data extracted from a customer device that includes information on application usage.

[12] For a review of the state of the market in 2017 see Francis, Blumenstock, and Robinson (2017)

product, and the fact that they need a partnership with a financial institution which makes them subject to regulatory approval. The cost of development and maintenance of these products can be large. As reported by Björkegren, Blumenstock, and Knight (2020), there is evidence that customers strategically change their behavior to manipulate the algorithm in their favor, requiring constant updating of the credit scoring algorithm to avoid increments in the default rate.[13] Yet, as airtime loans prove the feasibility of providing uncollaterilized credit to customers, algorithms improve, and competitive pressure increases, we expect that MNO will become more conformable, and willing, to expand credit access using digital loans that can be converted into cash.

1.2.2 Cellphones in Haiti

Haiti is the poorest country in the Americas with a quarter of the population making less than 1.90 dollars a day (The World Bank, 2020).[14] In terms of phone ownership, the country lags behind the rest of the Western Hemisphere with only 60% of households owning a mobile device in 2018. Still, this represents a large increase from only 20% eight years before, and makes cellphones the second most commonly owned asset, only behind beds (72%), and above radios (52%), TVs (37%), and fans (20%) (FinScope, 2018). There is reliable cellphone service in the whole country, with operators offering additional services such as mobile money and airtime loans.

In a typical month, there are 3.5 million active subscribers to the network operated by our collaborating MNO.[15] As is the norm of the cellphone market in developing countries, the majority

[13] We do not know of any evidence that customers change their usage patterns to get more favorable access to the airtime loans. This is not surprising given the relatively low stakes and the simple rule granting access to the product. We can not discard that customers defaulting an airtime loan do it strategically to get a new number soon after. However, the data suggests that the problem, if exists, is a minor concern.

[14] For comparison, the poverty rate in the Latin America region is close to 3.5% and has been on a downward trend for several years. In contrast poverty rate in Haiti has been stagnant for most of the decade (The World Bank, 2020)

[15] There is a large number of lines that are active for short periods of time. We focus on numbers that have been active for at least four consecutive months.

of cellphone customers are prepaid.[16] Postpaid plans are available but there are several reasons that hinder their adoption. First, they are expensive with lower end plans costing more than the monthly expenditure of 96% of prepaid customers. Second, they create a financial commitment that most households prefer to avoid given the volatility and uncertainty of their income. Finally, lack of proper documentation and financial information makes that most users would simply not classify for postpaid billing.[17]

Panel A in Table 1.1 shows summary statistics for a typical month of usage both for recharge and communication transactions. A key characteristic of prepaid plans is that they do not restrict the amount or schedule of recharges. As documented in other studies, payment flexibility induces a pattern of transactions characterized by small and frequent purchases of airtime, with recharges that tend to coincide with the timing of cash-payments, (Attanasio and Frayne, 2006; O'Donoghue, 2020; Jack and Smith, 2020). We observe a similar pattern with the median customer spending around 3.8 dollars in over eleven different recharges during a typical month. Individual recharges are small, with an average amount of only USD$0.30. Most customers are active everyday, and have, on average, 8 unique contacts in a regular week. This is a similar level of usage to what other studies report in the developing world, see Khan and Blumenstock (2016).[18]

[16]As an example, Bharadwaj, Jack, and Suri (2019) reports that prepaid connections account for about 95% of total subscribers in India, as are 97% in Kenya, 98% in Tanzania, and 74% in Brazil.

[17]Acquiring a postpaid plan requires customers to approach one of the companies' offices and show proof of identification and financial documents. However, only 75% of Haitians have an official ID card. Once a customers is approved, he must leave a one-month deposit. Limited postpaid plans start at 1,800 HGT (25 USD), with more expensive plans ranging between 3,500and 5,000 HGT (50-70 USD) per month.

[18]An entry level postpaid plan costs around 25 USD, Figure 1.5 shows how this more than the monthly expenditures of 95% of active customers. Figure 2A shows that there is a strong preference for small and frequent recharges that takes place along the whole distribution of total expenditure.

Table 1.1: Network statistics: Active customers April 2019

Panel A	mean	std	1%	10%	50%	90%	99%
Recharge activity							
Total recharge (monthly USD)	6.98	15.77	0.0	0.45	3.9	15.23	50.14
Average recharge (USD)	0.55	1.01	0.13	0.17	0.32	0.96	4.43
Number recharges	14.21	15.61	0	2	11	30	61
Communication Traffic							
Contacts called per week	9.91	20.65	0.0	1.5	7.25	21.75	43.25
Days a week with activity	5.09	1.92	0	2	5	7	7
Total Calls	98.12	121.64	0	6	56	241	574
Total SMS	126.99	408.45	0	0	2	327	2072
Panel B: If used loan only							
Number of loans	2.9	2.6	1	1	2	6	13
Total amount borrowed (USD)	2.0	2.8	0.1	0.3	1.2	4.4	11.8
Share of expenses financed	0.28	0.2	0.02	0.08	0.24	0.52	1.05

Note: Includes customers that in April had been active for four consecutive months. Unless otherwise noticed, values correspond to month-aggregates.

1.2.3 Airtime loans in Haiti

In Haiti, most prepaid customers have access to airtime loans. The only eligibility restriction is that a phone number must have been in the network for four weeks, and report at least one recharge in the previous month. In practice these conditions make access to the product almost universal, with 97% of active numbers being eligible. Airtime loans are very popular with 40% of eligible customers requesting at least one loan each month; this percentage increases to 65% when considering loan demand over a two-month period. To understand the magnitude of the reach of the product, it is worth considering that 46% of the adult population does not have access to any financial service, with two out of three loans coming from informal lenders, friends, and family (FinScope, 2018). The high demand for airtime loans is intrinsically linked to high dependence on prepaid plans. When a customer runs out of balance, he cannot new initiate transactions unless more balance is added to the account. Finding places to recharge is not difficult, every street vendor offering recharges, as well as the possibility to buy airtime on formal shops, or electronically using mobile money or a web application using a debit card. On a typical day 89% of all recharges are made with street vendors and 8% and using mobile money.[19]

Loans can be requested directly from any handheld device.[20] When a customer requests a loan, the system provides a single loan offer that ranges between 0.13 and 2 dollars. Figure 1.1 shows most of the loan transactions are less than one dollar, with a median loan size of USD$0.39 dollars (mean USD$0.56). After accepting the offer, but before the loan is disbursed, the customer must read a menu that explain the loan conditions. The customer agrees to pay the loan principal and a 10% facilitation fee before thirty days. The total facilitation fee does not change in the case of early, or even immediate repayment. We only see transactions where a customer accepts these conditions. The total amount can be paid in multiple installments but the full amount must be paid before additional credit can be obtained.[21] As panel B in Table 1.1 shows, the median borrower takes two loans each month. This adds up to 1.2 dollars, or 20% of his total expenditure. The amount the system offers varies depending on the customer's recharge history and correlates with the average amount deposited in the past.[22]

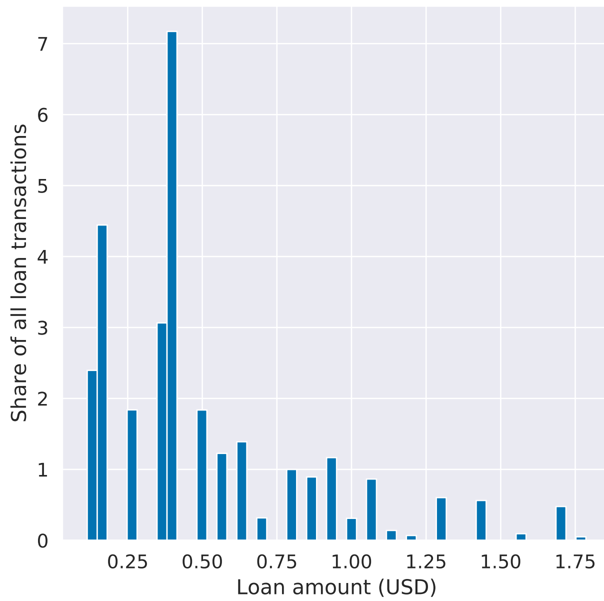
[19]Recharging using a mobile money account requires to have a positive balance. It is not possible to take Mobile Money loans.

[20]The system works both over Unstructured Supplementary Service Data (USSD) and a proprietary mobile app. USSD is an interactive, menu-based technology, supported on most mobile devices. It is similar to SMS with the main difference that messages travel directly to the mobile network provider, creating a two-way exchange of data between users and the network. An additional advantage is that it works on any phone without the need to install any app, or the need of mobile data.

[21]One exception is balance transferred from another customer. In the case that balance transferred is less than the principal plus the origination fee, the customer receives 80% of the amount sent, and the rest is used to pay the loan.

[22]In practice, there are two types of loans available. One credit line provides smaller loans that can be used in network-activity only, while the second can be used in any services. As they are accessed using the same platform, have the same service fee, and can be used simultaneously, we treat them as a single product. When requesting a loan, it is not necessary that a customer has zero balance. However, we observe that loan requests occur when balance is approaching zero. It is not possible to access to the precise algorithm that provides the loan offer. However, the loan offer is linear with the average recharge

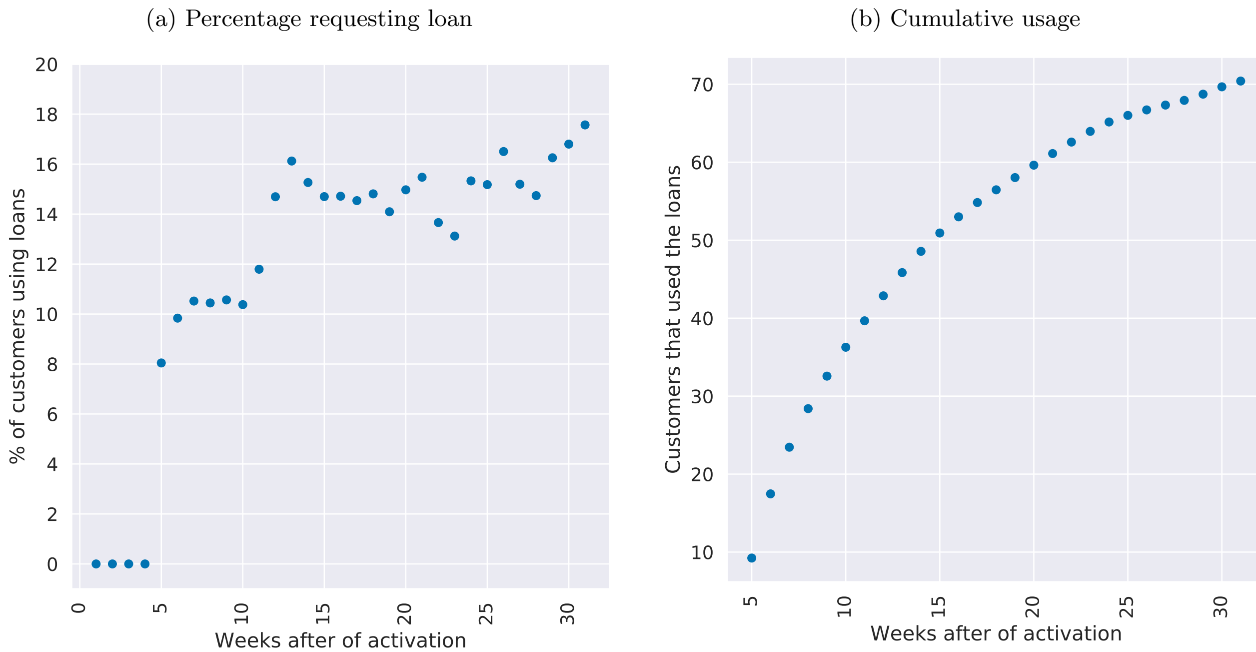
Figure 1.1: Airtime loan transactions amount (May to November 2019)



Note: Loan amounts are discrete values

The only effective collateral for an airtime advances is the customer's phone number. Focus groups we conducted among the working poor in peri-urban Haiti revealed just how attached people become to their phone numbers given the potential costs of changing numbers, including the disruption to one's social network. These costs increase the longer a person has own the number and when the number is used for work. Most loans are fully repaid in less than 5 days with a very low default rate, a result that seems to validate the high valuation people place on their numbers that we found in the qualitative work. Revealing the popularity of airtime loans, Figure 1.2(a) shows that in the week they became available almost 8% of eligible numbers use the service. This percentage increases overtime, with the majority of customers using the loans, at least once, by the time they reach 30 weeks in the network, see Figure 1.2(b).

Figure 1.2: Loan demand by week relative to line activation



Note: New customers become eligible for airtime loans five weeks after activation. Includes only customers who were ultimately active for at least 7 months in the network.

1.3 Theoretical Model

In this section, we introduce a simple model to capture in a stylized manner the intra-day tradeoffs agents face between communication and food expenditure. A key feature of the model is that the demand for cellphone communication occurs at higher frequency than income (e.g., wage) payments. This feature introduces the possibility of an intra-day liquidity constraint (i.e., a mid-day cash crunch). When this intra-day liquidity constraint binds in the absence of a credit market, consumption decisions become sensitive to the timing of income. With this structure, the model is intended to reflect in the abstract the daily financial dilemmas faced by unskilled workers, day laborers and the self-employed in the informal economy, of which there are many in Haiti.

This model has two periods, which roughly correspond to mid-day and end-of-day. During the first period, $t = 1$, the agent has the possibility to consume cellphone minutes m_1, at a price of P, which provides utility of $u(m_1)$. Function $u(\cdot)$ is twice continuously differentiable with $u(0) = 0$, $u'(m_1) > 0$, $u''(m_1) < 0$ for all $m_1 \geq 0$. The total consumption of minutes during $t = 1$ depends on the cash-on-hand the agent has available, which we denote D. Additionally, to consume minutes the agent must recharge his balance account by visiting an airtime vendor. Finding a vendor distracts him from working by an amount of time e, representing a transaction cost of manually recharging one's airtime balance. While this effort cost depends on several factors, including the agent's location relative to airtime vendors and time of day, these are beyond the scope of this simple model. Similarly, we abstract from any uncertainty about this effort cost and, for simplicity, assume it is fixed at $\bar{e}$.

The second period, $t = 2$, represents the end of the working day when the agent receives his earnings and makes additional consumption decisions.[23] Total earnings equal $E(1) = \bar{E}$ if the agent does not spend time looking for an airtime vendor, and $E(1 - \bar{e})$, with $E(1 - \bar{e}) \leq \bar{E}$ if he bought airtime in $t = 1$. The agent consumes a bundle good c (e.g., food) for a price of P_c. The agent can consumer additional cellphone minutes, m_2, at the same price P. However, reflecting that during this period the agent is not working and has to acquire good c by going to the market, we assume there is not penalty for buying airtime. Utility from consumption of c is given by $v(c - \gamma)$ with $v'(m_1) > 0$ and $v''(m_1) < 0$ for all $c \geq \gamma$. The parameter γ reflects that the agent must guarantee a minimum consumption level, effectively making some consumers too poor to buy minutes.

[23] A more complicated model could include the agent receiving a fraction of his earnings discreetly over the day, or having uncertainty over the final value of his earnings.

Given this setup, the agent face the following intra-day maximization problem:

$$\max_{m_1,m_2,c} u(m_1) + \beta\left[u(m_2) + v(c-\gamma)\right] s.t. \tag{1.1}$$

$$Pm_1 \leq D \tag{1.2}$$

$$Pm_2 + P_c c \leq D - Pm_1 + E(1-\bar{e}) \tag{1.3}$$

A solution set m_1, m_2, c is such that:

$$\frac{u'(m_1)}{u'(m_2)} = \beta \tag{1.4}$$

$$\frac{u'(m_1)}{v'(c-\gamma)} = \beta\frac{P}{P_c} \tag{1.5}$$

$$Pm_2 + P_c c \leq D - Pm_1 + E(1-\bar{e}) \tag{1.6}$$

An interior solution is guaranteed by $P_c\gamma > D + E(1)$. We focus on that case since, otherwise, the agent is too poor to buy cellphone minutes.[24] In the case the solution to the maximization is not binded by condition 1.2 the agent has no need for credit markets.

In the absence of credit markets, an agent with limited cash-on-hand D is constrained in his consumption of cellphone minutes at $t = 1$, implying that the marginal rate of substitution between the consumption of minutes at $t = 1$ and $t = 2$ is larger than β, and $\beta\ \frac{P}{P_c}$ for the case of good c (conditions 1.4 and 1.5). In practice, this implies that an agent that starts with a low level of D would be better-off if he was allowed to borrow against his future earnings. The share of D over total income that determines when an agent is constrained depends on the relative prices of minutes, the price of good c, and the relative valuation of minutes consumed in the first versus the second period.

[24]When $P_c\gamma \in [D + E(1-\bar{e}), D + E(1)]$ a solution with $m_1 = m_2 > 0$ and $c > 0$, as well as $m_1 = 0$ and $m_2 > 0$ $c > 0$ is possible depending on how large the cost e is. As we assume that e is relatively small, we rather focus on the case where $P_c\gamma > D + E(1)$.

Airtime loans allow the agent to acquire unlimited balance in period $t = 1$ without the need to incur in the effort $\bar{e}$. By taking a loan, the agent agrees to pay in $t = 2$ the full amount plus a fee r: $P(1 + r)$. This makes minutes used in the first period more expensive and creates two sources that explain loan demand. For people that are constrained at $t = 1$ airtime loans allow them to equalize the marginal rate of substitutions between m_1, m_2, and c, leading to an increase in the total expenditure on cellphone minutes. Second, for agents that are not constrained in $t = 1$, the loan eliminates the effort costs associated with finding an agent. If the cost e is relatively large compared with the increase in the cost of minutes, the agent finds more convenient to pay the cost of the loans instead of recharging through an airtime vendor.

We provide a simple example that shows the existence of these two effects. To illustrate this, we fix total income and vary the share that cash-on-hand at $t = 1$ represents of the total daily income.[25] A characteristic we impose on $u(\cdot)$ and $v(\cdot)$ to reflect the nature of the two goods, but that is not necessary for the solution of the model, is that $u'(x) << v'(x)$, implying that the marginal utility of an additional minute decreases fast compare with the marginal utility of additional consumption of food for the same level of expenditure, making the share of cellphone expenditure small.

When D is small, condition 1.2 is binding such that the total amount of minutes consumed at $t = 1$ equals when $m_1 = \frac{D}{P}$. In this case, condition 1.4 no longer holds. This causes the agent to trade phone consumption in the second period for additional phone consumption in the first period. In Figure 1.3(a), we see that the total utility for an agent with the same level of total income, but that is not constrained in his purchase of m_1 (solid line) is higher than the constrained solution (dashed line). This leads to a consumption of minutes below what otherwise the agent would demand if he could borrow, see Figure 1.3(b).

[25]As the searching cost of airtime vendors is a percentage of E, an agent that starts with a high D will have a searching cost as a percentage of his total income that is lower in absolute terms.

When borrowing is allowed, we see that constrained agents increase their consumption of minutes by a large amount, also increasing their total utility. In models that conform to the canonical Permanent-Income Hypothesis, the timing income payment does not affect consumption decisions, and any increase in the credit limit should not generate significant changes in the debt level. This is not the case when the agent borrowing capacity is limited.

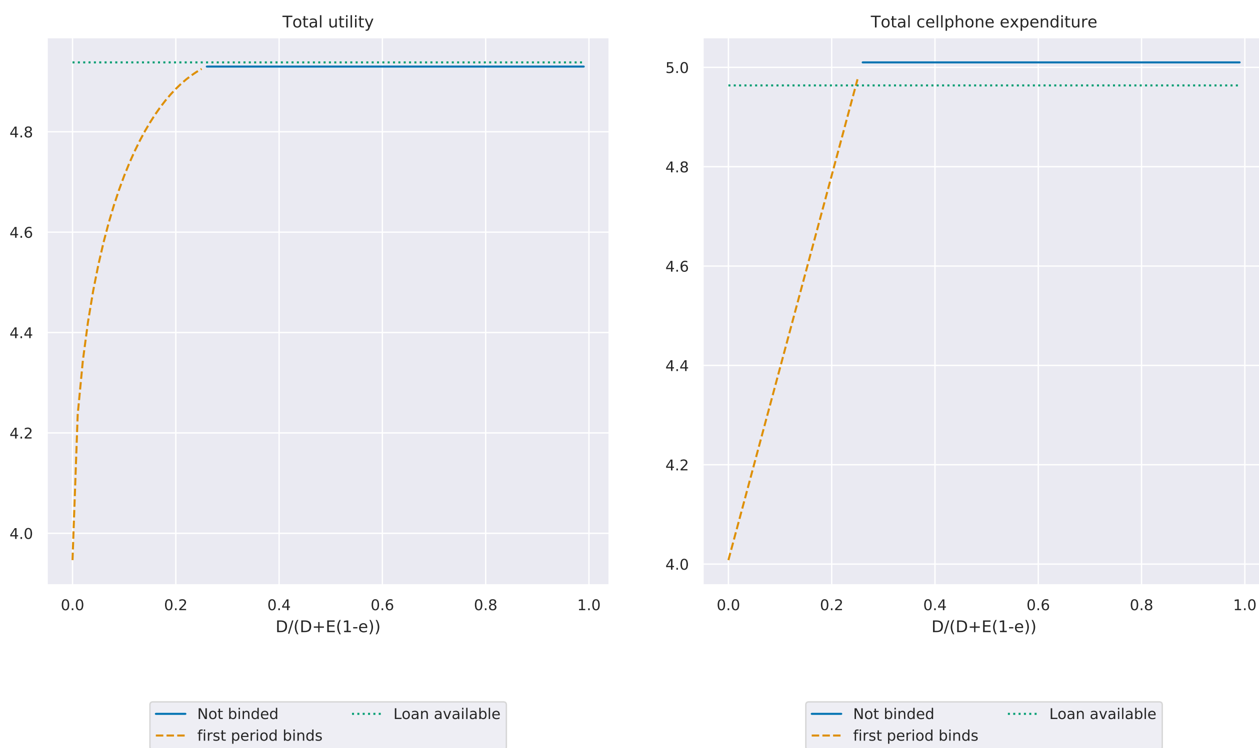
An agent who is not cash constrained at $t = 1$ will use airtime loans because of the convenience they bring by avoiding the need for looking for airtime vendors. Minutes become more expensive, but when the costs associated with recharging during $t = 1$ are large enough the agent prefers to take loan over the consumption bundle that is purely financed with the agent own resources. It is worth mentioning that in this case, the change in total consumption tends to be small because it is only driven by the transaction costs of recharging. If they were absent, a non-constrained agent would not use the loans. An additional channel that could induce demand for loans from unconstrained agents in the absence of searching costs is the presence of uncertainty on the second period earnings, as it can induce precautionary saving motives concerning the possibility that income might bind in the future (Gross and Souleles, 2002).

Finally, consider two simple extensions to this model. Without modifying the model to make these extensions, we discuss why they might make sense given our research question and how they might change the patterns in Figure 1.3. First, it would be reasonable to introduce heterogeneity in daily earnings that is correlated with cash-on-hand D (e.g., poor with low D and low E versus non-poor with high D and high E). Obviously, this would amplify the effort costs associated with airtime recharges in $t = 1$ for the non-poor relative to the poor and, with it, the convenience of airtime loans for the non-poor. Second, to reflect the fact that many self-employed workers in the informal sector now rely on their phone to find and coordinate jobs, we could include minutes as an argument in the earnings function as well as in the utility function. This would naturally amplify the effect airtime loans have on crowding-in additional communication expenditure. These two extensions would only sharpen the core results of the model that access to airtime loans (i) prompts differential responses from the cash-constrained poor and the unconstrained non-poor and (ii) crowds-in additional communication expenditure for the former but not the latter.

Figure 1.3: Utility and communication expenditure

(a) Utility level

(b) Total phone expenditure



Note: Model simulations varying the share of income the agent has access when the day starts.

1.4 Empirical Strategy

1.4.1 Identification

To identify the impact of credit access on cellphone expenditure and network behavior, we leverage the eligibility rule that grants access to airtime loans four weeks after a line is activated. This allows us to implement an event study design (Athey and Imbens, 2018).[26] Proper identification depends on several assumptions. The first is parallel trends in the absence of treatment. A weaker version of this assumption, that is more likely to be satisfied, only requires parallel trends conditional on covariates, see Callaway and Sant'Anna (2018). The second assumption is no anticipatory behavior. As stated by Sun and Abraham (2020), this is more plausible when participants do not have private knowledge about the treatment path that might change their behavior in anticipation of the treatment. In the setting we study, it is possible customers are aware that after a certain period they will have access to airtime loans. Evidence from other digital credit products shows customers are willing to take costly actions, like changing their network patterns, or buying pre-used a SIM card, to gain access to loans (Björkegren, Blumenstock, and Knight, 2020). We argue that the value of airtime loans is sufficiently low to deter such behavior; a fact that the low default rate seems to support. Moreover, even in the case people increase their expenditure prior to eligibility to classify to larger airtime loans, our results would only under-estimate the true impact.

The final assumption imposes no variation across cohorts. These requires that each cohort experiences the same path of treatment effects, in particular, that the composition of individuals does not differs over time in characteristics that affect how they respond to treatment. Additionally, we need that the treatment effects are the same across cohorts in every relative period, that is,

[26]This is a special case of a general Difference-in-Differences strategy that has been applied empirically to a wide range of contexts. Such designs are sometimes also referred to as Staggered Adoption Design (SAD). For a complete review of the studies implementing a similar methodology see Clarke and Schythe (2020)

that the type and intensity of treatment does not vary due to calendar time-varying effects. Given that we rely on administrative-data, it is not possible to test for differences in the characteristics of the customers entering the sample each week. We argue, however, that once we center the analysis on customers that stay for a significant period in the network most of the differences between customers disappear.[27] The eligibility criteria and size of the loans does not change over the period studied, and any calendar effects that seasonality might introduced are controlled by their respective dummies.

As our objective is to understand changes on the levels of the transactions, we aggregate all entries for the same customer at a weekly level.[28] Working with data aggregated at the week-level has the advantage that it filters most of the noise created by both inter and within-day fluctuations. Additionally, aggregating each customer's transactions at a week-level facilitates the estimation, since the data in its original form has more than two billion entries. We first estimate a standard two-way fixed effects model as described by equation 1.7:

$$y_{i,week} = \alpha + \beta_1 Eligible_i + \mu_i + \lambda_{week} + u_{i,week} \tag{1.7}$$

Our main variable of interest is total weekly expenditure. This variable aggregates all the recharge transactions a person makes during the week using any of the recharge methods available. We also explore the impact of credit access on different network features that include the number outgoing contacts, average call duration in seconds, and number of outgoing interactions. We find that total expenditure is a better indicator of network behavior as people strategically changes how many people they call, how often, and for how long, in the presence of low balance. The estimates for

[27] As described by Roth (Working Paper), restricting the sample to only customers that do not drop from the sample can induce selective survival bias. Our results are robust to lowering the inclusion criteria by allowing customers that drop early from the sample, see 5A

[28] Each week contains Monday to Sunday. Week of the year is defined according to the International Organization for Standardization (ISO)

this and all subsequent models use standard errors clustered at the individual and week levels.

This model uses a single post-eligibility indicator $Eligible_i$ for all periods after airtime loans become available.[29] We also include μ_i and λ_{week} to capture individual and calendar-week fixed effects. As a control group, we include a random sample of well-established lines. Due to data limitations we cannot say precisely when these lines were first activated, but they were active at least five months before the period we study. We chose a random sample twice as large as the number of lines that we observe as treated, we implemented several tests and our results are robust to the size and the sample draw for the control group. The eligibility of these phone numbers does not change during the period studied and they act as a counterfactual. Additionally, including these numbers allow us to properly estimate the calendar week fixed effect.

A growing literature shows that the presence of heterogeneous treatment effects in a specification like equation 1.7 is problematic as customers who are treated first are weighted more in the estimation of the coefficients with the weights proportional to the size of each treatment unit and the number of periods treated (Goodman-Bacon, 2018).[30] To account for this, we also estimate a model with coefficients for each week a subscriber in the sample:

$$y_{i,week} = \alpha + \sum_{j=-4}^{-2} \beta_k (Lag\, j)_{i,week} + \sum_{k=0}^{7} \gamma_k (Lead\, k)_{i,week} + \mu_i + \lambda_{week} + u_{i,w} \tag{1.8}$$

Lag and lead are dummies defined with respect to the number of weeks a number is from gaining access to airtime loans. We identify the fifth week after activation as week zero and define the lags and lead accordingly. As the notation shows, we omitted the lag for the week before a customer

[29] As the first week in the network is very noisy, we only use weeks two to four to estimate the baseline of the pre-eligibility indicators.

[30] Goodman-Bacon (2018) shows that the difference-in-difference estimator is a weighted average of all 2x2 estimators in the data, This makes that this estimator can easily change between specifications as controls can induce additional identifying variation.

became eligible. It is common practice to use the first lag as baseline. However, we decided to use the last lag because the first week of activity tends to be noisy and not to represent the activity levels in the following three weeks. The eligibility rule only allows us to observe a numbers' activity during four weeks before access to the loans is granted. As numbers can be activated during any week of the year, this event takes place at different dates depending on the calendar-week of activation. The limited window of pre-loan eligibility also drives our decision to measure impacts only eight weeks after loan access is obtained; therefore each new customer is in the sample for only twelve weeks.

For completeness, we implement as a robustness check, two additional specifications. First, we implement a simple event study design where we do not include the random sample of well-established subscribers as a control group. Second, we take advantage of our relative large sample and use the subscribers that joined the network during calendar weeks 18 to 21 as a control group for the new subscribers entering the network during weeks 20 to 30. During this period, the eligibility status of the control group does not change. Results for key variables under the two additional specifications are provided in section 1.A.1, and are qualitatively the same as in our preferred specification.

1.4.2 Data

For this study, we used an anonymized database from the largest cellphone provider in Haiti, between January and December 2019. For each transactions, we observe the day, time and duration, as well as the cellular towers that connected the subscribers. Additionally, for each subscriber we have a daily register with information containing the time and amount of airtime purchases, balance transfers, and airtime loans usage. Similar data have been previously used to study population movement (Gething and Tatem, 2011; Lu, Bengtsson, and Holme, 2012; Zagatti et al., 2018), risk sharing in the face of natural disasters (Blumenstock, Eagle, and Fafchamps, 2016), and forecasting socioeconomic trends (Blumenstock, Cadamuro, and On, 2015; Frias-Martinez et al., 2013).

We aggregate the transactions of each customer at the week-level. This makes that, in calendar-year terms, our data covers week 18 to 48 of 2019. To observe how network behaviour changes before and after access to the airtime loans, we focus on new lines activated between May and July 2019 (calendar week 18 to 31), and follow their activity until the end of November of the same year (calendar week 48). During this period, a total of 278,697 new lines were activated, with new number entering the sample at a relative constant pace (Figure 1.4a). In the Appendix, Figure 1A shows the data coverage in terms of calendar-weeks in 2019.

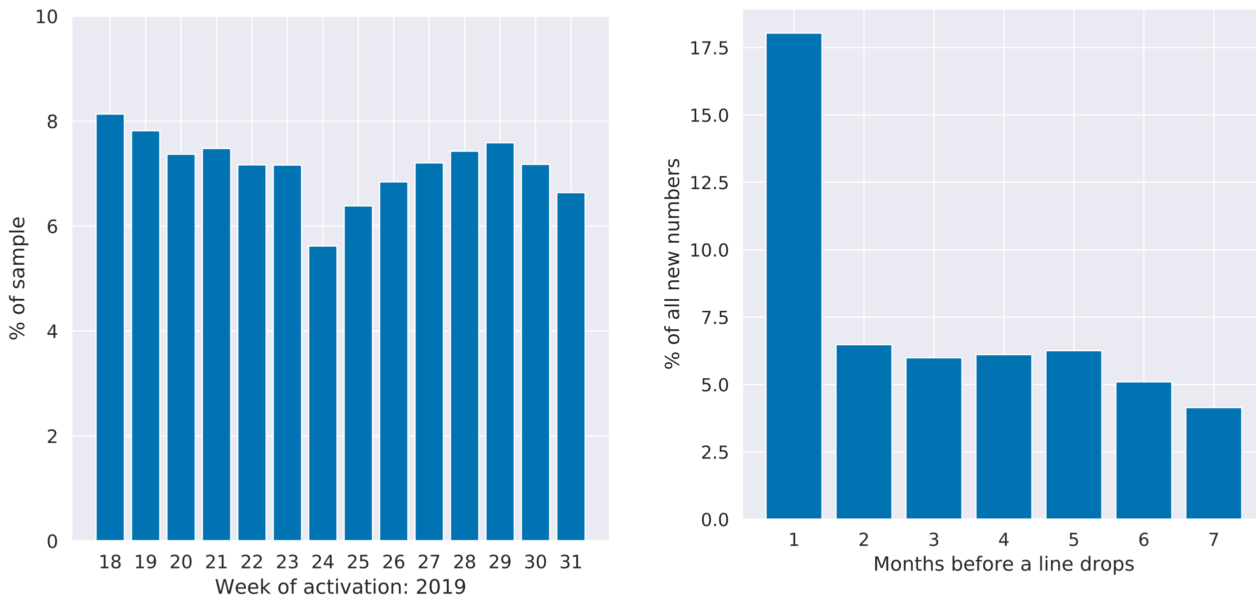
Similar to the experience in other settings, there is a large level of subscribers churn (Roessler et al., 2018). Only 39% of the lines activated during this period remained active when our records stop. We call these lines the long-term customers. The largest attrition occurs during the first month, when almost 18% lines stop registering activity. After this initial drop, attrition continues at a slower pace over the following months (Figure 1.4b). Customers that stop using their numbers are free to obtain a new number without any penalty, however, the eligibility condition still imposes a waiting period of four weeks before the new number can obtain loans.[31] We do not find evidence that a number dropping from the sample correlates with having outstanding loan balance.

[31] A customer losses ownership of the number if he does not recharge during four consecutive months; in that case, the number can be reassigned. We assign as the last date we stop observing transactions as the day the line was dropped. One additional group that we identify are lines that we classified as sparse activity but still active (10%). These lines have gaps in activity for more than four weeks, but then register additional activity so it is not possible to classify them as inactive numbers, see Figure 3A.

Figure 1.4: Activation new lines (May and July 2019)

(a) Percentage of new lines activated weekly

(b) Attrition over time

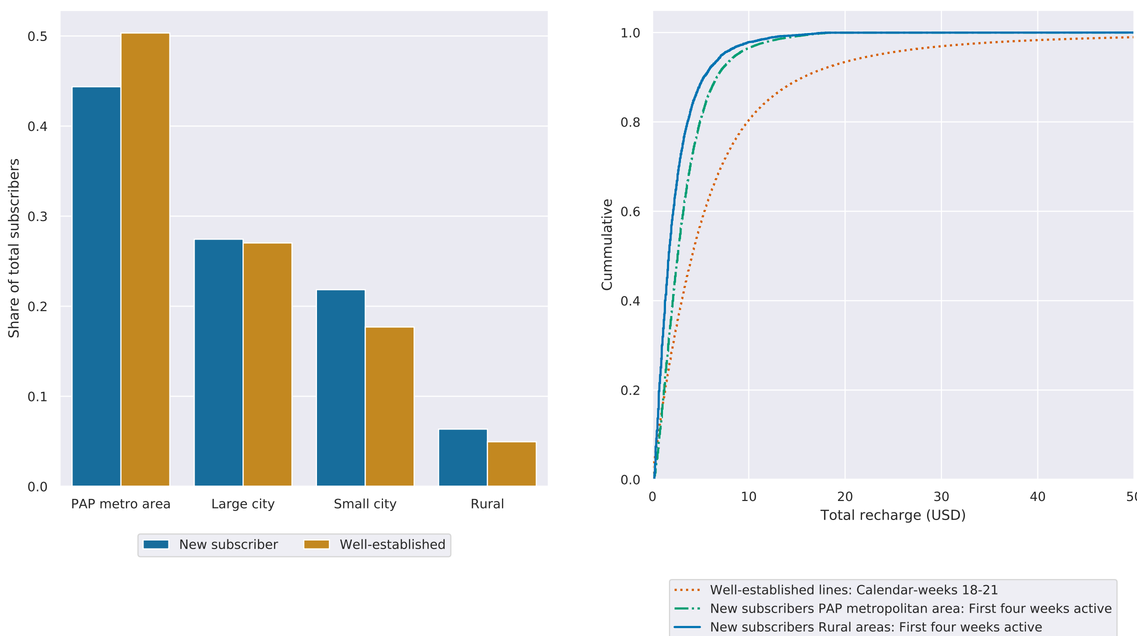


Note: Week of activation makes reference to the calendar-year. Week 18 corresponds to May first.

We do no have personal information to compare the individual characteristics of subscribers in the long-term customers group with the set of well-established lines. However, evidence suggest that early adopters of cellphone technology are male, concentrated sector of the urban areas with higher income. Based on this, we should expect that the marginal new customers belongs to less favored groups. From the administrative data we see that, in fact, new lines have a slight higher probability to be locations in rural areas and outside the Port-au-Prince metropolitan area (Figure 1.5(a)). Still, phone subscribers are concentrated in the metropolitan area of the capital, where 65% of lines are located but only 37% of the population live, see Table 1A for details. Additionally, we observe that the total expenditure of the long-term customers is systematically lower than for the whole universe of well-established line (Figure 1.5(b)), a pattern that suggests that early adopters represent a wealthier segments of society.[32] We add a longer discussion on who new customers are with respect of the overall network in section 1.6.1, and account for differential effects by location as a robustness check in section 1.A.2.

Figure 1.5: Key network metric activities for long-term customers

(a) Home tower location (b) CDF cummulative expenditure



Note: Well-established lines includes numbers that were active at the time when data first became available and remained active during the period of study. Total expenditure includes the calendar weeks 18 to 21 during 2019. New subscribers includes lines activated between May and July that remain active long-term. Location was assigned using the tower that managed most of night activity of each subscriber.

[32]For reference, new subscribers in the top 10% of the expenditure distribution are below the expenditure of the top 10% of well-established lines (Figure 6A).

1.5 Results

In this section, We showing the results from equation 1.7 and then shift to graphical depictions of the lead and lag specification in equality 1.8. Table 1.2 shows the impacts in monetary and as a percentage change with respect to the baseline values before loan availability. Loan access increased the expenditure of new subscribers by 16%. The additional expenditure comes with a marginal increase in the number of recharges from an average of 2.6 to 2.8 recharges per week. In terms of network activity, it is difficult to point to a single metric that explains the additional expenditure; instead, this increase captures the net effect of several simultaneous changes in network usage. Overall, subscribers end up having shorter, but more frequent interactions, while maintaining the same number of total unique contacts. Specifically, after loan eligibility the average customer makes 3.7 more calls that are on average 3 seconds shorter and increases total communication on the network such that their total expenditure increases by 16%.

Table 1.2: Impacts of airtime loans in post-eligibility period

	Expenditure (USD)	Average recharge (USD)	Number of recharges	Outgoing contacts	Outgoing transactions	Average call duration	Gambling expenditure (USD)
Baseline	0.92	0.31	2.58	6.74	41.07	74.93	0.01
Effect	0.15***	0.01***	0.19***	0.13*	3.72***	-2.65***	0.0
Δ in percentage	16.07	3.51	7.52	1.97	9.07	-3.53	1.33

Note: The effect variable shows the results of a difference-in-difference where the pre eligibility period includes the three weeks before eligibility and the post period the 7 weeks that follow.

One concern about facilitating access to digital credit is that it can be used to finance gambling, which is particularly relevant in our setting because the Haitian lottery is exceptionally popular (Dizon and Lybbert, 2021) and is offered by our collaborating MNO as a gaming service that can convert airtime balance to lottery credit.[33] While there are no official data on the number

[33] There is not an official register of how many Haitian regularly gamble. However, it is telling that more than $1.5 billion dollars are spent per year, and it lottery stalls are a easy to see in most streets, with over 35,000 independently

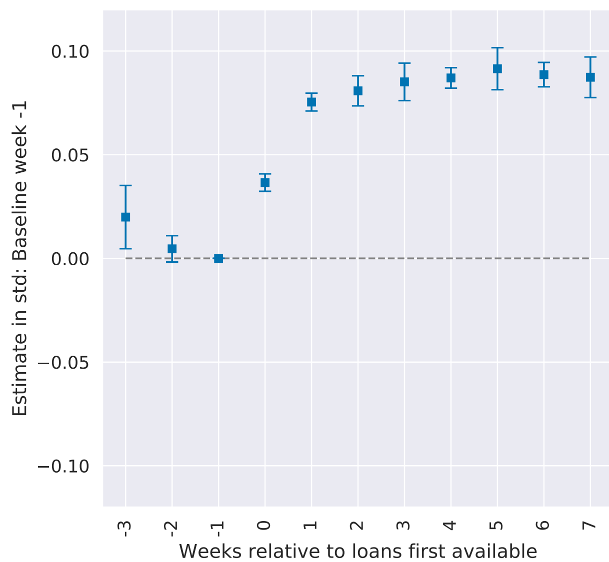
of lottery players nationwide, we find that based on cellphone records around 33% of subscribers on the network play the lottery on their phones at least once per month. We test the impact of airtime loans on total expenditure on lottery gambling in Table 1.2 and find no effect. This suggests reassuringly that airtime loans are not systematically redirected to lottery wagers even though they could be.

We now transition to graphical depictions of the lead and lag specification in equation 1.8 that shows how these effects change by week relative to eligibility. Unless explicitly noted, we present these graphical results in terms of standard deviations of the dependent variable. We find that loan access increases weekly expenditure in a magnitude equivalent to 0.09 standard deviations, a magnitude that is in line with the aggregated result that we provide in Table 1.2. This impact starts in the week after loans are available and reaches (and maintains) this level by the second week after customers are eligible (Figure 1.6). Figure 1.7 provides an overview of the key network metrics. The patterns present large variation before and after eligibility. These results hold when we also include in the sample numbers that dropped 3 months after they were activated (Figure 5A)

Results suggest that access to credit crowds-in additional expenditure. Loan fees only account for 1% of total weekly expenditure, with the loan's principal financing 5% of total expenditure in the week airtime loans become available, and 12% a couple of weeks later (Figure 1.8). To provide a definite answer on the welfare effects of this result, it would be necessary to have information on the return to calls financed by loans, the effect on consumption of other goods, and the extend that airtime loans replace or complement other credit sources. We lack the data properly answer this question. However, as overall expenditure increases over several week, this is indicative that there

owned lottery stalls in the country Bhatia (2010) must be a reduction in consumption of other goods, or in the levels of savings.

Figure 1.6: Total weekly expenditure



Note: Includes only long-term customers and Well-established lines. Loan access is provided at week 0 and the week before is used as baseline.

The pre-loan transactions we observe emerge as the equilibrium between the need for cellphone usage and each customer's capacity to prepay for the service. To explain why credit access increases total expenditure in a magnitude above repaying the loan's fees, three mechanisms can be at play. First, credit access relaxes binding liquidity constraints. As loans are quickly repaid and overall expenditure increases, these effects are consistent with subscriber perceiving the value of holding cash, and not spending in airtime, as high. This can be the case for poor individuals who depend on the informal economy and earn a living during the day, making them extremely sensitive to the timing of their income. Second, airtime loans introduce a behavioral component that affects the salience of the costs of calls as they eliminate the need to pay upfront for airtime using cash. From a theoretical perspective, a sophisticated but present-biased agent faces clear incentives to limit the amount of airtime balance available at any given time to prevent future self from over-consuming. (Laibson, 1997; O'Donoghue and Rabin, 1999). By allowing instantaneous access to airtime, this self-control mechanism embedded in prepaid disappears. This mirrors the impact of switching to prepaid electricity billing, where the change led to a reduction of total consumption as it made explicit the cost of electricity for those consumers that switched (Jack and Smith, 2020).[34] A third mechanism is that the transaction costs of looking for a vendor during certain times of the day are perceived as very high, deterring a subscriber from recharging even when he has cash available. Under this mechanism, access to credit eliminates the cost of finding an agent and allows customers to modify their recharge patterns across the day towards times when they find it more convenient.

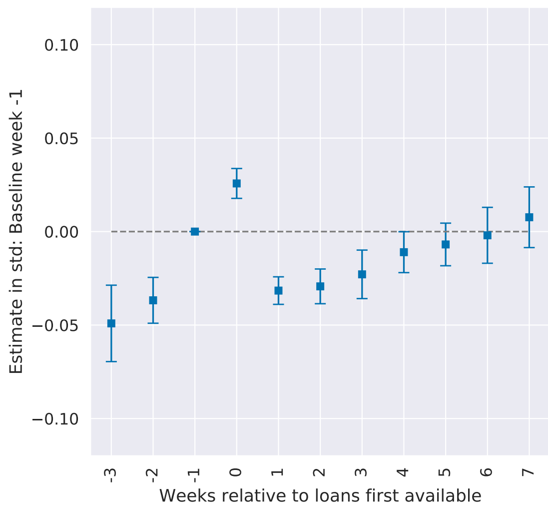
1.6 Heterogeneous Effects by Income

While the average impacts of airtime loans as presented in the previous section are important, they may reveal distinctly heterogeneous effects that reflect the extent to which different mechanisms and motivations shape individual reactions to airtime loan offers. In this section, we provide empirical evidence of differential impacts on communication expenditure for poor and non-poor consumers. As our main analysis relies on administrative data, we do not have a direct measure of income to use in this analysis. Instead, we use baseline network expenditure as a proxy for income after showing that total airtime expenditure correlates strongly with income for a subsample of customers who responded to a survey that elicited income (a stylized fact that multiple studies have corroborated (Gutierrez, Krings, and Blondel, 2013; Blumenstock, Cadamuro, and On,

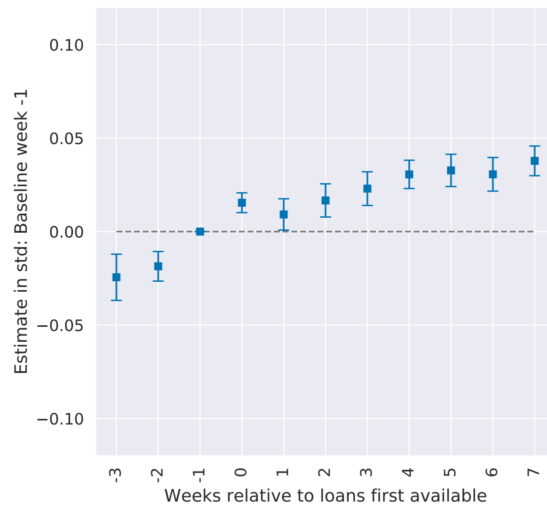
[34]The literature that explores the expenditure patterns of the poor also finds storage costs as a limiting factor on keeping large airtime balances (O'Donoghue, 2020). This does not apply as airtime take a month to expire, and can be transferred to other subscribers or renewed.

Figure 1.7: Key network metric activities of long-term customers

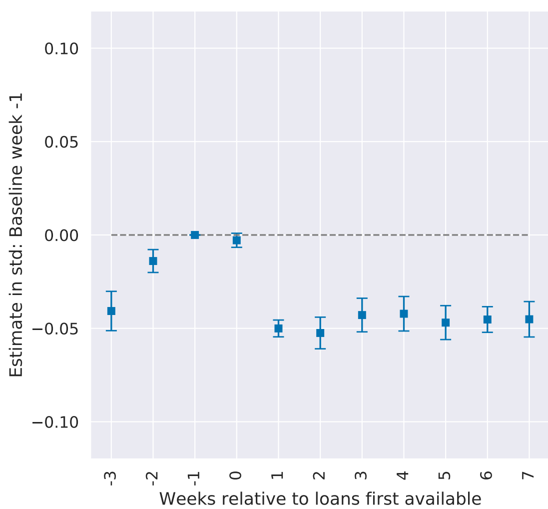
(a) Outgoing contacts



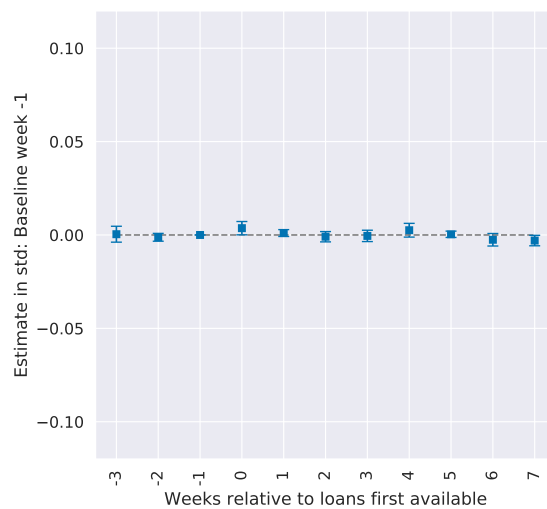
(b) Total transactions (calls+SMS)



(c) Average call duration



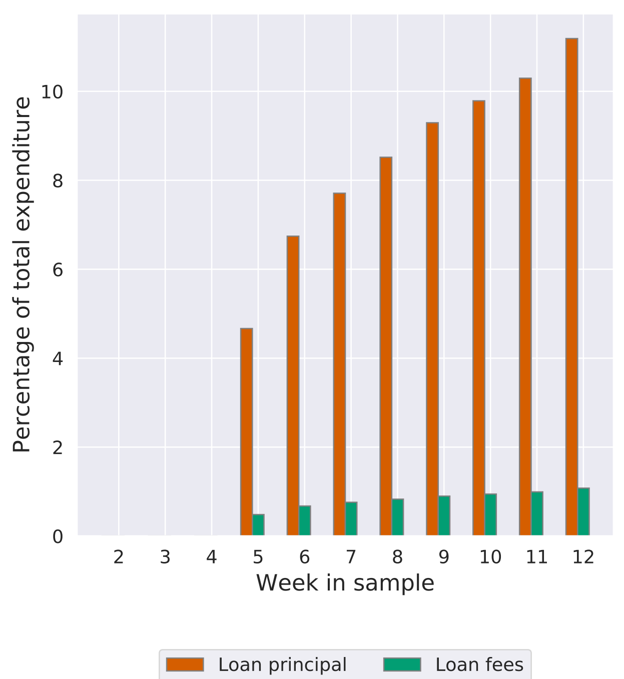
(d) Gambling expenditure



Note: Includes only long-term customers and Well-established lines. Loan access is provided at week 0 and the week before is used as baseline.

2015; Blumenstock, 2018b)). Results show that poorer customers in the lowest tercile of initial communication expenditure more than double their mobile communication spending when airtime loans become available, while access to loans leaves expenditure of the highest tercile unchanged. These pronounced differences exist despite relatively uniform patterns of loan usages between the two groups. Mapping our results to the mechanisms described above, we find suggestive evidence that poorer customers appear to use loans to relax liquidity constraints at critical communication times whereas non-poor customers primarily use loans for their convenience, as it gives them more discretion in when to visit airtime vendors.

Figure 1.8: Share of total expenditure financed by loans



Note: Includes only long-term customers. The x-axis shows the number of weeks a customer has been active.

1.6.1 Income level and cellphone expenditure: A simple approximation

Several studies show that it is possible to predict individual socioeconomic indicator using the data-trail created by cellphone usage. These applications are useful to obtain economic indicators in data-poor setting, and to update existing data at a lower cost.[35] These applications rely on statistical methods to detect a relation between a socioeconomic indicator and cellphone usage patterns.[36]

There are still several open questions about how to best exploit cellphone records and the properties of the predictive models. Two specific limitations of these methods prevent us from using them to predict income for individuals in our data. First, these methods use relatively long series of retrospective cellphone data.[37] In our case, the eligibility period provides a window of only one-month to collect the survey, which implies that we have very limited cellphone metadata to work with. Second, it is unknown how fast a model's predictive capability decays over time and when it is applied to different samples and different time periods. Given the complexities in our sample in these respects, we are unsure how much we could trust income predictions in our case. We therefore opt for a simpler approach that uses total communications expenditure to proxy for unobserved income. Other studies show that this relationship exists, even if it is not as predictable as machine learning methods (Gutierrez, Krings, and Blondel, 2013).

[35] A regular LSM survey requires the National Statistical Institutes to hire and train a large numbers of enumerators. An expensive task relative to the budget of emerging countries.

[36] These patterns include, among other, the number and average size of recharges, the of number of calls, the reciprocity of the calls, and the average distances travelled by citizens (Frias-Martinez et al., 2012; Gutierrez, Krings, and Blondel, 2013; Blumenstock, Cadamuro, and On, 2015; Blumenstock, 2018b). State of the art models do not depend on a single variable, and in several cases the use feature engineering to create features that can not be easily interpreted. Due to widespread data limitations Haiti has a long history in the usage of these methods. In the past, cellphone Detail Records, similar to the ones we use, have been helpful to understand the impacts of natural disasters on population displacement (Gething and Tatem, 2011; Lu, Bengtsson, and Holme, 2012; Zagatti et al., 2018)

[37] In order to implement these methods using individual predictors, it is necessary to be able to link individual characteristics of subscribers with their own cellphone metadata. A process that, by law, requires that the number owner agrees.

We use a phone surveys with 600 respondents that are representative of the universe of mobile money users.[38] As part of the informed consent process, we received authorization to link their answers with the mobile phone transaction database. We match survey answers to each participant cellphone records during the four weeks prior the survey in order to capture both weekly and monthly communication patterns.[39]

The majority of participants were male and head of their households (61% and 55% respectively), and had higher wealth levels than the average Haitian. Still, we observe high levels of food insecurity, with 61% reporting skipping meals or reducing portion sizes. Revealing the potential of MFS, all of them had a mobile money account but only 10% any type of banking product. Demand for airtime loans is high, with only two people not using the product in the 30 days before the survey. As a form of credit, airtime loans are used with a higher frequency than other credit products by survey respondents. Descriptive statistics on the survey participant characteristics are shown in Table 2A.[40]

We only have weekly income for survey respondents who were employed. For this group, we estimate equation 1.9 to understand how income correlates with total cellphone expenditure, average amount and number of recharges. Total cellphone expenditure and the average size of recharges increase with the reported income, while we do not see any effect in the total number of recharges (Table 1.3).[41]

[38]The information was collected in July 2019 as part of a related project. The universe of mobile money holder is 939,315. Restriction on the levels of activity and time in the network left us with a sample of 36,879 potential subscribers to survey. Two levels of compensation (50 and 150 HGT) in the form of airtime were offered as a participation incentive. We do not find evidence that the response rate was different depending on the compensation level.

[39]By using a shorter period, we run the risk of omitting the recharges of consumers that recharge a single large deposit per month.

[40]With the exception of bank loans, most people have debts that are less 2 dollars, and amount that is not far from the credit provided by airtime loans. However, for debts owed to family and neighbors less than 5% accrue interest

[41]A similar story can be seen in Figure 9A, where we see that the recharge terciles map to higher levels of income and larger average recharges. We also checked the relation between income stability and income. We find that people with more predictable income spend more on their cellphones, both in total and per recharge (Table 3A)

$$y_i = \beta_0 + \beta_1 ln(income)_i + \beta_2 age_i + \beta_3 men_i + \beta_4 Household\,Head_i + \beta_5 Day\,survey_i + u_i \quad (1.9)$$

Table 1.3: Network transactions and income for phone survey respondents

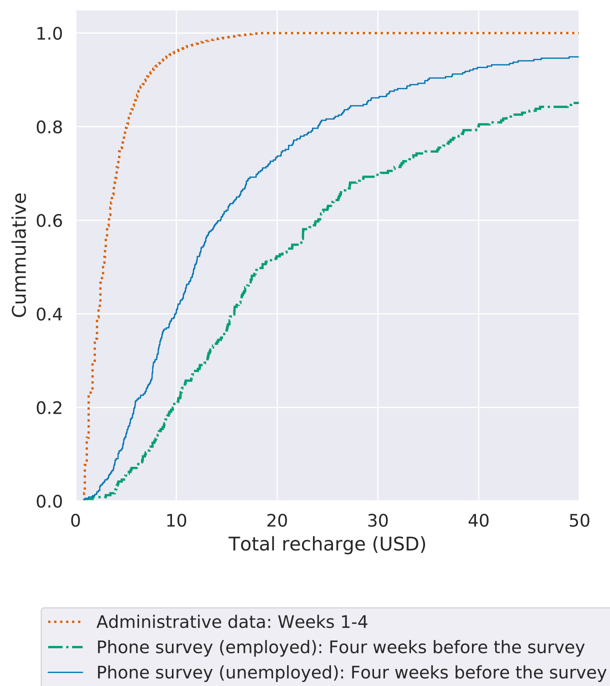
		Recharges	
	Total Recharge (USD)	Average recharge (USD)	Number of recharges
ln(income)	7.74***	17.58**	1.29
	(2.64)	(7.49)	(0.97)
Household head	6.06	20.02**	-2.42
	(4.68)	(9.34)	(2.52)
gender	4.8	-8.87	6.06***
	(5.86)	(14.47)	(1.94)
age	0.09	0.45	0.05
	(0.27)	(0.56)	(0.11)
const	-43.75**	-73.35	13.17
	(21.42)	(61.53)	(8.66)
Observations	306.0	306.0	306.0
R2	0.07	0.07	0.08
Adjusted R2	0.02	0.02	0.03

*p<0.1; **p<0.05; ***p<0.01

Note: Includes only respondents that had labor income in the week prior to the survey.

To show how cellphone expenditure levels compare between the survey and the administrative data, Figure 1.9 shows the cumulative distribution of total expenditure for the long-term customers and the participants in the phone survey. These distributions suggest that the survey sample contains wealthier individuals, which likely has two explanations. First, the sampling process was done on mobile money users, a population that several studies show tend to be younger, more urban, and wealthier (Khan and Blumenstock, 2016). Second, the survey sample contains more well-established lines, which tend to have higher levels of expenditure.

Figure 1.9: Cumulative distribution of total expenditure.
Long-term customers and survey respondents



Note: Administrative data contains only long-term customers. Weeks 1-4 represent the first month after activation when airtime loans where not available.

1.6.2 Heterogeneous effects of airtime loans

We explore the role of economic status on creating heterogeneous responses to credit access. As we do not have economic information for all the individuals in the cellphone transaction data, we leverage the discussion in the previous subsection and divide customers into three groups depending on their expenditure in the four weeks before airtime loans were available. Based on the premise that expenditure levels provides a simple proxy for income, we proceed to explore how access to credit affect customers in a differential manner. Table 1.4 shows that there are important differences in the expenditure levels between the groups, with the median person in the high expenditure group spending five times more than the median customer in the low-expenditure category.[42]

Table 1.4: Total Expenditure before loans are available (USD)
Long-term customers

	count	mean	std	min	25%	50%	75%	max
Low Expenditure	32,598	0.89	0.40	0.13	0.58	0.91	1.23	1.56
Medium Expenditure	32,027	2.40	0.52	1.56	1.95	2.34	2.82	3.38
High Expenditure	31,717	6.11	2.89	3.38	4.09	5.13	7.01	18.56

Notes: Includes all the recharge transactions during the first four weeks. As we use terciles of the total expenditure the number of people in each group is very similar.

Estimating equation 1.8 for each individual group reveals a high heterogeneity of impacts. Results show that for low expenditure group, credit access leads to a large increase in total expenditure and sizable growth in the level of network transactions. Panel (a) in Figure 1.10 shows that the increase in expenditure takes place soon after loans became available. Weekly expenditure more that doubles, with individuals in this group spending, on average, 1.4 times more each week, gaining one more contact and making fourteen more transactions.[43] Similarly, the group with in the second tercile of initial expenditure experienced an increase of 30% in their weekly expenditure. In contrast, people in the high expenditure group maintain similar levels of expenditure and we observe little difference in their level of network transactions.[44] Still, the expenditure and level of transactions of the low expenditure group remains below the levels of more affluent groups, see Table 1.5.[45]

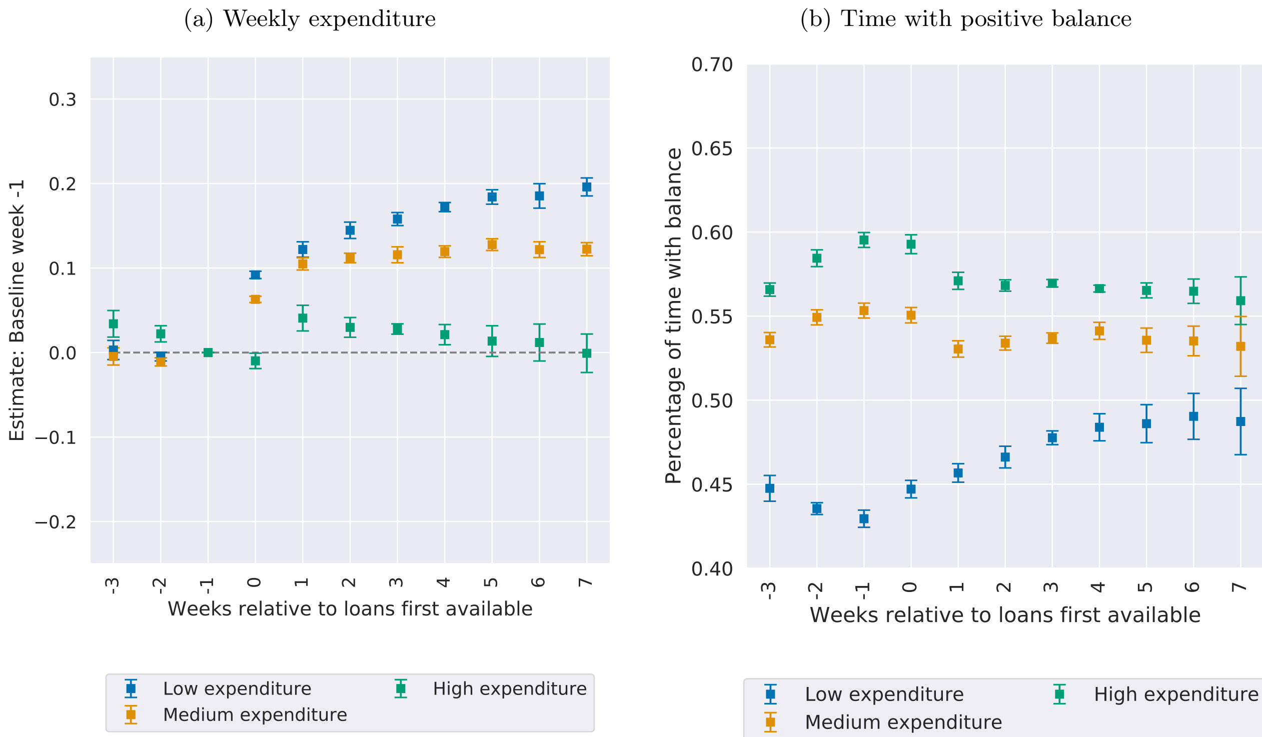
[42] Total expenditure also presents differences in terms of the average amount of airtime bought in each group. As in (Gutierrez, Krings, and Blondel, 2013), we find that people with higher levels of expenditure tends to make larger average transactions (10A).

[43] In monetary terms, the total weekly expenditure went from 0.20 to 0.55 dollars.

[44] We do not find that the location of subscribers significantly affects the magnitude of our results, with people living in urban and rural areas reacting on a similar way to the introduction of airtime loans (Figure 14A).

[45] Figure 7A shows estimations at the week-level for key network variables.

Figure 1.10: Heterogeneous impacts of airtime loans



Note: Includes only long-term customers and well-established lines. Loan access is provided at week 0 and the week before is used as baseline.

The different size of the impact of access to airtime loan between low and high expenditure individuals exists despite relative uniform patterns of loan usage. Specifically, airtime loans finance, on average, 9% of the expenditures of the low expenditure group in the eight weeks after they first become available; only two percentage points more than the share financed for the group with a higher initial expenditure (Table 1.6).[46]

One possibility is that airtime loan access affects the incentives to keep a positive balance. A subscriber using a pre-paid account has incentives to keep enough credit to cover (at least) his short-term call needs in the absence of the loans. The optimal balance that a subscriber holds depends on

[46]Table 4A shows the share of total expenditure financed with loans each week. For details on the probability of borrowing each week see Figure 4A in the Appendix.

several factors, including the utility from calls, the transaction costs associated with buying airtime, and the perceived value of holding cash. Holding constant the transaction costs of recharging and the utility from calls, we expect that a low-income subscriber is less willing to substitute cash for airtime balance than a subscriber with higher means. Additionally, the willingness to buy airtime goes down if we introduce uncertainty about future cash flows.[47]

We reconstruct the balance available on each subscriber's account during each week. We consider that an account has a positive balance if it has enough credits to cover a one-minute call. Considering that most calls last less than 60 seconds, this is enough to cover the most common call in the network.[48] We find that the average pre-paid account spends half of the time with no balance. High income subscribers spend more time with positive balances, but even this group does not spend more than sixty percent of a given week with enough credits to cover a one-minute call. Panel (b) on Figure 1.10 shows that access to airtime loans has little effect on how much time subscribers have a positive balance in their accounts.

Table 1.5: Heterogeneous impacts

	Low expenditure			Medium expenditure			High expenditure		
	Baseline	Effect	Δ in percentage	Baseline	Effect	Δ in percentage	Baseline	Effect	Δ in percentage
Expenditure (USD)	0.22	0.33***	148.77	0.63	0.24***	38.13	1.62	-0.0	-0.26
Avg. recharges (USD)	0.14	0.07***	49.79	0.25	0.03***	10.42	0.47	-0.03***	-6.82
Number of recharges	0.95	0.74***	77.7	2.48	0.2***	7.93	4.02	-0.27***	-6.7
Outgoing contacts	4.39	1.2***	27.42	7.16	-0.3***	-4.12	8.62	-0.61***	-7.13
Outgoing transactions	17.25	14.2***	82.32	40.59	3.01***	7.41	62.19	-5.34***	-8.59
Avg. call duration	57.92	2.41***	4.17	77.68	-6.52***	-8.39	89.13	-5.67***	-6.36
Gambling expenditure (USD)	0.0	0.0***	54.25	0.01	-0.0	-3.48	0.01	-0.0**	-7.88

Note: Baseline levels show the average weekly expenditure during the three weeks prior to access to credit and compares it with the average outcome in the eight weeks that follow.

[47] This is what a model with precautionary saving motives predicts, see (Gross and Souleles, 2002; Lang, 2020)

[48] For reference, the most common recharge amount can cover about a five-minute call. This estimate ignores VoIP calls. A person can call other numbers using a messaging app. However, using those applications requires that the person called also has a smartphone and internet service available.

The previous results show that for the group with a low initial expenditure access to airtime loans crowds-in additional network expenditure. On the other hand, those with higher income have a similar demand for loans but their total expenditure remains unchanged. In the next subsection, we argue that these distinct patterns by income are the result of different motivations driving these individuals to tap airtime loans as a form of credit.

Table 1.6: Loan demand by group

	Borrowed	Average weeks with loans	Total loans	Average expenditure financed
Low Expenditure	0.45	2.46	3.31	0.09
Medium Expenditure	0.43	2.26	2.77	0.08
High Expenditure	0.41	2.26	2.73	0.07

Note: Groups were defined using the terciles of total expenditure in the four weeks before eligibility. Only long-term customers.

1.6.3 Heterogeneous Motivations for Using Airtime Loans

We revisit the mechanisms that explain how credit access increases cellphone expenditure. Our results align with the existence of a liquidity constraint for poor customers. In the context we study there are large imperfections in the credit market that make consumption patterns extremely sensitive to the cash available at any point in time. If, additionally, there is uncertainty on future income, the sensitivity of expenditures to cash patterns increases. We observe this precise pattern in the case of poor customers who more than double their expenditure once the introduction of airtime loans allows financing cellphone usage. We cannot dismiss that, to some extent, there is also a reduction in the salience of the costs of calls since taking a loan avoids paying up front in cash. However, given the magnitude in the increase of expenditures, and that it lasts for several weeks, we believe this second mechanism is marginal, with the main impact of airtime loans being on reducing the extend liquidity constrain limit cellphone expenditure.

In the case of subscribers with a higher income, they present similar demand for loans but do not change their expenditure. This group faces less binding liquidity constraints, but seem to value the convenience of airtime loans as a way to top-up their prepaid balance. To understand the role convenience on credit demand, we explore for changes in the recharge patterns across the day. For this, we build on the fact that we are able to observe the precise timing of every recharge transactions. With this intra-day timing, we can test the convenience value of airtime loans. These loans offer an easy alternative to physical recharging at times when it is difficult to find a vendor or agent or in a pinch when one unexpectedly hits a zero airtime balance and urgently needs to make a call.

The first type of convenience is the possibility to use loan to recharge at times outside business hours. Despite most of the airtime loans are taking during business hours, their share of the total airtime expenditure is highest precisely at the time when finding a street vendor is more difficult. Since that street vendors manage almost 90% of recharge transactions, we consider that the hours between midnight and six in the morning as periods where finding a place to buy airtime is difficult. Figure 1.11 panels (a) and (c) shows that demand for loans increases once economic activity starts around 6am, and remains relatively stable until the end of the working day. This is similar to the demand for airtime and the total volume of calls.[49] Figure 1.11 panels (b) and (d) shows the total amount of airtime bought using airtime loans over actual recharges. We see that the share of total transactions, airtime loans represent most of the airtime used precisely at the time when finding

[49]Recharge transactions follow a similar pattern across the day. Figure 11A shows recharges transactions per hour across the day, both as a share of total transaction (panels (a) and (b)), and as a share of total monetary value (panels (c) and (d)).

a street vendor is difficult. This pattern is mirrored by poor and non-poor customers without statistically significant differences between the two groups.[50]

The second, and more interesting, type of convenience is when airtime loan are used to strategically change the recharge pattern. We test for differential changes in hourly recharge patterns between poor and non-poor consumers. The high frequency of the data provides the precise timing of every recharge transaction. We estimate changes after customers become eligible using equation 1.10.

$$\begin{aligned} recharge_{i,day,hour} = \alpha + \sum_{h=1}^{24} \beta_h hour_{i,day} + \gamma Eligible_i + \sum_{h=1}^{24} \beta_h hour_{i,day,hour} \times Eligible_i + \\ \mu_i + \lambda_{week} + u_{i,day,hour} \end{aligned} \tag{1.10}$$

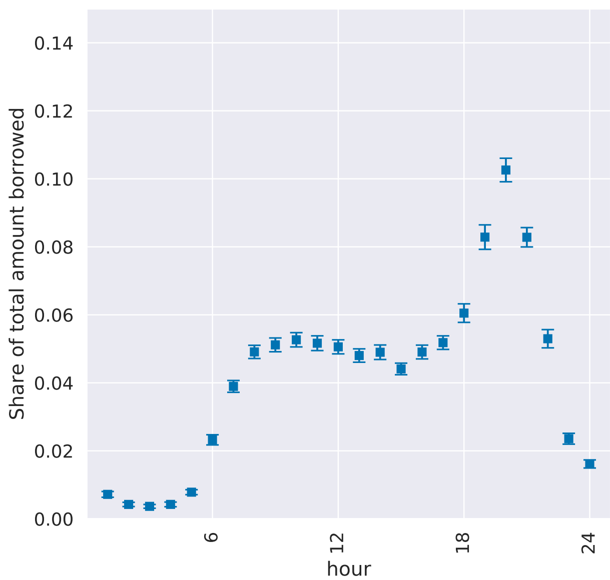
Our coefficient of interest is the interaction between the hour dummy and the indicator if a customer is eligible for the credit product.[51] To understand the results from Figure 1.12, it is important to remember that airtime loans eligibility reduces the number of recharge transactions for the non-poor, with the opposite effect on the poor. We see that the reduction in the number of recharges for the non-poor is particularly marked in the recharges that happen after 7pm, a time of the day when it is more likely that the transaction costs of recharging are higher. Poorer customers, on the other hand, increase their expenditure during those hours. A results consistent with the idea that, poor customers, who more likely to only have certainty over their daily incomes at the end of the day, wait until then to decide how much to recharge.

[50] Figure 12A presents the same results but by number of transactions per hour

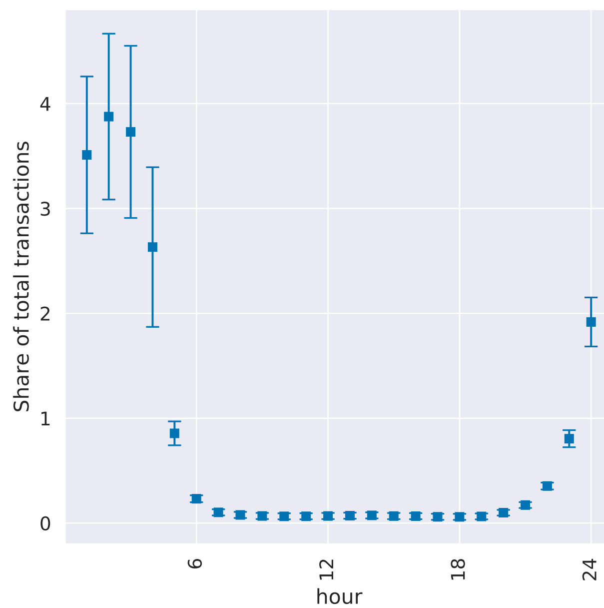
[51] This estimation requires that we have a dataset that indicates for each customer if a recharge transaction happened at any hour of the day. As the memory requirement grows exponentially, we use a random sample with a third of the original subscribers, and aggregate the transactions between midnight and 5am, and 10 and 11pm into a single dummy. After several iterations we do not find that the results change significantly if we draw a different random sample.

Figure 1.11: Share of amount borrowed (hourly)

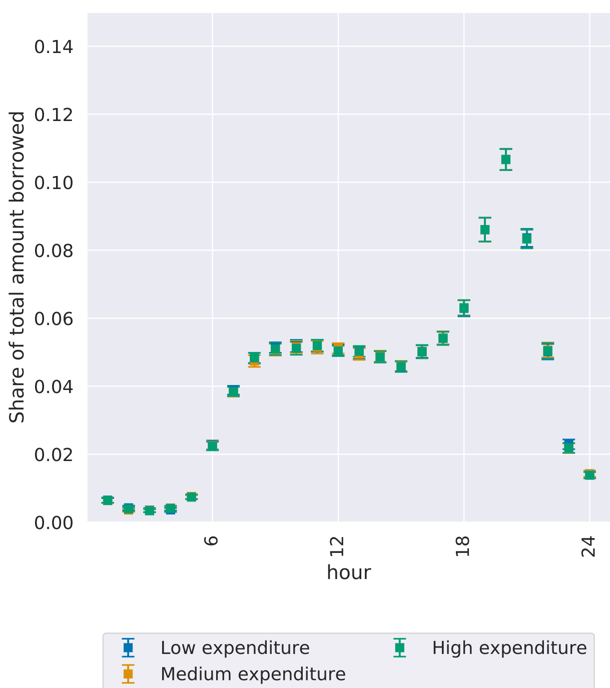
(a) Amount borrowed



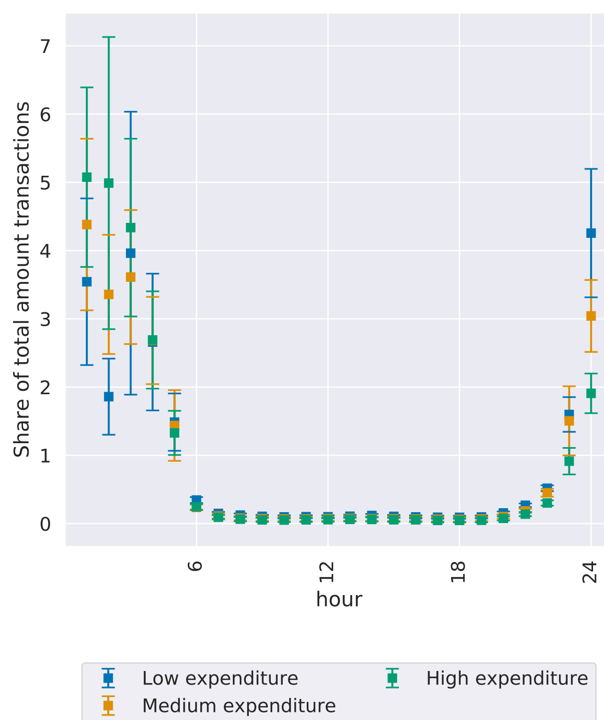
(b) Amount borrowed over amount recharged



(c) Amount borrowed
by initial expenditure tercile



(d) Loan transactions over recharges transactions
by initial expenditure tercile



Note: Includes only customers that are eligible for the loans. The estimation of daily demand patterns includes controls for day of the week and calendar week.

Figure 1.12: Changes in recharge probability (hourly)

Note: Includes only long-term customers. Transactions before 6am were aggregated

1.7 Heterogeneous Effects by Gender

Mobile phones can enhance welfare and enrich living standards through the communication, financial and other services they offer, but whether and how an individual benefits from these network services varies widely based on a host of factors. In the previous section, we explored how the impact of airtime loans is distinctly heterogeneous by income (proxy) terciles. In this section, we turn to a second primary dimension of heterogeneity: gender.

1.7.1 Mobile phones and gender in Haiti

Globally, mobile access has increased dramatically since 2010, but women are not equally represented among these new subscribers. The mobile gender gap varies significantly across regions, but on average women are 10 percentage points less likely to own a mobile phone (GSMA, 2019). In South Asia and Africa, this mobile gender gap is particularly high (28 and 15 percentage points, respectively). By contrast, the mobile gender gap in Latin America is much smaller or even vanishes altogether.

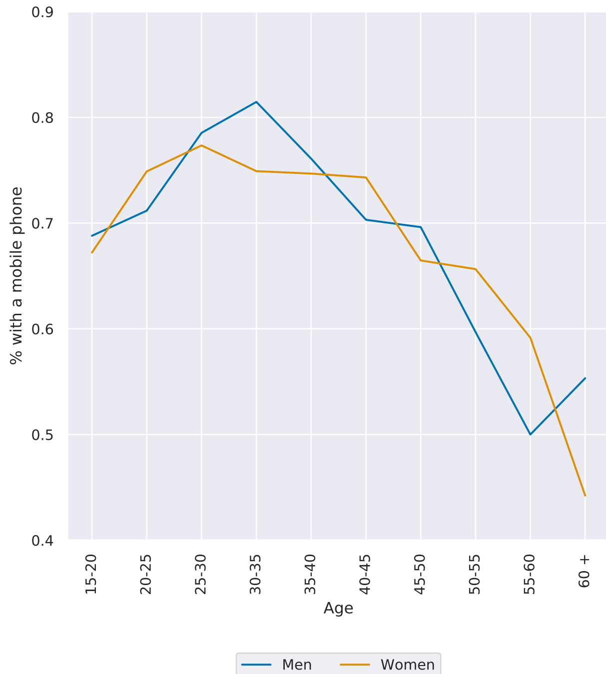
The case of Haiti is interesting: While its high poverty rate and low cellphone ownership rate are on par with some Sub-Saharan Countries, its gender gap looks much more like Latin American than Africa. The small gender gap holds across age groups and regions. As Figure 1.13 shows, most of the variation comes from older individuals who are 25% less likely to own a phone than a person in his thirties.

Relative parity in access to cellphones does not ensure that women can benefit fully (equally) from the communication and financial services offered by mobile networks. Other structural gender gaps - e.g., education, occupational choice, and income - continue to shape how women use these services and to what effect. Given these persistent and structural differences between the economic realities of men and women in Haiti, the usage and impacts of airtime loans could be heterogeneous by gender in important ways.

To explore this dimension of heterogeneity, we combine a representative phone survey of cellphone users in the country with administrative data. The survey took place between January and April of 2021 and included 2,361 participants, 40% of them women.[52] We matched survey respon-

[52]The lower participation rate of women can have several causes, among them, lower activity levels by women, and a lower predisposition to provide information to strangers over the phone. This lower engagement of women is a factor that we have experience before conduction phone surveys in the country.

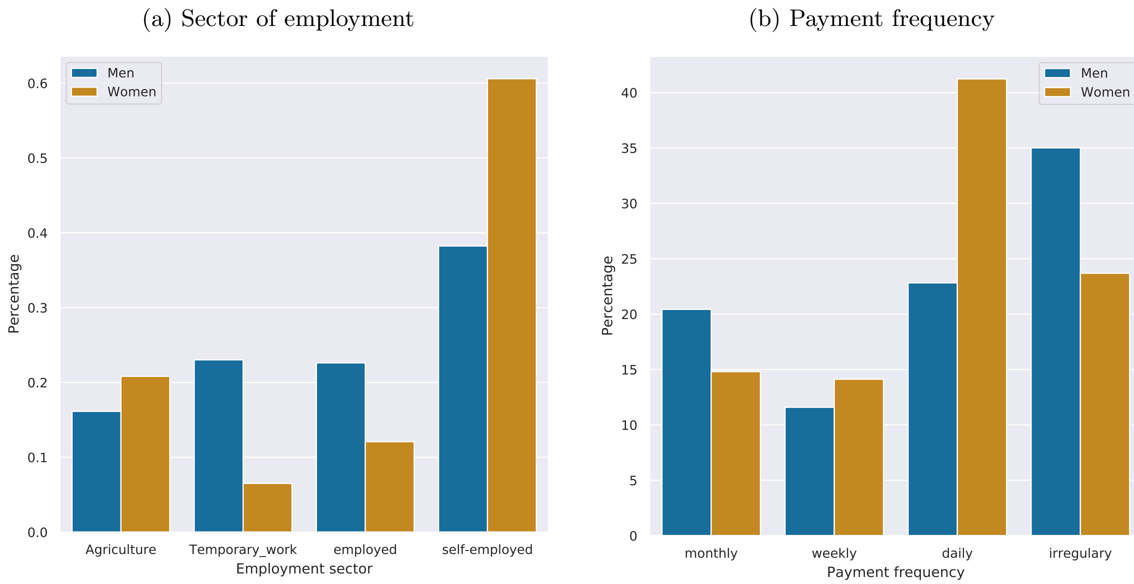
Figure 1.13: Mobile phone gender gap in Haiti by age group



Note: Author's calculation using Finscope (2018) data.

dents, including data on employment, food security, and wages, with their cellphone records to explore how financial insecurity drives cellphone usage. Fewer women participate in the labor force than men (74% versus 80%), and the employment profile of working women differs from that of men's. While both are much more likely to be engaged in informal work than employed in formal sector jobs, women depend much more on self-employment (typically, as street vendors or service providers) with wages that are less predictable and that are earned in small (daily) installments, see Figure 1.14. Men are more likely to work temporary jobs with irregular payment. While neither work and remuneration arrangement is particularly conducive to financial stability and planning, it is especially challenging for women: 66% of female respondents report being frequently short of cash to fulfill their financial obligations compared to 55% of male respondents.

Figure 1.14: Employment profile



Note: Author's calculations using information from phone survey participants.

Although the combination of cheap devices and prepaid plans made cellular service accessible to almost everyone, high levels of financial constraints can still limit mobile usage. With matched survey responses and CDRs by individual we can better understand how men and women manage their prepaid accounts. As shown in panel (a) of Table 1.7, women spend about 25% less than men on network communications on average, a difference that is driven by fewer recharges in a month rather than by recharge amounts. Panel (b) shows how payment frequency correlates with cellphone expenditure patterns. People paid on monthly installments spend higher amounts, while people with irregular incomes spend the lowest amounts on a typical month.

Table 1.7: Gender differences in cellphone expenditure and loan demand

	Total recharge (monthly USD)	**Average recharge (USD)**	**Number of recharges**
Panel (a): Expenditure patterns by gender			
Men	10.2	0.7	16.2
	(23.3)	(0.9)	(15.1)
Women	7.7	0.6	14.3
	(10)	(0.9)	(13.8)
Difference	2.5***	0.1	1.9***
Panel (b): Expenditure patterns by payment frequency			
Monthly	11.2	0.8	16.3
	(18.7)	(0.8)	(18.3)
Weekly	9.1	0.6	15.6
	(13.5)	(0.5)	(13.2)
Daily	9.9	0.6	16.7
	(23.7)	(0.9)	(15.5)
Irregularly	8.6	0.6	15.5
	(14.1)	(1.1)	(14.1)

Note: Author's calculations using information from phone survey and their CDR transactions for November 2020.

In unreported analysis, we find that women tend to have fewer contacts than men, but interact with them more frequently. Additionally, we find women are active fewer times in any given week and with a lower geographical dispersion of their activity, but their transactions tend to have a longer average duration. Exploring intra-day variation in network usage, we find no evidence that women recharge at different times of the day than men.

1.7.2 Heterogeneous gender effects of airtime loans

Since we do not directly observe gender for the new numbers we use in our eligibility analysis, we use our survey data combined with the matched CDR data as training data to predict gender. We implement a Random Forest Classifier with a 10-fold cross-validation. We achieve an AUC of 0.65, a results in line with previous literature results (Frias-Martinez, Frias-Martinez, and Oliver, 2010; Al-Zuabi, Jafar, and Aljoumaa, 2019). For the new numbers that are activated between May and July, our model classifies 48% of them as women with no significant differences in terms of initial expenditure tercile. The model predicts a lower percentage of female owners (34%) for well-established lines, which is consistent with evidence that early adopters are disproportionately male with the gender gap closing over time.

We use these predicted gender classifications to disaggregate total communication expenditure by gender and across the terciles of initial expenditure in Table 1.8. These terciles, as before, are constructed for the all long-term customers, so the breakdown by gender is not forced to be balanced, but the gender composition of each tercile is remarkably constant. This suggests, again, that despite systematic employment and remuneration differences by gender, the distribution of total initial expenditure is very similar for men and women.

Table 1.8: Total communication expenditure by gender
Before airtime loan eligibility

	Female	Percentage	mean	std	min	25%	50%	75%	max
Low Expenditure	0.0	51.17	0.90	0.40	0.13	0.58	0.91	1.23	1.56
Low Expenditure	1.0	48.83	0.90	0.40	0.13	0.58	0.91	1.23	1.56
Medium Expenditure	0.0	51.34	2.41	0.52	1.56	1.95	2.36	2.86	3.38
Medium Expenditure	1.0	48.66	2.40	0.51	1.56	1.95	2.34	2.79	3.38
High Expenditure	0.0	51.63	6.11	2.86	3.38	4.09	5.13	7.01	18.55
High Expenditure	1.0	48.37	6.09	2.88	3.38	4.09	5.13	6.96	18.56

Notes: Includes all the recharge transactions during the first four weeks for long-term customers only. Gender predicted from cellphone transaction data.

With CDR-based gender predictions in hand, we return to our primary specifications above to test for heterogeneous effects by gender. We start by estimating equation 1.7 separately for men and women. The baseline averages in Table 1.9 suggest that women initially spend slightly less than men, but otherwise show similar recharge patterns. Against this baseline, we see no statistical differences by gender in terms of impact of airtime loans on total communication expenditure, recharge size or frequency.

Table 1.9: Heterogeneous impacts of airtime loans by gender
Loan post-eligibility period

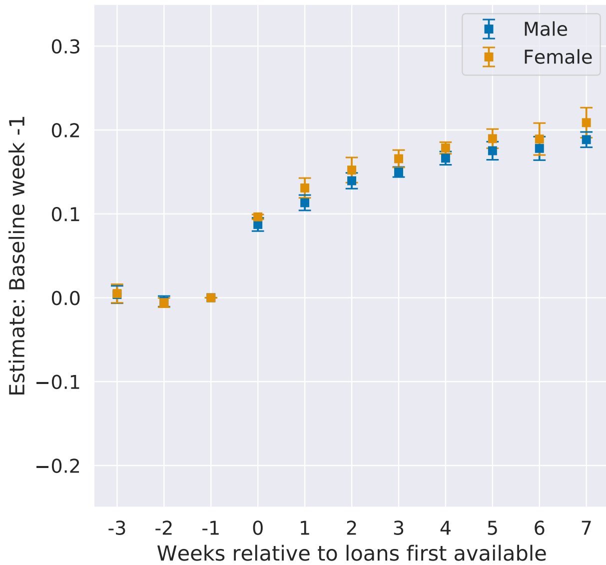
	Expenditure (USD)		Average recharge (USD)		Number of recharges	
	Male	Female	Male	Female	Male	Female
Baseline	0.92	0.88	0.31	0.3	2.62	2.55
Effect	0.15***	0.16***	0.01***	0.01***	0.18***	0.2***
Δ in percentage	16.7	17.8	4.12	4.21	7.06	7.94

Note: The effect variable shows the results of a difference-in-difference where the pre-eligibility period includes the three weeks before eligibility and the post period the 7 weeks that follow.

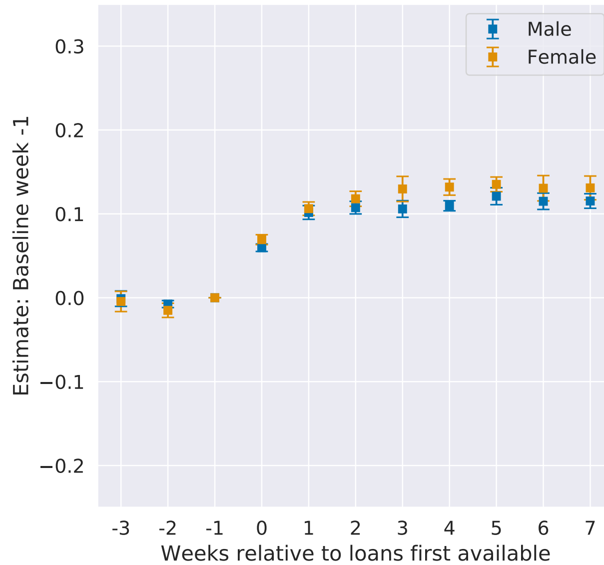
Since the similarity in average post-eligibility in airtime loan effects may mask important differences in the time path of these effects, we estimate the lead-lag specification in equation 1.8 separately for men and women. Here again, we see no strong evidence of heterogeneous effects by gender (see Figure 1.15) and Table 1.10).

Figure 1.15: Impacts credit access by gender and income level

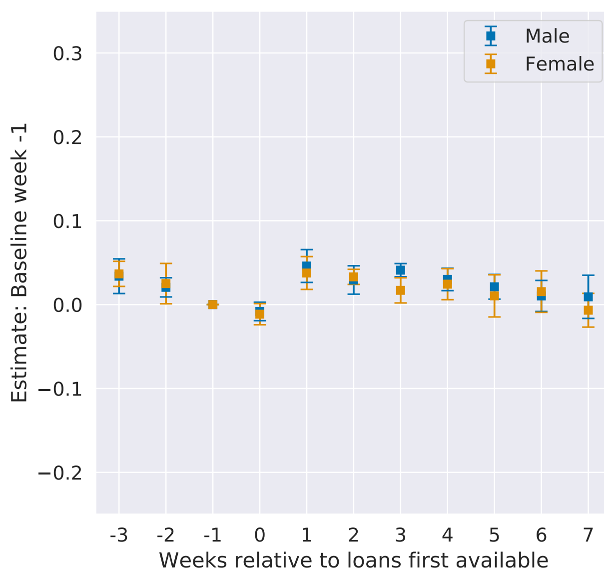
(a) Low expenditure



(b) Medium expenditure



(c) High expenditure



Note: Gender predicted using cellphone usage patterns. Includes only customers that are eligible for the loans. Long- term

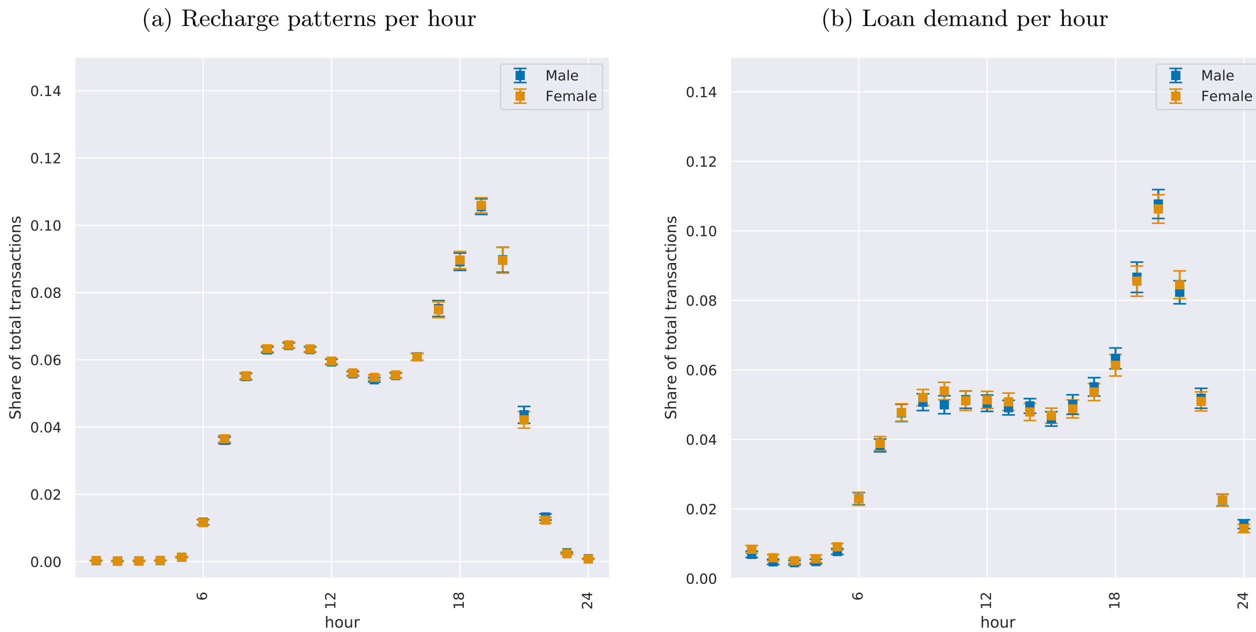
Table 1.10: Heterogeneous impacts by gender

	Low expenditure			Medium expenditure			High expenditure		
	Baseline	Effect	Δ in percentage	Baseline	Effect	Δ in percentage	Baseline	Effect	Δ in percentage
Expenditure (USD)									
Male	0.22	0.33***	148.99	0.63	0.24***	37.28	1.62	0.01	0.58
Female	0.22	0.31***	142.7	0.63	0.24***	38.25	1.62	-0.01	-0.68
Average recharge (USD)									
Male	0.14	0.07***	49.31	0.25	0.03***	10.15	0.47	-0.03***	-5.58
Female	0.14	0.06***	47.05	0.25	0.03***	10.07	0.47	-0.04***	-7.8
Number of recharges									
Male	0.96	0.75***	78.23	2.51	0.19***	7.45	4.04	-0.28***	-6.92
Female	0.96	0.73***	75.68	2.48	0.2***	8.1	4.05	-0.26***	-6.44

Note: Baseline levels show the average weekly expenditure during the three weeks prior to access to credit and compares it with the average outcome in the eight weeks that follow. Gender predicted using cellphone usage patterns of long-term customers.

Finally, we test whether airtime loans affect intra-day recharge patterns and the proportion of time in a week customers carry a positive prepaid balance differently for men and women. We find no such differences. Airtime recharges spike between 5pm and 8pm in a near identical pattern for men and women (see Figure 1.16). Time with positive prepaid balance is also nearly indistinguishable for men and women within initial expenditure terciles (see Figure 1.17).

Figure 1.16: Demand for physical recharges and loans by gender

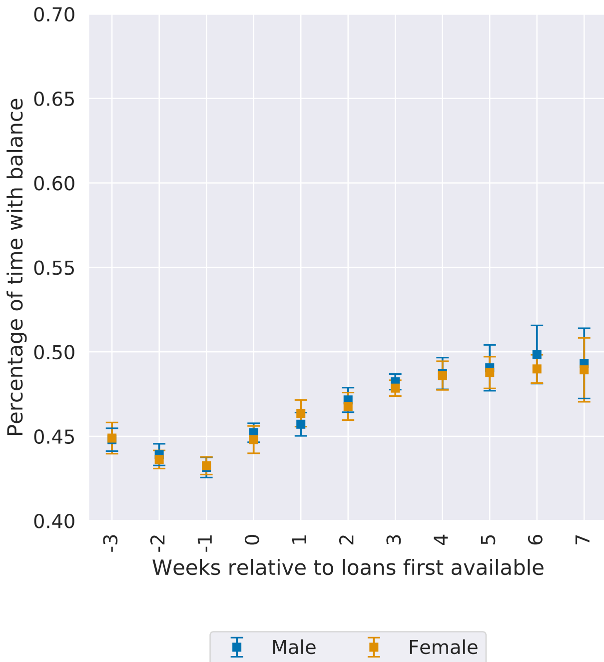


Note: Includes long-term customers. The estimation of daily demand patterns includes controls for day of the week and calendar week. Gender predicted using cellphone usage patterns.

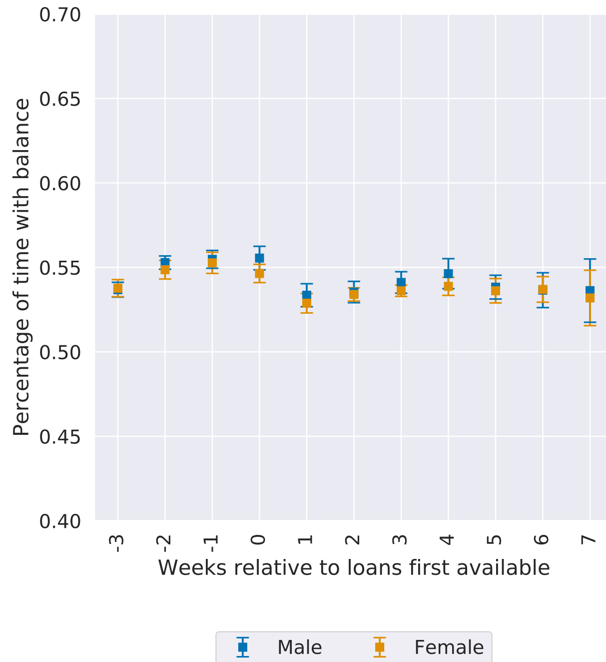
In sum, our exploration of potential heterogeneous effects of airtime loans in the context of Haiti generates clear insights: While heterogeneous effects by income (as proxied by initial expenditure) are pronounced and persistent in ways that suggest differential motivations for using this form of digital credit, we find no similar evidence for heterogeneous effects by gender either overall or within expenditure tercile.

Figure 1.17: Time with positive balance by gender and income level

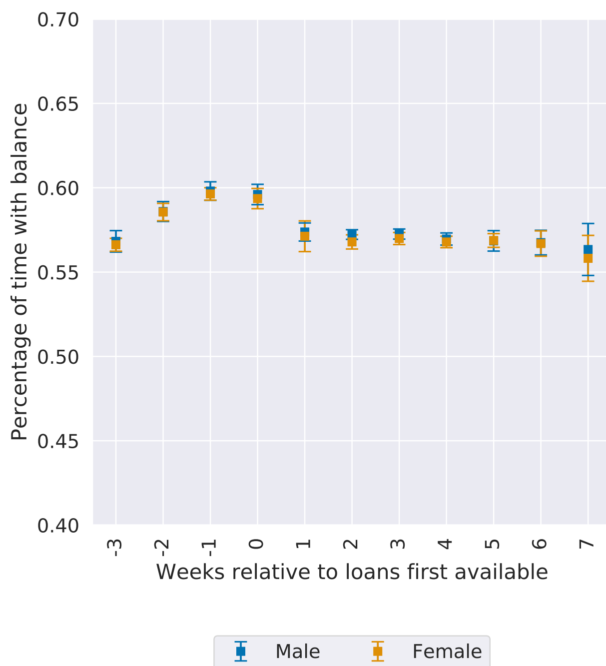
(a) Low expenditure



(b) Medium expenditure



(c) High expenditure



Note: Includes long-term customers. The estimation of daily demand patterns includes controls for day of the week and calendar week. Gender predicted using cellphone usage patterns.

1.8 Conclusions

There are several challenges to provide credit access to segments of the population that are not served by traditional formal financial institutions. Digital credit has the potential to reach these part of the market by solving the information asymmetries and bypass large transaction cost. We do not expect that a single product can serve all the diverse financial needs that exists in this segment of the market. Nevertheless, it is difficult to think that products that depend on physical locations can easily become attractive in areas where banking and other infrastructure is limited, and very expensive to develop.

We show that short-term financial constraints impact the consumption decisions of low-income individuals. The widespread availability of airtime loans and the high demand that exists for the product show the potential for the introduction of new products. The level of risk and know-how of each new product determines how fast it can be introduced in the market. Airtime loans are leading the way to make MNO more comfortable with providing fully digital loans. Experience with airtime loans can make MNO to build the knowledge base necessary to launch new digital credit products to the market. Additionally, from the consumer perspective, airtime loans offer the possibility to gain experience with their first formal financial product.

Our findings show that low income populations are not able to adjust their communication consumption using the saving and informal credit methods available to them. A situation that makes their consumption decisions extremely sensitive to the timing of their income, and can potentially have negative welfare effects. This sensitivity can have serious ramifications in their capacity to manage negative shocks that extend beyond the realms of cellphone expenditure.

Several questions remain. First, we cannot provide a definite answer on the welfare effects of increasing cellphone expenditure. Fully answering this question would require to know the role cellphone communication plays in the income generating process, as well as its interactions with other credit sources. Evidence suggests that cellphone communication plays a role in the income generating process by providing information on market prices, and insurance through a long-distance risk sharing network that is key to manage covariate shocks (Jensen, 2007; Blumenstock, Eagle, and Fafchamps, 2016).

Second, more research is necessary to understand to what extend convenience and liquidity constrains contribute to the demand of digital credits, together with its implications on the welfare effects of increasing credit access. Easy access to credit can have positive affects for groups that are highly liquidity constrained, a results that seems confirmed by Bharadwaj, Jack, and Suri (2019), where the authors find that digital loans reduce vulnerability to shocks. However, the internal validity of their results makes that there is no evidence on how less vulnerable groups react to an increase in their credit limit. Our results suggests that loan demand in less vulnerable groups is driven by factors other than liquidity.

Third, airtime loans is a first step towards using MFS for financial inclusion. The small size of the loans is an advantage that allows for their introduction is most markets. However, as the market moves towards products with higher risk levels, more research must be placed to find the the best way to leverage the experience of airtime loans, and on the optimal system of repayment incentives. Research shows that there is an optimal loan-size that encourages repayment, and that larger than optimal loans make customers more likely to default (Carlson, 2018).

In summary, after adding 700 million new users in the past decade the cellphone market still has enough room to add new customers, with a vast potential to develop new products. Most of these new customers have low incomes, and completely skipped the ownership of a bank account and even a landline GSMA (2019). The experience of microfinance and payday lending offers a cautionary of the risks of providing credit to people that have binding liquidity constraints, high marginal returns to capital, and difficulty copying with unexpected shocks. Properly managing these risks depends on a constant investment on better credit scoring algorithms, and an environment that fosters competition between providers reducing fees. Above all, research on digital credit must continue in order to improve the methods to screen customers, better tailor consumer protection policies, and create better channels to guarantee that customers are fully informed of the costs of credit and their conditions.

Chapter 2

The potential and limitations of big data in development economics: The use of cell phone data for the targeting and impact evaluation of a cash transfer program in Haiti

This chapter is co-authored with Joshua Blumenstock (Associate Professor, Professor, University of California, Berkeley), Travis J. Lybbert (Professor, University of California, Davis), and Daniel Putman (Postdoctoral fellow, Innovations for Poverty Action)

A series of recent papers demonstrate that mobile phone metadata, in conjunction with machine learning algorithms, can be used to estimate the wealth of individual subscribers, and to target resources to poor segments of society. This paper uses survey data from an emergency cash transfer program in Haiti, in combination with mobile phone data from potential beneficiaries, to explore whether similar methods can be used for impact evaluation. A conventional regression discontinuity-based impact evaluation using survey data shows positive impacts of cash transfers on household food security and dietary diversity. However, machine learning predictions of food security derived from mobile phone data do not show statistically significant effects; nor do the predictions accurately differentiate beneficiaries from non-beneficiaries at baseline. Our analysis suggests that the poor performance is likely due to the homogeneity of the study population; when the same algorithms are applied to a more diverse Haitian population, performance improves markedly. We conclude with a discussion of the implications and limitations for predicting welfare outcomes using big data in poor countries.

2.1 Introduction

Understanding whether a given product, program, or intervention improves livelihoods is as important as it is challenging. While established impact evaluation techniques can provide credible evidence of impact, they can also be expensive because they typically require active survey-based data collection. As individuals – rich and poor – generate data through their use of information and communication technologies, new opportunities emerge to leverage these passive records to understand behavior, preferences, and well-being. Blumenstock, Cadamuro, and On (2015) first demonstrated such an opportunity. Using metadata associated with mobile phone usage, the study predicted wealth levels with surprising accuracy and triggered a cottage industry of similar machine learning-based approaches to extracting meaningful signals from these Call Detail Records (CDRs). This empirical success raised a host of intriguing possibilities. Two such possibilities, here phrased as questions, motivate our analysis in this paper: Can CDRs identify potential beneficiaries of means-tested programs and thereby enhance the cost-effectiveness of targeting? Can CDRs also capture *changes* in wealth or well-being and thereby enable near real-time evaluation of the impact of new programs, products or policies?

In 2016, the World Food Program (WFP) responded to a third consecutive year of drought in Haiti with an emergency unconditional cash transfer program to protect household food security. These transfers, which targeted the poorest households in drought-hit areas, consisted of three consecutive monthly disbursements via mobile money. The value of these transfers was significant, each representing 130% of monthly average per capita income.[1] We collaborated with WFP to evaluate the impact of the cash transfers on household food security, dietary diversity and consumption. Additionally, we partnered with the mobile network operator that facilitated these mobile money-

[1]Based on the 2012 ECVMAS households survey.

based transfers to obtain access to cellphone transactions data of the people that participated in the targeting and evaluation phase of the program. In this paper, we provide a conventional survey-based impact evaluation of the program and then use these results as a benchmark for an alternative CDR-based impact evaluation of the same program.

As in the benchmark evaluation, we leverage the targeting threshold used by the WFP to implement a regression discontinuity (RD) design that estimates the impacts of the unconditional cash transfer program using survey data. We find the program increased household food consumption and food expenditure in 0.35 and 0.31 standard deviations, respectively. For the CDR-based alternative, we use feature engineering and a structured combinatorial method to generate several hundred CDR metrics from the volume, intensity, timing, direction, and location of communication, as well as the household's position in the cellphone network.[2] With this rich set of features in hand, we deploy machine learning algorithms to predict the targeting status of beneficiaries, as well as household food security outcomes, which we then use to replicate the RD impact evaluation.

At the outset, this cash transfer program seemed to provide an ideal setting in which to test the viability of a CDR-based targeting and evaluation. On the targeting side, WFP uses a standard and simple scorecard to quickly rank would-be beneficiaries in terms of wealth, and CDRs have shown promise in predicting precisely such outcomes. On the evaluation side, the monthly transfers were large relative to household income and were sent for three consecutive months, so it is reasonable to expect the impacts to readily detectable. The use of a scorecard threshold for targeting beneficiaries enables RD identification of these impacts for both survey- and CDR-based measures of well-being. Finally, the transfers were distributed via mobile money in collaboration with Haiti's largest mobile network operator, which enjoys a dominant 80% market share in the country. This facilitates our

[2]For our main results we use Bandicoot, an open-source Python toolbox. See section 2.4.1 for details.

ability to link scorecard responses and cash transfers to specific mobile network users.

Despite these advantages, we were not able to replicate the survey-based RD results using these metatdata or the targeting. Our failure to replicate these impact evaluation results runs deep as all of our the predictions of the household-level food security outcomes using our CDR features and machine learning algorithms present low levels of accuracy. That is, in the horse race we set out to run between conventional survey-based and novel CDR-based impact evaluation, the CDR horse stumbled out of the gate, which ended the race before it really began. There is, however, much to learned from this failed prediction attempt about the limits of CDR-based analyses such as the targeting and impact evaluation applications that motivate this work.

In a postmortem assessment, we discuss why CDR predictions failed in this seemingly-ideal setting. First, given high levels of poverty and vulnerability, cellphone ownership and, in some regions, access to the cellphone network is low. Only 34% of the participants in the scorecard survey reported owning a cellphone, with cellphone ownership concentrated among less vulnerable individuals. Solely relying on cellphone data limits the population we can observe (target) and can introduce bias as higher vulnerability levels correlate with lower levels of cellphone ownership. Furthermore, the practice of sharing devices, more common in low income households, creates additional challenges as it dilutes any potential signal about wealth levels in the data. Second, the targeted geographic areas in this program were pre-selected based on drought-induced and general vulnerability. Within these pre-selected communes, the WFP administered the scorecard to identify specific beneficiary households. While this two-tiered approach can improve targeting, it restricts the statistical variation we can observe in household wealth and outcomes and limits our ability to predict outcomes using CDRs. Third, the exogenous eligibility cut-off is ideal for RD impact evaluation, but its primary identifying assumption is that those on either side of the cut-off are statistically indistinguishable, limiting the variation in the CDR usage patterns on both sides of the eligibility cut-off. Finally, the primary outcomes of interest in this study are flow variables (e.g., food consumption) rather than stock variables, and the promising evidence of CDR predictions has so far concentrated on the prediction of stock variables such as assets and wealth.

We empirically explore these postmortem considerations from a variety of perspectives. We compare our data to nationally-representative data from Haiti to understand the implications of tight (effective) targeting. The statistical variation in the primary outcomes for our WFP sample is indeed significantly lower than for the general Haitian population. The scorecard sample, unsurprisingly, is more vulnerable with food insecurity levels double the national average and the total food expenditure distribution below the 70th percentile of the overall distribution. These factors limit the variation that machine learning models can use to recreate both the targeting and the outcome variables. With additional tests, we find that CDR predictions improve as we increase the variability in outcomes. To explore our ability with greater variation in the data to predict flow variables such as food consumption, we obtained informed consent from a subsample of the participants in the nationally representative survey and predicted their wealth levels and food expenditures. We show that while we can predict wealth levels relatively well in this broader sample, the predictions are much less reliable when it comes to flow outcomes like consumption.

In the next section, we provide a broader introduction to the use of passive data in development economics and then describe additional details about the WFP cash transfer program. In the methods section, we present the RD design and the various data sources we use in our analysis. Section 5 provides the survey-based impact estimates that serve as a benchmark for the alternative CDR-based impact estimation. In section 6, we present our CDR-based predictions and discuss the associated implications for CDR-based targeting and impact evaluation. We conclude with a detailed postmortem discussion and broader reflections.

2.2 Passive Data in Development Economics

Reliable and up-to-date data is a key factor in the effective implementation of public policies. The absence of official data is more acute in poor and developing countries, forcing governments to implement public programs with limited information (Blumenstock, 2016a). Collecting these data is expensive in both monetary and administrative terms, and in many situations, it cannot be produced with the necessary speed to attend to extraordinary demands such as relief programs following a natural disasters.

The last few years have seen a large increase in the amount of data produced daily by private digital transactions. This digital footprint contains information on billions of individuals, including those living in poverty. One of these novel sources of information is the transaction data that mobile phone subscribers create every time they make or receive a call. Users do not create this information to contribute to any policy or research objective; instead, the data are passively created as part of the regular network operation. While these datasets may be a step removed from the on-the-ground outcomes policy-makers care about, advances in feature engineering and machine learning allow us to use this information to circumvent data limitations.[3]

With more than 95% of the global population with mobile-phone coverage, CDR data create a unique opportunity to address a major challenge policy-makers and researchers face in contexts where reliable quantitative data are scarce (Blumenstock, Cadamuro, and On, 2015; Blumenstock, 2016a). The first applications of these methods studied how regional socioeconomic conditions

[3] There are plenty of remote sensing applications that rely on data other than cellphone records. For example, Jean et al. (2016) uses satellite imagery, nightlights to infer poverty in Nigeria, Tanzania, Uganda, Malawi, and Rwanda, and Goldblatt et al. (2018) use satellite imagery and remote sensing data to characterize land cover and urbanization in India, Mexico, and the United States.

create signals that can be detected on aggregated the CDR data trail, in particular regional poverty levels (Hernandez et al., 2017) and food security indicators (Decuyper et al., 2014). In areas where official statistics are not recent, CDR data and machine learning methods provide an tool for updating indicators in-between household survey rounds.

Building on the previous results, several papers demonstrate that mobile phone data can be use to estimate outcomes at the individual level. The logic behind this approach is that phone usage captures many behaviors that have some intuitive link with socioeconomic indicators, allowing researchers to differentiate the most vulnerable households. As described by Björkegren and Grissen (2019), a phone account is a financial account that captures part of a person's expenditure, with most of the calling behavior being an indicator of how a person manages his expenses. For example, individuals with different income streams are likely to have call patterns indirectly linked with socioeconomic status. These include when a person uses his phone, geographic mobility, and the diversity of the calling network. Blumenstock, Cadamuro, and On (2015) use CDRs and data from a nationally representative survey in Rwanda to demonstrate how an individuals' socioeconomic status can be inferred from CDR transactions, while Blumenstock (2018b) demonstrates similar techniques can also be used to accurately predict wealth levels in an Afghan sample. In both applications, the authors find their behavioral features can explain about 46% of the total variation of a wealth composite index.[4]

The success in predicting individual-level outcomes potentially opens the door to new applications to improve public policy.[5] The first application is in complementing/replacing conventional methods used to target social programs. The importance of targeting to make anti-poverty pro-

[4]Applications of these methods already exist for commercial applications, for example, providing credit scores in settings where credit bureau are not present (Björkegren and Grissen, 2019).

[5]Blumenstock, Cadamuro, and On (2015) provides a complete list of potential applications of CDR data for social research.

grams more cost-effective has been widely studied (Alatas et al., 2012; Coady, Grosh, and Hoddinott, 2004; Brown, Ravallion, and Van de Walle, 2016). Common targeting protocols rely on administrative and survey-based measures of assets or consumption. These information is not available in many developing, and if it is available, usually has reliability problems. Moreover, for most practical applications to small and medium scale programs collecting this information adds a large administrative cost. The passive nature of CDR data may provide an additional tool to improve targeting efforts with shorter deployment times. Aiken et al. (2020) studies the extent that mobile phone data can be used to identify ultra-poor households in the context of an anti-poverty program in Afghanistan. In their study, a community wealth ranking and a deprivation survey provide the ground-truth data that classifies a household as ultra-poor.[6] An advantage of this study is that a household survey was collected independently of the ultra-poor classification survey, providing a much richer set of indicators to compared the CDR-based method accuracy. Using six months of cellphone data, the authors find CDR-based methods have an accuracy of 70%, very similar to standard survey-based consumption (73%) and asset-based measures (70%). Combining the information from the CDR data and the household survey into a single classification problem provides the best results, with an accuracy of 79%. However, as discussed in the paper, using several data sources might not a possibility in most real-life applications.

The second potential application is the use of mobile phone transaction records to enable new approaches to impact evaluation and program monitoring. An initial application of CDR data to complement impact evaluation studies used the data as part of the identification strategy. One example is (Olivieri et al., 2020) who uses cellphone records to identify the geographical distribution

[6]The ranking used the following questions 1. Household is financially dependent on women's domestic work or begging. 2. Household owns less than 800 square meters of land or is living in a cave. 3. Targeted woman is younger than 50 years of age. 4. There are no active adult men income earners. 5. Children of school age are working for pay. 6. The household does not own any productive assets.

of Venezuelan migrants in Ecuador to understand their impact on labor market outcomes. Other papers have used the mobility patterns captured by the usage of mobile phones to understand how local-level policies affected refugees' mobility in Turkey (Beine et al., 2019), population movement in the wake of natural disasters (Wilson et al., 2016), and the spread of infection disease (Wesolowski et al., 2012; Milusheva, 2020). We go one step further by trying to use cellphone data to estimate changes in welfare over time in the context of an impact evaluation study.

2.3 The EMOP program

In 2016, in response to a third consecutive year of drought in Haiti, the WFP conducted an Emergency Food Assistance Program (EMOP), with the objective to improve food security of households in the affected areas. During the program's lifespan, it provided 46,163 households with three monthly transfers of 3050 HTG (around 50 USD) delivered via mobile money; a large sum with each payment roughly equivalent to the minimum wage or 130% of monthly average per capita income.[7]. The rapid-response nature of the program guided key decisions on the targeting and implementation protocols, including the use of a simple scorecard mechanism to measure vulnerability and relying on the existing mobile money infrastructure to deliver funds.

The EMOP used a three-step targeting process to select beneficiaries. During the first stage the program's geographical targeting was determined. The WFP, together with the Haitian government and other organizations, conducted a nationwide emergency food security assessment to determine the areas with the largest concentration of food-insecure households.[8] To avoid duplication of aid

[7] Based on the 2012 ECVMAS households survey.

[8] The WFP carried out the data collection together with the Haitian National Coordination for Food Security and in collaboration with its partners (FAO, FEWS NET, OCHA and others) between November and December 2015. The assessment measured the impact of the drought on the number of moderately and severely food insecure households. The analysis also entailed assessing the functioning of markets and their response capacity in case of providing aid in the form of cash. See WFP targeting

initiatives, the WFP focused on areas where no other actors were intervening at the time. This led to the prioritization of 21 rural communes across the country.[9]

In the second stage, the WFP engaged with local stakeholders to construct a list of the families living in the area, build trust in the program's implementation, and tailor the scorecard questionnaire to local conditions. The result of this process was the sampling frame used to conduct the scorecard survey, and slight modifications to the questionnaires, tailoring them to the specific farm animals present in each region.[10]

The selection of beneficiaries took place during the third stage of the program's targeting process. The WFP administered a short survey, that official documents call scorecard survey, and used the information to calculate the value of a vulnerability index for every household interviewed. The objective of the vulnerability index is to measure how susceptible a household is to suffer from food insecurity and hunger. A higher number corresponds to higher levels of vulnerability, and therefore, greater need of the program's assistance.[11] The overall score of a household is an integer number based on responses to questions (most of which were simple "yes" or "no" questions) about land, animal ownership, and the presence of vulnerable people in the household, with possible scores ranging from zero to eight.[12] The scorecard exercise is a standard practice for the WFP operations worldwide, especially when rapid delivery of funds is necessary to provide aid.

The scorecard vulnerability index follows a proxy means test approach towards identifying the most vulnerable households in the absence of consumption based poverty indicators (Grosh and

[9] These communes are located in prioritize the Nord-Est, Artibonite, Nord, Centre, Ouest, Nippes, Grande-Anse, Sud and Sud-Est Departments. Full details on the program's implementation are available at: WFP information

[10] Depending on the region, questions about animal ownership included different combinations of cows, horses, sheep, goats, and pigs.

[11] The scorecard exercise is part of the Household Economic Analysis (HEA) used by the WFP. The HEA arose from a collaboration in the early 1990s as a tool to improve the FAO ability to predict short-term changes in a population's access to food, see (Holzmann et al., 2008).

[12] This includes children, pregnant and lactating women.

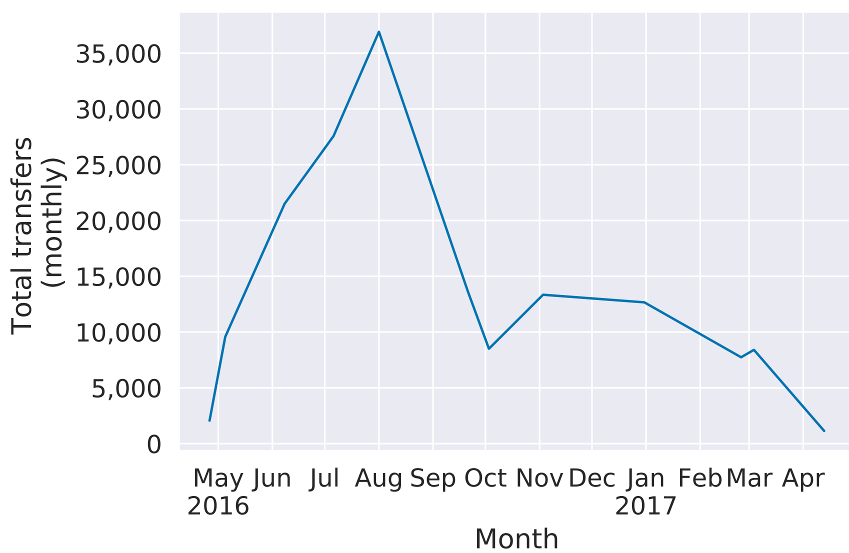
Baker, 1995). As explained by Gazeaud (2020), a complete proxy means test requires a two step process. First, an in-depth survey is administered to a sample of households to collect data on consumption as well as some easily observable and verifiable correlates of consumption. These data are used to estimate a regression of log consumption per capita on correlates of consumption. Second, a short survey is administered to all potential beneficiary households to collect information on the same correlates of consumption, and compute a new score using short survey information. Unfortunately, the EMOP implementation lacks the necessary data to implement the first stage of the identification and weighting of the correlates of the indicators, relying instead on previous WFP experience in to choose the variables used in the construction of the index.

The implementation of the program used the vulnerability score to target aid. At the commune level, the cut-off to be eligible for the program depended on the total funds assigned to the region.[13] Since the eligibility cut-offs were selected after scores were assigned taking into account the total funds assigned to each area, households had not opportunities to try to manipulate their vulnerability score and precise treatment status. This constitutes the basis for the evaluation of the program using a Regression Discontinuity design, see section 2.5. Under the selection criteria, there are several instances where a person with a specific score did not receive aid given the commune where he lived but would have been eligible with the same score in another area. Most communes required a score between three to five to grant eligibility (81% of the covered population), with some communes requiring a vulnerability score as high as 6 (17% of the covered population). We provide more details on how this affects our replication of the eligibility criteria using CDR data and machine learning methods in section 2.6.

[13]The total funds assigned to each communed where the results of the total population in an area, with communes with higher population levels receiving more funds independently of the distribution of the vulnerability scores in the commune.

The use of mobile money to distribute funds makes every transaction linked to a cellphone number. If beneficiary household did not have a SIM card to receive transfer payments, WFP and Digicel provided one.[14] Records show a total of 63,201 phone numbers receiving a transfer during the period. As Figure 2.1 shows, transfers occurred at different times over the period, reaching their peak in August 2016; however, within geographical locations (e.g., commune), transfer patterns tend to be similar between households.

Figure 2.1: WFP Transfers



Note: Author's calculations using mobile money transaction logs.

2.4 Methods

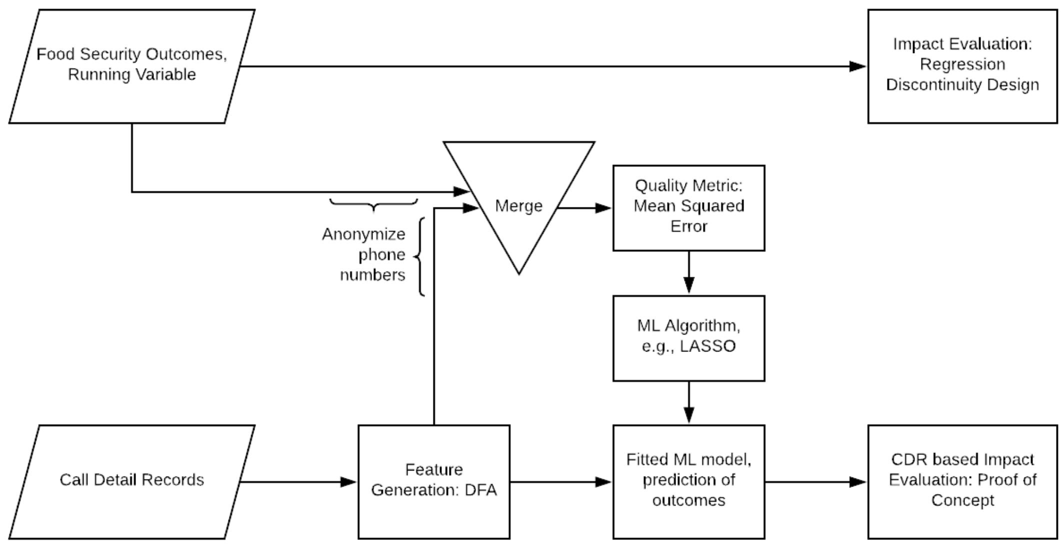
In this section, we discuss the validity of the RD design and explain how we combine the CDR data with the formal evaluation of the program. Figure 2.2 provides a diagrammatic representation of our approach towards using CDRs combined with the formal evaluation of the EMOP program. The

[14]We have access to all the mobile money transaction records, where the EMOP transfers are just one of the many transactions.

first part consists of the traditional approach to impact evaluation. The program's implementation relied on identifying households with the highest levels of vulnerability using the scorecard exercise described above. The use of a vulnerability score to determine eligibility allows us to evaluate the program's impact using an RD design. We measure the outcomes using an additional in-person survey administered seven months after the start of the program. We use both surveys as the ground-true data to feed the prediction models. The second part of Figure 2.2 shows the steps we follow to combine the CDR data with the information from the surveys. We start by processing the CDR data to create hundreds of behavioral features for each cellphone number and match them to individuals participating in the surveys. We face several challenges as a large percentage of participants lacked a cellphone and, in several cases, several household heads reported the same number as their own. After connecting the survey information with the behavioral features, we test different algorithms to try to replicate the program assignment determined by the scorecard vulnerability score and the program's main outcomes as captured during the in-person survey. We use cross-validation to limit the possibility of overfitting these models and choose the one with the best out-of-sample performance.

The rest of this section is organized as follows. Section 2.4.1 explains the feature engineering process we implement to extract useful information from the raw cellphone data. Section 2.4.2 describes the different surveys available for the impact evaluation of the program and as ground-truth data for the prediction. Additionally, this section explains the process to integrate individual survey responses with the corresponding the CDR data. Finally, section 2.4.3 discusses the identification strategy using an RD design.

Figure 2.2: Our Approach to Running a Call Detail Record Based Evaluation



2.4.1 Feature engineering

Our objective with the cellphone transaction data is two-fold. First, we want to predict the eligibility status from the scorecard survey. Second, we want to predict the program's main outcome variables and used the prediction to replicate the results from the impact evaluation of the program. In its raw format, CDR data provides a detailed account of each cellphone number activity. We want to extract information about each user's behavioral patterns that correlate with his socioeconomic characteristics. For this, we use feature engineering on the cellphone transaction data to compute behavioral indicators that capture aggregate aspects of each individual's mobile phone usage. In essence, for each phone number, we construct a vector of values that represent usage patterns and link them back to the survey responses of individual i. We generate the features using Bandicoot, an open-source toolbox for CDR analysis. The program creates metrics similar to other feature engineering methods used in the literature, such as Blumenstock, Cadamuro, and On (2015). Its main advantage is that it provides a ready-to-use and computationally optimized method to extract features from cellphone metadata.[15]

Indicators include information about an individual's overall behavior (average call duration and percent initiated conversations, percent of nocturnal interactions, inter-event time between two phones calls), spatial patterns based on cell tower locations (the number of unique antennas visited and the radius of gyration), social network (the entropy of their contacts and the balance of interactions per contact), and recharge patterns (including the average amount recharged and the time between recharges) (De Montjoye, Rocher, and Pentland, 2016). Each feature is calculated as a week-level mean, standard deviations, median, min, max, kurtosis, and skewness, as well as additional statistics disaggregated over weekdays, weekends, and working and night hours. Figure 1B explains the process to calculate individual features.

Since the features represent week-level statistics, we must decide how many days of cellphone transaction data to include in the feature generating process. Two forces are at play. First, a wider time window provides a more diverse set of transactions from which to extract information. On the other hand, we must consider how much the outcome we want to predict varies during this time window. In principle, outcomes with little variation over time benefit from using long series of cellphone data as the different features can capture more information. Applications that predict individual-level wealth levels, a variable that during normal times presents little variation over time, extract features from one year to six months of CDR data (Blumenstock, Cadamuro, and On, 2015; Aiken et al., 2020). There is evidence on how changes in the size of the ground-true data affect the prediction capacity of CDR-based models and how a models' accuracy decays over time (Blumenstock, Cadamuro, and On, 2015; Lazer et al., 2014). However, to the best of our knowledge, there is no evidence on how the length of the time series of CDR data affects predictive

[15]For a full description of the method, see De Montjoye, Rocher, and Pentland (2016).

capacity, and how that relates to the variability over time of the predicted outcomes. Considering that we are interested in food security outcomes from individuals who live in regions affected by a natural disaster, we expect that the observed levels at the time of the survey were affected by both the shock and coping strategies. We have no prior about the optimal number of months of pre-survey CDR data to create the features. Instead, we opt to compute features using three different time windows preceding the date a person was surveyed: Fifteen days, one month, and six months. To calculate the time window, we omit the week preceding the survey to avoid changes in calling patterns in expectation to be interviewed.

For each individual, we extract a total of 2,148 behavioral features in each of the three time windows. We implement two initial filters on the features that make each time window includes a slightly different set of variables. First, we drop any feature with more than 50% of missing values. Second, we eliminate those with a variance below 0.02. Finally, to avoid models that contain highly correlated features that do not provide additional information, we calculate a correlation matrix for all the features and eliminate all but one for those with a correlation above 0.98.[16] Table 2.1 shows the number of features available for each time window and survey combination.

2.4.2 Data: Description and integrated sample

Since our main application uses CDR to replicate the program's targeting and impact evaluation, we are constrained by our ability to match the survey responses with the cellphone data. To explain this process, we first describe the available CDR data. Then for each survey, we detail the sample we can match and provide descriptive information to understand how the remaining sample compares with the overall population. Table 2.1 summarizes the samples of the scorecard

[16] This approach is similar to Aiken et al. (2020)

and in-person surveys, and shows the specific samples we are able to match to the cellphone data.

Table 2.1: Data sources directly collected from participants

	Survey	
	Scorecard	**In-person**
Sampling		
Unit of analysis	Household	Household
Data collection period	Apr. - Sept 2016	Dec. 2016
Location	21 communes	2 communes
Observations	58,881	1,137
Beneficiaries	46,163 (78.4%)	697 (61.3%)
Cellphone related data (% of total sample)		
Owns a phone	20,190 (34.3%)	1,076 (94.6%)
Phones matches	13,780 (68.3%)	872 (81.0%)
Numbers that match (% of matched numbers)		
Beneficiaries	10,491 (76.1%)	506 (58.0%)
Non-beneficiaries	3,289 (23.9%)	366 (42%)
Sample with features		
Fifteen-days	12,739	797
Thirty-days	13,142	825
Six-months	13,682	855
Features available		
Fifteen-days	1,355	1,265
Thirty-days	1,457	1,394
Six-months	1,315	1,303

Note: Features available after filtering those with more that of 50% missing values, a variance of less that 0.02, and correlations above 0.98

Call Detail Records

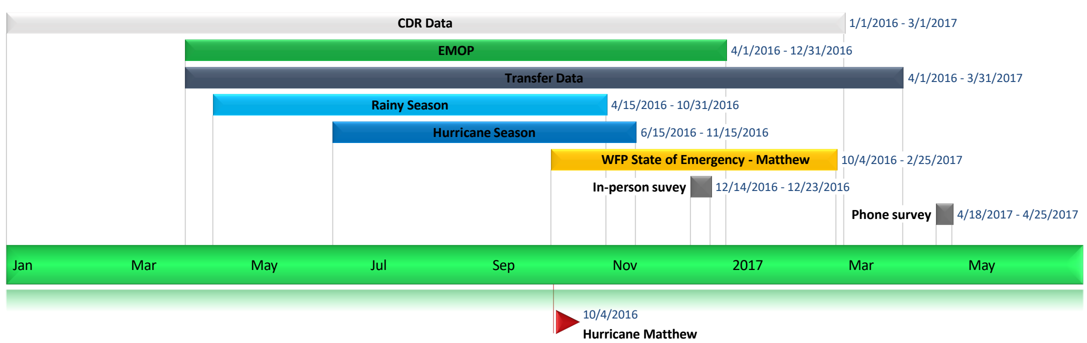
Digicel is the largest cellphone company in Haiti. Besides providing cellphone services, the company operates the largest mobile money platform in the country. In 2016, it partnered with the WFP to deliver funds to beneficiaries using its mobile money platform. As part of the partnership with the WFP, Digicel provided cellphone records containing all the transactions made, in the network, between August 2015 and May 2017. All the information is anonymized, and we cannot see the content of calls or text messages. The CDR data covers all the evaluation cycle, including several months before each survey took place. Figure 2.3 shows how all the data sources overlap. For each type of transaction, we can observe:

- Calls: Anonymized numbers for the caller and the receiver, time of the call, duration of the call, and cellphone towers connecting the call.

- Text message: Anonymized numbers for the sender and the recipient, time of the text message, and the closest tower to the party that sent the text message.

During the period, a total of 29,907,850 transactions took place, broken down into 12,523,717 calls, 17,384,133 text messages. From this transaction, we extract behavioral features for three different time windows.

Figure 2.3: Program's timeline: Data availability and implementation



Scorecard

This survey provides the basis for eligibility for the program. The WFP conducted the scorecard exercises between April and mid-September 2016.[17] It constitutes the final step of the WFP three-step targeting strategy. The survey took place in regions the WFP had previously identified as

[17] Data collection was planned from April until August. The schedule was extended to account for initial delays and the inclusion of additional areas.

the hardest hit by the 2016 drought, with only households living in the pre-screened candidate communes participating in the survey. Eligibility depends on a commune-specific score. Due to the rapid response nature of the survey, the scorecard survey is not as thorough as an LSMS survey and lacks a full module on assets and consumption. [18]

A total of 58,881 households participated in the scorecard survey. Using the vulnerability score and the program's financial constraints, the WFP determined 46,163 (78.4%) eligible for the program. Considering that cellphone ownership in Haiti was below 50%, it is not surprising that in a vulnerable population, such as ours, only 20,190 (34.3%) of the participants reported having a cellphone number.[19] Since our application relies on matching individual survey responses to cellphone records, we are limited to participants with a cellphone number that matches our records. We are able to recover the CDR data for a total of 13,780 (68.3%) of the respondents with a phone, a number that is close to the market share of Digicel. As Table 2.2 shows, the share of beneficiaries in the sample of cellphone owners is marginally lower than in the overall sample (76.1%), as a consequence of cellphone ownership being positively correlated with wealth and negatively correlated with the vulnerability score. Additionally, we find a statistically significant difference in the vulnerability score within the group of cellphone owners. In particular, numbers that match our records have the lowest average vulnerability levels; see last two columns of Table 2.2. We are not aware of the cause of this apparent self-selection into different carriers. However, this highlights the intrinsic limitations of CDRs as a targeting tool.

We extract behavioral features for the 13,780 individuals with matched cellphone records. As Table 2.1 shows, the number of subscribers with features increases with the length of the time

[18]Similar sampling methodology and questionnaires are standard for the WFP operations worldwide, especially when rapid delivery of funds is necessary to provide aid. The WFP website provides more details about how interventions are implemented and current programs: https://www.wfp.org/emergency-programming

[19]A minority of participants in the survey (174) reported a phone number that was also reported by another household. We omit their answers from the study.

window, and it is always below the total number of numbers in the network. In the presence of sparse cellphone usage, short time windows are less likely to capture any activity.[20] For the six-month window, the only reason for a number not presenting activity is that the respondent activated the number in the week before participating in the survey, and therefore is not included in the feature extraction process.[21]

Table 2.2: Vulnerability score by cellphone ownership status

	Cellphone Ownership			**Sig. diff.**	
	(i) No	(ii) Other operator	(iii) In network	(i) -(iii)	(ii)-(iii)
Panel (a): Scorecards survey					
Vulnerability Score	0.74	0.41	0.32	***	***
	(1.55)	(1.28)	(1.27)		
Beneficiary (%)	79.19	78.52	76.13		
Households	38,691	6,410	13,780		
Panel (b): In-person survey					
Vulnerability Score	0.54	0.17	-0.11	***	***
	(1.18)	(1.25)	(1.28)		
Beneficiary (%)	85.25	68.14	58.03		
Households	61	204	872		

Note: A higher vulnerability score makes a household more likely to participate in the program. We only have access to the transaction data for the number in the participating network.

In-Person Survey

To measure the impact of the program, we collected an in-person survey in December of 2016. The survey collects information for a total of 1,137 people in two regions in the south of Haiti: Aquin (801) and Fonds-Verrettes (336).[22] The sampling of this survey targeted people around the eligibility threshold as it is the main tool to evaluate the program's impact using a RD strategy. The original scorecard survey in those two communes included 16,637 people. In contrast to the

[20] Sparse activity (low demand for communication) correlates with low-income levels that limit cellphone usage.
[21] Ownership of a number is lost if no recharge takes place during three consecutive months.
[22] Aquin is on the southwestern peninsula, and Fonds-Verrettes, which is inland, southeast of Port Au Prince.

scorecard survey, this survey includes detailed questions on food consumption and expenditures, occupation, livestock holdings, and family composition. The share of beneficiaries in the sample is 61.3%, which is 17 percentage points lower than in the general population. The rate of cellphone ownership stands at 94.6%; a number that represents a large increment from the 44% captured by the scorecard survey in the same communes.[23] We are able to recover the CDR data for a total of 872 of the respondents (81.0% of the respondents with a phone). In this sample, the share of beneficiaries is 3 percentage points lower than in the general population (58.0%). We also observe that respondents with a cellphone have lower vulnerability levels, see Panel (b) of Table 2.2. As expected, the number of individuals with features increases with the time window, see Table 2.1.

Additional Data: Nationally representative survey

As mentioned before, the emergency nature of the intervention implies that the survey data is concentrated in areas affected by the 2016 drought. We use a nationally representative survey of 4,267 households interviewed between May and October 2018 (FinScope, 2018). These data set provides a sample with greater socio-economic variation, which allows to put the participating population in context, and test the extend that a narrow sampling limits our capacity to predict the main outcomes of the program. [24] The questionnaire is close to an LSMS survey but has several modules focused on access to financial services. Using the questions available, we harmonized food consumption and food security variables to reflect as close as possible the WFP surveys.

A key advantage of using this survey is that for a subset of participants we are able to link

[23]The rate of phone ownership is close to 95% for both beneficiaries and non-beneficiaries. This larger than average rate of cellphone ownership is caused by the sampling strategy and not the result of the WPF providing SIM cards to eligible households.

[24]The sample can be disaggregated at the urban/rural level and seven regions, including the Port-au-Prince metropolitan area. Given the limited data sources available in Haiti, this is the survey closest to the time of the program. The latest LSMS survey in Haiti took place in 2011.

their responses with their individual CDR data. For this, we conducted a follow-up call where participants provided 'informed consent' to access their cellphone records. Since participation in the Finscope survey was in person, we only contacted those who provided a phone number at the time of the survey. Between January and March 2021 we made four attempts to contact each number. Out of 2,870 participants with a phone number, we can use the CDR data for 1,132 cellphone lines.[25]

Table 2.3 shows differences in household composition, location, wealth, and food security by phone ownership status (see columns 2 and 3). As expected, people who own a phone tend to be younger and are more likely to live in urban areas. Phone ownership is correlated with higher levels of wealth and lower levels of food insecurity.[26] When we look for differences across the population with a phone an interesting dynamic appear. First, active lines at the time of our follow-up survey, but whose data we lack permission to use in this study, present highest levels of wealth (columns 5 and 6). These lines include people who were contacted and explicitly declined to participate and numbers that did not answer in any of the three contact attempts. Second, data suggests there is a correlation between joining a particular cellphone network and wealth, with cellphone owners who belong to a network other than Digicel (column 3) presenting lower wealth levels and being more likely to live in rural areas. There is an apparent self-selection by location and wealth levels into a person's decision to contribute his data and choosing a particular network. This raises several questions about who we are able to predict socioeconomic indicators using cellphone data, specially when most studies only gather data from one cellphone provider.

[25]The Finscope survey has information for a 4,267 people, but only 2,870 (67%) provided a phone number had a phone, of which only 1,960 were part of the Digicel network. The IRB approved that we could link survey answers with CDR data in the case where the original survey participant was contacted and provided informed consent (519 cellphone numbers), and when a line had been disconnected for more than six months, making impossible to re connect with the original survey respondent (613 cellphone numbers). A total of 167 people refused to participate, and 661 active numbers did not answer to any of our contact attempts.

[26]In this case, the PCA to compute the wealth index omitted cellphone ownership from the calculation.

Table 2.3: Descriptive stats Finscope sample

	Owns phone			Owns phone: Number in Network-					
	(1)	(2)		(3)	(4)	(5)	(6)		
	No	Yes	Sig. diff.	(i) No	Yes			Sig. diff.	
			Yes-No		(ii) Matched	(iii) No consent	(iv) No answer	(i)-(ii,iii,iv)	(ii)-(iii,iv)
Age HH head	49.62	44.64	***	43.34	44.4	47.07	46.19	***	***
	(16.24)	(14.61)		(14.07)	(14.48)	(12.93)	(15.73)		
HH size	3.34	3.88	***	3.82	3.93	3.66	3.96		
	(1.86)	(1.99)		(2.08)	(1.92)	(1.99)	(1.97)		
Urban	0.62	0.86	***	0.83	0.89	0.93	0.85	***	
	(0.49)	(0.34)		(0.38)	(0.31)	(0.25)	(0.35)		
Food insecure	0.54	0.49	***	0.5	0.49	0.41	0.49		
	(0.5)	(0.5)		(0.5)	(0.5)	(0.49)	(0.5)		
Wealth index	-0.42	0.2	***	0.05	0.19	0.64	0.32	***	***
	(0.7)	(0.89)		(0.84)	(0.73)	(0.92)	(1.11)		
Obs.	1,398	2,869		909	1,132	167	661		

Note: Author's calculations using Finscope 2018 and CDR data. Phone ownership is determined at the household head level. Columns 3 to 6 include only those with a cellphone. Column 3 includes cellphone owners with a line in a different network. Column 4 includes numbers that agreed to participate as well as those that were deactivated in the six months prior to our follow-up survey; we can only match the survey answers of these numbers with their CDR transaction data.

2.4.3 Regression discontinuity

The WFP conducted its standard scorecard survey to calculate a vulnerability score for each household in the areas affected by drought. The running score provides us with the basis to evaluate the impact of the intervention using a RD evaluation. The WFP's scorecard surveys were administered before the cutoffs were determined, elimination the possibility that households could precisely manipulate their eligibility status. In this way, we achieve randomization around the cut-off to assess the impact between beneficiaries and non-beneficiaries using a traditional regression discontinuity design. Due to the running variable being discrete, it is relatively difficult to run the regular diagnostics to detect precise manipulation of scorecard responses. This type of targeting is one of the bluntest used by WFP – used when information needs to be quickly gathered so fund or food assistance can be distributed. While it is manipulable in that those interviewed can misrepresent their asset holdings, household demographics, and occupations, doing so precisely would be extremely difficult without either the cut-off score and question set (which differed from area to area) or the locally committed budget of the program, the per household expenditure, and a prior about the responses of all others taking the same survey. Figure 2.4 shows the score distributions for each commune.

Our Regression Discontinuity design takes the form:

$$Y_i = f(S_i) + \rho_1 T_i + \varepsilon_i \tag{2.1}$$

where Y_i is the outcome variable $f(S_i)$ is some function of the running variable (in this case, S_i is the normalized targeting score), and T_i is an indicator variable which tracks whether S_i is above or below the threshold which determines treatment status. Conditional on local randomization at the boundary between beneficiary and non-beneficiary status, $\hat{\rho}_1$ estimates the Local Average Treatment Effect (LATE) of those at the margin between beneficiary and non-beneficiary status (Imbens and Lemieux, 2008; Lee and Lemieux, 2010).[27]

We are limited in the forms f can take due to the discrete nature of the running variable. We commit to a linear functional form as our preferred specification.[28]

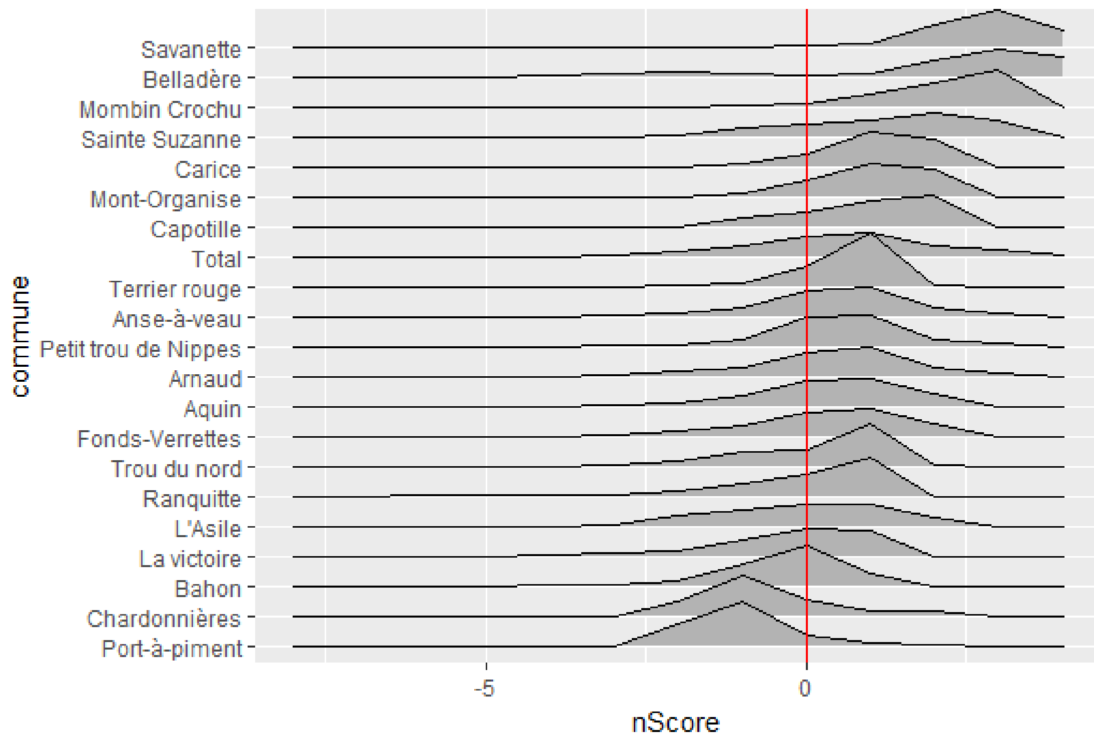
$$Y_i = \mu_1 + \delta_1 S_i + \gamma_1 S_i T_i + \rho_1 T_i + \varepsilon_{1i} \tag{2.2}$$

The main assumption for the RD design is that those who are just beneficiaries should have similar observable characteristics to the those who just miss the cut-off for beneficiary status. A common approach to falsifying regression discontinuity designs is to use a density test for manipu-

[27] In this case, we estimate the LATE for a household with a score of zero. A possible robustness check would instead estimate the LATE for a household with a score of -0.5 or 1. Given the linear specification described below, we expect this choice should not matter much.

[28] Despite the advantages of splines and local linear regressions, we suggest that these usual choices for f will not be worth exploring. In fact, even third order polynomials may be subject to issues at the boundary point, due to a limited number of values of the running variable.

Figure 2.4: Distributions of Normalized Scores by commune



Note: Author's calculations using scorecard survey. Cut-off (at zero) is marked with red. The heterogeneity of normalized scores by location around the cut-off does not suggest a consistent pattern of bunching near the cut-off.

lation of the running variable. These allow us to test for discontinuities in the distribution, which would suggest that beneficiaries have bunched, for example, directly above the cut-off (Mccrary, 2008; Cattaneo et al., 2017). Boilerplate falsification tests for manipulation of the running variable tend to over-reject in the case of a discrete running variable. These tests tend to depend on the assumption that as sample size increases; the sample size adjacent to the cut-off will also increase. With coarse variables, however, the area local to the cut-off does not "fill in." Thus the density estimation tends to yield precise but misspecified density estimates on either side of the discontinuity. Moreover, the issues caused by the violation of this assumption become worse as sample sizes increase and the running variable coarsens. Hence, these tests are of little use in a falsification exercise and in our context will tend to reject even when manipulation is not present (Frandsen, 2016).

We use a test proposed by Frandsen (2016) which adjusts the McCrary test for use with a discrete running variable. The intuition here is the same as in other similar tests – changes in the rate that probability mass is accumulated at a cut-off will indicate bunching in or out of treatment. We test for continuity of baseline covariates around the treatment threshold. Baseline characteristics were collected only for treated households, making them of little use. Therefore, we utilize plausibly invariant characteristics of households from the in-person survey. Balance tests on the selected variables can be found in Table 2.4.[29] We are not able to reject the null hypothesis that there is no discontinuity in the distribution at the cutpoint. This result gives us confidence that households could not precisely target their locations in terms of normalized score.[30]

[29]Specific animal assets contribute to the treatment status. These depend on cut-offs about exactly how many animals of a given type a household owns. Hence, we include in the balance test a continuous animal stocks aggregated into tropical livestock units. Poultry is not used to determine eligibility so it is also included in balance test.

[30]This test is controlled by a parameter $k \geq 0$ which scales the bound the second order finite difference of the probability mass function. A larger value of k allows for more deviation from linearity in point local to the cut-off, while $k = 0$ requires the distribution to be linear at the cut-off. Too low of a k will tend to result in overrejection and too low will be underpowered. For arbitrarily small values of $k > 0$ (e.g., $k = 1e - 64$), We fail to reject the null hypothesis ($p = 1.000$)

Table 2.4: Individual characteristics by elegibility status

Variable	Normalized Score < 0	Normalized Score ≥ 0	t-stat on diff	RD estimate	t-stat on diff
Respondent					
Age	39.050	41.444	2.674	0.195	0.11
Gender	0.695	0.706	0.421	0.015	0.26
Household					
Size	5.862	6.100	1.446	0.452	1.38
Children	1.013	0.941	-1.156	-0.148	-1.14
Pregnant or nursing	0.335	0.322	-0.3907	-0.068	-1.03
Monoparental	0.283	0.364	2.839	0.078	1.34
Livestock					
TLU	0.478	0.357	-1.213	-0.051	-0.23
Poultry	2.357	2.449	0.385	0.199	0.39

Note: Outcomes variables are part of the in-person questionnaire collected seven months after the cash transfer implementation. The vulnerability score comes from the scorecard survey collected before the program started and used as targeting instrument. First two columns are means by beneficiary status. Third column is the t-statistic on difference in means. Fourth column is the coefficient on beneficiary status from our preferred RD specification. Fifth column is the t-stat related to this coefficient. Gender is coded 1 if female, 0 if male. Children are age 5 or younger. Poultry is number of birds, TLU is in cattle equivalent units.

2.5 Conventional Survey-Based Impact Evaluation as Benchmark

Based on the program's goals of improving food security outcomes, the evaluation considers five main outcomes. A full description of how each outcome variable is constructed can be found in Appendix 1.B.1.

1. Food expenditure: Food purchased by the household (in HTG).

2. Food consumption: Food purchased plus home production and food received through informal or NGO assistance (in HTG).

3. Food Consumption Score (FCS): The index measures the nutritional content of food eaten in

the past week. Food is weighted by the nutritional content.

4. Dietary Diversity Score (DDS): Number of food categories consumed in the past week. Categories with no nutritional value have a weight of zero.

5. Coping Strategy Index (CSI): This index assesses a household's food security status. It is possible to see a positive impact in FCS or DDS which is the product of coping strategies i.e., households consuming food across more days, but restricting meal sizes or number of meals.[31]

Results of the RD estimations are presented in Table 5. We find an increase in the expenditure and consumption of food (measured in HTG) as well as the nutritional intake of beneficiary households in the week prior to the survey. In particular, the program increase food expenditure in 224.5 HTG, and food consumption (including donations and food consumption) in 282.1 HTG. However, we do not just see more spending by beneficiary households; we also see greater nutritional intake over the week before the survey as measured by the FCS. In particular, we see a 5.9 unit increase in the FCS of a base FCS of 40.6 units, significant at a 95% confidence level. To get a sense of magnitude in terms of food consumption, we can think about this in terms of food categories over the course of the week. This should be equal to an additional day and a half of proteins or dairy, two days of pulses, three days of cereals, or six days of fruits and/or vegetables. This increase in food consumption and nutritional intake is coupled with an increase in the diet diversity. We see an approximately 0.50 unit increase in the DDS, significant at the 99% confidence level. This magnitude corresponds to every other household consuming an additional category of food during the week, and a 5.8pp decrease in the share of consumption going to cereals, see Table 2.5.

[31] Additionally, only some of the improvement in household welfare is captured by the FCS or DDS because cash transfers might serve as a substitute for coping strategies otherwise used by the household (Maxwell and Caldwell, 2008).

We do not find evidence that the transfer program reduces the usage of coping mechanisms. For the CSI, beneficiary households have a 0.80 unit reduction in CSI (not significant at any standard level). Similarly, we see little evidence of impact on disaggregated coping strategies. We suspect our measurement of the impact on coping mechanisms is limited by the occurrence of hurricane Matthew. The acute stress may have wiped out access to certain strategies for coping with risk, while ensuring that accessible strategies are used almost universally. Figure 2B presents a graphical validation of the RD results at the score cut-off.

Table 2.5: Effect of Cash Transfer on Indices of Food Security
RD specification

	(1) FCS	(2) DDS	(3) CSI	(4) Food Expenditure	(5) Food Consumption	(6) Share Cereal
Beneficiary	5.880*	0.499**	-0.796	224.5*	282.1**	-0.0582*
	(2.494)	(0.164)	(1.483)	(91.78)	(96.36)	(0.0269)
Normalized Score	-1.844	-0.155	0.938	-48.36	-66.84	0.0393**
	(1.432)	(0.0973)	(0.840)	(47.69)	(48.41)	(0.0146)
Beneficiary X Normalized Score	1.411	0.0549	-1.153	62.42	54.99	-0.0418*
	(1.753)	(0.117)	(1.040)	(62.72)	(65.93)	(0.0182)
Observations	1132	1132	1132	1132	1132	1130
R^2	0.206	0.192	0.100	0.224	0.208	0.161

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Outcomes variables are part of the in-person questionnaire collected seven months after the cash transfer implementation.

2.6 CDR-based Targeting and Evaluation

2.6.1 Predicting eligibility status

An emergency response intervention requires a fast mechanism to identify potential beneficiaries. In a context with limited resources, the necessity for speed leads to the implementation of rapid surveys that identify the most vulnerable with a restricted set of questions. Our objective is to replicate the program's assignment of eligibility status using CDR data. We assume the scorecard exercise represents a "gold-standard" against which we evaluate the capacity of CDR-based methodology to replicate the identification of beneficiaries. The vulnerability score acts a proxy means test for the consumption level of participant households, providing a cost-effective way to target aid (Del Ninno and Mills, 2015).[32]. As the program lacks any other targeting mechanisms, it is not possible to assess any miss-classification of potential beneficiaries either by inclusion or exclusion[33] For our purposes, the scorecard exercise provides the ground-truth data for our analysis, and we assume it perfectly identifies different vulnerability levels, with households with higher scores being objectively more vulnerable.

There are two main concerns about the usage of the scorecard survey as a targeting tool. First, this methodology assumes the underlying regressions are error-free or measured with random error. The program's implementation of the scorecard survey did not validate the scores against food consumption –the proxied outcome–and relied instead on previous WFP experience to choose the variables included in the scorecard survey. This makes it impossible to quantify how effective the scorecard survey is in distinguishing vulnerability levels across households. We recognize that the validation of the targeting tool is a relevant concern to guarantee the proper implementation of social programs. However, assessing the targeting tools is outside the scope of this paper since our objective is to explore how CDR-based methods, can complement existing strategies. Second, the WFP used commune-specific eligibility thresholds that were determined by the availability of resources for each region, allowing for cases where two households with the same vulnerability score have different eligibility status. Considering that the differences in the cut-offs are small and

[32]There is ample evidence that similar methods are more efficient and cost-effective than a universal allocation (Houssou et al., 2019).

[33]Systematic reviews of the methodology reveal that it tends to yield relatively low inclusion errors but high exclusion errors (Brown, Ravallion, and Van de Walle, 2018).

households could not precisely target their eligibility status, we assume this to be an additional form of random miss-classification error that introduces noise into the CDR-based models.[34]

To evaluate the extent that cellphone usage patterns differ by eligibility status, we perform a two-sided t-test comparing the means across each sample. We follow (Khaefi et al., 2019) and classify features into five groups that reflect similar information content. Table 4B provides more details of the classification:

1. Basic phone usage, e.g., call, text, interactions

2. Active user behavior, e.g., call duration, percent initiated conversations, response delay

3. Spatial behavior, e.g., frequent antennas, number of antennas, the radius of gyration

4. Regularity, e.g., inter-event time, percent nocturnal, the entropy of contacts

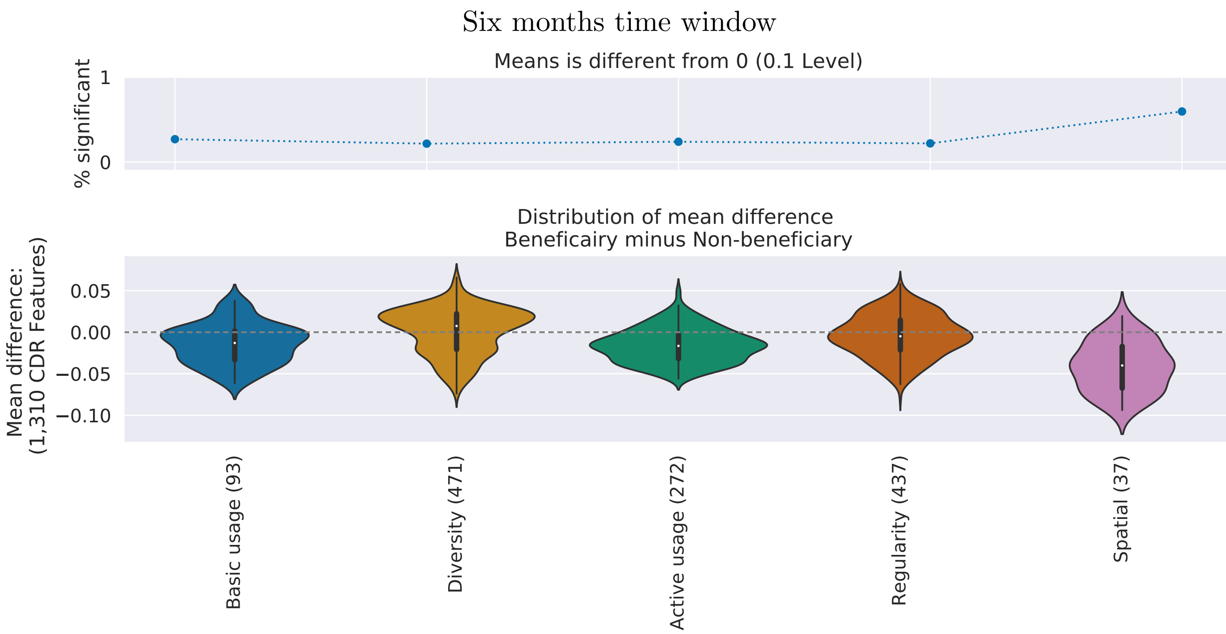
5. Diversity, e.g., number of contacts, the balance of contacts, interactions per contact

First, we want to check if the eligible population presents different patterns of behavior that the CDR can capture. For this, we compute the average difference between the two groups for each individual feature. In order to provide a similar scale we normalized all the features. Figure 2.5 shows the distribution of the mean difference for each group, as well as the percentage of features that present statistically detectable differences for the six-month time window. A negative value indicates that non-beneficiaries have a higher average for a given feature. It is not easy to characterize a general pattern since the size of the coefficients and the percentage that are significant varies depending on the time-window. Overall, results suggest beneficiary households are active

[34] To reduce the miss-classification error from the community-specific thresholds, we implement additional tests where we only use the information from the households at the tails of the vulnerability index distribution. Since these households have the highest (lowest) levels of vulnerability, they would have been eligible (ineligible) independently of the commune they live. We do not see significant differences in our results.

fewer days, with fewer contacts and lower mobility; however, they interact more frequently and are more likely to start the calls.[35] We find that using a six-month window provides the highest number of statistically significant features (458), with this number going down with the length of the time window (316 for the one-month window, 172 in the two-week window). Figure 3B shows the results for the one-month and two-week time windows. Additionally, Figure 4B shows only the statistically significant coefficients.

Figure 2.5: Mean difference CDR features by eligibility status



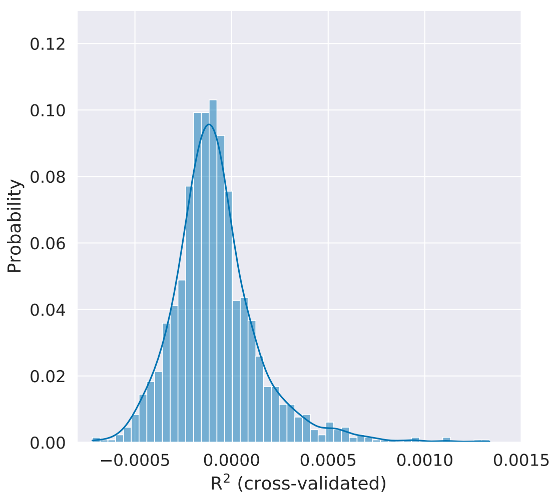
Note: Author's calculations using scorecard survey sample with records matching CDRs. Violin plot shows the distribution the mean differences between the average of beneficiaries and non-beneficiaries for each features' category. All variables normalized. A negative value indicates that for a feature non-beneficiaries have a higher average. Number of features on each group in parentheses.

[35] Location is captured at the tower level. Mobility is detected only when a person uses his phone and switches the towers connecting the transactions. Therefore, usage and mobility are are deeply interconnected.

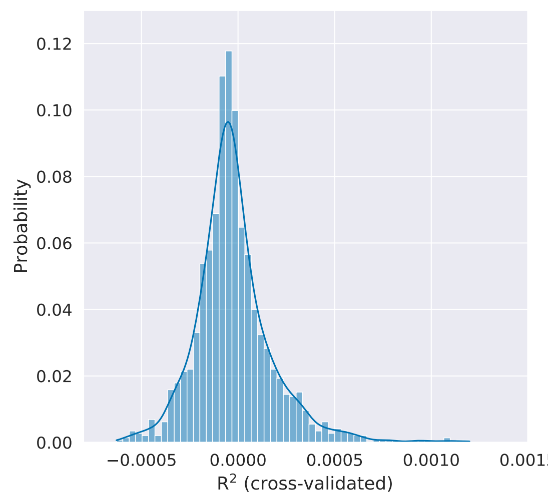
The previous results show the program's participants present different network usage patterns. Despite this, once we account for the individual features out-of-sample predictive capacity using cross-validation, no feature reaches an R square above 0.015. Figure 2.6 shows the distribution of R squares for each time window.

Figure 2.6: Beneficiary status: individual features predictive power

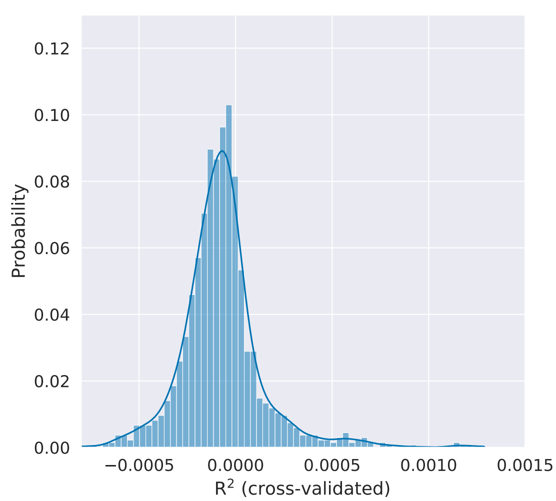
(a) Six month



(b) One month



(c) Two weeks



Note: The distribution of R^2 values from separate regressions of the beneficiary status on each available feature, showing average accuracy on the test set after 5-fold cross validation

We evaluate the model performance of the different machine learning methods using 10-fold cross-validation sets. Considering that a large percentage of the sample is eligible for the program, we stratified each fold to preserve class balance. For each fold, we train different machine learning models that include Random Forest Classifier, XGBoost, and Elastic Nets. Technical Appendix 1.B.2 provides details about each model the model's hyperparameter tuning. To account for class imbalance, we rank individuals using their predicted probability to be in the beneficiary class and use a cut-off such that the model identifies the correct proportion of beneficiaries. To capture the trade-off between inclusion and exclusion errors for varying values of this threshold, we consider the area under the curve (AUC) score to measure targeting quality.

Results are not encouraging. As Panel A of Table 2.6 shows, predicting the beneficiary status using the sample in the scorecard survey produces an AUC no greater than 0.52, with little differences independently of the time window used to calculate the features. In practical terms, these results imply that if we select two random phone user from the network, with only one of them being eligible, our model will rank the eligible person as more likely to be a potential beneficiary with a probability of 52%. This means our models are not better than randomly assigning eligibility. Results are no different if we try to predict beneficiary status on the in-person sample, in which case our best performing model reaches an AUC of 0.52.

Table 2.6: Predicting binary outcomes

	Random Forest		XGBoost		Elastic Nets	
	Full sample	Restricted	Full sample	Restricted	Full sample	Restricted
Panel A: Scorecard survey						
Six months	0.51 (0.01)	0.52 (0.02)	0.50 (0.01)	0.51 (0.02)	0.51 (0.01)	0.51 (0.01)
One Month	0.49 (0.01)	0.49 (0.02)	0.50 (0.00)	0.52 (0.02)	0.50 (0.00)	0.51 (0.01)
Two weeks	0.49 (0.01)	0.48 (0.02)	0.50 (0.01)	0.51 (0.02)	0.50 (0.01)	0.51 (0.02)
Panel B: In-person survey						
Six months	0.50 (0.04)	0.50 (0.05)	0.50 (0.03)	0.49 (0.06)	0.52 (0.02)	0.51 (0.05)
One Month	0.50 (0.03)	0.54 (0.06)	0.52 (0.04)	0.49 (0.04)	0.51 (0.02)	0.53 (0.04)
Two weeks	0.52 (0.03)	0.52 (0.05)	0.51 (0.04)	0.50 (0.05)	0.51 (0.04)	0.52 (0.04)

Note: Accuracy for predicting beneficiary status and food expenditure. Binary measures are evaluated using the AUC score. Results are averages over 10-fold cross validation, with standard deviations in parentheses. Restricted sample contains a subset of the survey participants at the extremes of the vulnerability score.

Table 2.7 shows the exclusion errors (a beneficiary classified as non-beneficiary), inclusion errors (a non-beneficiary classified as beneficiary), and different classification metrics like accuracy, precision and recall for the best performing model. We see that low AUC translate into large errors of inclusion and exclusion. For example, in the six-month time window, out of a total of 10,189 beneficiaries identified by the scorecard survey, only 75% were incorrectly classified (false negatives); at the same time, 21% of the non-beneficiaries were deemed eligible by the model (false positive). Even if accuracy seems high, we see that the large number of false positives and false negatives produce a relative low precision and recall.[36] Depending on how the WFP weights in-

[36]Accuracy denotes the percentage of correct predictions, in a dataset where the samples in one class are highly skewed (like ours), a high accuracy not necessarily reflects a well performing model since it might not be classifying the class with few observations correctly but still display a high accuracy level. Formally, accuracy is described by $\frac{TP}{TP+TN+FP+FN}$. Precision is defined as the proportion of beneficiaries that are correctly classified divided by the total number of predicted beneficiaries; a higher precision implies the classification returns more correctly classified outcomes than incorrect one. Formally, it is described by $\frac{TP}{TP+FP}$. Recall shows the proportion of beneficiaries that are correctly classified divided by the total number of individuals beneficiaries; a high recall implies the classifications identifies a higher proportion of the actual beneficiaries.Formally, it is described by $\frac{TP}{TP+FN}$.

clusion and exclusion errors, the metric guiding the classification performance can rely on these additional evaluation metrics. A similar situation happens when we predict beneficiary status for the participants in the in-person survey.[37]

Table 2.7: Classification metrics for beneficiary status.

Survey	True Pos.	True Neg.	False Pos.	False Neg.	Accuracy	Precision	Recall
Scorecard	2,452	2,450	7,572	708	0.372	0.245	0.776
In-person	209	208	275	131	0.507	0.432	0.615

Note: Results using the best performing model for each sample.

Two elements of the WFP targeting process can affect out capacity to predict a person's beneficiary status. First, the WFP multi-step targeting process only interviews people living in the regions hardest hit by the 2016 drought. This makes that the sample only includes very vulnerable households, limiting the variation in the levels of vulnerability we can use to train our models. We do not have information for people outside of the targeted areas to increase the available information. Second, since the eligibility cut-off changes depending on the fund available for each region there are instances where two people with the same vulnerability score present different beneficiary status. We assume this constitutes an additional source of random noise in the eligibility criteria that should not invalidate out approach.

With the information at our disposal, we try to solve these two problems by restricting the sample to respondents at the extreme tails of the vulnerability score. In this sample the amount of variation in the vulnerability levels is maximized, while at the same time the overlapping eligibility cut-offs are eliminated. To make the results comparable, we reweight the new sample to represent the same number of observations and share of eligible households. The sample we can use for this greatly reduced as in both the scorecard and in-person surveys a large share of respondents are

[37]Figure 5B shows the ROC for each model

concentrated around the eligibility threshold.[38] It is important to clarify that this exercise does not rule out that lack of variability in the outcome variable drives our low predictive capability. It can still be the case that, even after subsampling in the tails of the vulnerability distribution, the resulting sample has no detectable differences in their true underlying vulnerability. As Table 2.6 shows, the changes in the AUC are marginal at best; with still large levels of false positives and negatives, see Table 2.8.[39]

Table 2.8: Classification metrics for beneficiary status: Restricted sample

Survey	True Pos.	True Neg.	False Pos.	False Neg.	Accuracy	Precision	Recall
Scorecard	215	212	1,587	20	0.210	0.119	0.915
In-person	109	107	68	126	0.527	0.616	0.464

Note: Results using the best performing model. Sample restricted to the tails of the vulnerability levels.

2.6.2 Predicting food consumption and expenditure

We follow a similar approach to predict the main two outcomes from the program: Food consumption and food expenditure. These outcomes are only present in the in-person survey and, in particular, to participants who had a valid cellphone at the time of the survey.[40] We evaluate model performance using the cross-validated correlation (r) between true and predicted outcome. As Table 2.9 shows, in all cases our models perform poorly, with correlation close to zero. As a point of reference, CDR-based prediction of a wealth index in Rwanda show performances around 0.68 (r) (Blumenstock, Cadamuro, and On, 2015).

[38] In the scorecard sample, out of the 12 levels of vulnerability, we keep we keep scores from [-8,-3] below and [2,4] above the eligibility cut-off, representing 15% of the original sample. The in-person survey only contains people with vulnerability scores between -3 and 2. We keep scores between [-3,-2] and [1,2], representing 49% of the original sample.

[39] Figure 6B shows the corresponding ROC.

[40] Valid means a cellphone that was part of Digicel's network and that had activity covered during the time window used to extract behavioral features. Section 2.4.2 provides details on this sample.

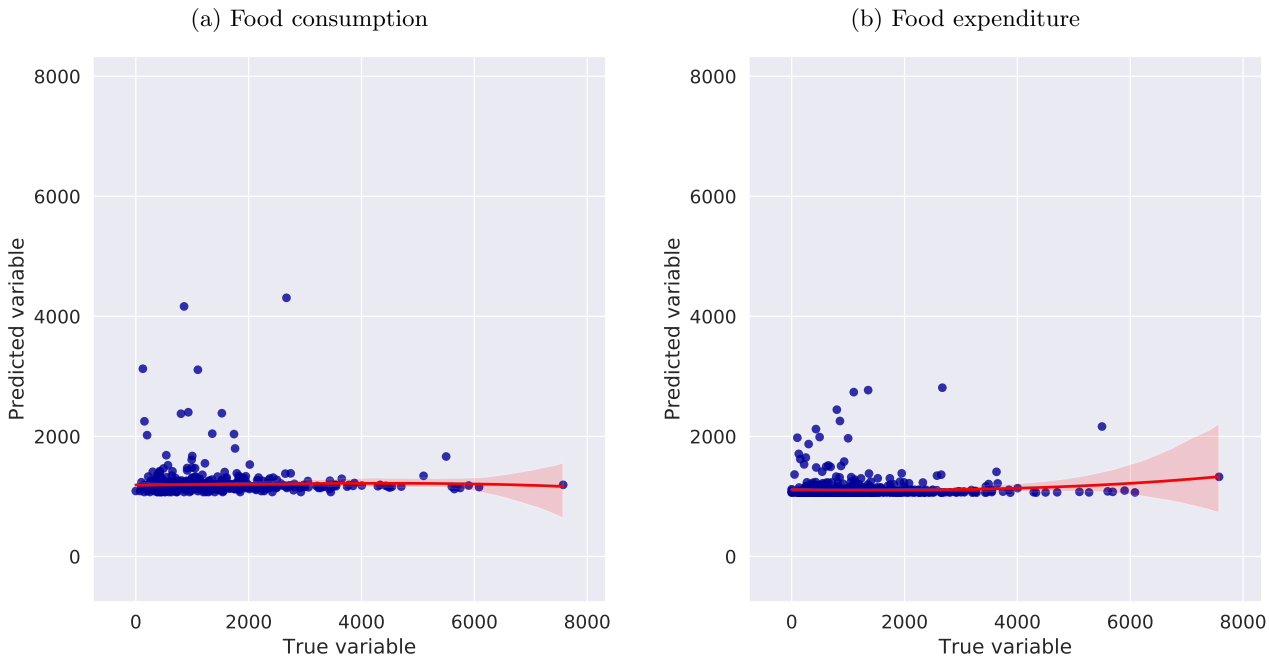
Table 2.9: Predicting continuous food outcomes

	Random Forest Full sample	**XGBoost** Full sample	**Elastic Nets** Full sample
Panel A: Food expenditure (r)			
Six months	0.08 (0.10)	0.08 (0.14)	0.00 (0.00)
One Month	0.05 (0.14)	0.06 (0.09)	0.00 (0.00)
Two weeks	0.07 (0.10)	0.06 (0.12)	0.00 (0.00)
Panel B: Food consumption (r)			
Six months	0.05 (0.09)	0.07 (0.13)	0.00 (0.00)
One Month	-0.01 (0.08)	0.00 (0.09)	0.00 (0.00)
Two weeks	0.09 (0.07)	0.12 (0.06)	0.00 (0.00)

Note: Data form in-person survey. Models fit measures using correlation coefficient between true and predicted values. Results are averages over 10-fold cross validation, with standard deviations in parentheses. Restricted sample contains a subset of the survey participants at the extremes of the vulnerability score.

Using the best performing model, in Figure 2.7 we plot the true and predicted values for the food consumption and expenditure outcomes. As we can see, the low predictive power of our models makes that there are no observable differences in the predicted values for households with high and low levels of consumption. This constrains any attempt to used the results to replicate the program's impact, as identified by RD results.

Figure 2.7: Predicting food consumption with phone data



Note: Relation between actual weekly consumption (expenditure) on food (as reported in the in-person survey) and predicted consumption (expenditure) as inferred from mobile phone data for survey participants with a valid phone. We use the best performing model using features extracted for the six-month time window.

2.6.3 Replicating RD results

The most compelling application of mobile phone data to predict socioeconomic indicators is to enable new approaches to impact evaluation and program monitoring (Blumenstock, Cadamuro, and On, 2015). However, complementing well-understood impact evaluation and monitoring methods presents several challenges. First, even if a large body of work indicates cellphone data present fingerprints that are unique to certain socioeconomic indicators, at the individual level, prediction tends to work best for outcomes with little variation over time, limiting its usage in the context of impact evaluation and monitoring. Second, mapping digital data to welfare outcomes is both population and time-period specific, with evidence suggesting that predictive capacity deteriorates quickly (Lazer et al., 2014; Blumenstock, Cadamuro, and On, 2015). Third, the quality of the prediction affects the value of the outcomes of interest reducing the power of any test run using the predicted outcomes.

Our previous results show that in a setting where we use data from a real-world impact, cellphone data does not replicate the program's main outcome. Despite these unimpressive results, we use the predicted food consumption and food expenditure to test if we can detect similar average differences between beneficiaries and non-beneficiaries as the impact evaluation results by estimating equation 2.2. For this, we use our best-performing model on the six-month time window features. Considering the power analysis at the end of this section, we provide all the impact effects in standardized units. Columns (1) and (2) of Table 2.10 shows effect of the program for the whole sample, including households both with and without a cellphone. We can see the program increase the food consumption and food expenditure among participants. In the next two columns, we replicate the same regression but only including households with a cellphone we were able to match to the CDR data. As we have discussed throughout the document, households with a cellphone present better baseline indicators. Using this sample, the program's impact on food consumption goes down while disappearing for food expenditure. Finally, using the predicted outcomes the best performing model, columns (5) and (6) detect no impact for the program. This is not surprising given that out low predictive capacity eliminates most variation across the food expenditure (consumption) space.

Table 2.10: Replicate RD results using predicted outcomes

	Full sample		Cellphone in network			
			Observed		Predicted	
	Food Consumption	Food Expenditure	Food Consumption	Food Expenditure	Food Consumption	Food Expenditure
	(1)	(2)	(3)	(4)	(5)	(6)
Beneficiary	0.354***	0.318**	0.254*	0.210	0.058	0.058
	(0.128)	(0.128)	(0.149)	(0.148)	(0.149)	(0.149)
Normalized Score	-0.088	-0.067	-0.042	-0.015	-0.094	-0.094
	(0.076)	(0.076)	(0.090)	(0.090)	(0.090)	(0.090)
Beneficiary X Normalized Score	0.039	0.038	0.048	0.046	0.181*	0.181*
	(0.092)	(0.092)	(0.109)	(0.109)	(0.109)	(0.109)
Observations	1,137	1,137	823	823	823	823
R^2	0.011	0.011	0.010	0.011	0.004	0.004

*p<0.1; **p<0.05; ***p<0.01

Note: Full sample includes all the participants in the in-person survey. Numbers in cellphone network restricts sample to participants with a valid phone number. Result in standard deviations.

We want to understand how the minimum detectable impact changes as the precision of the machine learning models deteriorates. When we include all the survey participants, we are able to detect a change of 0.17 standard deviation over the mean of the control group.[41] Only including household with a valid phone number increases the minimum detectable effect to 0.21. Under these circumstances, it is not surprising we are less likely to find significant impacts in the restricted sample, as it is the case in the results in columns (3) and (4). The sample of individuals with a valid phone number represents the ground-truth data we use to measure the predictive capacity of any machine learning exercise. Using different combinations of feature extraction and machine learning algorithms will create an unique combination of predicted values. Each combination will produce its own R^2, with the predicted values of the different algorithms converging as the R^2 approaches one.

[41] Assuming an alpha level of 0.05, and a power of .8

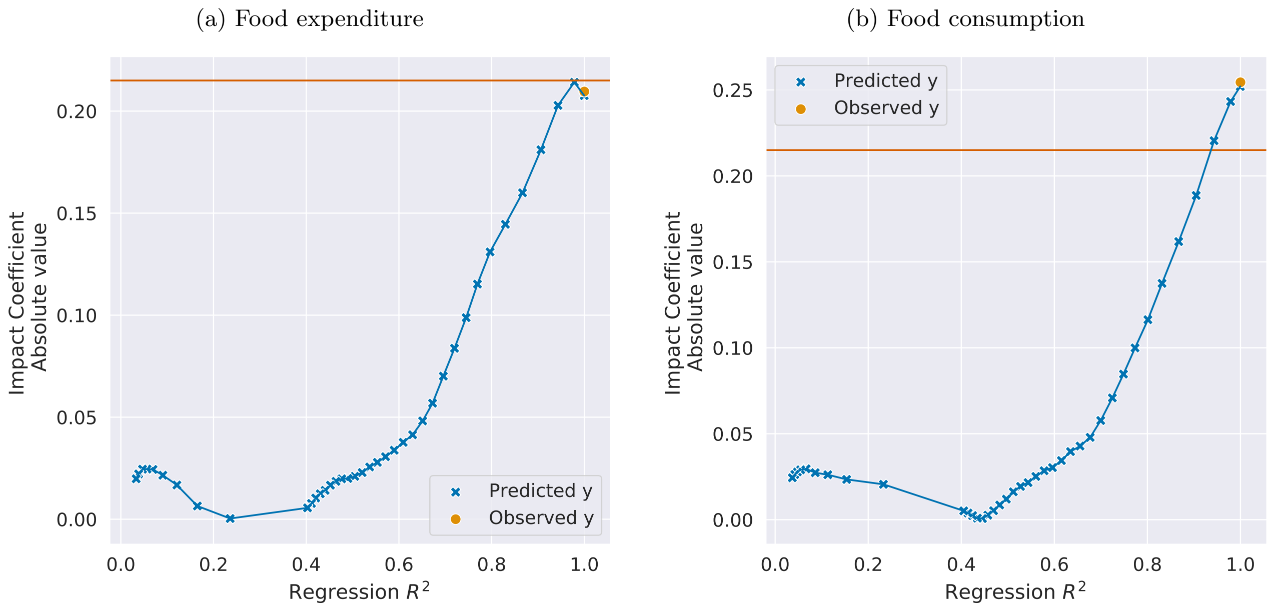
Figure 2.8 shows how prediction quality affects our capacity to detect statistically significant impacts. To overcome that our predictive models have a very low R^2, we use a Lasso model with different regularization levels to predict the different food consumption measures. A penalty of 0 will produce a model that perfectly (over)fits the observed data and provides the same regression coefficients as the original sample. In contrast, increasing the penalization reduces the number of features entering the model and reduces the R^2. Using the the predicted values, we proceed to estimate regression 2.2, showing how they change as the quality of the predictions goes down. The horizontal line provides the minimum detectable effect in our sample. Therefore, we cannot identify as significant regression coefficient below this level.

Two main themes appear. First, at least in this case, the estimated impact using predicted outcomes rapidly declines with reduction of the R^2. Of course, the minimum detectable impact can be reduced by predicting values on a large sample. This is often the case in social research, where the number of observations in the ground-truth data is small compared with the total population. However, this does not address the apparent underestimation of the true impacts. Second, for R^2 values below 0.4 we see that the predicted data signals a decline in the food expenditure and consumption levels of the programs' beneficiaries. Considering that high-performing models tend to have an R^2 around 0.4, this results raises questions about how to properly benchmark machine learning predictions in the context of an RD design.

2.7 Postmortem: What went wrong?

Why were we unable to replicate the success of others in using CDRs to predict household-level well-being? There are several insights that emerge from exploring this question. In this section, we discuss these insights and associated implications for predicting welfare outcomes using big data

Figure 2.8: Changes on minimum detectable impact for different levels of R^2



Note: The horizontal line shows the minimum detectable impact for a sample of a size equivalent to the number of in-person survey respondents with a valid phone number. The figure plots the impact coefficients from regression equation (2.2) using the predicted outcomes from a Lasso regression with different penalty level (α). Orange dot represents the effect observed in the actual data.

in poor countries. We use data from a nationally representative sample (Finescope, 2018), which includes many individual responses that can be matched to our CDR data, as the empirical basis for conducting this postmortem.[42]

We identify three reasons our prediction ability is limited in this specific context. First, the nature of cell phone ownership and usage among the rural poor in Haiti can undermine the signal value of CDRs for this population. Second, predicting flow variables like food expenditure and consumption is likely more challenging than predicting stock variables like wealth and asset indices. Finally, several features of this cash transfer program conspire to reduce statistical variation of outcome variables, which complicates prediction. We empirically explore each of these in turn.

[42]Section 2.4.2 provides details on the protocol to obtained informed consent for this sample.

2.7.1 Cellphone Penetration among Rural Poor Limits Value of CDRs

A natural limitation of using cellphones for social research is that having a phone in itself indicates socioeconomic status. Despite the impressive growth in the number of users in developing countries during the last decade, at the time of this cash transfer program less than 60% of Haitian households possessed a cellphone number (with large disparities across age groups and locations). The sample that participated in the WFP scorecard process lives in the poorest rural areas of Haiti where cellphone ownership is particularly low: Only 34% of scorecard respondents reported having a cellphone number. This lack of cellphone penetration into these households raises two primary challenges.

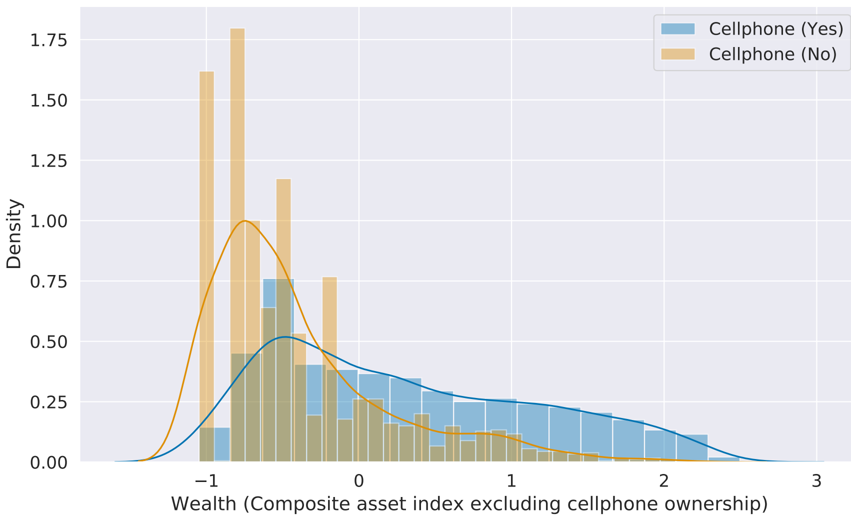
First, households without a cellphone are systematically poorer than those with cellphones. In the WFP data we use in this paper, we see that households with cellphones are less vulnerable on average (see Table 2.2).[43] This pattern unsurprisingly holds with nationally representative data: As well as shown in Figure 2.9, households without a cellphone are concentrated in the bottom of the wealth distribution.[44] While the observation that cellphone ownership is correlated with wealth is not surprising, it has important implications for the use of CDRs among the poor populations that are the intended beneficiaries of development or assistance programs.[45]

[43] Figure 7B shows the difference in the distribution of the vulnerability scores for households with and without a phone.

[44] Wealth is represented by a composite wealth index calculated using principal component analysis with cellphone ownership omitted from the calculation.

[45] It is also worth noting that relying on self-reported cellphone ownership as the basis for eligibility for an assistance program like the WFP cash transfer may introduce under-reporting incentives since cellphones are easy to hide as assets. Much like some face incentives to strategically manipulate their cellphone usage to gain access to resources allocated using CDR-based algorithms (e.g., nano-loans) (Björkegren, Blumenstock, and Knight, 2020), those whose access to benefits hinge on the kind of scorecard targeting used in this program face very well-documented incentives to under-report assets.

Figure 2.9: Comparison wealth index for households with and without a phone



Note: Author's calculations using Finscope 2018. Wealth index is the first principal component of household assets. We exclude cellphone ownership form the calculation.

Second, relatively low cellphone penetration rates among the rural poor make the resulting CDR data systematically less plentiful and less useful as a source of statistical signal. Naturally, where cellphones are sparse so too will be CDR data, which obviously undermines CDR-based prediction. Even if a meaningful prediction could be coaxed from these sparse CDRs, using them for targeting and evaluation would be complicated by the fact that CDRs are most likely to be missing for the poorest households. Low penetration rates also change the way cellphones are used in these rural settings as they are much more likely to be shared within households and even among households. Such sharing introduces considerable noise into resulting CDRs and renders these data less informative as the basis for prediction.[46]

[46]We find evidence of cellphone sharing in all the surveys available. In the case of the scorecard survey, 205 out of 13,780, and in the in-person survey 56 out of 872 households reported the same phone number as their own. Considering that we can only identify shared numbers when two or more households report the same line, these percentage of shared numbers represents a lower bound on the real level of number sharing.

2.7.2 Flow Variables Are Harder to Predict than Stock Variables

The main objective of the WFP intervention was to improve food security by enhancing households' ability to purchase food. We therefore used total consumption and expenditure on food as leading impact indicators for the program evaluation. Our use of CDR-based methods for predicting these outcomes was motivated in part by encouraging evidence from the literature, which suggests that CDRs can be used to predict not just wealth but also food security indicators at the census-tract level (Blumenstock, Cadamuro, and On, 2015; Hernandez et al., 2017; Decuyper et al., 2014). The sharp contrast between these prior prediction successes and this analysis suggests that estimating individual consumption levels is far more difficult. As a final reason for this failure to predict, we explore important differences between predicting consumption flow variables and stock variables such as wealth.

We lack the necessary information to replicate a wealth index based on the WFP survey data. Instead, we use the nationally-representative.[47]. This survey of diverse individuals includes measures of wealth, which span the entire wealth distribution. Results in Table 2.11 show that, at least in the case of Haiti, CDR-based behavioral features perform a far better job at predicting wealth than food and total expenditure. We see that across our models the predicted correlation between true and predicted wealth is above 0.4, in line with previous findings in the literature, see (Blumenstock, Cadamuro, and On, 2015). In the case of food expenditure our model clearly underperforms. Even when presented with a sample that covers the whole distribution of expenditures in Haiti, we are unable to properly predict either food and total expenditure with any degree of confidence. Although not definitive, this evidence suggests that flow variables are indeed harder to predict than stock variables.

[47] See Section 2.4.2 for details on this survey

Table 2.11: Predicting wealth and consumption using a nationally representative sample

	Random Forest	XGBoost	Elastic Nets
Panel A: Wealth composite index (r)			
Six months	0.45 (0.03)	0.45 (0.02)	0.49 (0.03)
Panel B: Food expenditure(r)			
Six months	0.09 (0.09)	0.14 (0.04)	0.16 (0.04)
Panel C: All expenditure (r)			
Six months	-0.02 (0.03)	0.02 (0.06)	0.00 (0.00)

Note: Models fit measures using correlation coefficient between true and predicted values. Results are averages over 10-fold cross validation, with standard deviations in parentheses.

For reference, figure 2.10 compares predicted and actual outcomes for the wealth asset index, food expenditure, and total expenditure. Here, we can see the degree that of variance in the outcome our model is able to capture when predicting wealth levels.

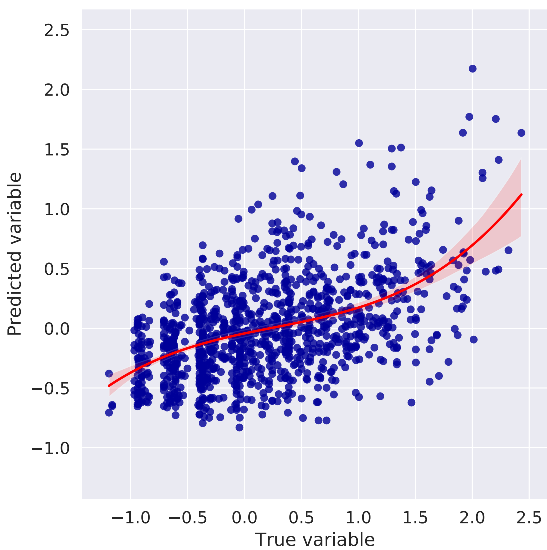
2.7.3 Effective Targeting Restricts Variation in Outcome Variables

WFP used a standard multi-stage process to identify beneficiaries for this cash transfer program. After pre-selecting rural regions and communities most directly affected by drought in the years prior to the program, the WFP staff constructed lists of vulnerable households in collaboration with local authorities. These potential beneficiaries then participated in a scorecard survey described previously. WFP set a cut-off vulnerability score based on the overall budget for the program and included all households with vulnerability scores greater than this threshold in the cash transfer program.

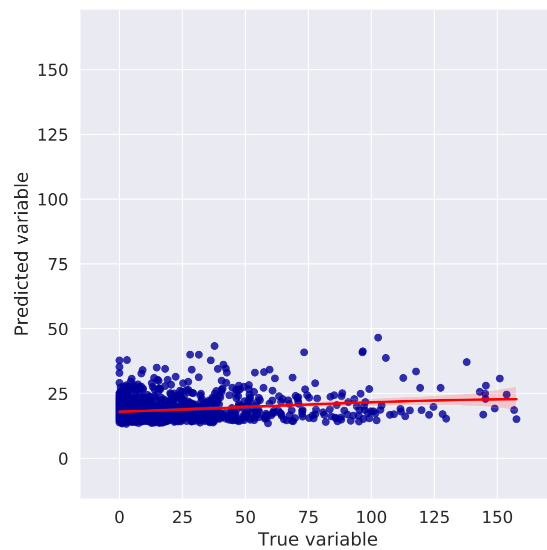
While this targeting process can be cost-effective as part of program implementation, it can

Figure 2.10: Observed and predicted outcomes: Finscope data model

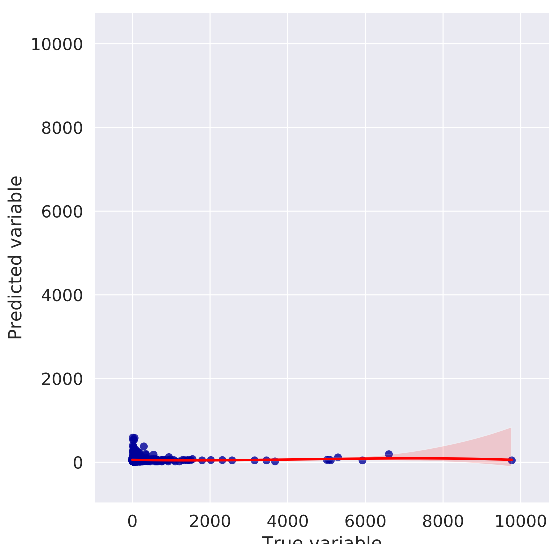
(a) Wealth index



(b) Food Expenditure



(c) Total expenditure



Note: Results using individuals from the Finscope, 2018 survey with a valid phone number who agreed to participate in the study.

simultaneously restrict the statistical variation of key outcome variables. Indeed, the more effective the approach targets potential beneficiaries for inclusion in the scorecard survey, the less useful the resulting scorecard data are for training CDR-based algorithms to predict these outcome variables. Compared with nationally-representative data, these pre-screened households are worse off in every aspect. For example, among these households per-capita food expenditure levels are one-fourth of the national average, food insecurity levels reach almost 100%., and food deprivation is higher (see Table 2.12). As Figure 2.11 shows, the cumulative distribution of food expenditures for the in-person survey is far below both the urban and rural distributions from the nationally-representative data. This is compelling evidence of the effectiveness of the early stages of the WFP targeting process, which is good news for programmatic operations but implies a very narrow statistical basis for predicting outcomes among this population.

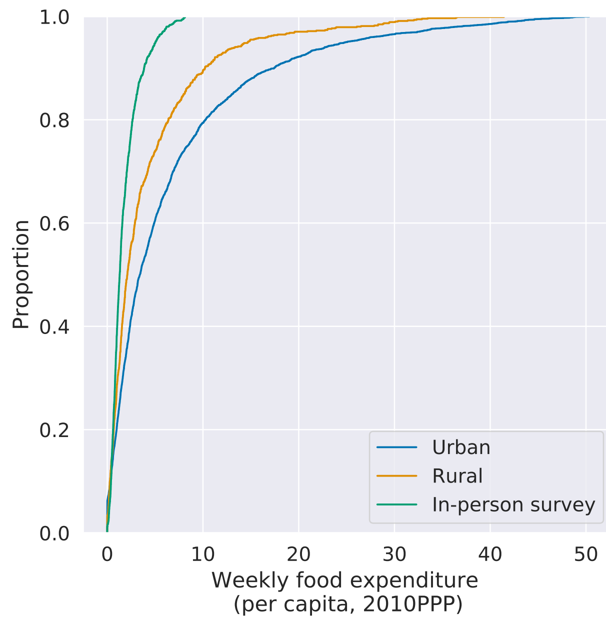
Table 2.12: Comparing in-person survey with a nationally representative sample

	In person	National survey	Sig. diff.
Food expenditure per capita	1.95	8.06	***
	(2.05)	(20.89)	
Food insecure	0.97	0.5	***
	(0.17)	(0.5)	
If food insecure (Last seven days with:)			
Smaller meals	3.12	1.74	***
	(1.91)	(2.18)	
Adults ate fewer meals	2.47	1.37	***
	(2.27)	(2.04)	
Fewer meals	3.01	1.77	***
	(2.0)	(2.23)	

Note: Monetary values are expressed in 2010 USD PPP

A similar prediction problem emerges with the scorecard cut-off that determines final eligibility status. Pre-screened households for which vulnerability scores are available are much poorer than the average Haitian, as already demonstrated. Thus, the eligibility cut-off differentiates between those who are very vulnerable and those who are just slightly less vulnerable. This is precisely the appeal of the RD design to evaluating impacts of the cash transfer. But as greater similiarity between households on either side of the threshold strengthens the RD case, it simultaneously (and potentially dramatically) restricts the statistical basis for machine learning algorithms to predict outcomes and therefore their performance.

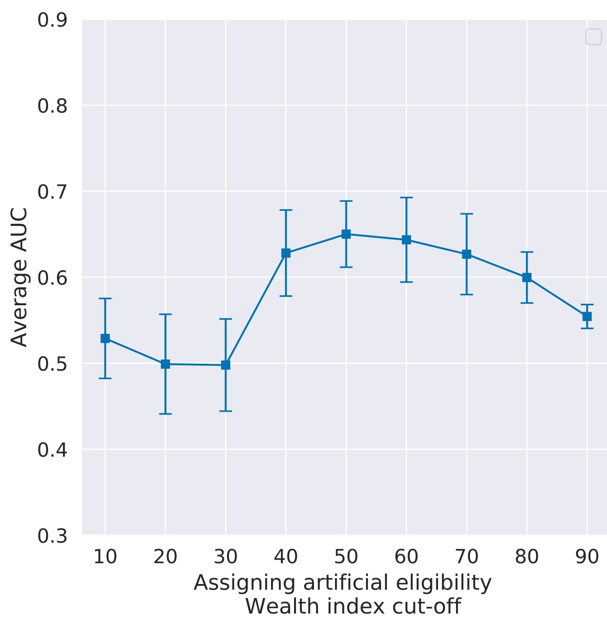
Figure 2.11: CDF food expenditure



Note: Author's calculations using the in-person survey. Information for the urban and rural population comes from Finscope 2018. All values are deflated to 2010 prices.

We want to provide some evidence about how highly targeting the data collection process affects the model's predictive ability. For this, we use the nationally representative and exploit that we have a higher capacity to predict an asset wealth level index. We implement two exercises. Our first simulation explores how selecting different eligibility thresholds changes the classification's AUC. We use the full sample and artificially identify as "eligible" people with a wealth index below several cut-offs. For each cut-off value, we estimate a classification model using elastic nets and obtain the corresponding AUC. Figure 2.12 shows how the model performance does not behave linearly with respect to the eligibility cut-off. At least in our case, the signals from the CDR behavioral features appear to be better at identifying who are the wealthy individuals –wealth index above the 50th percentile–, than discerning who are the poorest –cut-offs below the 30th percentile. If we were able to extrapolate these results to the classification of WFP beneficiaries, a sample concentrated on poor individuals, this could further explain our model's low capacity to identify beneficiaries.

Figure 2.12: Classification AUC: Assigning an artificial eligibility threshold



Note: Author's calculations using Finscope data; only numbers that provided consent to match their responses with their CDR data. Classification uses an elastic nets model. Bars show the standard error across folds. Figure shows how using different eligibility thresholds on the asset wealth index affect the model classification AUC.

Our second simulation looks at how the classification's error changes at different levels of the wealth distribution. We use the 50th wealth percentile as the artificial eligibility threshold and calculate for each decile the percentage of individuals that were correctly classified.[48] Under perfect classification, any individual left (right) of the threshold should be classified as eligible (non-eligible). We see the percentage of individuals correctly classified increases as we move away from the eligibility cut-off; a result that holds both for eligible and non-eligible individuals, see Figure 2.13. The previous result suggests that collecting a sample that represents a large dispersion in the wealth distribution facilitates the classification exercises.

2.8 Conclusions

In recent years, the field of development economics has been invigorated by the prospects, potential and promise of big data and machine learning. Empirical breakthroughs using these novel data sources and methods have pushed research frontiers and deepened our understanding of key questions, many with direct and compelling policy relevance. This paper is specifically motivated by impressive and exciting results that suggest that machine learning techniques can extract surprising insights about mobile phone users from the CDR metadata they generate as they engage with the cellular network. Building on evidence that CDRs can predict wealth levels, we explore the limits of CDR-based analysis in the context of programmatic targeting and impact evaluation as second-generation applications of these methods. Such uses of these data and methods are especially intriguing as they would enable cost-effective and near real-time targeting and monitoring, evaluation and learning. Testing the viability of these applications is our objective in this paper.

[48] We use that specific eligibility threshold since it provides the highest performance, see Figure 2.12.

Figure 2.13: Classification error around the eligibility threshold.

Note: Author's calculations using Finscope data; only numbers that provided consent to match their responses with their CDR data. Classification uses an elastic nets model. Artificial eligibility threshold at the 50th percentile of the wealth distribution: individuals to the left (right) of the threshold assumed eligible (non-eligible). The figure shows for each decile level the percentage correctly classified. Individuals with low (high) wealth levels have a higher probability of being correctly assigned to their simulated eligibility status. Bars denote standard error.

Current targeting methods require that every potential beneficiary participates in a survey to inform his eligibility for the program. Imagine a scenario where CDR-based features can predict a vulnerability score. Once a model has been trained on a passive data source, such as cellphone records, out-of-sample predictions of the targeting score allow for the immediate scoring of large swaps of the population at almost no additional cost. Building in these methods can provide real-time monitoring of socioeconomic indicators, opening the door to complement the impact evaluation of social programs. After training a model to predict a specific outcome, these methods make it possible to observe how an intervention changes the outcome of interest in populations for which the follow-up information was not collected.

The training of CDR-based models will require careful calibration for every context, and there is ample evidence that well-performing models tend to decay quickly over time. We provide evidence of how there are inherent limitations in using these CDR data when a program has been highly targeted. By showing the limitations of these methods under the specific circumstances we study, we hope to inform the design of future applications.

In collaboration with the WFP and the major mobile network operator in Haiti, we first conducted a conventional survey-based impact evaluation of a large, emergency cash transfer program administered and implemented by WFP. Using an RD design based on a threshold used to determine eligibility for receipt of the transfers, we find that the program indeed achieved its goal of improving food security, increasing food expenditure and expanding dietary diversity in the wake of a punishing drought. We then turn to the CDR-based analysis and assess how well these passive metadata and machine learning techniques can replicate the program's targeting and our survey-based impact evaluation. This alternative CDR approach fails in this case on both the targeting and the evaluation front for one simple reason: Even advanced machine learning methods are unable to generate useful predictions of phone users' well-being.

We identify and discuss several explanations for this prediction failure. First, both CDR-based targeting and prediction of food security outcomes apply only to households with mobile phones. Despite rapid growth in cellphone ownership rates, the most vulnerable, and therefore the intended beneficiaries of programs like the WFP cash transfer, have the lowest cellphone ownership rates and often share phones, which undermines the usefulness of CDR data. Second, the WFP program effectively targeted the poorest and most vulnerable households in the rural areas hardest hit by drought, which severely limits the statistical variation in key outcome variables and therefore undermines our ability to predict these outcomes. Third, we find that CDR-based methods perform better in predicting asset-based wealth measures (stock variables) than consumption (flow) variables. Although we cannot directly test it in our setting, this finding may further imply that using CDRs to predict *changes* in wealth over time may be deceptively more difficult than predicting wealth levels in a given cross-section.

The fact that we fail to replicate the survey-based impact evaluation using CDR-based predictions - and the reasons behind this failure - suggest that the excitement generated by CDRs as a data source in development economics should be tempered by some very real limitations. A vision of passive data taking the place of more costly conventional survey data is irresistible in many ways, but it is also likely unrealistic. While information and communication technologies, combined with administrative and big data, may continue to improve the data collection process and the quality of data collected, active data collected through these improved survey techniques will certainly continue to play a central role. This is as true for researchers as for program managers.

More specifically, the analysis in this paper reveals a fundamental empirical tension between algorithmic predictions using machine learning methods, on the one hand, and causal identification, targeting and program evaluation on the other. This tension is multi-faceted, but based on the notion that effective outcome prediction requires statistical variation across a broad support of the underlying outcome distribution. This requirement runs counter to causal identification, which hinges on counterfactuals for comparable individuals or households. This is most clearly on display with RD identification strategies that leverage similarities between treated and untreated units on either side of a threshold and within a narrow bandwidth. The narrower the bandwidth and the more similar the units are, the more compelling the RD estimates, but the more challenging outcome prediction becomes. In the context of targeting, the same pattern emerges: The more effective targeting is at identifying a sub-population with relatively homogeneous needs or characteristics, the less effective outcome prediction will be among this sub-population. Navigating this empirical tension will require a better appreciation for and understanding of its nuances - something to which we hope this paper can contribute.

Chapter 3

Can information drive adoption of mobile money? A field experiment in Haiti

Despite the potential of mobile money to allow vulnerable populations access to financial services there is little information on the determinants of adoption. Existing research explores the determinant of early adoption, but little is known about what drives later adoption decisions, specially when market adoption has stagnated. In the case of Haiti, after several years in the market presents a stagnant level of subscriptions. We test if lack of knowledge about how to use the services explains low levels of new subscriptions. Combining a survey with randomized component and mobile money transaction logs, we measure the extend that informational videos can induce adoption (extensive margin) and increase the number of services customers use (intensive margin). There are three main takeaways. First, awareness of mobile money services is high and, even if having an account is free, many people use mobile money indirectly by asking others to make transactions for them. Second, the intervention increased adoption by 5.4%. However, a large share of new users came from the group that declined the opportunity to watch the videos, indicating that there is more than simply information driving adoption. Third, we do not find any impact of the videos on the usage of additional services by people with an account at the time of the survey, a result connected with the few opportunities to use services beyond buying airtime. Taken together, these three factors signal that further growth of the mobile money platform requires increasing the number of services available to attract additional customers and incentivize the daily usage of mobile money for economic transactions.

3.1 Introduction

Mobile money (MM) offers a new avenue for financial inclusion. Using the existing cellphone infrastructure, MM overcomes the high transaction costs that have inhibited the development of financial services for the poor and those living in remote areas (Suri, 2017). Besides providing a channel to send, received, and store funds, MM creates new development opportunities by increasing the capacity of households to manage risks, and save.(Suri and Jack, 2016; Economides and Jeziorski, 2017). Despite the potential of MM, there is little information on what makes a product successful. Current research centers on characterizing which groups are the first to adopt, but less attention has been placed on understanding what determines the long-terms success of MM.

With one of the lowest densities of formal financial providers in the world, Haiti provides an ideal setting for the development of MM services. However, after ten years in the market, and a relaunch of the service in 2015, MM adoption in the country is stagnant, with only 29% of mobile subscribers using the product. Industry reports often cite information and reliability problems as the main main barrier for further product growth (GSMA, 2017). In this paper, we formally test if providing information about how to use the service can promote: (i) Adoption of the services, measured as the probability of opening an account; and in the case of people who were already subscribers (ii) usage of additional services, measured as using a service not used in the past.

We implement a survey with a randomized control trial component that granted access to the informational videos to a subset of participants. A total of 2,261 people from two intermediate cities participated in the experiment. After collecting information on MM usage, a randomized component allowed individuals in the treatment group to watch as many videos as desired. We created a series of four videos showing (i) how to open an account, (ii) buy mobile airtime, (iii) funds transfer, (iv) store payment. Embedded in the video, there is a short story where the service is useful and all the necessary codes to use the service. As part of the survey, participants provided informed consent to match their responses to their cellphone and MM records; this information allows me to measure how the intervention affected adoption and product usage.

The information videos tackle two of the main challenges industry reports identify as constraints for greater MM usage. First, people find it challenging to use MM, in particular on feature phones. In Haiti, most users, including those with a smartphone, access MM using Unstructured Supplementary Service Data (USSD) in Haiti. Videos provide step-by-step instructions about the codes a person should dial to complete a transaction. However, their main takeaway is to show that once a person opens the USSD menu, the system can guide them through the necessary steps without memorizing the codes for the transaction. Providing a short video that shows menus are self-explanatory could be enough to demystify the usage of the service, especially when data reveal that up to 32% of cellphone users without an account have asked others to make transactions for them. Considering that having an account has no cost, this is, at least, suggestive that people find registering and making transactions difficult. Second, short videos represent a much more scalable solution to provide information, making that direct interaction with MM agents is left only for the most complicated usage questions.

The survey shows that most people know about MM, and in general, people perceived it as a secure mechanism to transfer funds. Additionally, the survey reveals that only counting people with an account as MM users omits the large share of individuals that use the service by asking others to perform transactions for them. We find two main impacts from the intervention. In terms of adoption, people in the treatment group were 5.5% more likely to use MM in the seven weeks following the intervention. Most of the new subscribers joined in the two weeks after the intervention. Since data limitations constrain the study to observe individuals for only seven weeks after the survey, we cannot assess long-term adoption. However, we find that new subscribers used MM for more than two weeks, revealing, at least partially, that adoption was not a single usage of the service. The majority of new subscribers used mobile money to buy mobile airtime, and, interestingly, we do not find a correlation between the videos a person watched and the products used. In fact, a large percentage of individuals in the treatment group who did not watch any of the videos still adopted. We do not find differences in the usage levels between new subscribers who did not watch the videos but still adopted. These mixed results signal that people watching the videos learned how to used the product, but that for a large group, opening a MM account required only a small incentive in the form of questioning their knowledge about the product.

For survey participants with an account at the time of the survey, we do not find any impact of the videos on the usage of additional services. We argue that once a person has an account, lack of information about how to use the products is not an important factor. Moreover, out of the three services mobile money offers, only paying for services has a small user base, with less than 10% of the subscribers using the service. Considering only formal business accept MM as a payment method, it is not surprising it has such low usage levels in the highly informal Haitian economy.

The paper continues as follows. Section 3.2 explains how MM works and how access to the service affects development indicators. It follows with a review on the determinants of adoption, and finishes with a brief history of the mobile money services in Haiti. Section 3.3 and 3.4 outline the experimental design and data. Section 3.5 shows results on adoption and new usage of services, and section 3.6 concludes.

3.2 Mobile money: A development economics perspective

The first mobile money service in the world, M-PESA, was launched in Kenya by Safaricom in 2007. It was extremely successful; after seven years in the market 90% of households were active users (Jack, Ray, and Suri, 2013). Since then, similar services have been launched across many countries in the developing world with varying degrees of success. In this section, we explain how mobile money works and how it affects development outcomes, following a literature review on the determinants of adoption, and finish with a brief history of the mobile money market in Haiti.

3.2.1 Inner workings and its impact on households welfare

In many developing countries, fewer than fifty percent of the population has access to financial institutions. Two elements give MM a competitive edge. First, it can serve people and regions where high costs deter more traditional brick-and-mortar institutions. Managing transactions over the the cellphone network and using as agents already established business allows MM to greatly reduce the cost of offering services (Bank, 2012).[1]. Additionally, the business model of MM is different from other approaches to financial inclusion. In particular, it departs from the "credit and saving first" approach and focuses on creating a payment and transfer system, on top of which other financial services can operate (Mas and Radcliffe, 2010).

A network of mobile money agents manages all the cash entering and leaving the network. Agents tend to be small retailers that sell airtime and phones, but this is not a requirement. The growing popularity of mobile money makes becoming an agent attractive for larger retailers and even microfinance institutions. Mobile money allows any cellphone owner to deposit, transfer, and withdraw funds without the need of a bank account. To join the service, a customer must

[1] In developing countries, more households own a mobile phone than have access to electricity or improved sanitation (World Bank, 2016a)

open a mobile wallet. The whole activation process takes a few minutes and is done directly over the phone.[2] Opening an account usually requires the subscriber to provide a government-issued ID. However, several regulators have lifted this requirement and allowed special accounts, usually called mini-wallets, which have restrictions on the amounts they can manage. Basic services include holding carrying cash (with no interest), pay for services, and transfer funds between users. In more mature platforms, subscribers can perform additional transactions that include paying at stores, saving accounts that yield interest, receiving wages, and receiving government-to-person (G2P) transfers. With the usual exception of cash deposits, all transactions are subject to a fee that pays for the agents and the Mobile Network Operator (MNO) service.

Beyond providing a dramatic reduction in transaction costs and convenience, MM has positive impacts on socioeconomic indicators.[3] In particular, there are three main channels through which MM can improve poverty and vulnerability. First, MM reduces the cost of transaction and facilitates sharing resources via risk-sharing networks. In Kenya, Jack and Suri (2014) find that in the presence of shocks, mobile money users can better maintain their consumption levels as they can receive more frequent remittances from a more diverse group of senders.[4] These results are similar to Riley (2018), who finds residents in villages affected by rainfall shocks are better able to smooth their consumption in the presence of mobile money. The impacts she finds extend to non-mobile money users, as remittances received via mobile money are shared within villages, creating wider benefits to the community. Facilitating the transfer of funds has positive impacts even in the absence of

[2]Usually over Unstructured Supplementary Service Data (USSD). USSD is an interactive menu-based technology supported on most mobile devices. The main difference with SMS is that messages travel directly to the mobile network provider, creating a two-way exchange of data between users and the network. An additional advantage is that it works on any phone without installing any app or the need for mobile data.

[3]For example, in Kenya, the average transfer traveled 200 km and incurred a \$0.35 fee; a much more secure and convenient alternative of paying \$5 for the equivalent bus ride (Suri, 2017).

[4]The reach of transactions is, on average, 100 km greater for M-PESA users, and reciprocity is also greater for M-PESA users: They are 13.2 percentage points more likely to engage in at least one reciprocal transfer

shocks. In Bangladesh, (Lee et al., 2021) experimentally introduces mobile money to very poor rural households where some members have migrated to the city. Their results reveal mobile technology increased total transfers by 26 percent, with rural family members increasing their consumption by 7.5 percent and reducing their borrowing.

Second, MM can increase savings. This is true even when funds deposited on a MM account do not yield interest. Gurbuz Cuneo (2019) finds that in rural Kenya, where mobile money presents more convenient and competitive deposit opportunities than existing banks, the possibility to store cash on a mobile money account increases the probability a household saves between 16 to 22%. On an RCT setting, Dizon, Gong, and Jones (2020) randomly assigned a mobile money account, labeled for saving, to women in Kenya. They find the treatment increased savings while reducing risk-sharing. However, the reduced risk-sharing was more than compensated for by the increased savings improving women's ability to manage risk, resulting in an overall improvement in women's ability to manage shocks.

Finally, MM increase the security of transactions. Even if withdrawing funds incur a fee, in many settings this is a safer alternative to carry cash. Mas and Morawczynski (2009) finds that mobile money usage increases during periods of violence following the 2007 post-election unrest. (Economides and Jeziorski, 2017) shows that consumers deposit cash into their mobile money accounts extremely short-term storage (less than 2 hours). Exploiting a natural experiment where transaction fees were raised in Tanzania, their results reveal that costumers are willing to pay up to 1% of the total amount transaction to avoid carrying money as cash and up to 1.1% to avoid keeping money at home for an extra day. They show suggestive evidence that this willingness to pay to avoid carrying cash is related to the presence of high levels of street crime and burglaries. However, not all crime is created equal. Blumenstock et al. (2015) finds mobile money users in

Afghanistan increase their cash holdings (reduce mobile money balance) in the presence of fear of future targeted violence. In particular, a one standard deviation increase in individual forecasts of violence is associated with holding 20% less mobile money and 20% more cash.

3.2.2 What drives mobile money adoption?

Access to mobile money depends on cellphone ownership and network coverage (Baumüller, 2015). Like the adoption patterns of cellphones, MM adoption happens first among younger, more educated, and economically better-off individuals (Kirui, Okello, and Nyikal, 2012; Masocha and Dzomonda, 2018). For example, in Kenya, MM only popularized among low-income customers once the service was well-established on high-income groups (Jack and Suri, 2014). Cross-country evidence shows there are no simple metrics that predict who adopts first. Using Khan and Blumenstock (2016) data from cellphone records from Ghana, Zambia, and Pakistan shows the usage characteristics of first adopters are heterogeneous, reflecting that local conditions determine which individuals will find the MM services useful.

Overall, three elements affect how fast (or if at all) MM is adopted. The first element is that, as a network-good, the individual benefit from becoming a user depends on how widespread the adoption is in the subscriber's social circle. The literature that explores the adoption of network goods shows individuals are unlikely to internalize all the benefits their adoption generates, leading to suboptimal adoption rates (Björkegren, 2019). Several studies show that cellphone subscribers with more mobile money users in their network are more likely to adopt (Khan and Blumenstock, 2016).[5]

[5] To the best of our knowledge, we do not know of any study on the welfare costs of suboptimal mobile money adoption.

Second, the quality and density of the network of mobile money agents play a critical role in adoption. Creating and quickly expanding a large network of agents presents several challenges. The network operator must set transaction fees large enough to motivate businesses to become agents, but that do not disincentives customer adoption. At the same time, it must screen agents to guarantee they can manage the liquidity needs of customers and meet minimum standards in terms of service quality (Heyer and Mas, 2009). Available evidence shows that agent density increases accessibility and the likelihood of adoption, even in the presence of traditional financial providers (Baumüller, 2015; Kirui, Okello, and Nyikal, 2012). Few studies experimentally introduce mobile money to assess how MM agents affect adoption. The introduction of agents tends to be accompanied by additional promotional campaigns that make identification of the effect of agents difficult. On a field experiment, (Batista, Vicente et al., 2018) recruited and trained mobile money agents in half of 102 villages in rural Mozambique. Community meetings and dissemination activities accompanied the opening of the agent's store. The authors find that nearly 87% of individuals in treated villages used the service in the three years after the experiment compared to only 1.8% individuals in control villages. However, as it is common in the studies of mobile money adoption, only 53% of treated individuals became long-term customers.

Finally, the product's perceived characteristics affect the pace at which people adopt mobile money services. The impact of these elements is difficult to quantify, but evidence suggests that branding (Mas and Morawczynski, 2009), trust in the security of the system, product design usability, simple and transparent retail pricing (Karlan and Zinman, 2010; Baumüller, 2015; Athey and Imbens, 2013; GSMA, 2017; Kirui, Okello, and Nyikal, 2012)

3.2.3 A brief history of mobile money in Haiti

Haiti has one of the lowest levels of financial infrastructure. The banking system has few branches, and it is heavily concentrated in the capital's metropolitan area.[6] In 2018, 46% the population had no access to any formal financial service, and for those with access, one-fifth relied on informal providers (FinScope, 2018). This situation made the country an ideal candidate for the introduction of mobile money. Mobile money started in the country in the aftermath of the 2010 earthquake when USAID and the Bill & Melinda Gates Foundation funded the Haiti Mobile Money Initiative to speed the development of mechanisms to facilitate the delivery of cash aid. Besides transfers of international aid, the ecosystem included services to send, receive, and store money.[7]

After eight years in the market, adoption in Haiti has become stagnant with only 22% of the population ever using it.[8] This adoption rate is far from the observed in Kenya (97%), and behind products with less time in the market such as the case in Uganda (35%) and Tanzania (32%) (Suri, 2017). It is fair to mention that despite its challenges, MM in Haiti is still present, in contrasts with unsuccessful deployments. For example, between 2009 and 2015 adoption rate of MM in Nigeria was only 2.3% (Demirguc-Kunt et al., 2018)

The services currently available include buying mobile airtime, store payment, funds transfer.[9] As of 2018, the service does not include savings (money on the system does not yield interest), borrowing, lending, or receiving transfers from abroad. There are two types of subscriptions: Mini

[6] Only 2.69 bank branches per 100,000 inhabitants, the country ranks 205 out of 218 countries in financial infrastructure (Demirguc-Kunt et al., 2018). The country's banking system only has 157 branches and 53 ATMs; 66% of them in metropolitan Port-au-Prince. Cooperatives and MFIs operated a total of 270 branches across the country (Stahl and Coetzee, 2018).

[7] The mobile money industry started, and remains, with two service providers that were key players in the bulk disbursement of international aid. Tcho Tcho Mobile, a Digicel and Scotiabank product, and next T-Cash, a Voila and Unibank product. In 2015 Digicel rebranded the service as MonCash (Stahl and Coetzee, 2018).

[8] Between 2015 and 2018, the active 90-day mobile money customer base rose from 83,000 to 795,000, representing an increase of 860% (https://www.gsma.com/latinamerica/digicel-haiti-moncash/).

[9] Funds transfer includes person-to-person (P2P), as well as transfers from employers and aid from the government and international organizations.

and regular wallets. Mini wallets have lower Know-Your-Customer requirements and do not an ID to be opened, albeit with caps on the amount and number of monthly transactions. Considering that only 75% of Haitians have an official ID card, it is easy to understand that this is the most common type of wallet (FinScope, 2018).[10]

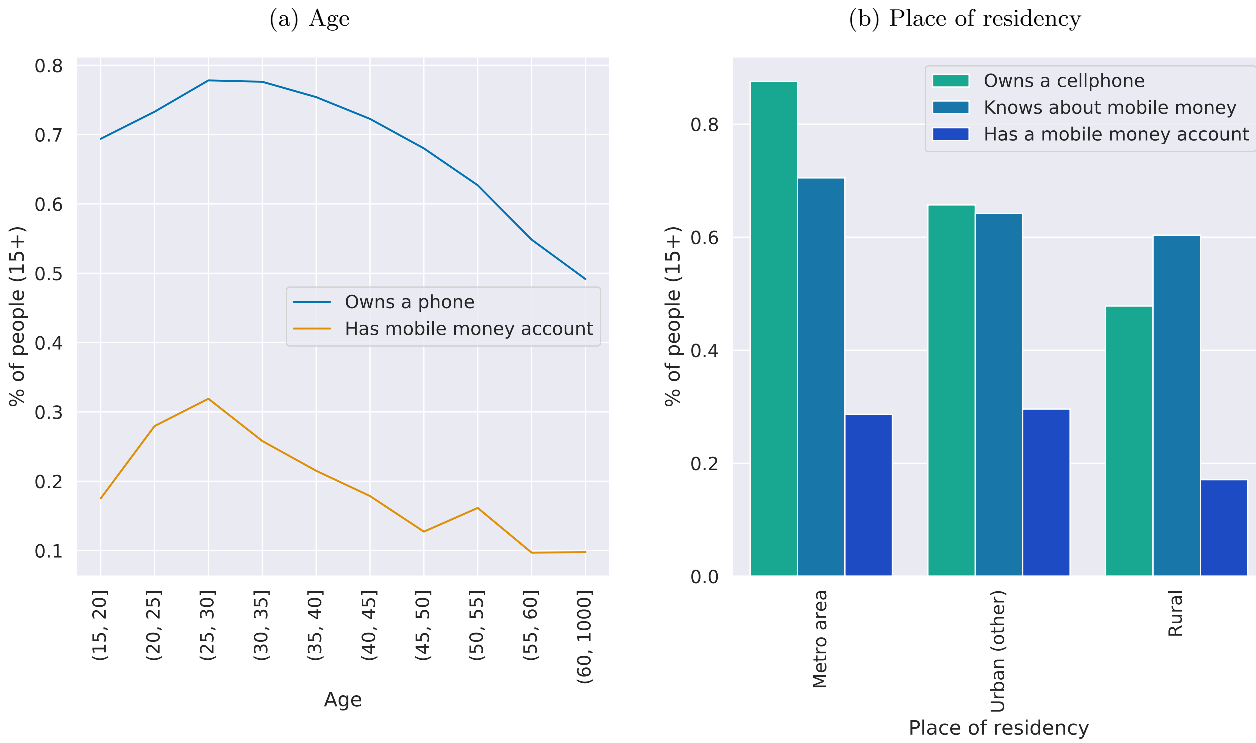
The potential market for mobile money is intrinsically linked to phone ownership since a cellphone is required to open and manage an account. Currently, 67% of Haitians have a mobile phone (60% of households) with little differences by gender (FinScope, 2018).[11] Overall, mobile money usage and cellphone ownership follow a similar pattern across age groups, with younger individuals being twice as likely to own a phone and have a mobile money account, see Figure 3.1(a). The highest heterogeneity on cellphone ownership comes from where individuals live.[12] More than 80% of the residents of the capital's metropolitan area have a cellphone, compared to 65% of residents in other urban areas and 47% in rural areas. Despite these differences, all urban areas have a similar percentage of cellphone owners that use mobile money (around 29%), with rural areas experiencing lower adoption levels (around 17%). Importantly, we do not find evidence that the different levels of MM adoption arise from lack of knowledge about the existence of the product. In particular, the percentage of respondents that know about the product consistent across geographical areas (around 65%), Figure 3.1(b).

[10] Information available does not identify the type of wallet, except when the restriction on the number of transactions is reached and the system puts a hold in the account.

[11] The gap on cellphone ownership between men and women is close to one percent, and there are no apparent significant gender differences in mobile money account ownership, see Figure 1C

[12] The metropolitan area concentrates 28% of the country's population.

Figure 3.1: Mobile money account ownership



Note: Author's calculations using Finscope 2018. Nationally representative sample of people for individuals 15 and above. Knowledge about mobile money includes cellphone owners that answered affirmatively about recognizing any of the two services in the market. Percentage of mobile money users was calculated with respect to the population currently owning a cellphone.

The agent network serves the whole country, with agents in every city and most small towns. Data shows agents are accessible to most of the population and are, in general, closer to customers than banks and other financial providers. In urban areas, the average travel time to an agent is almost half the time needed to visit a bank branch, and a third of the time to visit other financial institutions (Table 3.1). Most people are aware of the location of mobile money agents. For example, only 3.5% of respondents do not know where to find a mobile money agent. This percentage is well below the 4% of respondents who are unaware of the location of the nearest bank branch and the 65% who do not know where to find a micro-financial institution (FinScope, 2018).

Table 3.1: Average time from home (in minutes)

	Work	Store	Bank	Other financial	Mobile money agent
Metro area	29	8	16	24	9
Urban (other)	35	18	54	47	23
Rural	52	46	111	95	72

Note: Author's calculation using Finescope 2018. Other financial services includes Credit Unions, MFI, and Mutuelles. Time to mobile money agents only considers those that recognize any of the two brand names in the market.

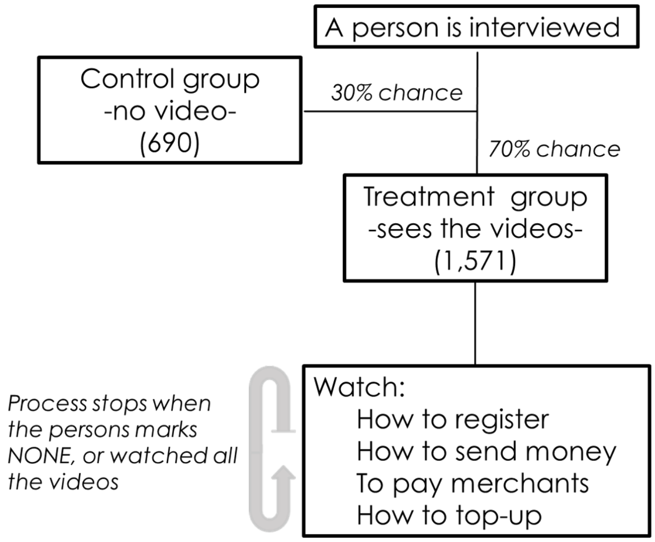
What are the remaining challenges for greater mobile money adoption? New subscribers can come from two sources. First, people without a cellphone represent 40% of the Haitian population and concentrates those most likely to have little access to alternative financial services. Second, individuals who own a mobile phone but have not yet open a MM account represent a more immediate target for growth. For those who own a cellphone, the results show there are high levels of brand recognition access to the network of agents. Current evidence on the reasons deterring mobile subscribers from using mobile money shows the most commonly cited reason is lack of knowledge about how to use the products (25%), followed by lack of interest (13%), and lack of trust in the platform (9%). Interestingly, very few people report the difficulty of finding providers or the cost of the services as a reason not to use it (3 and 0.7%m respectively). These results are in line with industry reports show that customers find the USSD menus difficult to use.

3.3 Experimental design

The research is part of a partnership with the largest cellphone and mobile money provider in Haiti. It combines a survey with a randomization design. The survey consisted of three stages. During the first stage, enumerators confirmed that participants had a cellphone. Only people with a number from the participating cellphone can have a MM account and were eligible to participate. There are no hard technical limitations that impose this restriction, but interoperability is not yet available in the country. As part of the informed consent process, participants provided access to their cellphone and mobile money records. In the second stage, enumerators asked questions about the participant's knowledge on how to use mobile money, perceptions about its security, and if they had asked other people to make mobile money transactions for them. The piloting of the survey revealed that including questions about income or food security reduced participation rates, so they were not included.

In the final stage, the survey introduced a randomization design where each participant had a 70% probability of seeing a series of videos that explained how to (i) register, (ii) transfer money between account holders, (iii) pay for services, and (iv) buy additional airtime balance. Subscribers were free to choose the order of the videos, if any, using the tablets the enumerators had available. The original experimental design envisioned sharing the videos with participants who had a smartphone; however, participant's concerns about data usage were high, and this part of the intervention was not implemented. Figure 3.2 shows the experimental design on detail.

Figure 3.2: Experimental design

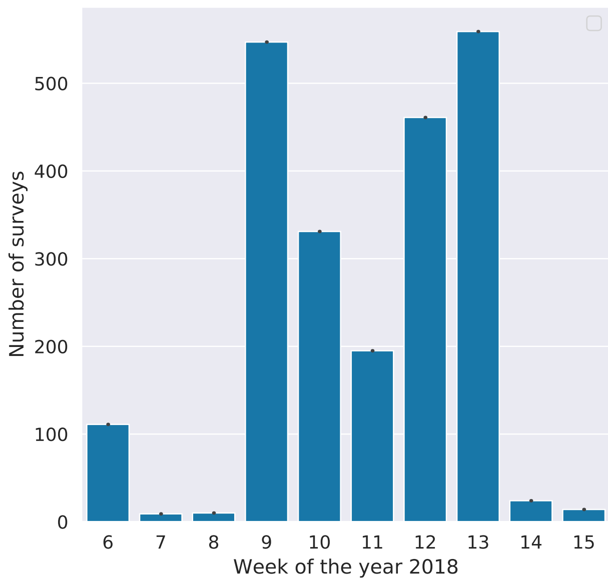


Note: Participants were assigned randomly into the control or treatment groups. Their treatment status was not revealed to them, or the enumerators, until the end of the survey when the option to watch the videos was offered.

3.4 Data

The empirical work uses two data sources. First data, we use the information from the survey, collected between February and April 2019. The questionnaire included basic demographic information and perceptions about how mobile money worked. Enumerators were located in central intersections in two of the main cities outside the capital. For security considerations, data collection stopped early during several days, with most of the data collected in the last week of February and March, see Figure 3.3. Our final sample consists of a total of 2,261 respondents. A total of 301 were not eligible for the mobile money product as their cellphone number is part of a different network, and therefore they were excluded from the study. The second source of data comes from administrative cellphone and mobile money records provided by the MNO. The data identify which participants were mobile money users at the time of the experiment, the adoption patterns after the intervention, and some basic descriptive statistics about the mobile money transactions of all participants in the experiment. The data cover all transactions between December 2018 and June 2019.

Figure 3.3: Total surveys per week



Note: Total number of surveys per week. Data collection stopped in several occasions due to civil unrest.

Table 3.2 shows the individual characteristics of participants in experiment. In the final sample, 59% of participants were men, with an average age of 30 years old. In terms of mobile money usage, 42% of the sample had an account at the time of the survey. This usage level, almost one-third above the average in urban areas outside of the capital, is related to most participants being between 25 and 30 years old; an age where MM usage level spikes. The survey lacks information about income levels, but the high percentage of smartphone users and higher than average education levels suggest participants have better economic indicators than the rest of the population. We find statistically significant differences between those with and with a MM account, but they are in line with previous results from the literature. In particular, MM users tend to be younger, with higher education levels and, reflecting higher purchasing power, are more likely to own a smartphone.

Table 3.2: Individual Characteristics

		Treatment status			Mobile money user		
		Control	Treatment	p-value	Yes	No	p-value
Demographics	Male	59.13	59.07		58.6	59.4	
		(0.5)	(0.5)		(0.49)	(0.49)	
	Age	29.33	29.68		29.0	30.0	**
		(9.0)	(9.4)		(8.49)	(9.78)	
	Works	45.8	43.73		43.3	45.1	
		(0.5)	(0.5)		(0.5)	(0.5)	
	Smart phone	60.87	57.73		63.9	54.9	***
		(0.5)	(0.5)		(0.48)	(0.5)	
Education	None	3.33	5.22	**	2.5	6.2	***
		(0.2)	(0.2)		(0.16)	(0.24)	
	Primary	28.55	29.03		26.8	30.4	*
		(0.5)	(0.5)		(0.44)	(0.46)	
	Secondary	17.68	18.65		19.3	17.7	
		(0.4)	(0.4)		(0.4)	(0.38)	
	Technical	50.43	47.1		51.3	45.8	***
		(0.5)	(0.5)		(0.5)	(0.5)	
Knows mobile money		87.54	88.42		100.0	79.6	***
		(0.3)	(0.3)		(0.0)	(0.4)	
Uses mobile money		42.75	41.5		100.0	0.0	***
		(0.5)	(0.5)		(0.0)	(0.0)	
MM is safe	Yes	64.29	61.31		70.9	55.3	
		(0.5)	(0.5)		(0.45)	(0.5)	
	Do not Know	34.84	37.39		27.8	43.7	
		(0.5)	(0.5)		(0.45)	(0.5)	
Asked others for MM transactions	Transfer (sent)	26.38	26.99		35.2	25.3	***
		(0.4)	(0.4)		(0.48)	(0.44)	
	Transfer (received)	12.46	10.5		14.3	10.9	***
		(0.3)	(0.3)		(0.35)	(0.31)	
	Recharge	1.88	2.61		3.3	2.2	**
		(0.1)	(0.2)		(0.18)	(0.15)	
	Bill payment	0.29	0.32		0.2	0.5	
		(0.1)	(0.1)		(0.05)	(0.07)	

Note: Author's calculations using own survey.

A total of 1,571 people, 70% of participants, were chosen to be part of the treatment group and had the opportunity to select what videos to watch. A large percentage of treated individuals, around 44%, passed the opportunity to watch any video, with no significant differences between those with and without a mobile money account. There are no apparent differences in the percentage of people that watched each video, with 53% of participants watching one video, 13% two, and 33% watching all of the videos available. Figure 4C shows the first video participants decided to watch. Assuming the first video correlates with how the appeal of the service, people without an account were interested on understanding how to transfer funds (20%), followed by the instructions on how to register (14%), pay at stores (10%), and buy mobile airtime (10%).

3.5 Results

Results are divided into three sections. First, we rely on the survey responses to provide an overview of how people use MM and the perceptions about the platform's security. Second, we study the impact of the information campaign on the probability that a person opens a mobile money account and the services used. Finally, for the survey participants who had an account at the time of the survey, we verify if the videos increase the probability that people use new services, and increase the number of transactions in the platform.

3.5.1 Mobile money usage

Results from the survey show most people know about the existence of MM. Despite low adoption levels, around 80% of the people without an account know about the service. Moreover, the system is considered to be quite secure, with both subscribers (71%) and non-subscribers (50%) considering it a safe alternative to transfer funds. These percentages are well above the same perception for banks, at only 36%.

Traditionally, a person is only considered a MM user if he has opened account in the past. However, the survey shows that 32% of subscribers without an account use the service indirectly by asking subscribers to make transactions for them. The most common type of transaction people other to do for them is the transfer of funds.[13] One-quarter of the respondents without an account have asked a person to send funds for them, while ten percent have asked help receiving funds. Interestingly, despite being the most common transaction in the MM platform, we do not see that people ask others to buy airtime for them.[14] Considering that opening an account has no cost and can be done directly from a phone, it is puzzling that a large percentage of people ask others for MM transactions but does not join the service.

The administrative data shows that in an average week, around 40% of accounts have some level of activity, with no apparent bunching around periods with more economic activity, see Panel (a) Figure 3.4. There is high heterogeneity in the usage levels. We find that 14% of accounts show no activity during the first two months of the year. For those accounts with transactions activity seems sporadic, with only one (23%) or two weeks (16%) with any activity (Panel (b) Figure 3.4). Sporadic usage of the services and the fact that the accounts yield no interest make that there are little incentives to leave money in the accounts. We calculate the end of day balance for each account during February 2019, finding that money is quickly spent, with the medium subscribers not having enough balance to buy the cheapest airtime recharge available.[15]. To provide context of the amounts MM manages, we also provide the end-of-day balance for the 90th percentile. We can see that this amount is never above USD$3, see Panel (c) 3.4.

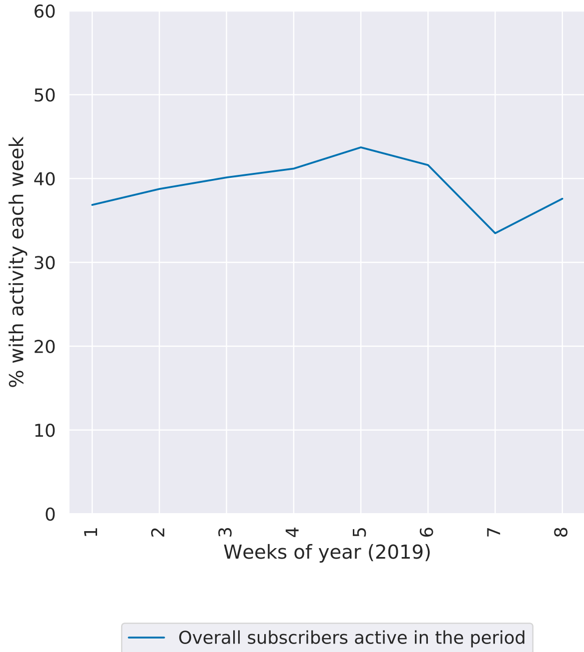
[13]A similar behavior exists for those with an account, but it is not possible to identify when requests occurred relative to when a subscriber opened his account.

[14]This is probably caused by the existence of a service to send airtime that does not require the usage of MM.

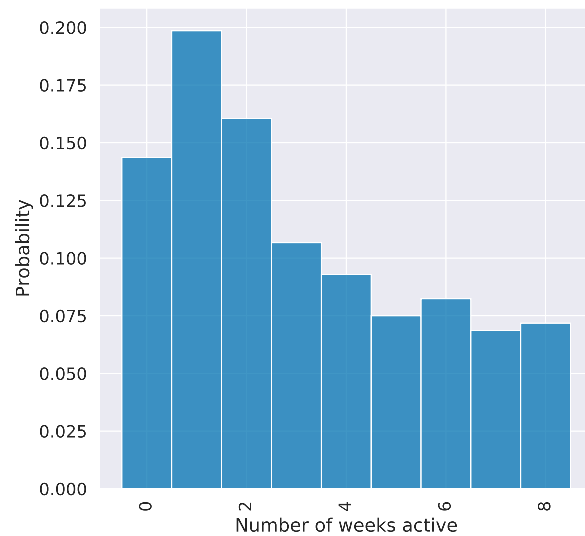
[15]The lowest airtime purchase is USD$0.15, with medium balances below that amount.

Figure 3.4: Mobile money usage: Pre-intervention

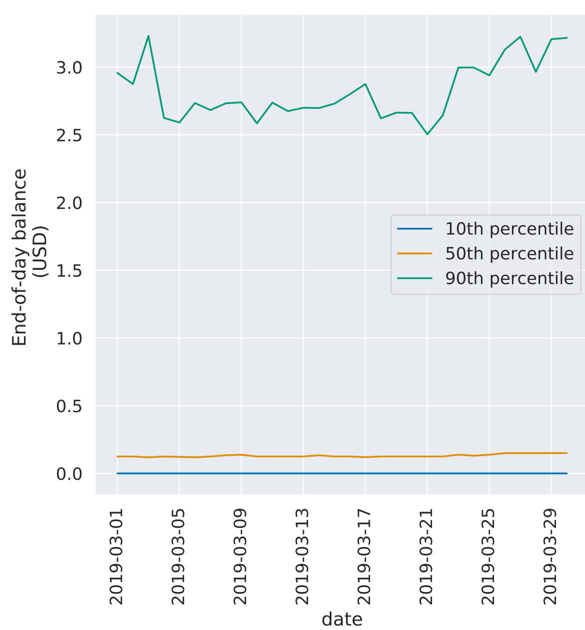
(a) MM users active per week



(b) Number of active weeks



(c) End of day balance



Note: Includes information for mobile money transactions eight weeks prior to the survey. Subscribers with an account at the time of the survey only.

3.5.2 Impact of informational videos on adoption

This section studies how receiving information on the different MM services increases the adoption. Data limitations make that we only observe transactions in the MM network but not action of opening an account. Instead, we consider a person became a MM user if we detect any MM transaction after the survey. There is the possibility that a person opened an account but made no transactions, in which case we would not consider that person 'adopted' the service. We do not believe this is problematic as our interest is to measure the capacity of the videos to create active customers. Additionally, MM logs only covers transactions until the end of June 2019. Since the survey took several weeks to complete, in order to observe all participants for the same amount of time we restrict out adoption measure to seven weeks after the survey. Given the nature of the treatment, we expect impacts to take place soon after the intervention. However, this restricts our capacity to determined long-term adoption and usage patterns.

To measure the effect of the intervention on adoption we estimate equation 3.1. $Adopt_i$ is a dummy variable that equals one if a person became a mobile money user at any point during the seven weeks following the survey; $Treated_i$ indicates whether or not a survey participant was selected to watch the information videos, and λ_{week} control for week-specific fix effects. We also include a vector X of baseline controls including age, employment status, education level, and usage of a smartphone.

$$Adopt_i = \beta_0 + \beta_1 Treated_i + X^{'}\eta + \lambda_{week} + u_{i,week} \tag{3.1}$$

Table 3.3 shows how being eligible to watch any videos increases the probability of opening an account by 5.6%. In columns (2) to (4), we estimate equation 3.1 using as outcome the usage of any of the three services available. Buying mobile airtime concentrates most of the demand from new subscribers, without significant usage of other services such as transfer funds or pay at stores. A key question is the capacity of the intervention to create long-term adoption of the service. Table 1C shows new subscribers in the treatment group make their MM first transaction in the two weeks following the intervention. However, the data available only allow us to follow the usage of the service for seven weeks after the survey. With the data at hand, we explore for for the number of weeks with activity. We see that most of the new users made transactions on at least two different week. A result that indicates, even if only partially, that adoptions goes beyond a single test of the services MM offers.

Putting in perspective the effectiveness of the intervention is difficult. Private companies test their campaigns constantly but the results are not publicly available. However, to provide a benchmark of the impacts, in the case the 5.6% adoption rate holds among the general population of cellphone subscribers without MM accounts, a similar large-scale campaign would achieve the equivalent to the growth rate of new subscribers in the previous year and a half. A key difference between our videos and traditional campaigns is that we focus on how to use the services, instead of just promoting their existence. Information from the survey shows that traditional campaigns been successful at creating product awareness, but still lack of knowledge about how to access and use the products remain a key concern among the industry reports.

Zooming into product usage by new subscribers reveals that a large percentage (34%) did not use any service at all during the seven weeks after creating their accounts. Despite incurring on a transaction fee, people in this groups limited their transactions to deposit and withdraw money

Table 3.3: Effect on adoption of mobile money

	Open account	Service used		
		Mobile airtime	Funds transfer	Store payment
	(1)	(2)	(3)	(4)
Treatment	0.056***	0.030***	0.015	0.005
	(0.019)	(0.009)	(0.012)	(0.005)
Controls	Yes	Yes	Yes	Yes
Observations	1,314	1,314	1,314	1,314
R^2	0.021	0.034	0.010	0.008

*p<0.1; **p<0.05; ***p<0.01

Note: Includes only survey participants without an account at the time of the survey. Control variables include age, gender, working on the previous week, usage of smart phone, education level, and a dummy for the week the survey took place.

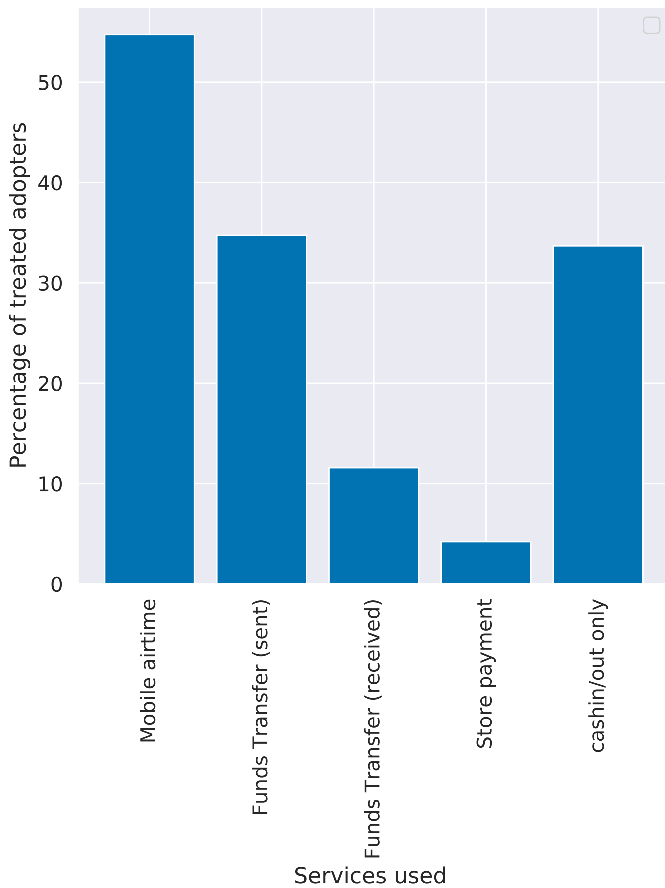
from their recently created accounts, see Figure 3.5. Whether this behavior is caused by people not being able to use any services or as a way to avoid carrying cash is difficult to determined. MM logs show that all but one of the people in this group made deposit and withdraw transactions over several weeks. This supports that people are relying on MM to avoid carrying cash, as has been observed in other settings (Economides and Jeziorski, 2017). However, it remains puzzling that we do not observe the usage of any other service.

Considering that an important percentage (40%) of participants in the treatment group decided not to watch any of the videos, we extend equation 3.1 to control for the videos each person chose to watch.[16] There is no apparent relation between the videos a person watched and usage the usage of the specific service. Table 3.4 shows evidence the aggregate adoption result combines both a higher salience of the existence of MM, via participation in the survey, and an information effect caused by the videos. This raises doubts about how difficult people actually find navigating the MM platforms. In the case of people in the treatment group who did not watch any video, they

[16]A person could decide to watch all of the videos.

did not gain any additional information but were still able to open an account and start using the MM services. On the other hand, people who did received information ended up using MM, almost exclusively, to buy mobile airtime.

Figure 3.5: Products used: New adopters only



Note: Includes treated individuals only

Table 3.4: Effect on adoption of mobile money: Videos watched

Watched how to:	Open account	Service used		
		Mobile airtime	Funds transfer	Store payment
	(1)	(2)	(3)	(4)
None	0.041**	0.029***	0.012	0.005
	(0.019)	(0.009)	(0.012)	(0.005)
Register	-0.056*	-0.044***	-0.010	0.003
	(0.033)	(0.011)	(0.022)	(0.012)
Transfer funds	-0.003	0.004	-0.005	-0.001
	(0.025)	(0.010)	(0.015)	(0.001)
Pay at stores	-0.018	0.026*	-0.014	0.000
	(0.026)	(0.014)	(0.019)	(0.001)
Buy mobile airtime	0.069**	0.006	0.042**	-0.002
	(0.028)	(0.012)	(0.020)	(0.001)
Controls	Yes	Yes	Yes	Yes
Observations	1,314	1,314	1,314	1,314
R^2	0.022	0.039	0.014	0.010

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Includes only survey participants without an account at the time of the survey. Control variables include age, gender, working on the previous week, usage of smart phone, education level, and a dummy for the week the survey took place.

3.5.3 Usage of new products

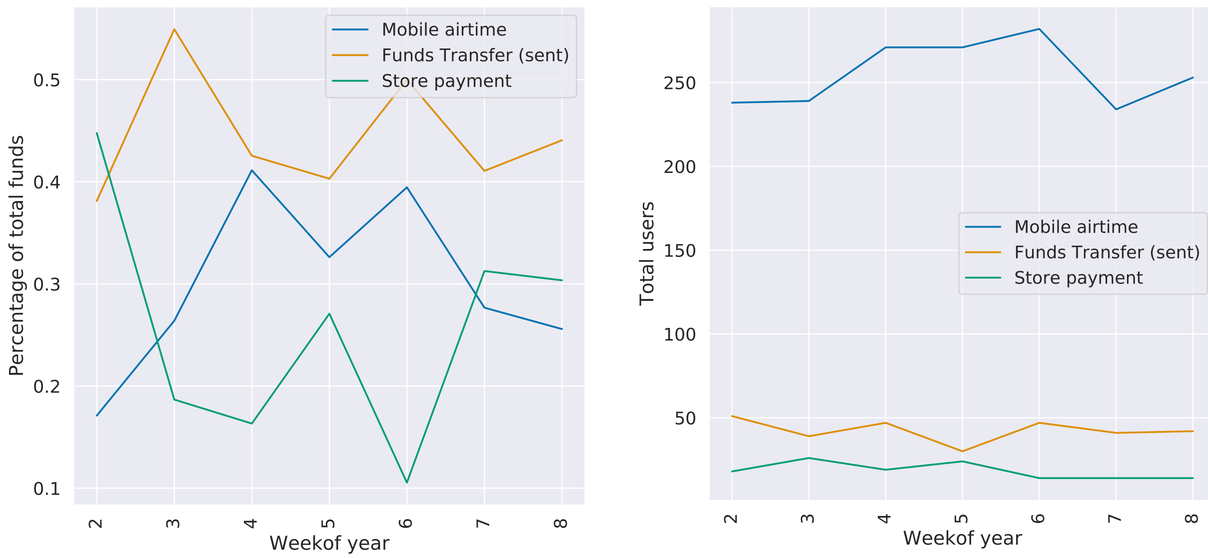
The administrative data shows the subscriber base and the total amount of funds each service concentrates greatly differs. Mobile airtime concentrates most of the subscribers, with more than 75% of subscribers at the time of the survey using the service, a level that is consistent across time, see panel (a) Figure 3.6. However, despite a much small user base, of around 52% in the case of fund transfer and only 9% in the case of store payment, these two services concentrate the lion share of the funds that go through the MM network, see panel (a) on Figure 3.6.[17]

[17] This is in line with activity patterns from the whole network. It is possible subscribers have used the a particular service before before the three-month period. However, we cannot observe these transactions.

Figure 3.6: Mobile money usage: Pre-intervention

(a) Share of total funds (weekly)

(b) Number users per service



Note: Include information for mobile money users prior to the survey.

The differences in the usage levels of each service provide the potential market each product has to grow its customer base. Given the large size of the transactions in the fund transfer and store payment services, increasing the number of people making these transactions can greatly increase the total amount of money the MM platform manages.

Table 3.5 shows the percentage of customers that have not used each product in the past for both treatment and control groups. As a baseline, we use product usage in the three months prior to the survey. We consider a person used a new service if we observe if following the intervention, we observe a transaction on a service not used before.

Table 3.5: Customers at the time of the survey not using each service (%)

	Mobile airtime	Funds transfer	Store payment
Control	25.0	48.3	92.1
Treated	24.5	46.3	89.0

Note: Includes mobile money subscribers at the time of the survey.

For each service, we estimate equation 3.1 with the sample of subscribers that have not used a specific product in the past. We do not find the possibility of watch the videos had any impact on the probability of people using a new service.[18] Two reasons, not completely independent, can explain these findings. First, subscribers are already using MM and understand the system. Therefore, providing additional information does not change their decision on what products to use. Second, the usage of the fund transfer and store payment services depend of particular circumstances. Fund transfers requires a person has the need to send money to a third party in the country. In the case of the store payment service, it can only be used in formal establishments; a limiting factor in the highly informal Haitian economy.

Table 3.6: Effect on adoption of mobile money

	Mobile airtime	Funds transfer	Store payment
	(1)	(2)	(3)
Treatment	-0.033	0.020	0.001
	(0.054)	(0.034)	(0.010)
Controls	Yes	Yes	Yes
Observations	287	524	865
R^2	0.059	0.038	0.018

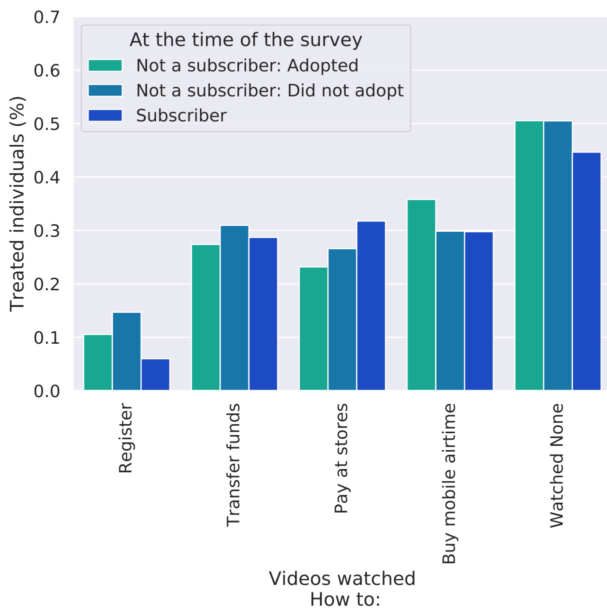
*p<0.1; **p<0.05; ***p<0.01

Note: Includes mobile money subscribers at the time of the survey that have not used the specific service.

[18] We also test for changes in the end-of-day balance and number of transactions. We do not find any impact from the intervention.

Lack of usage of new services by subscribers does not mean lack of interest. Figure 3.7 shows that MM users at the time of the survey present higher levels of interest in the videos. In particular, subscribers at the time of the survey were more likely to watch at least one video. Moreover, they show the highest interest in watching the video explaining how to use the store payment service.

Figure 3.7: Videos seen by adoption category



Note: Includes treated individuals only

3.6 Conclusion

Haiti has one of the lowest access to financial services globally, making it the perfect setting for a thriving MM ecosystem. However, after more than eight years in the market, MM has struggled to increase the number of people using it. At the time of this study, only 22% of cellphone users have an account despite the fact that joining has no cost and can be done directly over the phone. With widespread access to cellphone services and little competition from traditional financial providers, the stagnation of the MM ecosystem in the country provides an important case of study to understand what makes mobile money, and alternative financial service, successful.

Industry reports cite a lack of understanding about how to use the service, including difficulties with the codes to navigate the menus on feature phones, as the main constraint for further growth of the MM platform. However, there is no evidence about what works for teaching people how to use the service. We fill this gap by implementing an experimental intervention to assess if short videos, explaining how to use each of the services available, can attract new subscribers and nudge current users to try additional services. To properly measure the impact of the videos, we conducted a survey that provides access to information videos to a subset of participants. By matching survey responses with administrative logs from the MM money company, we are able to measure if the intervention led to opening an account (adoption) and the usage of additional services. These videos explain each of the services currently available: (i) buying mobile airtime, (ii) funds transfer, and (iii) store payment -paying at stores and online.

The survey component of the experiment reveals that having an account is not a good measure of MM usage. Even if opening and maintaining an account is free, 32% of people without an account still using MM services by asking subscribers to do transactions, mostly funds transfer, for them. This result can point to both difficulties opening an account as well as the current services not providing enough incentives to have a personal account. There is not enough information to identify the weight of each of these factors. However, the scope of the three services available can be somehow limited for many subscribers. In the case of using MM to buy airtime, this requires to have a balance above what the medium subscriber has on a typical day. For these subscribers, recharging using MM would require adding funds to their accounts first, making recharging with MM as practical as finding a regular airtime vendor in the street, and without the need to lose liquidity by maintaining a positive MM balance. The service of paying at stores using MM compounds the low balance problem with the fact that only formal stores accept MM payments, making it of little use in the highly informal Haitian economy.

From the randomized component of the survey, results show that being eligible to watch the videos increases the probability that a person starts using MM on 5.4%, with most new subscribers using MM to buy mobile airtime. However, we find evidence that the information component of the videos is not the main driver of the results. The intervention had a large percentage of non-compilers in the treatment group, with 40% of the eligible population choosing not to watch any videos, but still presenting a significant adoption rate. Moreover, there is no apparent direct link between the specific videos a person watched and the service used. Our results show that most of the new subscribers, independently of the video watched, only used MM to buy mobile airtime; a transaction pattern in line with the general usage patterns of MM. It is not possible to asses long-term adoptions due to data limitations. However, we find some evidence of repeated use among new subscribers. To benchmark the results, they are equivalent to the growth rate of new subscribers in the previous year and a half.

For participants who had a MM account at the time of the survey, we do not find any impact on the usage of new service or the number of transactions. We argue that, in the case of active users, they have few opportunities to use services other than buying mobile airtime. The most striking example is the possibility to use MM to pay at stores. This services has the largest room to grow since it is only used by 9% of MM users. However, only formal businesses accept MM as a form of payment, limiting how many people can use it.

From the previous results, two recommendations arise. First, people can learn on their own how to use the service; a result that the large adoption levels by people who watched no video seem to support. Despite the common assumption that lack of knowledge about how to navigate the in-phone menus constrains usage, the experiment shows evidence that not watching any video did not seem to constrain adoption. Even if we cannot determine if people learned by asking others, it seems that the menus' complexity is not a big factor in limiting usage. Second, MM wider success depends on launching additional services that promote individual accounts and encourage maintaining a large balance so MM can be used when the right situation arises. Launching additional services in a market such as Haiti is challenging; the country has high poverty levels, political instability and is very vulnerable to shocks (both economic and natural). With that said, experience from African countries signals the markets with the highest levels of MM users tend to be the ones with the most diversified set of services, including receiving international remittances, savings, digital credit, and paying informal vendors.

Bibliography

Agüero, Aileen, Harsha de Silva, and Juhee Kang. 2011. “Bottom of the Pyramid Expenditure Patterns on Mobile Services in Selected Emerging Asian Countries.” *Information Technologies International Development* 7 (3):19–32. URL https://itidjournal.org/index.php/itid/article/view/756/316.

Aiken, Emily L, Guadalupe Bedoya, Aidan Coville, and Joshua E Blumenstock. 2020. “Targeting Humanitarian Aid with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan.” *Mimeo* .

Aker, Jenny C and Joshua E Blumenstock. 2015. “The economic impacts of new technologies in Africa.” *The Oxford handbook of Africa and economics* 2:354–371.

Aker, Jenny C. and Isaac M. Mbiti. 2010. “Mobile Phones and Economic Development in Africa.” *Journal of Economic Perspectives* 24 (3):207–32. URL https://www.aeaweb.org/articles?id=10.1257/jep.24.3.207.

Al-Zuabi, Ibrahim Mousa, Assef Jafar, and Kadan Aljoumaa. 2019. “Predicting customer’s gender and age depending on mobile phone data.” *Journal of Big Data* 6 (1):18.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. 2012. “Targeting the poor: evidence from a field experiment in Indonesia.” *American Economic Review* 102 (4):1206–40.

Aron, Janine et al. 2017. “‘Leapfrogging’: A survey of the nature and economic implications of mobile money.” Tech. rep., Centre for the Study of African Economies, University of Oxford.

Athey, Susan and Guido W Imbens. 2013. “The power of social networks to drive mobile money adoption.” Tech. rep., CGAP.

———. 2018. “Design-based analysis in difference-in-differences settings with staggered adoption.” Tech. rep., National Bureau of Economic Research.

Attanasio, Orazio and Christine Frayne. 2006. “Do the poor pay more?” *MIMEO* .

Ausubel, Lawrence M. 1991. “The Failure of Competition in the Credit Card Market.” *The American Economic Review* 81 (1):50–81. URL http://www.jstor.org/stable/2006788.

Bank, World. 2012. *Information and communications for development 2012: Maximizing mobile*. The World Bank.

Batista, Catia, Pedro C Vicente et al. 2018. "Is mobile money changing rural Africa? Evidence from a field experiment." Tech. rep., Universidade Nova de Lisboa, Faculdade de Economia, NOVAFRICA.

Baumüller, Heike. 2015. "Assessing the role of mobile phones in offering price information and market linkages: The case of M-Farm in Kenya." *The Electronic Journal of Information Systems in Developing Countries* 68 (1):1–16.

Beine, Michel, Luisito Bertinelli, Rana Cömertpay, Anastasia Litina, Jean-François Maystadt, and Benteng Zou. 2019. "Refugee Mobility: Evidence from Phone Data in Turkey." In *Guide to Mobile Data Analytics in Refugee Scenarios*. Springer, 433–449.

Bharadwaj, Prashant, William Jack, and Tavneet Suri. 2019. "Fintech and Household Resilience to Shocks: Evidence from Digital Loans in Kenya." Working Paper 25604, National Bureau of Economic Research. URL http://www.nber.org/papers/w25604.

Bhatia, Pooja. 2010. "Haiti Puts Its Faith in the Lottery." *Mimeo* URL http://www.poojabhatia.com/DreamTicket.pdf.

Björkegren, Daniel. 2019. "The adoption of network goods: Evidence from the spread of mobile phones in Rwanda." *The Review of Economic Studies* 86 (3):1033–1060.

Björkegren, Daniel, Joshua E Blumenstock, and Samsun Knight. 2020. "Manipulation-proof machine learning." *arXiv preprint arXiv:2004.03865* .

Björkegren, Daniel, Joshua E. Blumenstock, and Samsun Knight. 2020. "Manipulation-Proof Machine Learning."

Björkegren, Daniel and Darrell Grissen. 2018. "The Potential of Digital Credit to Bank the Poor." *AEA Papers and Proceedings* 108:68–71. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20181032.

———. 2019. "Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment." *The World Bank Economic Review* URL https://doi.org/10.1093/wber/lhz006. Lhz006.

Blumenstock, J. E. 2016a. "Fighting poverty with data." *Science* 353 (6301):753–754. URL http://www.sciencemag.org/cgi/doi/10.1126/science.aah5217.

Blumenstock, Joshua. 2018a. "Don't forget people in the use of big data for development."

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting poverty and wealth from mobile phone metadata." *Science* 350 (6264):1073–1076. URL https://science.sciencemag.org/content/350/6264/1073.

Blumenstock, Joshua E. 2018b. "Estimating economic characteristics with phone data." In *AEA papers and proceedings*, vol. 108. 72–76.

Blumenstock, Joshua E, Michael Callen, Tarek Ghani, and Lucas Koepke. 2015. "Promises and pitfalls of mobile money in Afghanistan: evidence from a randomized control trial." In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*. 1–10.

Blumenstock, Joshua E., Nathan Eagle, and Marcel Fafchamps. 2016. "Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters." *Journal of Development Economics* 120:157 – 181. URL http://www.sciencedirect.com/science/article/pii/S0304387816000109.

Blumenstock, Joshua Evan. 2016b. "Fighting poverty with data." *Science* 353 (6301):753–754. URL https://science.sciencemag.org/content/353/6301/753.

Bond, Philip, David K Musto, and Bilge Yilmaz. 2009. "Predatory mortgage lending." *Journal of Financial Economics* 94 (3):412–427.

Brown, Caitlin, Martin Ravallion, and Dominique Van de Walle. 2016. *A poor means test? Econometric targeting in Africa*. The World Bank.

———. 2018. "A poor means test? Econometric targeting in Africa." *Journal of Development Economics* 134:109–124.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2018. "Fintech, regulatory arbitrage, and the rise of shadow banks." *Journal of Financial Economics* 130 (3):453 – 483. URL http://www.sciencedirect.com/science/article/pii/S0304405X1830237X.

Bulando, Alfredo, Michael Kuhn, and Silvia Prina. 2020. "Digital Credit Delivery Speed and Repayment Rates." Working paper, University of Oregon. URL https://static1.squarespace.com/static/57869cbb2e69cf3d5af77419/t/5f4fb8901c1cad15afc6288c/1599060122571/BKP_2020_09_01.pdf.

Callaway, Brantly and Pedro HC Sant'Anna. 2018. "Difference-in-differences with multiple time periods and an application on the minimum wage and employment." *arXiv preprint arXiv:1803.09015* .

Carlson, Stacy Stacy Lynn. 2018. *Essays in financial innovation and development*. Ph.D. thesis, Massachusetts Institute of Technology.

Carrell, Scott and Jonathan Zinman. 2014. "In harm's way? Payday loan access and military personnel performance." *The Review of Financial Studies* 27 (9):2805–2840.

Cattaneo, Matias D, Ann Arbor, Michael Jansson, and Ann Arbor. 2017. "rddensity : Manipulation Testing Based on Density Discontinuity." *The Stata Journal* :1–24.

Clarke, Damian and Kathya Schythe. 2020. "Implementing the Panel Event Study." *MIMEO* .

Coady, David, Margaret Grosh, and John Hoddinott. 2004. "Targeting outcomes redux." *The World Bank Research Observer* 19 (1):61–85.

Cook, Tamara and Claudia McKay. 2015. "Top 10 things to know about M-Shwari." *CGAP Blog* .

De Montjoye, Yves-Alexandre, Luc Rocher, and Alex Sandy Pentland. 2016. "bandicoot: A python toolbox for mobile phone metadata." *The Journal of Machine Learning Research* 17 (1):6100–6104.

Decuyper, Adeline, Alex Rutherford, Amit Wadhwa, Jean-Martin Bauer, Gautier Krings, Thoralf Gutierrez, Vincent D. Blondel, and Miguel A. Luengo-Oroz. 2014. "Estimating Food Consumption and Poverty Indices with Mobile Phone Data." *Mimeo* :1–13URL http://arxiv.org/abs/1412.2595.

Del Ninno, Carlo and Bradford Mills. 2015. *Safety nets in Africa: Effective mechanisms to reach the poor and most vulnerable*. Washington, DC: World Bank; and Agence Française de Développement.

Demirguc-Kunt, Asli, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess. 2018. *The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution*. The World Bank.

Dizon, Felipe, Erick Gong, and Kelly Jones. 2020. "The Effect of Promoting Savings on Informal Risk Sharing Experimental Evidence from Vulnerable Women in Kenya." *Journal of Human Resources* 55 (3):963–998.

Dizon, Felipe and Travis J Lybbert. 2021. "Leveraging the lottery for financial inclusion: Lotto-linked savings accounts in Haiti." *Economic Development and Cultural Change* 69 (4):000–000.

Economides, Nicholas and Przemyslaw Jeziorski. 2017. "Mobile money in Tanzania." *Marketing Science* 36 (6):815–837.

Fafchamps, Marcel. 1992. "Solidarity networks in preindustrial societies: Rational peasants with a moral economy." *Economic development and cultural change* 41 (1):147–174.

FinScope. 2018. "FinScope: Consumer Survey Highlights, Haiti." https://http://finmark.org.za/wp-content/uploads/2019/04/Haiti_French_17-04-2019.pdf. Accessed: 2020-09-21.

Francis, Eilin, Joshua Blumenstock, and Jonathan Robinson. 2017. "Digital credit: A snapshot of the current landscape and open research questions." *CEGA White Paper* :1739–1776.

Frandsen, Brigham R. 2016. "Party Bias in Union Representation Elections : Testing for Manipulation in the Regression Discontinuity Design When the Running Variable is Discrete." In *Advances in Econometrics vol 38. Regression Discontinuity Designs: Theory and Applications*. 1–42.

Frias-Martinez, Vanessa, Enrique Frias-Martinez, and Nuria Oliver. 2010. "A Gender-Centric Analysis of Calling Behavior in a Developing Economy Using Call Detail Records." In *AAAI spring symposium: artificial intelligence for development*, vol. 5. 1–6.

Frias-Martinez, Vanessa, Cristina Soguero-Ruiz, Enrique Frias-Martinez, and Malvina Josephidou. 2013. "Forecasting socioeconomic trends with cell phone records." In *Proceedings of the 3rd ACM Symposium on Computing for Development*. 1–10.

Frias-Martinez, Vanessa, Victor Soto, Jesus Virseda, and Enrique Frias-Martinez. 2012. "Computing cost-effective census maps from cell phone traces." In *Workshop on pervasive urban applications*. 1–8.

Gazeaud, Jules. 2020. "Proxy Means Testing vulnerability to measurement errors?" *The Journal of Development Studies* 56 (11):2113–2133.

Gelman, Michael, Shachar Kariv, Matthew D. Shapiro, Dan Silverman, and Steven Tadelis. 2014. "Harnessing naturally occurring data to measure the response of spending to income." *Science* 345 (6193):212–215. URL https://science.sciencemag.org/content/345/6193/212.

Gething, Peter W. and Andrew J. Tatem. 2011. "Can Mobile Phone Data Improve Emergency Response to Natural Disasters?" *PLOS Medicine* 8 (8):1–2. URL https://doi.org/10.1371/journal.pmed.1001085.

Goldblatt, Ran, Michelle F. Stuhlmacher, Beth Tellman, Nicholas Clinton, Gordon Hanson, Matei Georgescu, Chuyuan Wang, Fidel Serrano-Candela, Amit K. Khandelwal, Wan Hwa Cheng, and Robert C. Balling. 2018. "Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover." *Remote Sensing of Environment* 205 (November 2017):253–275. URL https://doi.org/10.1016/j.rse.2017.11.026.

Goodman-Bacon, Andrew. 2018. "Difference-in-differences with variation in treatment timing." Tech. rep., National Bureau of Economic Research.

Grosh, Margaret E. and Judy L. Baker. 1995. *Proxy means tests for targeting social programs*. The World Bank. URL https://elibrary.worldbank.org/doi/abs/10.1596/0-8213-3313-5.

Gross, David B and Nicholas S Souleles. 2002. "Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data." *The Quarterly journal of economics* 117 (1):149–185.

GSMA. 2014. "State of the Industry: Mobile Financial Services for the Unbanked." Tech. rep., GSMA. URL https://gsdrc.org/document-library/state-of-the-industry-2014-mobile-financial-services-for-the-unbanked/.

———. 2017. "Landscape Report: Mobile Money, Humanitarian Cash Transfers and Displaced Populations." Tech. rep., GSMA. URL https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/05/Mobile_Money_Humanitarian_Cash_Transfers.pdf.

———. 2019. "The Mobile Gender Gap Report 2019." Tech. rep., GSMA. URL https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/03/GSMA-Connected-Women-The-Mobile-Gender-Gap-Report-2019.pdf.

Gurbuz Cuneo, Alev. 2019. *The Economics of Rural Populations in Sub-Saharan Africa: Financial Inclusion and Agriculture*. Ph.D. thesis, Georgetown University.

Gutierrez, Thoralf, Gautier Krings, and Vincent D Blondel. 2013. "Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets." *arXiv preprint arXiv:1309.4496* .

Hernandez, Marco, Lingzi Hong, Vanessa Frias-Martinez, and Enrique Frias-Martinez. 2017. *Estimating poverty using cell phone data: evidence from Guatemala*. The World Bank.

Heyer, Amrik and Ignacio Mas. 2009. "Seeking fertile grounds for mobile money." *Unpublished paper* :1–22.

Holzmann, Penny et al. 2008. *Household Economy Approach, the Bk: A Guide for Programme Planners and Policy-Makers*. Save the Children UK.

Houssou, Nazaire, Collins Asante-Addo, Kwaw S Andam, and Catherine Ragasa. 2019. "How can African governments reach poor farmers with fertiliser subsidies? Exploring a targeting approach in Ghana." *The Journal of Development Studies* 55 (9):1983–2007.

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142 (2):615–635.

Jack, Kelsey and Grant Smith. 2020. "Charging Ahead: Prepaid Metering, Electricity Use, and Utility Revenue." *American Economic Journal: Applied Economics* 12 (2):134–68. URL https://www.aeaweb.org/articles?id=10.1257/app.20180155.

Jack, William, Adam Ray, and Tavneet Suri. 2013. "Transaction networks: Evidence from mobile money in Kenya." *American Economic Review* 103 (3):356–61.

Jack, William and Tavneet Suri. 2014. "Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution." *American Economic Review* 104 (1):183–223.

Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. "Machine Learning To Predict Poverty." *Science* 353 (6301):790–794.

Jensen, Robert. 2007. "The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector." *The quarterly journal of economics* 122 (3):879–924.

Jensen, Robert T. and Nolan H. Miller. 2008. "Giffen behavior and subsistence consumption." *American Economic Review* 98 (4):1553–1577.

Karlan, Dean and Jonathan Zinman. 2010. "Expanding credit access: Using randomized supply decisions to estimate the impacts." *The Review of Financial Studies* 23 (1):433–464.

Khaefi, Muhammad Rizal, Dharani Dhar Burra, Rio Fandi Dianco, Dikara Maitri Pradipta Alkarisya, Muhammad Rheza Muztahid, Annissa Zahara, George Hodge, Rajius Idzalika et al. 2019. "Modelling Wealth from Call Detail Records and Survey Data with Machine Learning: Evidence from Papua New Guinea." In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2855–2864.

Khan, Muhammad R and Joshua E Blumenstock. 2016. "Predictors without borders: behavioral modeling of product adoption in three developing countries." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 145–154.

Kirui, Oliver K, Julius Juma Okello, and Rose A Nyikal. 2012. "Awareness of mobile phone-based money transfer services in agriculture by smallholder farmers in Kenya." *International Journal of ICT Research and Development in Africa (IJICTRDA)* 3 (1):1–13.

Lang, Megan. 2020. "Consuming Perishable Goods in the Presence of Transaction Costs and Liquidity Constraints." *MIMEO* .

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (6176):1203–1205. URL https://science.sciencemag.org/content/343/6176/1203.

Lee, David S and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2):281–355.

Lee, Jean N., Jonathan Morduch, Saravana Ravindran, Abu Shonchoy, and Hassan Zaman. 2021. "Poverty and Migration in the Digital Age: Experimental Evidence on Mobile Banking in Bangladesh." *American Economic Journal: Applied Economics* 13 (1):38–71. URL https://www.aeaweb.org/articles?id=10.1257/app.20190067.

Lu, Xin, Linus Bengtsson, and Petter Holme. 2012. "Predictability of population displacement after the 2010 Haiti earthquake." *Proceedings of the National Academy of Sciences* URL https://www.pnas.org/content/early/2012/06/11/1203882109.

Mas, Ignacio and Olga Morawczynski. 2009. "Designing mobile money services lessons from M-PESA." *Innovations: Technology, Governance, Globalization* 4 (2):77–91.

Mas, Ignacio and Dan Radcliffe. 2010. "Mobile payments go viral: M-PESA in Kenya."

Masocha, Reginald and Obey Dzomonda. 2018. "Adoption of mobile money services and the performance of small and medium enterprises in Zimbabwe." *Academy of Accounting and Financial Studies Journal* 22 (3):1–11.

Maxwell, Daniel and Richard Caldwell. 2008. "The Coping Strategies Index : Field Methods Manual - Second Edition." .

Mccrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design : A density test." *Journal of Econometrics* 142:698–714.

McKinsey. 2017. "Fintech: Opportunities and challenges for banks and regulators." *Market Research* URL https://www.mckinsey.com/~/media/McKinsey/Industries/Financial%20Services/Our%20Insights/FinTech%20Challenges%20and%20Opportunities/FinTechChallenges%20and%20Opportunities.ashx.

Milusheva, Sveta. 2020. *Using Mobile Phone Data to Reduce Spread of Disease*. The World Bank.

Morse, Adair. 2011. "Payday lenders: Heroes or villains?" *Journal of Financial Economics* 102 (1):28–44.

O'Donoghue, Brian Dillon Joachim De Weerdt Ted. 2020. *Paying More for Less: Why Don't Households in Tanzania Take Advantage of Bulk Discounts?* The World Bank. URL https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-9167.

Olivieri, Sergio, Francesc Ortega, Ana Rivadeneira, and Eliana Carranza. 2020. "Shoring up Economic Refugees: Venezuelan Migrants in the Ecuadorian Labor Market." *Mimeo* .

Riley, Emma. 2018. "Mobile money and risk sharing against village shocks." *Journal of Development Economics* 135:43–58. URL https://www.sciencedirect.com/science/article/pii/S0304387818304413.

Roessler, Philip, Flora Myamba, Peter Carroll, Cornel Jahari, Blandina Kilama, and Daniel L Nielson. 2018. "Mobile-Phone Ownership Increases Poor Women's Household Consumption: A Field Experiment in Tanzania." In *Proceedings of the Evidence in Governance and Politics Meeting, Nairobi, Kenya*. 1–15.

Roth, Jonathan. Working Paper. "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends." Tech. rep., Department of Economics, Harvard University.

Skiba, Paige Marta and Jeremy Tobacman. 2019. "Do payday loans cause bankruptcy?" *The Journal of Law and Economics* 62 (3):485–519.

Stahl, Lisa and Gerhard Coetzee. 2018. "Case study: Digicel Mobile Money (MonCash), Haiti." Tech. rep., Consultative Group to Assist the Poor (CGAP). URL https://customersguide.cgap.org/sites/customersguide.cgap.org/files/resource/2018/05/CGAP_Digicel-Case-Study.pdf.

Sun, Liyang and Sarah Abraham. 2020. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." Tech. rep., Working Paper.

Suri, Tavneet. 2017. "Mobile money." *Annual Review of Economics* 9:497–520.

Suri, Tavneet and William Jack. 2016. "The long-run poverty and gender impacts of mobile money." *Science* 354 (6317):1288–1292.

The World Bank. 2012. "Mobile Usage at the Base of the Pyramid in Kenya." Tech. rep., iHub Research and Research Solutions Africa. URL https://www.infodev.org/infodev-files/final_kenya_bop_study_web_jan_02_2013_0.pdf.

———. 2020. "Haiti, Poverty and Equity Brief." https://databank.worldbank.org/data/download/poverty/33EF03BB-9722-4AE2-ABC7-AA2972D68AFE/Global_POVEQ_HTI.pdf. Accessed: 2020-09-21.

Wesolowski, Amy, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee. 2012. "Quantifying the impact of human mobility on malaria." *Science* 338 (6104):267–270.

Wilson, Robin, Elisabeth Zu Erbach-Schoenberg, Maximilian Albert, Daniel Power, Simon Tudge, Miguel Gonzalez, Sam Guthrie, Heather Chamberlain, Christopher Brooks, Christopher Hughes, Lenka Pitonakova, Caroline Buckee, Xin Lu, Erik Wetter, Andrew Tatem, and Linus Bengtsson. 2016. "Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake." *PLoS Currents* 8 (Disasters).

World Food Programme (FAO). 2008. "Food Consumption Score - Construction of the FCS: Interagency Workshop Report WFP - FAO Measures of Food Consumption - Harmonizing Methodologies Rome, 9 - 10 April 2008." Tech. Rep. April, World Food Program. URL http://documents.wfp.org/stellent/groups/public/documents/ena/wfp196627.pdf.

Zagatti, Guilherme Augusto, Miguel Gonzalez, Paolo Avner, Nancy Lozano-Gracia, Christopher J. Brooks, Maximilian Albert, Jonathan Gray, Sarah Elizabeth Antos, Priya Burci, Elisabeth zu Erbach-Schoenberg, Andrew J. Tatem, Erik Wetter, and Linus Bengtsson. 2018. "A trip to work: Estimation of origin and destination of commuting patterns in the main metropolitan regions of Haiti using CDR." *Development Engineering* 3:133 – 165. URL http://www.sciencedirect.com/science/article/pii/S2352728517300866.

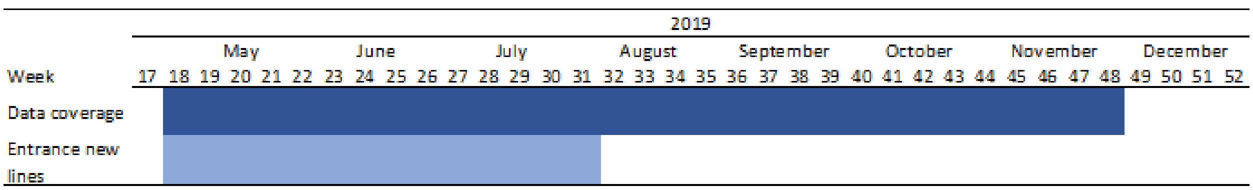
Zinman, Jonathan. 2010. "Restricting consumer credit access: Household survey evidence on effects around the Oregon rate cap." *Journal of banking & finance* 34 (3):546–556.

Appendix

Appendix for Chapter 1

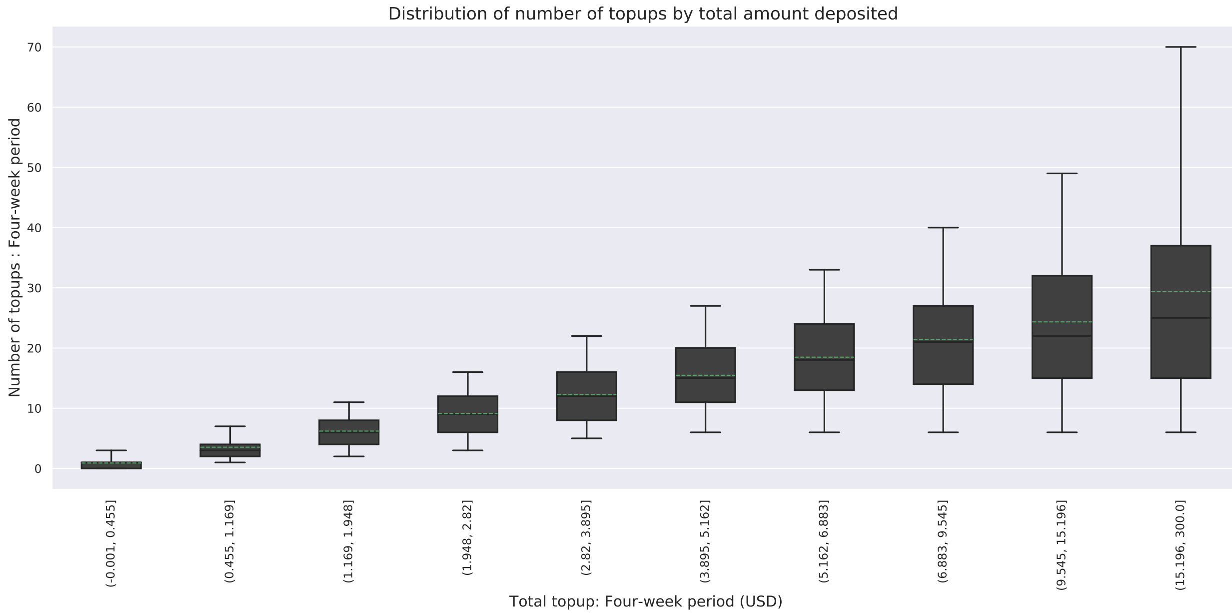
Additional figures and tables

Figure 1A: Cellphone data coverage



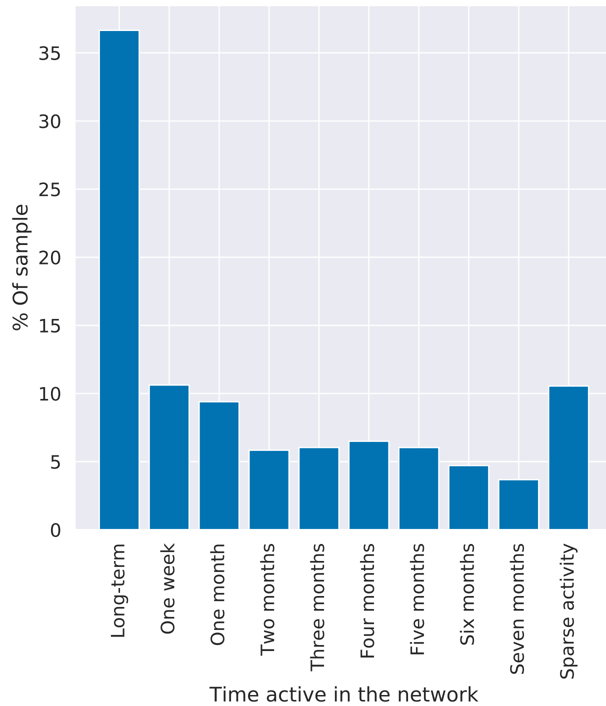
Note: Calendar-week includes Monday to Sunday and can overlap with the calendar month. Entrance of new lines defines the period we consider to identify the activation of new lines. The period after these weeks is used to observe the network patterns of the lines active, but do not consider any new line activated during the period.

Figure 2A: Active numbers April 2019
Total monthly expenditure and number of recharges



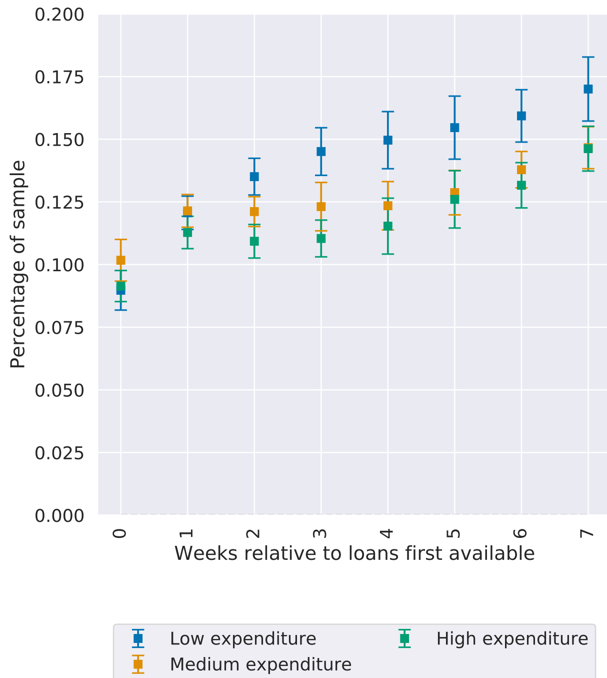
Note: Includes well-established lines only. The x-axis contains the deciles for total cellphone expenditure for April 2019.The y-axis contains the distribution of 95% of number of recharges for each subscriber during the month. Vertical lines show the minimum and maximum values.

Figure 3A: Time in the network
New numbers May and July 2019



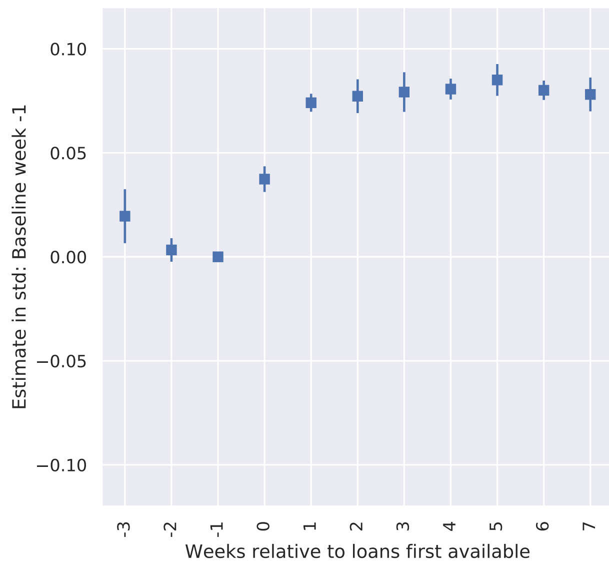
Note: Long-term customers remain active after 12 weeks after activation. We use the day with the last transaction in the network to determine when it dropped from the network. Sparse activity are numbers with large gaps in network activity.

Figure 4A: Percentage using airtime loans



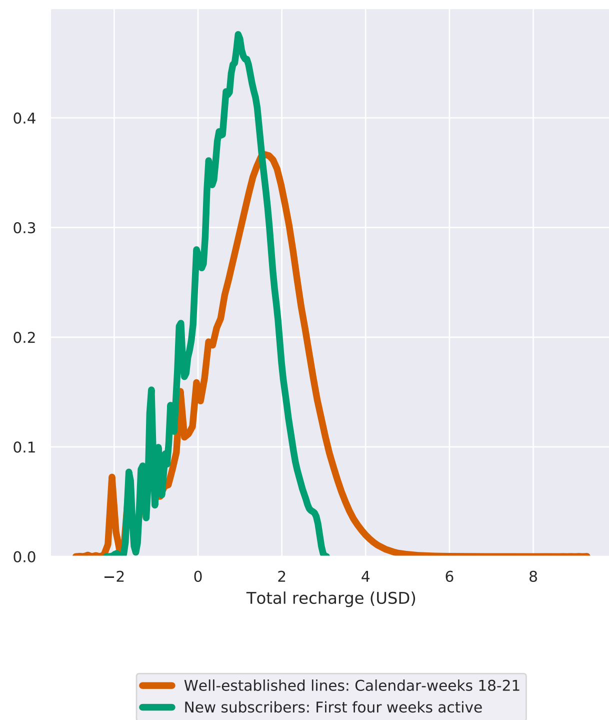
Note: Long-term customers only

Figure 5A: New customers total weekly expenditure (more than three months active)



Note: Includes only long-term and customers that were active for more than three months. Loan access is provided at week 0 and the week before is used as baseline.

Figure 6A: PDF cumulative expenditure



Note: Includes only long-term customers. Loan access is provided at week 0 and the week before is used as baseline.

Table 1A: Distribution of population and active cellphone subscribers

Department	**Population**	**Percentage of population**	**Well established lines**	**New customers**
Artibonite	1,727,524	15.8%	7.1%	9.0%
Centre	746,236	6.8%	2.2%	3.5%
Nord-Ouest	728,807	6.7%	2.1%	2.9%
Nord	1,067,177	9.8%	8.1%	8.3%
Nord-Est	393,967	3.6%	1.0%	1.7%
Grand'Anse	468,301	4.3%	1.0%	0.8%
Nippes	342,525	3.1%	2.6%	2.8%
Sud	774,976	7.1%	4.3%	4.2%
Ouest	4,029,705	36.9%	65.9%	60.4%
Sud-Est	632,601	5.8%	5.6%	6.5%

Note: Population information comes the 2018 National Population Projection

Table 2A: Descriptive statistics
Phone survey

	mean	std	min	10%	50%	90%	max
Demographics							
Age	31.59	9.13	16.0	22.0	30.0	43.0	87.0
Household Head	0.62	0.49					
Gender	0.55	0.5					
Labor Market							
Worked last week	0.41	0.49					
Income regular week (USD)	83.61	115.73	2.14	14.29	50.0	171.43	1071.43
Income last week (USD)	78.35	103.13	0.0	14.29	50.0	142.86	814.29
Income stability							
Unpredictable	0.58						
Somewhat unpredictable	0.28						
Very predictable	0.14						
Food security							
Small serving	0.64	0.48					
Borrowing							
Neighbor	0.27	0.44					
Amount (USD)	258.22	385.47	2.86	21.43	142.86	714.29	2857.14
Family	0.2	0.4					
Amount (USD)	413.29	1359.55	7.86	37.21	142.86	671.43	10742.86
Bank	0.07	0.25					
Amount (USD)	9824.17	48403.56	35.71	264.29	1428.57	4714.29	300000.0
Shopkeeper	0.04	0.2					
Amount (USD)	100.23	158.23	7.14	13.14	26.79	407.14	500.0
Informal	0.02	0.14					
nan	235.58	159.28	21.43	28.57	214.29	398.57	500.0

Note: Includes 589 survey participants with matched cellphone records.

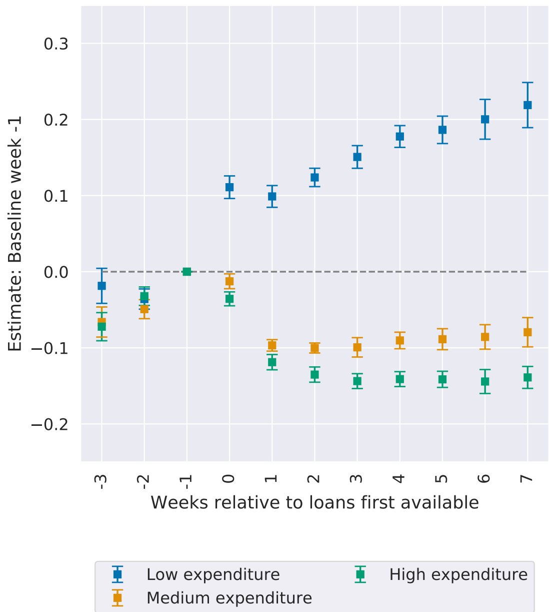
Table 3A: Network transactions statistics
Four weeks prior to survey

	mean	std	10%	50%	90%
Recharge Activity					
total recharge (USD)					
Very predictable	36.84	73.42	5.7	19.48	53.58
Somewhat unpredictable	21.88	20.77	5.09	15.26	46.12
Unpredictable	20.99	26.86	4.76	12.9	42.34
Number of recharges					
Very predictable	28.85	15.02	13.0	27.0	50.8
Somewhat unpredictable	27.88	16.54	11.0	23.5	52.3
Unpredictable	28.11	15.34	12.0	24.0	50.0
Average recharge					
Very predictable	85.88	112.82	18.67	53.75	162.69
Somewhat unpredictable	59.84	48.86	18.0	42.15	112.29
Unpredictable	57.75	81.97	15.4	36.63	108.44
Median recharge					
Very predictable	62.44	99.83	14.39	45.45	100.0
Somewhat unpredictable	47.63	46.1	14.25	25.0	95.46
Unpredictable	39.02	44.26	12.83	25.0	84.91
Number of loans					
Loan Demand					
Very predictable	3.76	3.1	1.0	3.0	8.9
Somewhat unpredictable	3.66	2.9	1.0	3.0	8.0
Unpredictable	3.44	3.41	1.0	2.0	7.0
Total amount borrowed (USD)					
Very predictable	5.29	6.78	0.39	3.1	10.61
Somewhat unpredictable	4.76	5.32	0.39	2.91	9.6
Unpredictable	4.3	5.81	0.27	2.59	11.6
Share of total expenditure financed					
Very predictable	0.21	0.17	0.02	0.19	0.44
Somewhat unpredictable	0.26	0.22	0.03	0.2	0.57
Unpredictable	0.21	0.17	0.02	0.18	0.45

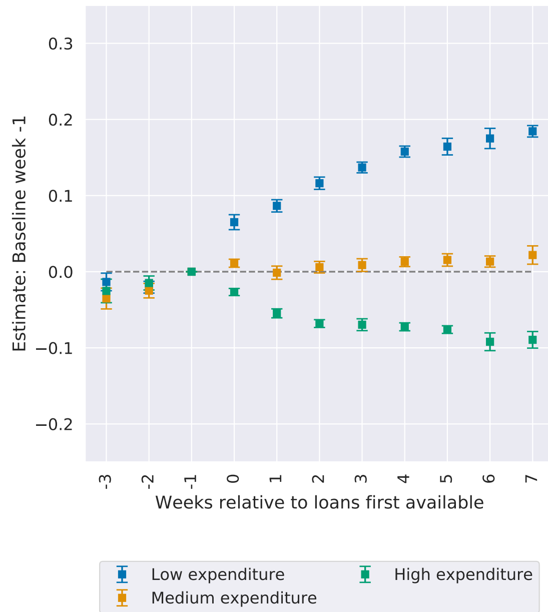
Note: Includes 589 phone survey participants that match cellphone records. Descriptive statistics on cellphone data include one month of mobile transactions.

Figure 7A: Heterogeneous impacts
Key network metric activities

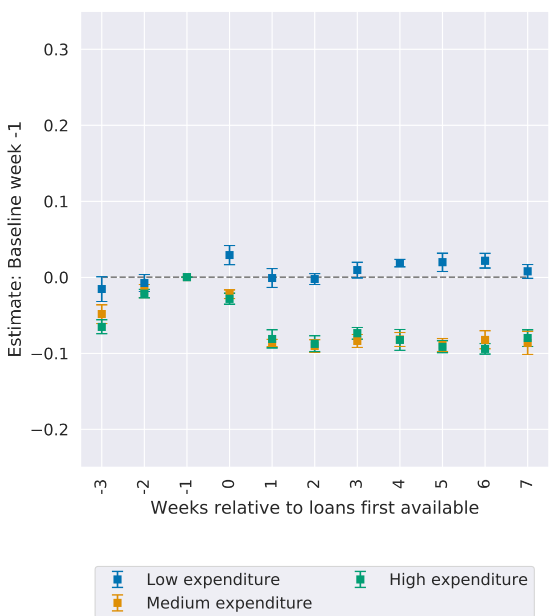
(a) Total number of contacts



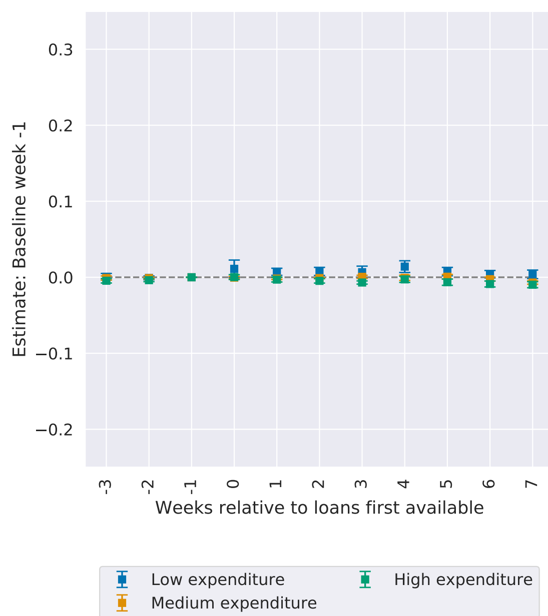
(b) Total outgoing transactions (calls+SMS)



(c) Average duration of calls

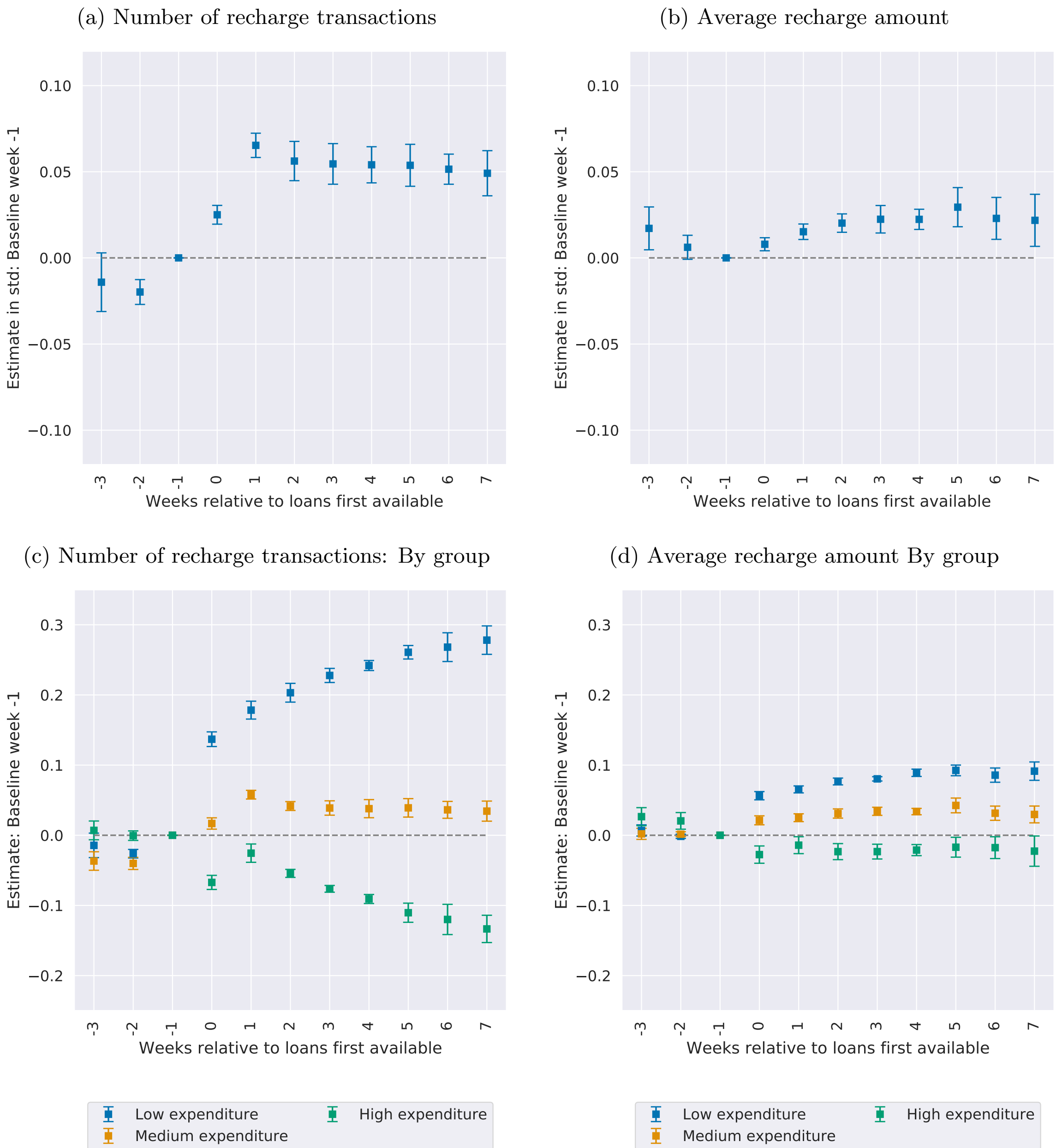


(d) Total expenditure in gambling



Note: Includes only long-term customers. Results are in standard deviations and use week -1 as baseline.

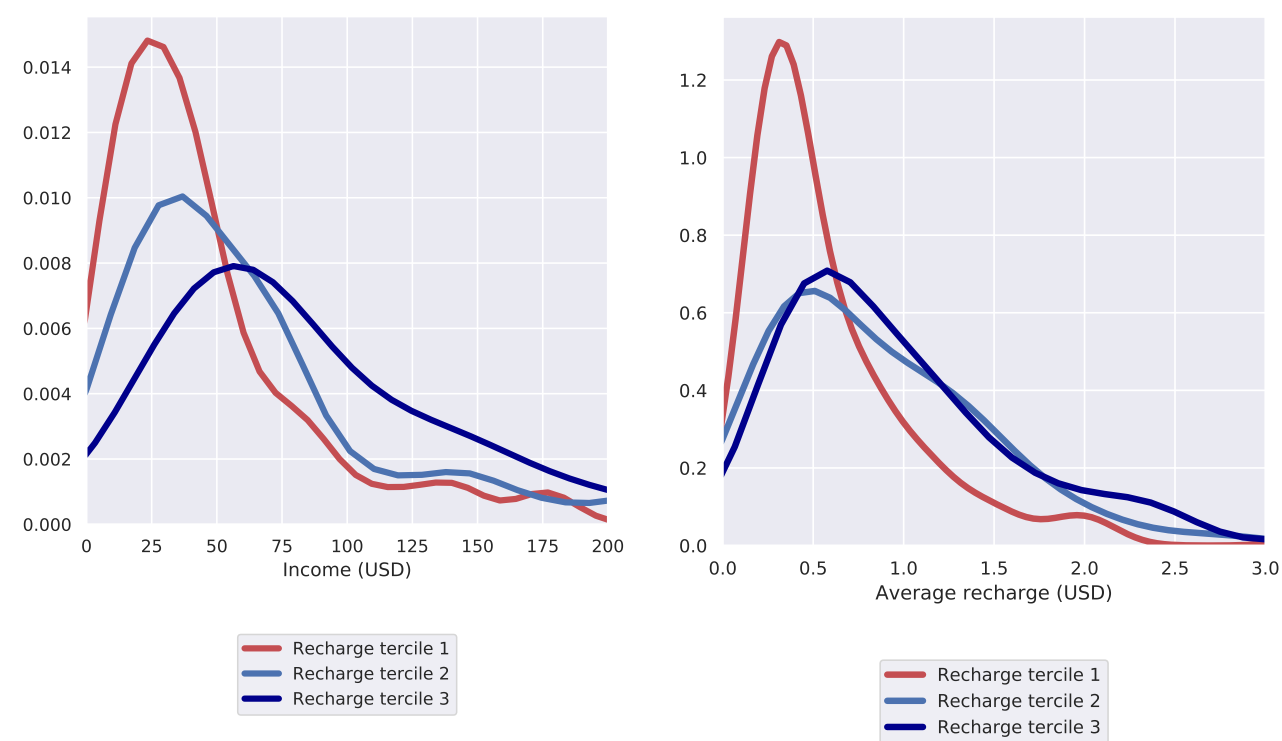
Figure 8A: Additional results on recharges: Frequency and amount



Note: Includes only long-term customers. Results are in standard deviations and use week -1 as baseline.

Figure 9A: Recharge by terciles and observed income
phone survey participants only

(a) Reported income (USD) (b) Average recharges



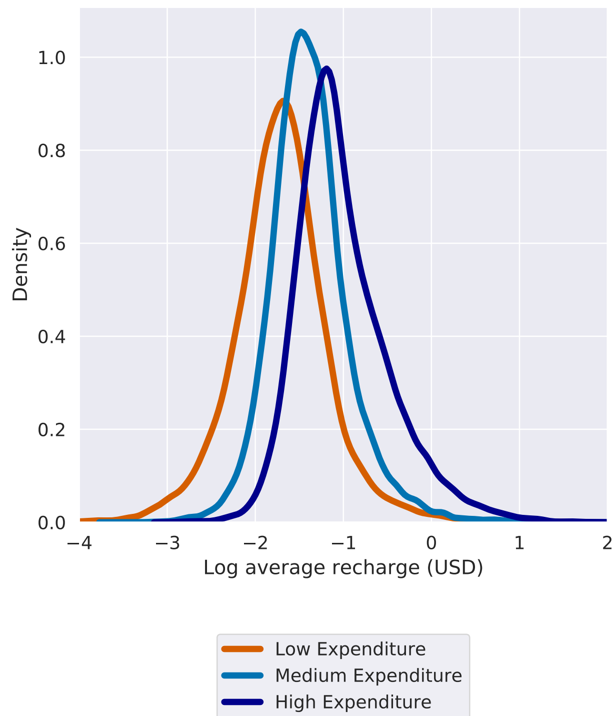
Note: Recharge terciles were constructed using the distribution of total cellphone expenditure in the month before the survey. Information on the reported income during the previous week is only available for those that were employed at the time of the survey.

Table 4A: Share of total expenditure financed

Week	Principal						Facilitation fee					
active	Low		Medium		High		Low		Medium		High	
5	5.71	(0.78)	5.25	(0.69)	4.5	(0.39)	0.69	(0.09)	0.57	(0.07)	0.43	(0.04)
6	8.69	(1.68)	7.45	(0.81)	6.29	(0.54)	0.98	(0.17)	0.76	(0.08)	0.6	(0.05)
7	10.6	(2.33)	8.57	(1.26)	6.79	(0.66)	1.14	(0.23)	0.85	(0.12)	0.64	(0.07)
8	12.2	(3.08)	9.44	(1.86)	7.29	(0.86)	1.26	(0.29)	0.92	(0.18)	0.69	(0.08)
9	13.37	(2.86)	9.86	(1.61)	7.97	(1.0)	1.35	(0.27)	0.95	(0.16)	0.75	(0.1)
10	13.93	(2.99)	10.17	(1.55)	8.49	(1.25)	1.4	(0.29)	0.99	(0.15)	0.8	(0.12)
11	15.13	(3.0)	10.94	(1.49)	8.45	(0.81)	1.5	(0.28)	1.06	(0.14)	0.8	(0.08)
12	16.12	(2.52)	11.82	(1.54)	9.34	(1.04)	1.59	(0.24)	1.14	(0.15)	0.88	(0.1)

Note: Long-terms customers only. Groups defined using total cellphone expenditure before loans are available

Figure 10A: Average transaction size (USD)
Long-term customers



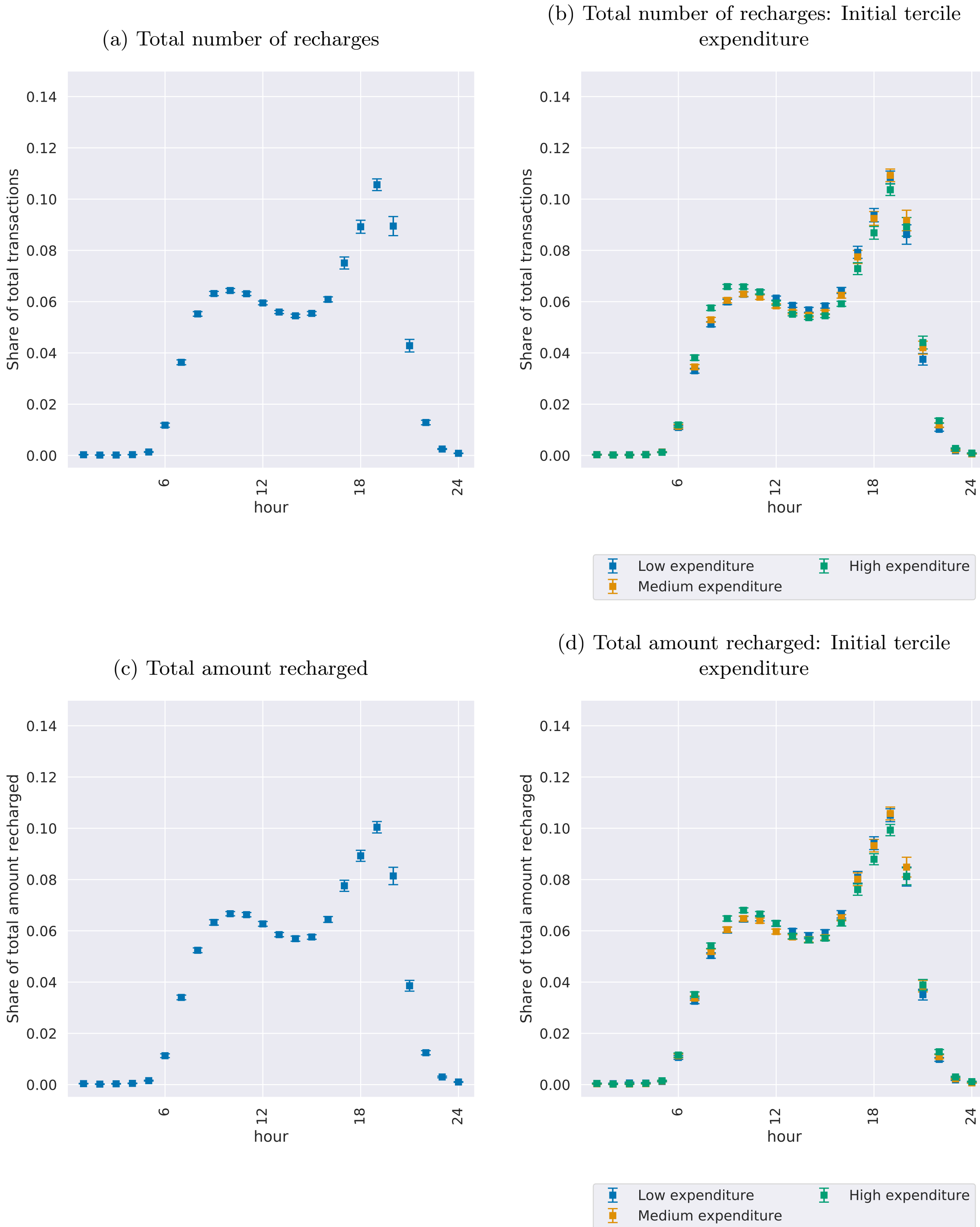
Note: Average transaction size during the initial four weeks

Robustness Checks

1.A.1 Robustness Checks: Alternative specifications

We implement two additional specifications were we modify the control group that we use in our preferred specification. The first one implements a pure Event Study Design and does not include a control group, while the second one uses the new subscribers that joined the network during wees 18 to 24 of 2019 as the control group for those that joined latter. Figure 13A shows our results are robust to changes in the control group.

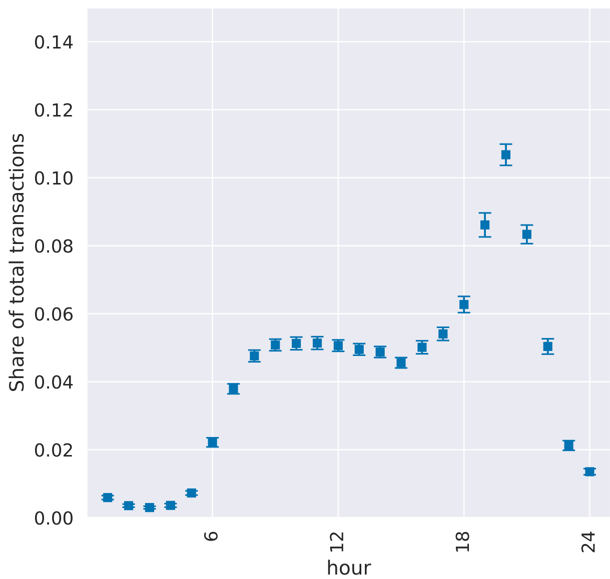
Figure 11A: Share of total recharge (hourly)
Long-term customers



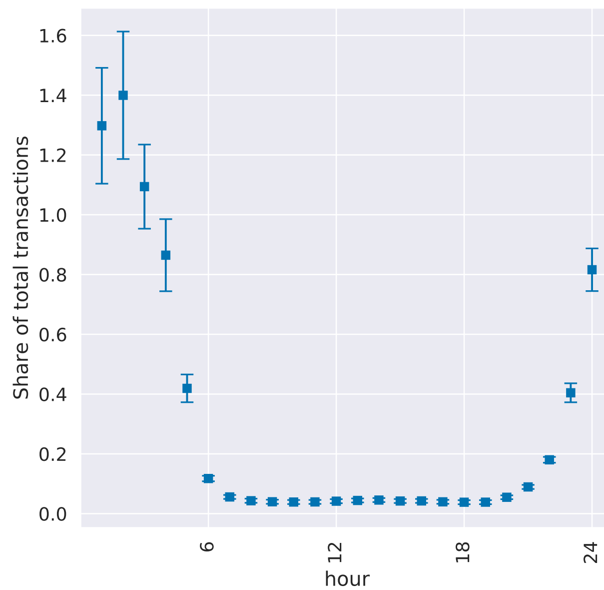
Note: Includes only customers that are eligible for the loan. The estimation of daily demand patterns includes controls for day of the week and calendar week.

Figure 12A: Share of loan transactions per hour

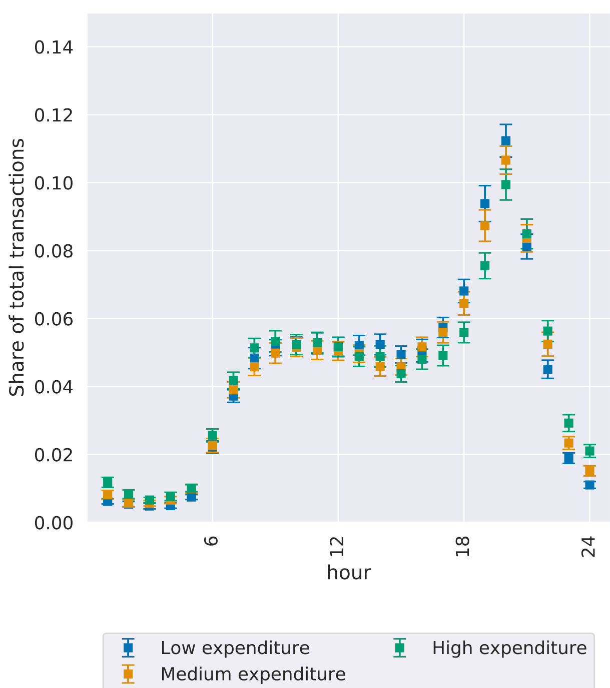
(a) Loan transactions



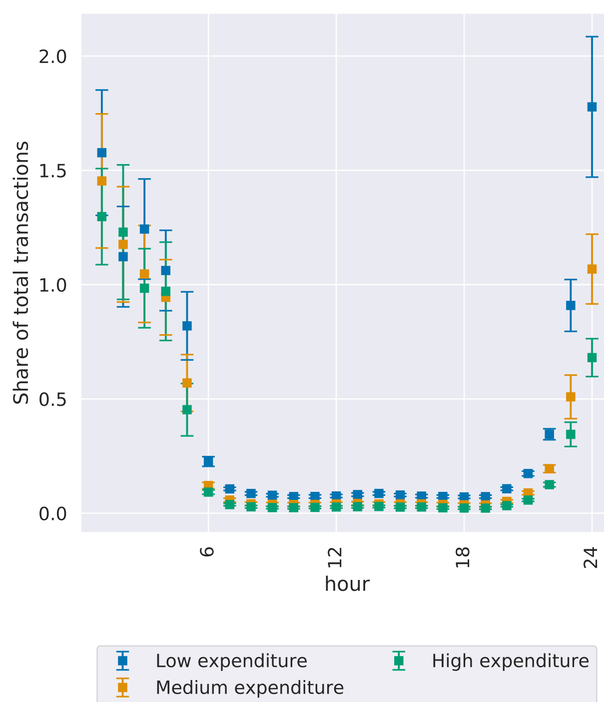
(b) Loan transactions over recharges transactions



(c) Loan transactions
by initial expenditure tercile



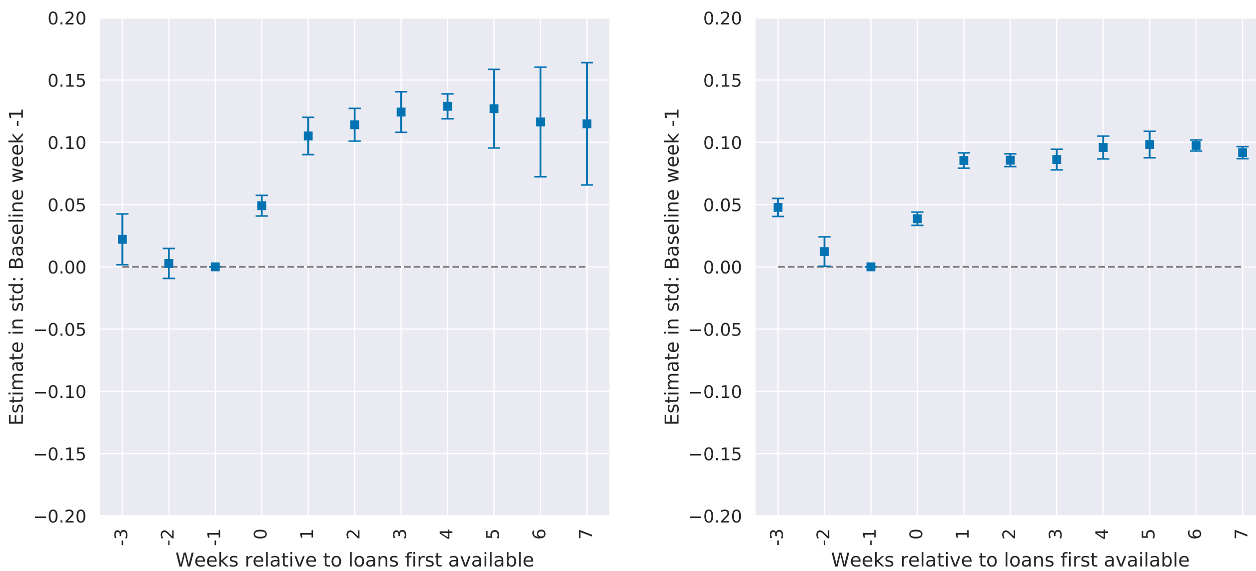
(d) Loan transactions over recharges transactions
by initial expenditure tercile



Note: Includes only customers that are eligible for the loans. The estimation of daily demand patterns includes controls for day of the week and calendar week.

Figure 13A: Event study
Additional specifications

(a) No controls: Simple difference (b) New lines weeks 18-31 as controls



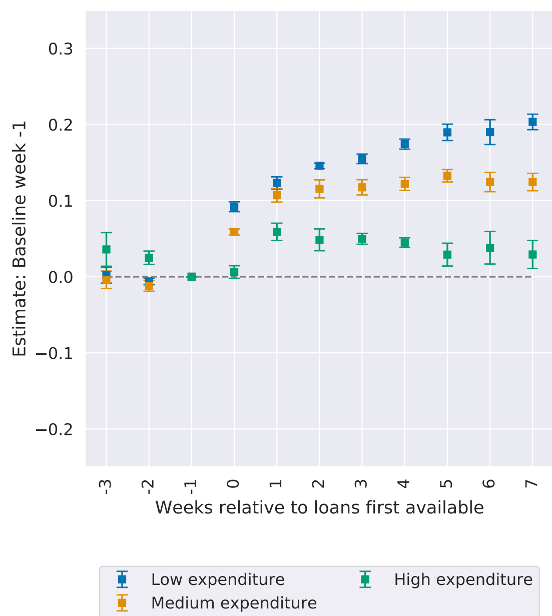
Note: Includes only customers that are eligible for the loans. Panel (a) uses the sample that enters the network between the week 18 and 31 of 2019. Panel (b) uses as control the lines that were activated between week 18 and 24, and checks for the impact of airtime loans eligibility on lines activated between 25 and 31. During this period, the eligibility of lines activated between week 18 and 24 does not change.

1.A.2 Robustness Checks: location of subscribers

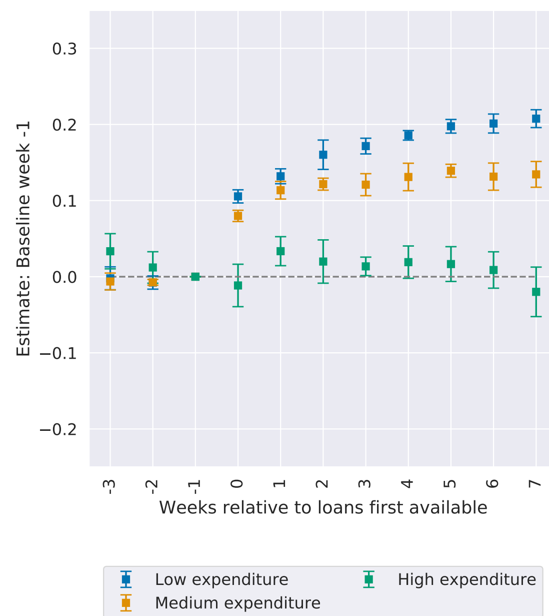
Descriptive analysis shows that expenditure levels differ depending on the location of subscribers, with new subscribers being more likely to be located outside of the Port-au-Prince metropolitan area. We test for differences in our parameter of interesting based on the location of subscribers. Overall, we find that credit access presents a similar affect across different locations.

Figure 14A: Impacts credit access
Results by location of subscribers size

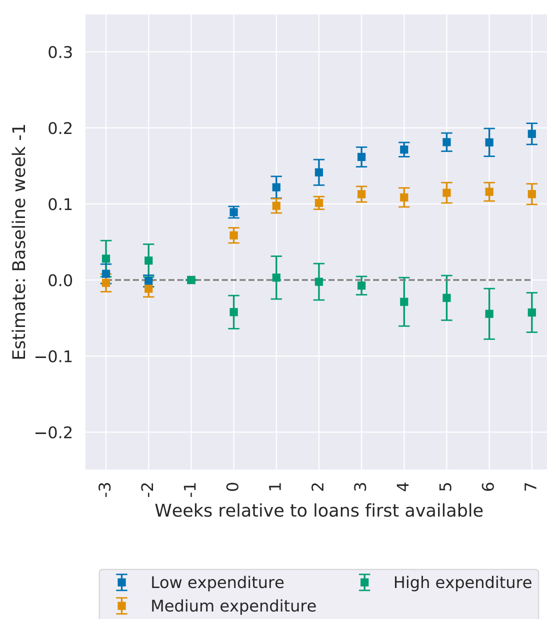
(a) PAP metropolitan area



(b) Large city



(c) Other



Note: Includes only customers that are eligible for the loans. Panel (a) uses only the sample that enter the network between the week 18 and 31 of 2019. Panel (b) uses as control the lines that were activated between week 18 and 24, and checks for the impact of airtime eligibility on lines activated between 25 and 31. During this period, the eligibility of lines activated between week 18 and 24 does not change.

Table 5A: Impacts credit access
Results by city size

	Low expenditure			Medium expenditure			High expenditure		
	Baseline	Effect	Δ in percentage	Baseline	Effect	Δ in percentage	Baseline	Effect	Δ in percentage
Expenditure (USD)									
PAP metropolitan area	0.23	0.35***	156.06	0.64	0.26***	40.68	1.61	0.04***	2.44
Large cities	0.22	0.36***	160.11	0.64	0.26***	40.86	1.67	-0.01	-0.65
Other	0.21	0.29***	139.3	0.62	0.21***	33.66	1.62	-0.08***	-4.93
Avg. recharges (USD)									
PAP metropolitan area	0.14	0.07***	47.32	0.24	0.02***	9.88	0.44	-0.02***	-4.03
Large cities	0.13	0.08***	56.19	0.24	0.04***	15.33	0.49	-0.04***	-7.72
Other	0.14	0.07***	50.95	0.27	0.02***	8.15	0.52	-0.06***	-10.8
Number of recharges									
PAP metropolitan area	1.0	0.9***	89.36	2.59	0.31***	11.91	4.25	-0.23***	-5.49
Large cities	1.0	0.8***	80.4	2.58	0.18***	7.1	4.01	-0.23***	-5.69
Other	0.87	0.56***	64.8	2.23	0.07***	3.04	3.56	-0.35***	-9.91
Outgoing contacts									
PAP metropolitan area	4.65	1.32***	28.39	8.06	-0.46***	-5.68	9.66	-0.8***	-8.32
Large cities	4.02	1.26***	31.42	6.27	-0.06	-0.94	7.57	-0.38***	-5.08
Other	4.28	1.11***	26.02	6.44	-0.17***	-2.61	7.62	-0.41***	-5.44
Outgoing transactions									
PAP metropolitan area	17.6	15.77***	89.57	44.27	2.85***	6.43	69.85	-7.11***	-10.17
Large cities	16.81	14.66***	87.19	37.82	4.44***	11.75	56.03	-3.57***	-6.37
Other	17.36	13.03***	75.08	37.76	2.37***	6.28	53.95	-3.15***	-5.83
Avg. call duration									
PAP metropolitan area	62.53	2.54***	4.07	84.62	-7.08***	-8.37	96.78	-6.69***	-6.91
Large cities	52.78	3.42***	6.48	68.86	-5.48***	-7.95	78.79	-3.94***	-5.0
Other	54.47	2.18**	4.0	71.77	-5.46***	-7.61	83.48	-5.05***	-6.04
Gambling expenditure (USD)									
PAP metropolitan area	0.0	0.0***	132.39	0.01	0.0*	7.96	0.01	-0.0***	-11.89
Large cities	0.0	0.0***	53.56	0.01	-0.0	-7.68	0.01	-0.0	-8.37
Other	0.01	0.0***	22.79	0.01	-0.0*	-10.21	0.01	0.0	0.27

Note: PAP metropolitan area includes the city of Port-au-Prince and adjacent cities including Croix-de-Bouquets. Large cities group include the urban centers with more than 100,000 people. Each group concentrate 28 and 30% of all the Haitian population.

Appendix for Chapter 2

Technical Appendix

1.B.1 Outcomes specification

Food Consumption and food expenditure

The survey asks respondents for seven day recall of spending across several categories of food, including total spending on fruits and vegetables, meat products (fish, offal, meat), eggs, dairy, legumes and tubers, cereals, sugars, oils and fats. Respondents are asked for spending over the past week first and then are asked for any other main sources of food consumed. If respondents report home production or donation as one of these sources, they are prompted to estimate the amount consumed but not purchased. From these data, we construct expenditure over the past week, as well as a measure of consumption, which includes food obtained by gifts, assistance, or home production. Finally, we construct the share of expenditure going to staples – in this case cereals – by dividing the past week's expenditure on these by total expenditure across all categories to proxy for proximity to subsistence (Jensen and Miller, 2008).

Food Consumption Score (FCS)

Let w_j be the weight for food category j and $days_j$ be the number of days food category j was consumed (see Table 1B for weights). Suppressing the household subscript, FCS is defined as

$$FCS = \sum_j w_j \times days_j. \tag{1.2}$$

When food subcategories are present (e.g., green v. orange vegetables), the number of days each food category is consumed is measured by summing food subcategories, capped at seven days. Let $days_{jk}$ be the number of days food subcategory k in category j was consumed (World Food

Programme (FAO), 2008).[19]

$$days_j = \min \left\{ \sum_k days_{jk} \right\} \tag{1.3}$$

Dietary Diversity Score (DDS)

The second index, the Diet Diversity Score (DDS) is usually presented alongside the FCS to give a better sense of extensive changes in the FCS. In particular, the DDS measures the number of food categories consumed in the past week. Categories are weighted zero in the computation if they are not nutritionally valuable – in this case sugars and condiments have weights of zero.

Coping Strategies Index (CSI)

Additionally, to further qualify increases in spending and our measures of nutritional quality, we look at the WFP defined Coping Strategies Index for Food, which is built from questions about "survival strategies."

To contextualize the food security indices and consumption outcomes above, we measure strategies used to cope when food insecurity arises. These outcomes encompass a set of decisions that affect future household welfare. Seven of fourteen of these outcomes look explicitly at food related coping strategies. These strategy questipns occur as a seven day recall of how many days the household ate smaller, fewer, or less preferred meals, borrowed food from friends or relatives, or restricted adult's food intake so that children could eat. From five of these outcomes we construct a Coping Strategies Index (CSI). For CSI specific strategies see Table 2B and for all additional strategies see Table 3B. Much like the FCS, the CSI is computed as a weighted sum of days over the past week that a certain coping strategy was undertaken.

$$CSI = \sum_j q_j \times days_j \tag{1.4}$$

Looking at various coping strategies as outcomes, the coefficient on treatment can be meaningful in various ways. We expect negative coefficients on treatment status as evidence of the impact of beneficiary status. In this case, a negative coefficient would suggest that the cash transfers would reduce the degree to which households needed to use coping strategies as part of an income effect. Income from beneficiary transfers would serve as a substitute for these strategies, in such a context.

[19] This method clearly gives a best case of the FCS where subcategories of food are substitutes day to day as opposed to complements day to day. That is, in this view, households consume fish or beef or chicken, but will not consume both on a given day unless every day already features a protein product. This should not matter for measurement of impact as long as the degree of substitutability of goods does not differ between beneficiaries and non-beneficiaries or generally, beneficiaries do not have some kind of taste for variety within the day. Given that these individuals are drawn from very similar populations, this would seem unlikely.

Because of the costly nature of strategies used to cope, such an impact could potentially still mark an improvement in welfare, even if substitution away from coping strategies reduced the overall impact of program on measured food security. To this end, coping strategies are chosen to be dynamically harmful to the household.

However, a null coefficient might arise in two scenarios. If household's budgets are strained and/or access to coping strategies is limited, households would operate at their "coping frontier." In this case, we would expect to see a negative impact on coping strategies only if transfer payments were sufficiently large. This case is consistent with a story where the measured increases in food security are not due to a simultaneous coping decision made at the household level for beneficiary households e.g., reduction in meal sizes or number of meals per day. Alternatively, if the transfer program was not effective in increasing food consumption, we would similarly see a null effect on coping mechanisms. However, in this case, we would not expect to see impacts on food consumption or nutritional intake. Finally, positive impacts on the use of coping strategies would be highly counterintuitive. This would suggest that households are made worse by receiving transfers, or perhaps that some other omitted variable is discontinuous at the boundary, i.e., that households that receive transfers are qualitatively different than those who do.

Table 1B: WFP Food categories, subcategories and weights for FCS and DDS Indices

Category	Subcategories	Weights FCS	Weights DDS
Cereals, Roots and Tubers		2	1
Vegetables	Orange, Green Leafy, Other	1	1
Fruits	Orange, Other	1	1
Protein	Meat, Offal, Fish, Eggs	4	1
Pulses, Nuts, Seeds, and Legumes		3	1
Dairy		4	1
Oil and Fats		0.5	1
Sugars		0.5	0
Condiments		0	0

Table 2B: Disaggregated strategies (weekly recall) and weights for CSI

Coping Strategy: How many days out of the last seven days did your household adopt the following strategies due to lack of food or money?	CSI weight
Eat less preferred or less expensive foods	1
Reduced number of meals per day	1
Reduce meal size	1
Restrict adult food consumption to feed young	3
Borrow food or rely on help from friends or relatives	2

Table 3B: Other Disaggregated strategies (weekly recall)

Days in the last week:	
Borrow food or rely on help from friends or relatives	Integer (0-7)
Didn't Eat	Integer (0-7)
Went to bed hungry	Integer (0-7)
Worried the household would not have enough food to eat	Integer (0-7)
In the past seven days have you:	
Used grain stock meant for the agricultural season, for food	Indicator
Withdrawn children from school	Indicator
Reduced health expenses	Indicator
Taken on debt to buy food or bought food on credit	Indicator
Had a household member migrate	Indicator
Sold livestock (cow, chicken, sheep, goat, etc.)	Indicator
Sold other productive assets (mill, agricultural land, etc.)	Indicator

1.B.2 Hyperparameter tuning

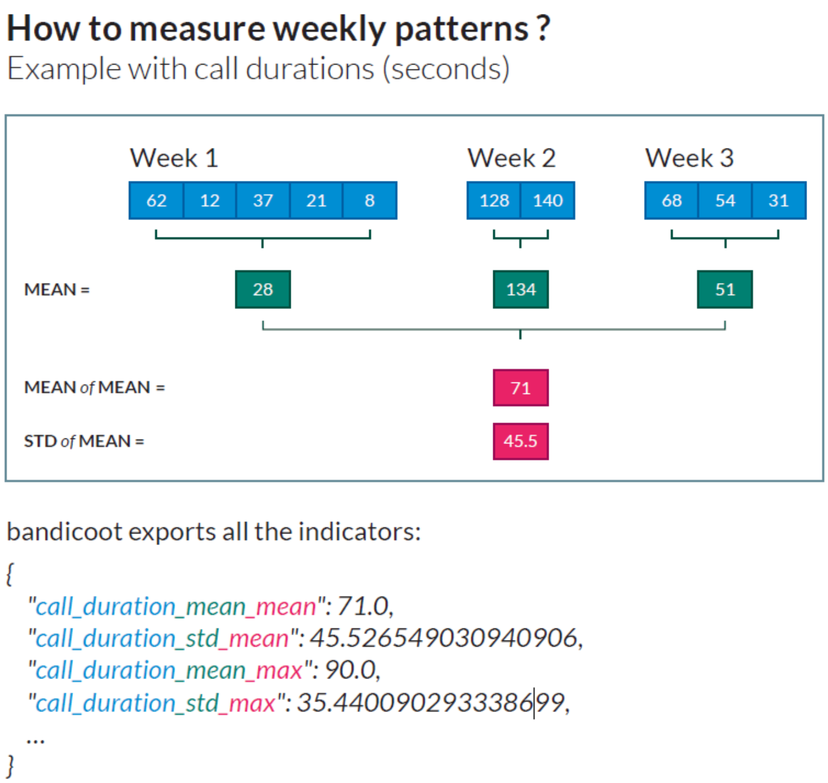
In the case of the random forest classifier (an ensemble of 100 decision trees) to predict the probability of beneficiary status on the training set and evaluate its out-of-sample performance on the test set. The maximum depth of the random forest is selected from $\{2, 4, 8, 16, 32\}$ via 3-fold cross-validation on the training set.

We also try to predict food expenditure using elastic net regression and three flexible two-based machine learning methods: a decision tree, a random forest, and XGBoost. For elastic net, the L1 penalty is chosen via 3-fold cross-validation from the set $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0, 1\}$ and the mixing parameter from the $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The maximum tree depth is selected for each tree-based model via 3-fold cross-validation from the set $\{2, 4, 8, 16, 32\}$. As in the previous exercise, we implement a 10-fold cross validation to account for over-fitting. Since the outcome is continuous, we use the R^2 coefficient to evaluate the predictions.

Additional figures and tables

1.B.3 Bandicoot feature engineering

Figure 1B: Bandicoot feature extraction process



Note: Taken from De Montjoye, Rocher, and Pentland (2016). The figure explains how information for each transaction type is calculated at the week level, and the computed as a single aggregate indicator.

Table 4B: Classification of features into similar information groups

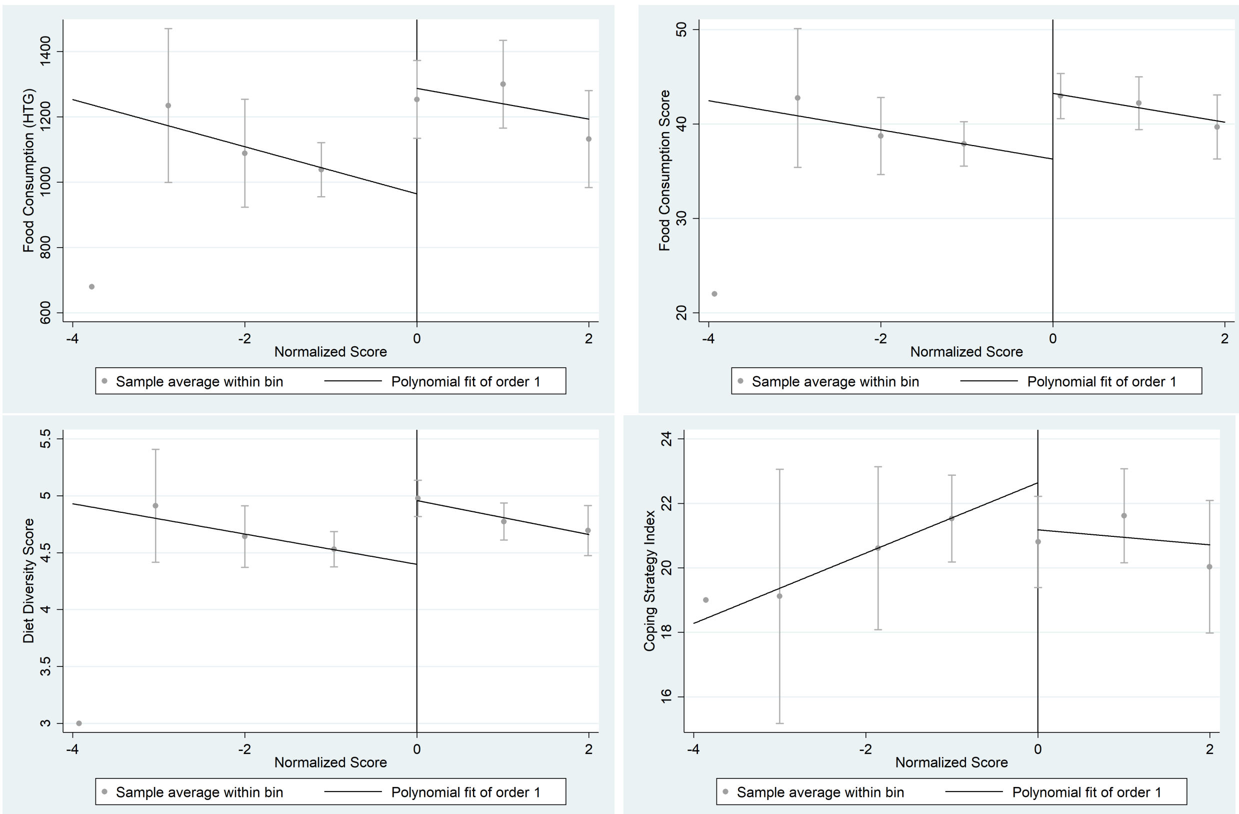
Category	Name	Description
B	Number of records	Number of actual records produced by user's mobile phone activity.
B	Active days	Number of days during which the user was active.
B	Number of interactions	Number of interactions by the user (incoming and outgoing).
B	Percent records missing location	Percentage of records different with home location
A	Call duration	Duration of the user's calls.
A	Percent initiated conversations	Percentage of conversations that have been initiated by the user.
A	Percent initiated interactions	Percentage of calls initiated by the user.
A	Response delay	Response delay of the user within a conversation (in seconds)
A	Response rate	Response rate of the user (between 0 and 1).
S	Percent at home	Percentage of interactions the user had while he was at home.
S	Radius of gyration	Returns the radius of gyration, the equivalent distance of the mass from the center of gravity, for all visited places.
S	Frequent antennas	Number of location that account for 80% of the locations where the user was.
S	Churn rate	Computes frequency spent at every tower each week, and returns the distribution of the cosine similarity between two consecutive weeks.
S	Number of antennas	Number of unique places visited.
S	User locations	LLG districts where the user resides, calculated using the tower geolocation (latlong).
R	Percent nocturnal	Percentage of interactions the user had at night.
R	Entropy of contacts	Entropy of the user's contacts.
R	Entropy of antennas	Entropy of visited antennas.
R	Interevent time	Interevent time between two records of the user.
R	Percent pareto interactions	Percentage of user's contacts that account for 80% of its interactions.
R	Percent pareto durations	Percentage of user's contacts that account for 80% of its total time spend on the phone.
D	Number of contacts	Number of contacts the user interacted with.
D	Balance of contacts	Balance of interactions per contact.
D	Interactions per contact	Number of interactions a user had with each of its contacts.

Note: Taken from Khaefi et al. (2019). The table shows the classification of features into different categories that reflect similar information content. B: Basic phone usage, A: Active users, S: Spatial behaviors, R: Regularity, D: Diversity.

Additional Results

1.B.4 Additional Impact Evaluation Results

Figure 2B: Effect of cash transfer on Food Consumption, Diet Diversity and Coping Strategies

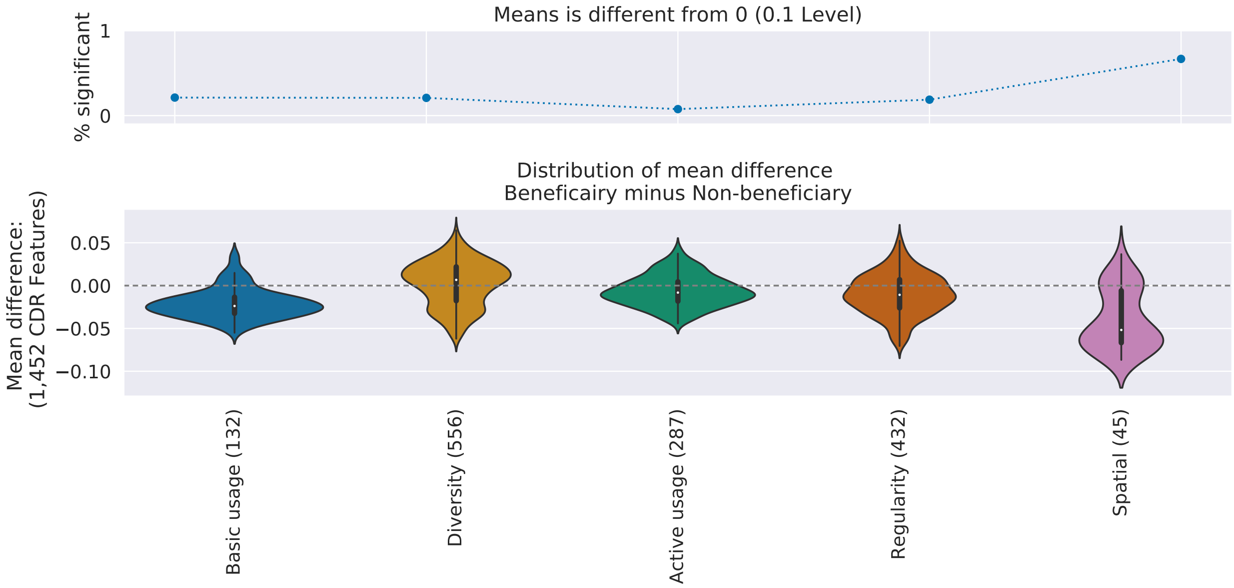


Note: Author's calculation using the in-person survey to measure outcomes, and the scorecard survey to create the vulnerability score. Figures show the RD impact at the bin level.

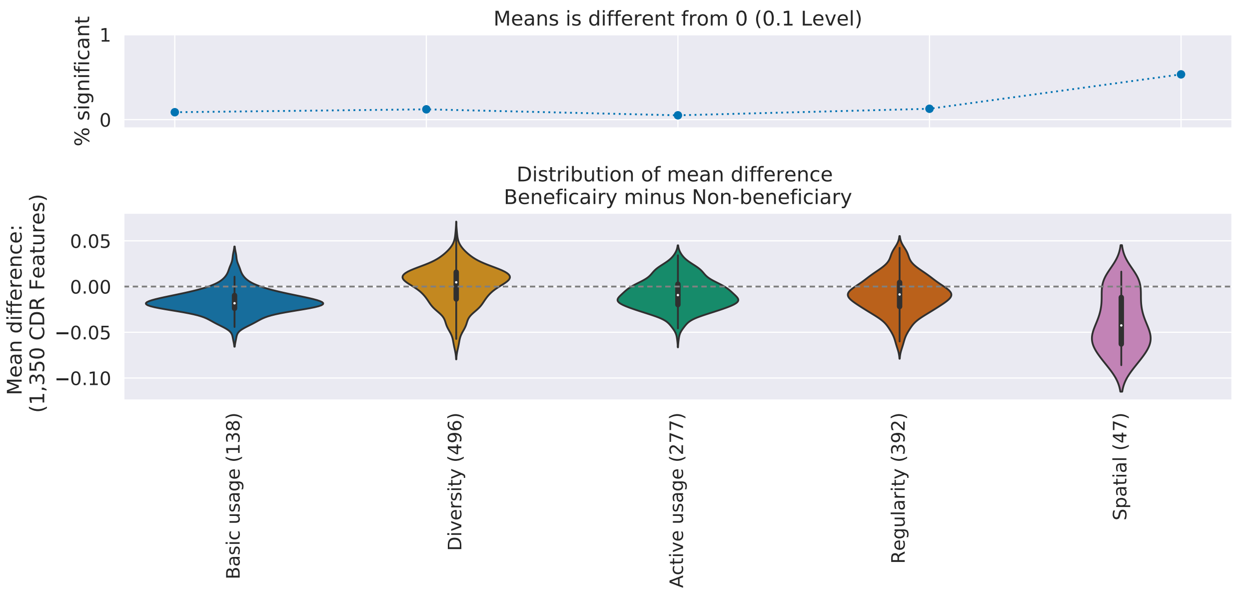
1.B.5 Additional Results Machine Learning

Figure 3B: Mean difference of CDR features by eligibility status:

(a) One month



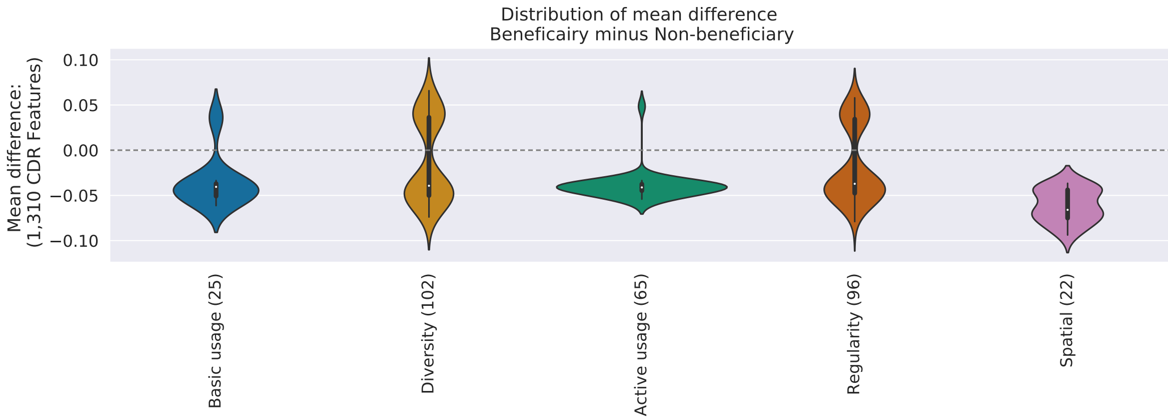
(b) Two weeks



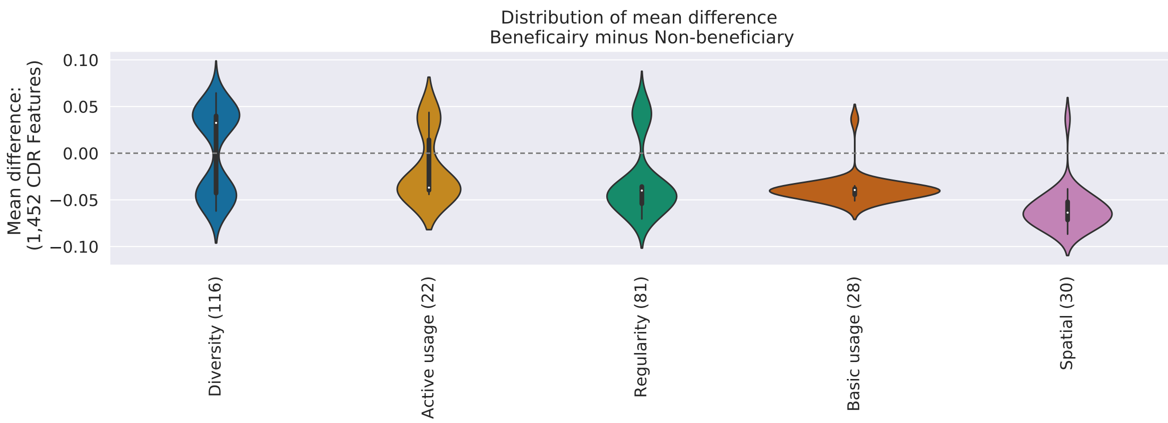
Note: Author's calculations using scorecard survey sample with records matching CDRs. Violin plot shows the distribution the mean differences between the average of beneficiaries and non-beneficiaries for each features' category. All variables normalized. A negative value indicates that for a feature non-beneficiaries have a higher average. Number of features on each group in parentheses.

Figure 4B: Mean difference of CDR features by eligibility status: Significant features only.

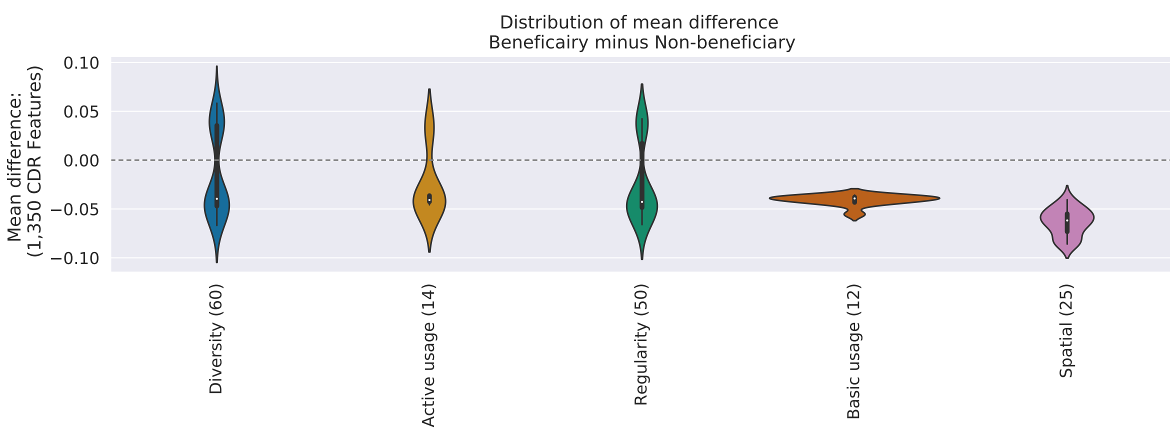
(a) Six months



(b) One month

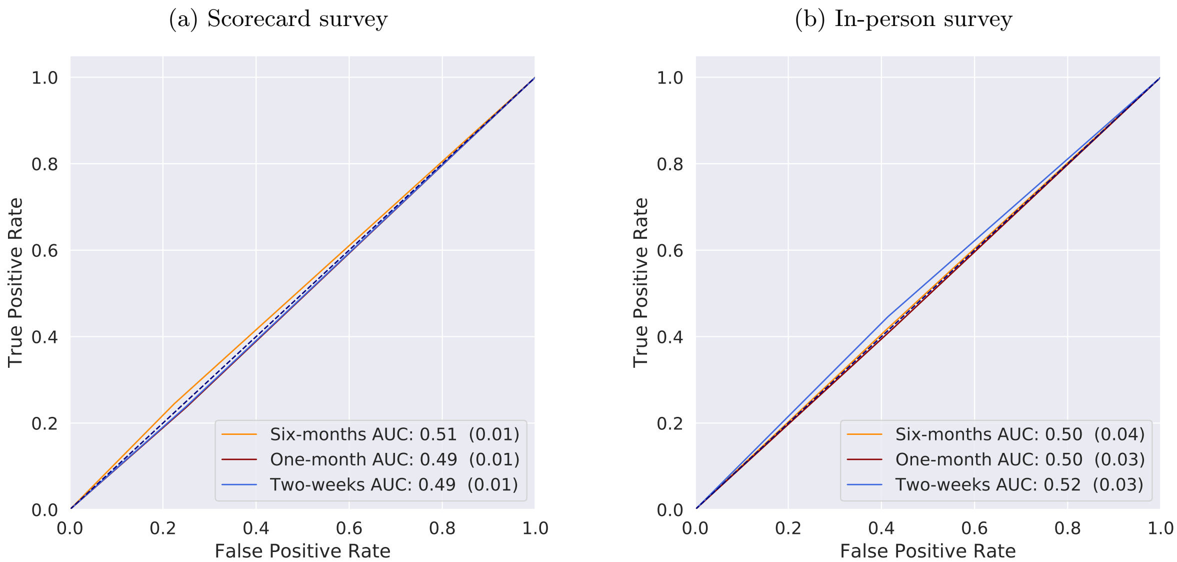


(c) Two weeks



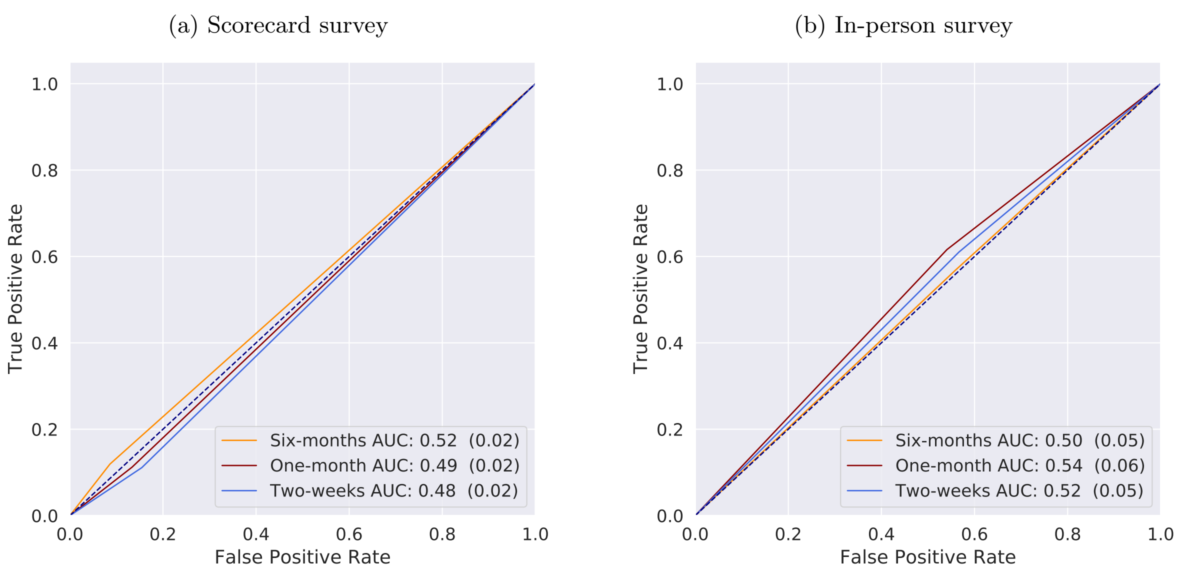
Note: Author's calculations using scorecard survey sample with records matching CDRs. Violin plot shows the distribution the mean differences between the average of beneficiaries and non-beneficiaries for each features' category. All variables normalized. A negative value indicates that for a feature non-beneficiaries have a higher average. Number of features on each group in parentheses.

Figure 5B: ROC Curves for CDR-based targeting of beneficiaries



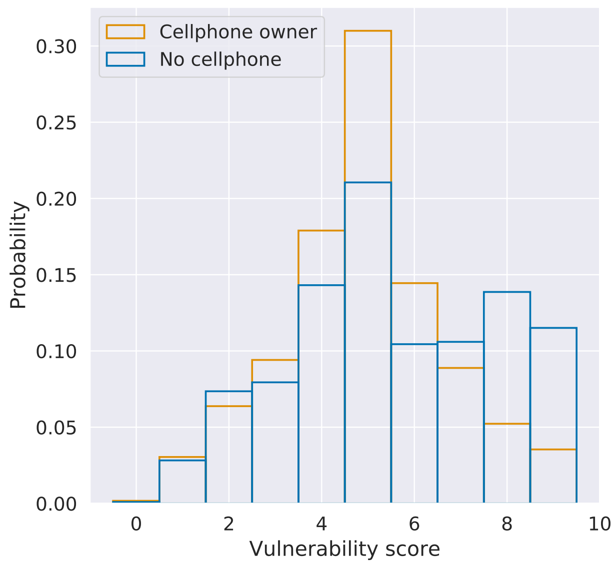
Note: ROC curves for classifying beneficiary status using CDR-data. Features extracted for the six-month time window for survey participants with a valid phone.

Figure 6B: ROC Curves for CDR-based targeting of beneficiaries: Restricted sample



Note: ROC curves for classifying beneficiary status using CDR-data for survey participants at the tails of the distribution of the vulnerability scores. Features extracted for the six-month time window for survey participants with a valid phone.

Figure 7B: Distribution of the vulnerability score by cellphone ownership



Note: Author's calculations using scorecard survey. A higher vulnerability score make a household more likely to be eligible for the program.

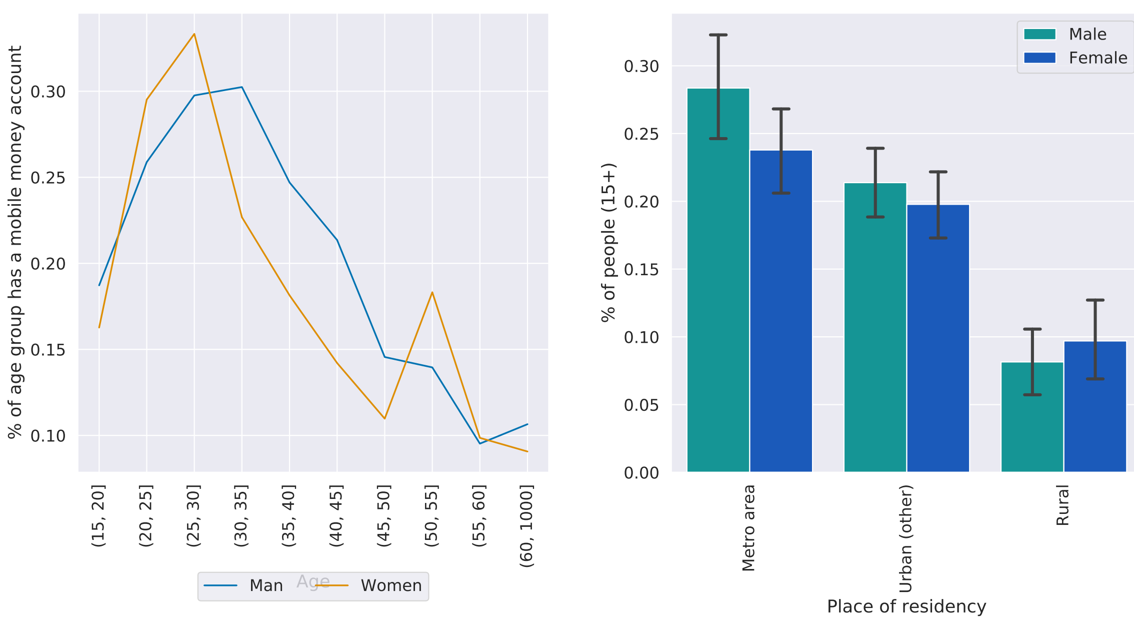
Appendix for Chapter 3

Additional figures and tables

Figure 1C: Mobile money account ownership

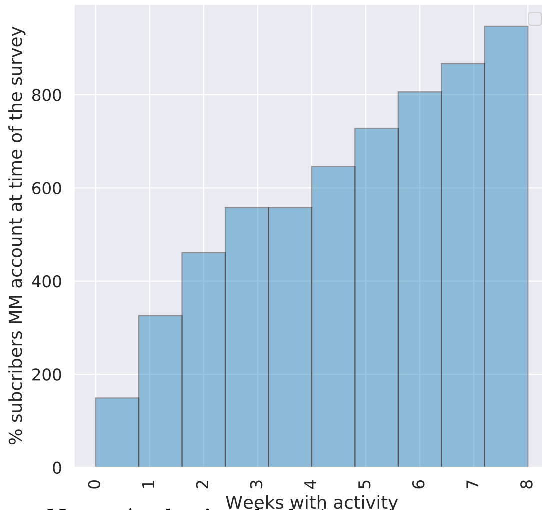
(a) By age

(b) By place of residency



Note: Author's calculation using Finscope 2018.

Figure 2C: Weeks active



Note: Author's calculation using own survey. For each participant in the survey who has a MM account, the figure includes activity during the eight weeks prior to the survey.

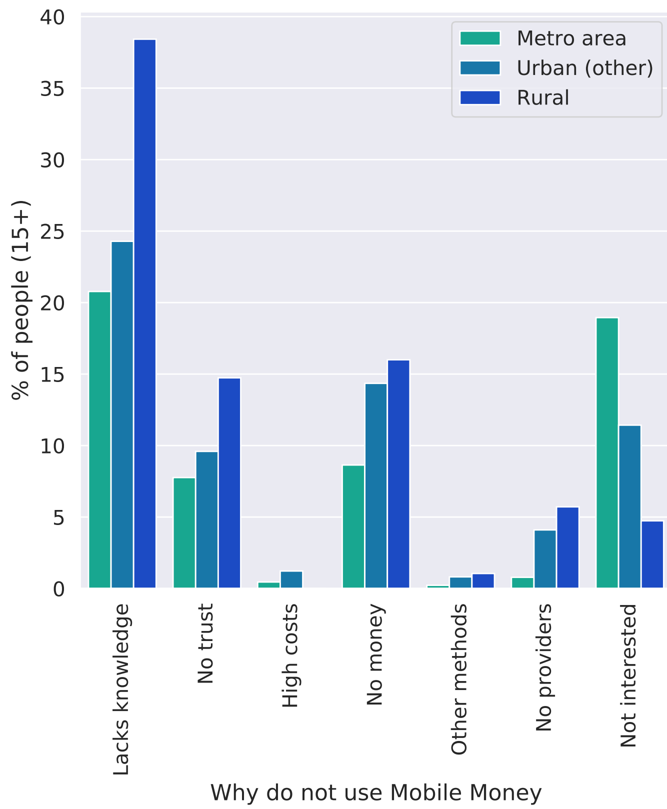
Table 1C: Effect on adoption of mobile money

	Open account	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.054***	0.014*	0.015**	0.009	0.006	0.009	0.001	0.000
	(0.015)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,314	1,314	1,314	1,314	1,314	1,314	1,314	1,314
R^2	0.021	0.013	0.016	0.018	0.007	0.008	0.013	0.010

*p<0.1; **p<0.05; ***p<0.01

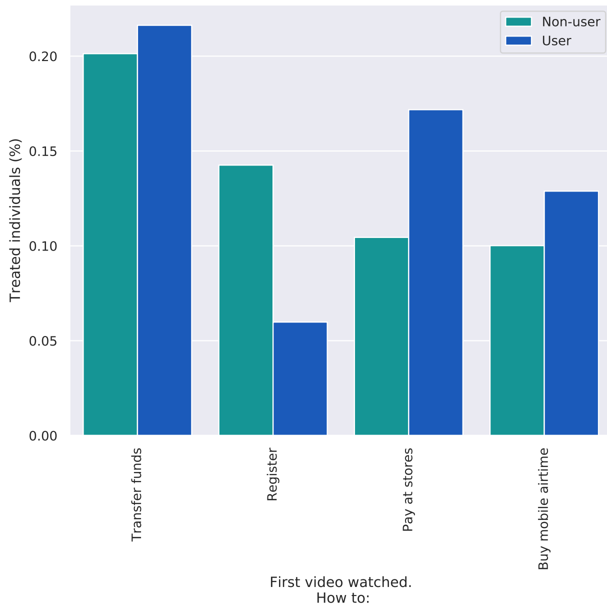
Note: Includes only survey participants without an account at the time of the survey. Control variables include age, gender, working on the previous week, usage of smart phone, education level, and a dummy for the week the survey took place.

Figure 3C: Why people do not adopt mobile money



Note: Author calculations using Finscope 2018. Includes answers from respondents that recognize any of the two brand names, currently own a cellphone, and at the time of the survey did not have an account. Lack of trust includes people that claim they do not use mobile money because they do not trust the cellphone provider or the electronic transfer.

Figure 4C: First video seen



Note: Includes treated individuals only. Percentages add up to the total of people that watched any video.