Title
Economic Indicators and Social Networks: New approaches to measuring poverty, prices, and impacts of technology

Author
Keleher, Niall Carrigan

Publication Date
2019

Peer reviewed|Thesis/dissertation

Economic Indicators and Social Networks:
New approaches to measuring poverty, prices, and impacts of technology

by

Niall Carrigan Keleher

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Information Management and Systems

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Dr. Joshua Evan Blumenstock, Chair
Dr. John Chuang
Dr. Jeremy Magruder

Fall 2019

Abstract

Economic Indicators and Social Networks:
New approaches to measuring poverty, prices, and impacts of technology

by

Niall Carrigan Keleher

Doctor of Philosophy in Information Management and Systems

University of California, Berkeley

Dr. Joshua Evan Blumenstock, Chair

Collecting data to inform policy decisions is an ongoing global challenge. While some data collection has become routine, certain populations remain difficult to reach. From targeting social protection programs in densely-populated urban areas to reaching the "last mile" of infrastructure coverage, data collection and service delivery go hand-in-hand. Understanding the populations that live in urban communities as well as remote villages can help to tailor the design, targeting, and implementation of development programs. New sources of information have the potential to improve awareness of the needs and preferences of individuals, households, and communities.

The goal of this dissertation is to provide multiple vantage points on the role that data, community input, and individual preferences can play in informing development policy. The empirical investigation presented in this dissertation covers two studies in Liberia and one in the Philippines. The unifying theme of the three chapters is the exploration of new sources of information about hard-to-reach populations.

In the first chapter, I seek to describe and explain how community members would prefer to see a cash grant program targeted. Cash grant programs are widely popular. However, targeting of these, as well as other social protection programs in densely-populated urban areas, is a challenging undertaking. I take a first principal's approach to assessing individual preferences for targeting. I find that individuals express a clear preference for selecting community members that are similar to themselves. This holds true in the first stage of the study when I asked people to nominate knowledgeable community members to target a social protection program. It also holds true when I asked community members to target a cash grant. The presence of homophily, widely observed in social networks, is an important factor to consider when leveraging private information from individuals.

In the second chapter, I present an empirical analysis of the determinants of cellular network adoption in the context of community cellular networks. The networks were installed in seven remote locations of the Philippines. Prior to the study, these locations had been overlooked by mobile network operators thus did not have reliable mobile phone service. I leverage a unique scenario where rich socio-economic data were collected prior to the installation of the cellular

networks. Using this data, I examine demographics, economic welfare, and access to information prior to network launch. To examine determinants of network adoption, I present an empirical investigation of the household characteristics that correlate with cellular network adoption on the extensive (any usage) and intensive margins (volume of calls and texts). I find that wealth is a key driver of network usage. Social network position, however, does not appear to influence cellular network usage. Taken together, the findings of Chapter Two present encouraging evidence for the potential for cellular networks in remote localities as well as a cautionary tale of the potential for cellular networks to advantage wealthy households via greater access to outside social networks.

The third chapter focuses on another angle of understanding hard to reach communities. The challenge of collecting high-quality, timely data on prices is at the forefront of assessing and responding to microeconomic and macroeconomic conditions throughout the world. By using high-frequency data collected through a mobile application, I analyze tens of thousands of individual price observations collected at hundreds of locations in Monrovia, Liberia. I show that these data can be used to construct composite market indices that mirror government price indices.

As a whole, the chapters of this dissertation are intended to push the edges of how researchers and policymakers approach the understanding of social networks and economic indicators in urban and rural localities.

Dedicated to:

my best friend and companion,

Jessica Ruth Kiessel;

my source of joy,

Myles Eanne Kiessel Keleher;

and

my motivators,

Kathleen Carrigan Keleher
&
Brendan Stephen Keleher.

Contents

Acknowledgments

Throughout my dissertation, I have benefited immensely from the advice and of my advisers, colleagues, and friends. I am especially grateful to Josh Blumenstock for his support throughout my studies. Josh pushed me to expand the scope of my research and focus on the quality of my analysis and writing. Jeremy Magruder and Lori Beaman have provided much-appreciated mentorship and advice, especially on the content of chapter 1 of this dissertation. I am grateful for colleagues like Jeremy and Lori. John Chuang was a reliable source of ideas and a broader perspective that undeniably improved my development as a researcher. Steven Weber provided valuable comments on my dissertation proposal. Colleagues and friends at the University of Washington Information School and the University of California Berkeley School of Information provided ideas, motivation, and encouragement that I will forever appreciate. I had the great pleasure of getting to know and learning from Muhammad Raza Khan and Guanghua Chi during our time in Seattle and Berkeley.

The research summarized in this dissertation would not have been possible without the work and dedication of colleagues at various other organizations and companies. I am especially indebted to Innovations for Poverty Action (IPA). Throughout the past 11 years, I have learned a great deal from IPA researchers of all levels of experience and concentration. The tireless and often thankless work of research associates, research managers, and field officers is the seed that makes high-quality field research possible. I am grateful to have had the opportunity to bring their efforts to a publicly-available document.

Finally, my work and this dissertation would not have been possible without the continuous support of family and friends. Without the encouragement of my wife, Jessica, I would have struggled to get out of various ruts along the way. She helped me to see a bigger picture and relevance to my work. My son, Myles, provided the best motivation to push forward towards my next stage. And, of course, he reminded me to play. My brother, Eamon, is a hero to me and our family. Without him, my upward vision would be unfocused. My parents, Kathleen and Brendan, taught me to explore ideas, approach questions with empathy, and to appreciate life. Their influence touches every aspect of my life and work.

Chapter-specific acknowledgements:

Chapter One:
Ellen Bates-Jefferys was integral to the successful implementation of an intricate research design. Ellen provided valuable suggestions and recommendations that undoubtedly improved the project and quality of data collected. Laura McCargo, Newton Toe, Saye Zawolo were instrumental in managing data collection in Monrovia. I am also grateful to Paloma Bellatin for her oversight of the cash grant implementation. A special thanks are owed to Arja Dayal and Osman Siddiqi who helped me to navigate project planning and administrative considerations. I am grateful that Arja and Osman didn't give up on the project even when we seemed to be nearing a dead end. Tala Ismail, Joseph Kamara, Erica Kuhlik also provided valuable input during field testing and piloting. Colleagues from the Ministry of Gender, Children, and Social Protection and Mercy

Corps provided important feedback on the research design to increase the broader relevance of the research. Data used in selecting study locations were provided by the Liberia Institute of Statistics and Geo-Information Services (LISGIS) and the World Bank. Funding was provided by the International Initiative for Impact Evaluation, International Growth Centre, and Innovations for Poverty Action. I collaborated with Lori Beaman and Jeremy Magruder on the research design for the project discussed in Chapter One.

Chapter Two:
The team from Innovations for Poverty Action were critical to the smooth implementation of the research discussed in Chapter Two. Allan Lalisan provided excellent leadership and recommendations during the baseline and implementation periods. I am grateful to Allan for his ability to find workable solutions to challenging problems that the remote nature of the project entailed. Ana Sheila Pacris, Yuna Liang, and Ann Mayuga provided incredible support during the endline survey. Nassreena Sampaco-Baddiri was a vital ally in the successful execution of research activities in the Philippines. Implementation of the project would not have been conceivable without the tireless efforts of Mary Claire Barela, Cedric Festin, Giselle Marie Dela Cruz, Philip Martinez, Ronel Vincent Vistal, and other team members at the University of the Philippines. Kurtis Heimerl and Eric Brewer provided intellectual leadership for the community cellular network technological innovations as well as funding for the cellular tower installations. Matthew Podolsky maintained a steady presence to ensure multi-institutional collaboration. I collaborated with Joshua Blumenstock, Arman Rezaee, and Erin Troland on the research design for the project discussed in Chapter Two. Funding for data collection was provided by USAID via NetHope and the Center for Global Action at UC Berkeley, and the Bill and Melinda Gates Foundation. Professor Blumenstock generously dedicated research funds from the University of California Berkeley Hellman Fellows Fund to the project.

Chapter Three:
Premise and the Liberia Institute of Statistics and Geo-Information Services (LISGIS) provided data for Chapter Three. Abe Tarapani and Joseph Reisinger provided valuable insight and guidance for understanding data collection and aggregation procedure employed by Premise. I collaborated with Joshua Blumenstock on the design, analysis, and writing of Chapter Three.

Introduction

Hard-to-reach populations often go unnoticed. This is especially true for the poor. Governments and NGOs face increasing challenges of identifying poor households in rapidly expanding urban slums. Global efforts to expand cellular and internet coverage grow increasingly difficult as the hardest-to-reach are the last to remain without connectivity. Building reliable economic indicators and statistics is also a challenge of access. Whether or not it is service delivery, infrastructure development, or government statistics, difficulty in reaching populations and obtaining information for policy decisionmaking is likely to be a persistent challenge for decades to come.

The challenge of collecting data that sheds light on hard-to-reach populations is one that I approach with great passion and urgency. I approach my research as a task to shed light on often-overlooked populations and ideas. Running through my research are the themes of social networks, economic indicators, and the design of development programs. I believe that more insights into the role that social networks play in economic behaviors will help researchers and policymakers to understand the interconnectedness of local communities as well as the potential for expanded connectivity. I believe that more insights into economic wellbeing not only document where things stand today but also help to focus priorities going forward.

As decisionmaking in economic and social development policy becomes more and more reliant on data, the excitement of new data sources is palpable. Nevertheless, the opinions of people, the constraints on households, and the limitations of various data sources should be taken into account.

This dissertation is intended to document and expound upon contexts where hard-to-reach populations are at the core of the policy problem. In the first chapter, I seek to describe and explain how community members would prefer to see a cash grant program targeted in Monrovia, Liberia. Cash grant programs are widely popular. However, targeting of these, as well as other social protection programs in densely-populated urban areas, is a challenging undertaking. In the second chapter, I present an empirical analysis of the determinants of cellular network adoption in the context of community cellular networks. The networks were installed in seven remote locations of the Philippines. Prior to the study, these locations had been overlooked by mobile network operators thus did not have reliable mobile phone service. The third chapter focuses on another angle of understanding hard to reach communities. The challenge of collecting high-quality, timely data on prices is at the forefront of assessing and responding to microeconomic and macroeconomic conditions throughout the world. I focus on urban locations in Monrovia, Liberia where collecting price data is particularly difficult.

In working on the research discussed in the chapters of this dissertation, I designed survey instruments to measure social network indicators in Monrovia and the Philippines. Taken together with survey modules on economic wellbeing and activity, I, along with colleagues, collected high-quality data that shed light on the relationship between social network position and economic wellbeing. Through descriptive analysis, I provide information about how economic poverty and social connectedness correlate. Often, I find that social connectedness is negatively correlated with economic welfare. That is, when we look at local social networks, the poor are well connected. This suggests the potential to leverage social connections. In the first chapter, I seek to test ways of leveraging social ties for targeting a cash grant. I find that individuals are able to identify poor households that they are socially connected to; however, the quality of information is biased towards over-estimating poverty of nearest neighbors. In the second chapter, I examine whether or not introducing a cellular network advantages those that are wealthy or those that are socially connected. I find that cellular networks may advantage the wealthy who were more likely to use the cellular network.

My research also leverages new sources of data. I have had the privilege of working with great colleagues, mentors, and collaborators on call detail records and high-frequency data collected through mobile phone applications. In the second chapter, I present analysis that relies heavily on call detail records from community cellular networks in the Philippines. In the third chapter, I include analysis of price data collected through a mobile phone application. The availability of these data makes detailed, temporal analysis possible. At the same time, approaching the data with humility and skepticism requires one to compare and contrast multiple sources of data. Throughout this dissertation, I seek to present data from multiple sources and contexts to help validate or invalidate what I find through new sources of data.

Chapter One

Targeting by Nomination

How homophily affects targeting of social programs

Abstract

Targeting beneficiaries for social programs in urban areas is increasingly important as urban populations grow. Many anticipate that poverty and emergency relief programs will become more prevalent in densely populated settings. However, current targeting strategies and tools may not be suited for dynamic urban environments. We provide evidence and recommendations for identifying beneficiary households/individuals for social programs in urban areas. We implemented a decentralized targeting mechanism through socially-knowledgeable members of urban neighborhoods of Monrovia, Liberia. We leveraged information from various types of community members to target an antipoverty program. The program implemented was a one-time unconditional cash grant to households. We sought to verify whether or not decentralized targeting of the unconditional cash grant is effective in reaching poor households and households that have experienced an economic or health shock. We find that community members are likely to recommend others with similar characteristics as knowledgeable agents and as ideal beneficiaries for the cash grant program. The presence of homophily in targeting preference is a key consideration that should be taken into account when eliciting information from community members. When asked about the welfare status of other households, agents provide accurate information for neighbors with whom they are socially connected.[†]

[†]The material in this chapter is based on joint work with Lori Beaman and Jeremy Magruder. Innovations for Poverty Action conducted the data collection and project management in Liberia.

1.1 Introduction

Effective targeting of beneficiaries for social programs, i.e., enrolling those who would benefit the most into any given social service or program, is a challenge around the world. In Sub-Saharan Africa and other nations in the Global South, targeting through existing social infrastructures, such as village chiefs and councils, and proxy means tests have become standard "best practices." However, as populations shift to dynamic urban environments, these strategies may no longer best address targeting needs. Social infrastructure may be more fluid or break down entirely and leaders may be less informed about their populace.

Targeting beneficiaries for social programs in urban areas is increasingly important as urban populations grow and poverty or emergency relief programs become more common in densely populated settings. Tools for targeting social programs in rural settings often leverage pre-existing social and political institutions to target beneficiaries. Without a better understanding of the social environment of urban areas, the effectiveness of community knowledge may break down in shifting, urban environments.

This paper aims to provide evidence and recommendations for identifying beneficiary households/individuals for social programs in urban areas. We implemented a decentralized targeting mechanism through socially-knowledgeable members of urban neighborhoods of Monrovia, Liberia. In doing so, we take a networks-based approach to targeting a social program. We seek to provide evidence to verify whether leveraging social connections leads to differential beneficiary selection.

Through an innovative, decentralized mechanism for reaching program beneficiaries, we investigated the channels through which needy households can be targeted for a social program. By collecting detailed and high-quality information at multiple stages of the study, we set out to examine whether decentralized targeting mechanism can be used more widely to leverage local community knowledge for identifying (a) key informants and (b) beneficiaries for social programs.

First, we elicited nominations from all households to identify socially-knowledgeable community members. Then, we leverage information from various types of community members to target an antipoverty program. The program implemented was a one-time cash grant to households. For the cash grant, we sought to verify whether or not this decentralized targeting of the cash grant is effective in reaching poor households and households that have experienced an economic or health shock and whether the decentralized targeting mechanism identifies households with a high return to the cash grant.

In designing this study, we worked with a sub-section of the Ministry of Gender, Children and Social Protection (MoGCSP), commonly called the "social protections pillar." This pillar has overseen several large-scale cash transfer programs that included monitoring components. UNICEF implemented the first large-scale unconditional cash transfer program in Bomi county. Beginning in 2009, it finished and completed a final evaluation in 2015, right as Ebola hit the nation.

A great deal of policy attention has shifted towards cash transfer programs, both in Liberia and internationally. Conversations with government agencies and donors active in Monrovia after the Ebola crisis suggest that cash transfers are an increasingly relevant policy option. The United Nations is currently implementing a cash transfer program in the country. Mercy Corps has expressed interest in reaching those directly affected by the Ebola crisis. GiveDirectly has established an office in Liberia. DFID has recently convened a panel on humanitarian cash transfers, suggesting that the timing for this project is very appropriate for testing variations of cash transfer programs after a humanitarian crisis. Nevertheless, little work has been focused on targeting of cash transfers in urban areas. As such, we feel that this research is timely and important to understanding how social programs—like cash transfers—can be designed, beneficiaries targeted, and cash delivered in urban locations.

At least 15 NGOs initiated unconditional cash transfer programs after the Ebola crisis. As the crisis ended, these cash transfer programs were brought to an end with no long-term replacement or policy solution. However, the emergency-response cash transfer programs provided valuable insight to the government about both cash transfer program's potential and challenges. These previous programs also established an experienced cash transfer apparatus within the government. The MoGCSP is actively planning for implementation of a large-scale cash transfer program that will build on top of the lessons from the UNICEF program. While the geographic focus of the program will be rural, the MoGCSP has expressed a desire to tackle challenges of social protection within urban areas of Liberia. Additionally, GiveDirectly is beginning to plan for urban cash transfer programs (initially in Kenya).

1.2 Social Networks and Targeting

Traditional community-based targeting methods are manifest through two primary mechanisms. Community members are called to meet in a central location to deliberate who should be included in a particular social protection program. Communities may identify beneficiaries to receive direct benefits while also enforcing some form of social insurance through which benefits are shared. Alternatively, or sometimes in tandem with community meetings, local leaders are asked to select beneficiaries. Where local leaders have an inordinate amount of discretion, idiosyncratic targeting of benefits may result. Community meetings have the benefit of transparency and inclusiveness. All members of the community are invited to the meeting. Such meetings require considerable coordination. Group decisions are not immune to capture by vociferous members of the community. Idiosyncrasies of community-based targeting can lead to differences defining who should be considered poor and eligible for social protection programs.

In most countries where cash transfers are implemented, proxy means tests are the primary method for targeting cash transfer beneficiaries (e.g., Fernald, Gertler, and Neufeld (2008) and Fiszbein and Schady (2009)). Rai (2002) presented early evidence of the role that community information plays in targeting social programs. Alatas, Banerjee, Hanna, et al. (2012) conduct an experiment in 640 Indonesian hamlets to compare targeting outcomes of community-based, proxy-means, and hybrid targeting schemes. They find that targeting through a proxy means test reduces error rates (combined errors of inclusion and exclusion). However, the difference in error rates does not fundamentally change the effect of the social protection program on poverty rates. Moreover, they find that communities view community-based targeting as more legitimate and transparent than the proxy-means test.

In a experiment that targeted a cash transfer program in Niger, Premand and Schnitzer (2018) compare three primary targeting methods: (1) Community-based targeting (CBT), (2) Proxy-means Tests (PMT), and (3) a food-insecurity metric (FCS). The authors find that PMT is more likely than the other two methods to direct program benefits to consumption-poor households. However, the methods are indistinguishable in targeting household based on assets, income, and subjective well-being. Legitimacy and perceived accuracy across the three methods were equivalent.

In a non-experimental study Basurto, Dupas, and Robinson (2017) show that community leaders in Malawi are prone to more errors in targeting relative to a PMT. Nevertheless, the social benefit of the program is enhanced by the local leaders by directing benefits to high-productivity households. Through an experiment in Zambia, Schüring (2014) find that individuals charged with the responsibility of determining cash transfer beneficiaries are no more likely to give to close relatives. Niehaus et al. (2013) caution that targeting through a proxy means test can be problematic when those charged with verifying the proxy means are corruptible.

This paper fits in with a broad literature on how targeting through specific individuals within a community alters the distribution of benefits. Identifying important nodes within a network is an important topic of research within the field of social and economic networks. Adamic and Adar (2005) discuss methods of searching a social network to identify specific nodes within the network. They test search strategies driven by high-degree individuals (people that communicate with many others are more likely to know target nodes), individuals with similar characteristics as the target nodes, and geographic proximity to the target nodes. Ballester, Calvó-Armengol, and Zenou (2006) posit that power, or as they term it "intercentrality," defines key nodes within a network. The importance of nodes is not only determined by their centrality but also by how much they alter the centrality of their neighbors. Relevant to community-based targeting, Fafchamps and Labonne (2019) show that benefits of a decentralized social protection program in the Philippines were more likely to go to households with high betweenness centrality, suggesting that benefits are used to form coalitions for electoral support.

This paper seeks to contribute to empirical evidence of the importance of homophily in economic behavior. By taking advantage of network structure for more efficient forms of search within a network, this paper builds off of fundamental contributions from Milgram (1967), Kleinberg (2000), and Watts, Dodds, and Newman (2002). These seminal works provide inspiration for efficient search strategies that leverage social connections and sub-graph clustering. Golub and Jackson (2012) present model that explains the importance of homophily in producing correct, aggregate information. Bloch and Olckers (2018) point to the importance of network structure in the ability to target in networks. They suggest that in the absence of complete ordinal rankings, incentive-compatible and efficient targeting mechanisms are possible when social networks are bipartite or defined by a triangle. Galeotti, Golub, and Goyal (2017) show that strategic behavior is important when targeting in networks. Baumann (2016) presents a valuable mechanism design for our paper in seeking to target a "prize" to the most valued node in a network.

Much of the research on the importance of social networks concentrate on the spread of behavior within the network (See Kitsak et al. (2010) and Aral and Walker (2012)). Within development economics, several studies have investigated the role of key nodes in the adoption of agricultural technologies (Beaman et al. 2018; T. G. Conley and C. R. Udry 2010; T. Conley and C. Udry 2001) and insurance (Cai, De Janvry, and Sadoulet 2015).

Most closely connected to this paper are a handful of studies that examine how network position influences targeted social protection programs. Banerjee et al. (2014) demonstrate that residents of 35 villages in India were able to identify highly central people in their community. The authors asked people to nominate others to help spread information about a loan product or an entertainment event. Banerjee et al. (2014) show that individuals that are nominated and leaders in their community are also highly central in the social network.

Alatas, Banerjee, Chandrasekhar, et al. (2016) surveyed nine households in each of 631 villages in Indonesia to assess the welfare of other households within the same community. Surveyed individuals were asked to rank their household and the other eight surveyed households based on household wealth. The researchers use survey data to quantify the actual wealth ranking of households, as measured through a 28-question proxy means test and a self-assessed welfare question. They also collected social network data to identify connections within the community. The authors show that the quality of the assessment deteriorates with social distance. Nevertheless, the authors show that knowledge of community members is relatively efficient in identifying poor households when compared to a proxy means test.

Utilizing social network data, Kim et al. (2015) test three different targeting methods to encourage adoption of multivitamins and a water purification treatment in 32 Honduran villages. In one set of villages, randomly-selected individuals were provided a health product (either multivitamins or water purification treatment). In the second group of villages, individuals with the highest in-degree centrality, as measured through a social network census, were approached for the health intervention. Finally, a third group was identified by asking members of the randomly-selected individuals to nominate a friend to receive the health product. The researcher then tracked purchased of the health products to assess which means of targeting led to the highest adoption rates. The researchers found that the most effective means of targeting the multivitamin supplement was the nominated-friends method. High-degree individuals were no more effective than random members of the community in promoting the adoption of the health products. The adoption rates of the water purification treatment were indistinguishable across the three targeting methods. (Also see Vera-Cossio (2017) and Shakya et al. (2017) for similar analysis in the same context.)

In a highly-relevant study by Hussam, Rigol, and Roth (2018), examine how well community members predict the marginal returns to capital of their neighbors. In an experiment with 1,345 entrepreneurs, the authors asked individual entrepreneurs to rank four to six other entrepreneurs within their neighborhood on predicted marginal returns to capital. Following personal interviews, the researchers distributed USD$100 grants to one-third of entrepreneurs. They find that entrepreneurs that are predicted to be high-performing yield two to three times the return to capital as the average entrepreneur.

1.3 Research Design and Data

We generated rich data sources on household welfare and social connections within 13 densely-populated neighborhoods in Monrovia. The study entailed five stages. First, we canvased thirteen neighborhoods in three selected communities of Monrovia. A team of surveyors from Innovations for Poverty Action (IPA) conducted a household listing of 2,656 households. We successfully completed 2,434 listing surveys (92% survey rate) in which all adult members of each household were listed through a household roster. Second, we returned to each household within a two-week period to conduct a comprehensive socio-economic baseline survey with heads of households or their partner. We completed 2,253 household surveys. Thus, we have baseline information on socio-economic information for 85% of households in the study area. The listing and baseline surveys were conducted between February and April 2018 (See Figure A.1.1 for the full project timeline). Third, we invited individuals from a fifteen-percent sample of households to participate in a "Targeting Survey" to assess knowledge of the welfare of households in each community as well as elicit nominations for a cash grant program. Fourth, IPA attempted to deliver a USD$80 cash grant to 280 households. Finally, we carried out an endline survey of the targeted cash grant beneficiaries as well as a matched-pair control household.

Baseline survey and social network census

Table 1.1 displays the number of households identified in each of the 13 blocks and the proportion of households surveyed in the household listing. The size of blocks ranges from 73 to 408 households (mean 204, s.d. 95.2). Survey completion rates for the baseline ranged from 78.1% to 91.2%. In eleven out of the thirteen blocks, we managed to complete a baseline survey with over 80% of households. As such, we believe that the baseline data provides a near-comprehensive perspective of household-level economic conditions and social networks within the 13 community blocks identified for the Leveraging Social Connections study.

The baseline survey included questions about subjective well-being, household expenditures, assets ownership, health and economic shocks in the past 12 months, and inter-household social interactions (i.e. a social network survey). From the asset information, we are able to calculate a standard proxy means score that mirrors the Poverty Probability Index and the Government of Liberia's PMT.

Among a ten-percent sample of households, we conducted a full household expenditures module. This module mirrored the expenditure module used by the Liberian central statistical office, LISGIS, in the 2014–15 Household Income and Economic Survey (HIES).[1] We carried out the full HIES expenditure module in order to assess and calibrate a reduced-length ("simple module") household expenditure module which was conducted among 100% of households in the baseline survey.

Through a social network census, we took painstaking efforts to identify intra-block social ties, even if a connected household was not available for the baseline survey. Our goal was to

[1] Survey instruments accessed from `http://microdata.worldbank.org/index.php/catalog/2563/study-description`.

Table 1.1: Count of Households by Neighborhood

Neighborhood	Households	Surveyed households	Adults listed	Prop. surveyed at baseline
C1	150	127	361	0.847
C2	251	222	536	0.884
C3	321	253	638	0.788
L1	408	372	1118	0.912
L2	227	186	514	0.819
L3	133	109	299	0.820
L4	190	160	351	0.842
W1	303	248	678	0.818
W2	190	160	401	0.842
W3	102	89	227	0.873
W4	159	138	361	0.868
W5	73	57	166	0.781
W6	149	132	347	0.886

identify current and meaningful social ties between surveyed household and other households within the community. With our full baseline sample, we elicited social network connections using the following question:

> *For the next questions I ask you, the answer can only be people who live in this same block with you.*
>
> *They also must be older than 18. I will ask you to name the 5 people to answer my question.*
>
> *Who did you spend time with the most because you wanted to spend time with them, in the last 14 days?*

We asked respondents to name up to five social ties within the community. Using this method, we successfully identified a total of 7,603 social ties within the 13 study blocks. The social network data allow us to perform detailed social network analysis within each block. In addition, we are able to compute measures of social network centrality for each household.

Nomination of knowledgeable community members

During the baseline interview, we asked each survey respondent to nominate one member of their community to assist with targeting a social assistance program. We elicited community member nominations by one of the following questions:

1. *If we want to spread information about a social assistance program, to whom do you suggest we speak?*

2. *If we want to identify which people would be best to help us identify which people in this block would benefit most from a social assistance program, to whom do you suggest we speak?*

3. *If we want to identify which people would be best to help us identify which people in this block would benefit most from a cash gift for social assistance, to whom do you suggest we speak?*

For each baseline respondent, we randomly selected one of the three questions listed above. The purpose of randomizing the elicitation questions is to quantify the extent to which nominations are influenced by the potential of a cash grant or other social assistance. 92 percent of households provided a nomination for the targeting assistance.

Targeting assistant survey

After the household census, we invited individual community members to complete a one-on-one interview with an IPA staff member. Invited community members were drawn from 15 percent of households interviewed in the household census. We refer to the individuals invited for one-on-one meetings as "Targeting Assistants" (TAs) as they were asked to provide input to help determine how the unconditional cash grant would be distributed among households living within the community block. We refer to the one-on-one interview as a "targeting survey" as the interview aims to elicit the beliefs and preferences of the Targeting Assistants.

TAs were selected through one of three methods. Households with the greatest number of nominations from baseline respondents were included as "Nominated TAs". We also invited leaders from the neighborhood as "Leader TAs". Finally, "Random TAs" were drawn randomly from adults within the neighborhood. In total, 394 TAs participated in the targeting survey. More than three-quarters of TAs were under the age of 45. The average age of randomly selected TAs (32.6 years) was lower than the average age of the nominated TAs (40 years). Gender split of random and nominated TAs was the same; 53 percent of all TAs were women.

Table 1.2: TA Types

	All N=394	Un-Nominated N=140	Nominated N=254	p-value
TA is female	0.53 (0.50)	0.54 (0.50)	0.52 (0.50)	0.715
Number of block members for TA Survey	30.9 (6.05)	30.8 (5.81)	30.9 (6.19)	0.801
TA understood the targeting survey	0.95 (0.23)	0.94 (0.25)	0.95 (0.21)	0.496

We randomized the way in which the targeting survey was introduced. One of the following two paragraphs was randomly selected to be read to the targeting assistant:

> **Poverty Framing (Option A):** *The goal of this program is to reach the poorest households in this community. We would like to request your assistance in deciding how to reach the poorest households in this community.*

OR

Shocks Framing (Option B): *The goal of this program is to reach the households affected by a major loss of wealth (fire, flood, theft) in this community. We would like to request your help in deciding how to reach households affected by a major loss of wealth (fire, flood, theft) in this community.*

Assessing targeting assistant knowledgeable of neighbors

During the targeting assistant survey, TAs were asked about a random sample of adults living in their neighborhood. We asked if the TA knew each of the individuals mentioned and, if so, we asked them for their subjective beliefs about the other individual's welfare. Specifically, we asked the TAs to express their belief about whether or not other households in the neighborhood were among the poorest 20 percent of households in the neighborhood.

Half of the targeting assistants heard the first framing (poverty) and the other half heard the second framing (shocks). The variation was important to understand whether or not framing the motivation changes the way in which targeting assistants assess the welfare of other members of the community.

In addition to questions to assess the TAs familiarity with other members of the community, TAs were asked the following two questions about the relative welfare of randomly selected households within their community:

Question 1: *Please imagine a 5-step ladder. On the bottom, the first step stand the poorest 20% of households in* community name*. On the highest step, the fifth stands the richest 20% of households in* community name*. Where do you think* name*'s household stands when you think of how poor or wealthy* name*'s household is compared to others in* community name*? (Ask the respondent to provide a number 1–5)*

Question 2: *If we asked other people in your community, what percentage would say that* name*'s household is on the first step of the ladder, that is in the poorest 20% of households in* community name*?*

We refer to the answers to the questions above as "votes" in the sense that TAs are providing informed beliefs about their neighbors. IPA surveyors facilitating the targeting survey were asked to assess the level of comprehension of each TA during the targeting survey. By in large, TAs understood the questions and tasks presented to them. We observe no difference between nominated and non-nominated TAs in their level of comprehension.

Eliciting votes for cash grant beneficiaries

We also asked each TA to nominate two households for a cash grant program. For each TA, we randomly varied the ordering of the following two questions:

> **Question 1:***Think of households within this block, which household would make the most use out of a cash grant?*

> **Question 2:** *Within this block, is there a household who recently fell on hard times and would benefit most from a cash grant?*

We required that a nominated beneficiary be identified within the baseline database of adults residing in the community. 135 households were nominated by a TA that said the household would make the most use of a cash grant. 144 households were nominated because the TA believed the household had fallen on hard times. 44 households received nominations through both questions.

Distribute cash grants

We conducted a matched-pairs design to establish a treatment and control group for the evaluation of a cash grant. First, we organized households that were nominated in the targeting survey to receive a cash grant into 8 strata. These strata were split by the type of targeting assistant (nominated leader (NL), randomly selected leader (RL), nominated non-leader (NTA), randomly selected non-leader (RTA)) and the type of beneficiary nomination (households that would make the most use out of a cash grant (Best) or households that recently fell on hard times (Hard)) among the nominated cash beneficiaries.

To give all poor households a non-zero probability of receiving a cash grant, we randomly selected households in the bottom 20% of the PMT score distribution to receive the cash grant. This random selection of household comprises the ninth strata.

Within each stratum, we identified matched pairs of households. We matched households on household size, the gender of household head, years in the community, in-degree centrality, betweenness centrality, per capita expenditures, PMT score, health and economic shocks, and the percent of TAs that knew the household in the targeting survey.

Our matching algorithm produced 560 pairs. Within each pair, we randomly selected one household to receive a cash grant. Beneficiaries of the cash grant drawing were eligible to receive USD$80. One member of each selected beneficiary household was invited by phone to collect their cash grant at the IPA offices in central Monrovia. Cash grant beneficiaries were required to verify their identity (name, age, gender, phone number). Cash grants were only be distributed to verified members of the urban neighborhoods included in the study. Households were eligible for a maximum of one cash grant.

Cash Transfers were delivered between July 18 and 25, 2018. The recipients were assigned specific slots of time within these days. We developed a SurveyCTO form to verify the information of beneficiaries when they arrived at the IPA office.

Upon arrival to the IPA office, a surveyor verified a photo ID of the recipient. Before receiving the cash grant, the recipients were informed of three important aspects:

1. The cash grant was a free gift. Beneficiaries would not have to pay anything back to IPA or anyone else; they could use the money however they wish.

2. It was a one-time cash grant. It would not be delivered every month or every year; only this time.

3. The cash grant amount and the beneficiary would remain confidential on our side. IPA would not tell anyone in their community nor the government that they have received the cash grant (though they were free to tell whomever they wanted).

In total, 269 of the targeted 280 households collected the cash grant. Of those that did not collect the cash grant, five had moved away from Monrovia or were deceased and six refused to or were unable to visit the IPA office within the allotted time (3 weeks) to collect the cash grant.

Follow-up data collection.

Table 1.3: Panel Sample, Response Rates

	Control N=280	Treated N=280	p-value
Baseline Response Rate	0.97 (0.18)	0.98 (0.15)	0.433
Endline Response Rate	0.96 (0.19)	0.97 (0.18)	0.650
Panel Attrition	0.04 (0.20)	0.03 (0.18)	0.632

Notes: For the full population across all 13 neighborhoods, the baseline response rate was 84.8%. Attrition indicates whether a household surveyed at baseline was not surveyed at endline.

Endline data collection took place between August 6 and 19. A total of 580 households were sampled for the endline survey: the 280 selected cash beneficiaries and their 280 matched pair households that would serve as a control group. Recipients of the cash grant were not obligated to participate in the endline survey. We completed surveys with 540 of the targeted 560 households (Non-response rate = 3.5%). Four households refused consent while 16 could not be tracked due to relocation outside of Monrovia or absence for the duration of the endline survey. Ultimately, we have complete endline data for 261 matched pairs (522 households).

Through the endline survey, we asked about labor activity of household members, subjective well-being, household expenditures, household enterprises, asset ownership and purchases, health and economic shocks in the past 12 months, and lending and borrowing outside of the household.

As shown in Table 1.3, we find no differential attrition across treatment groups. Column 1 is a check for any differential response rates at baseline. Column 2 shows that there was no differential response at endline by treatment group. Moreover, Column 3 reveals that assignment to the treatment and control group did not affect whether or not we were able to collect panel data.

1.4 Evidence

Welfare Measures, Shocks, and Social Networks

Machine learning approach to estimating expenditures

As outlined in Section 1.3, we listed all 2,656 households in 13 neighborhoods of Monrovia. We collected basic identifying information about all households, we collected demographic information on all adults in 2,434 households, and we completed a full socio-economic baseline surveys in 2,253 households. During the baseline survey, we implemented multiple survey modules to measure household welfare. A leading approach to measuring welfare is to conduct a detailed expenditure module. Among a ten-percent random sample, we conducted a the household expenditure module from the 2014–15 HIES with 257 households. Among the ten-percent sample, mean per capita expenditures using the full expenditures module were 244.39 Liberian Dollars (USD$2.36 in purchasing power parity (PPP)). We also implemented a reduced-form expenditure module in which we collected data on aggregate expenses across five expenditure categories: food, clothing, health, school, and energy. Using the reduced-form expenditure module, among the same ten-percent sample of households we estimate mean per capita at 230.05 Liberian Dollars. The reduced-form module under estimates per capita expenditures by an average of 6 percent. The "simple module" version of per capita expenditures is correlated with the full module estimate of per capita expenditures[2]; however, the simple module slightly over predicts per-capita expenses in poor households and underestimates expenditures in high-expenditure households. Using the simple module for the full sample, our estimate of per capita expenditures is 259.39 Liberian Dollars.

We sought to construct a more robust prediction of per capita expenditures by training a machine learning model to predict log per capita expenditures. Using the ten-percent sample that completed the full expenditure module, we used three-fold cross-validation to compare a linear model (elastic net) with a non-linear model (random forest). Our models included covariates for demographic, asset, subjective welfare, and expenditure categories. The random forest algorithm provided a lower RMSE, 0.759 compared to 0.792 from the elastic net algorithm.[3] We then imputed predicted per capita expenditures for the full sample. We will refer to the random forest prediction as "Predicted PCE". Figure 1.1 shows the relationship between the rank of households based on the Predicted PCE vs. the simple expenditure module. Rank is calculated as suggested by Athey and Imbens (2017), where households are ordered by their per capita expenditures then normalized to have zero mean.[4] We find that a one step increase in the rank of a household based on the simple module is associated with a 0.86 step increase in the rank of Predicted PCE. Our preferred measure of household welfare is the Predicted PCE since it captures long-run wealth differences through the inclusion of asset-based wealth and demographic characteristics.

[2]See Figure A.1.2.

[3]McBride and Nichols (2016) make the case for using the random forest algorithm to predict welfare.

[4]We use the following equation to construct the zero-mean ranks, $R_i = R(i;\ Y_i^{obs}, \ldots, Y_N^{obs}) = \sum_{j=1}^{N} \mathbf{1}_{Y_i^{obs} < Y_N^{obs}} + \frac{1}{2}(1 + \sum_{j=1}^{N} 1_{Y_i^{obs} = Y_N^{obs}}) + \frac{N+1}{2}$

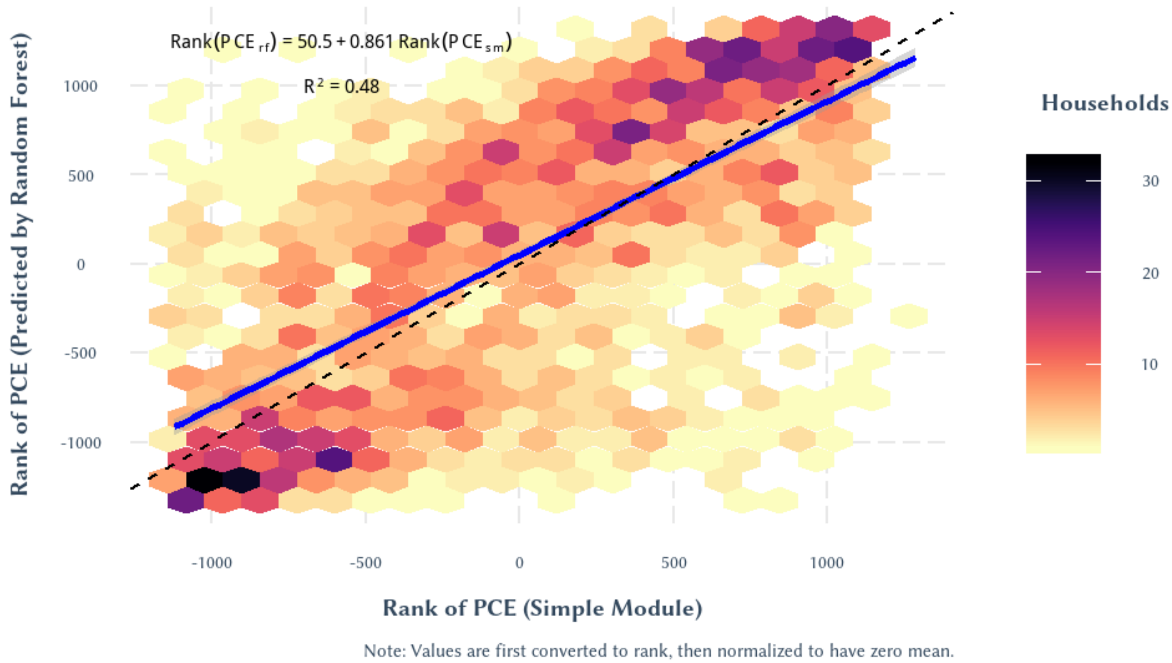

Figure 1.1: Per capita expenditure rank, simple module vs. predicted by random forest

In addition to our measures of household and per capita expenditures, we collected baseline data for three additional welfare metrics: an asset-based proxy-means test, a subjective welfare measure, and a measure of food security.

We collected asset information for the 10-question Proxy Means Test.[5] We worked with the research team at IPA to implement the method outlined in Kshirsagar et al. (2017) to construct a PMT specifically for urban Liberia.[6] Using a bootstrapped elastic net to select the combination of 10 variables that are consistently found to predict whether or not a household is below the national poverty line from the 2014–15 HIES, we arrived at a PMT module that asked about ownership of a telephone, chairs, wardrobes, computer, books, source of electricity, source of lighting, source of drinking water, household size, and region of the country. The PMT score for Monrovia ranges from 10 (the base score for the region in which Monrovia is located) and 100. Households at the lower end of the scale are predicted to be poorer than those at the upper end. The median

[5] Our proxy means test is based on the Poverty Probability Index (PPI), see `https://www.povertyindex.org`.

[6] We also construct a wealth index following the approach of Filmer and Pritchett (1999) and Kolenikov and Angeles (2009) we use the first principal component from a set of asset ownership as a proxy for household wealth. The resulting wealth index ranges from -3.60 to 2.92 with a mean of -0.60 and median of -0.64. The PCA is highly correlated with the PMT score generated by the PPI method. As such, we focus our discussion on the PPI-based PMT score.

household in our sample has a PMT score of 55.[7]

Additionally, we asked households for their subjective assessment of their wealth relative to other households in their community using a version of the Cantril Ladder method (Cantril 1966; Deaton 2008). We posed the following question to each household:

> *Please imagine a 5-step ladder. On the bottom, the first step stand the poorest 20% of households in Zone X, Block Y. On the highest step, the fifth, stands the richest 20% of households in Zone X, Block Y.*
>
> *Where do you think your household stands when you think of how poor or wealthy your household is compared to others in your Block?*

The median household considers itself to be on the second step of the Cantril ladder; 26 percent of households consider themselves to be on the lowest step of the ladder. Only 12 percent of households consider themselves to be on the top two steps of the ladder. We also asked about the number of meals per day that a household eats. The mean number of meals per day was 1.79.

In order to gain an assessment of existing social safety nets, we asked households about potential sources of financial support. We asked the following question about family members and friends separately:

> *If you needed money for anything (such as daily expenses, do business, fix house), how many FAMILY MEMBERS do you feel you could go to for help?* [8]

Very few households report that they could go to family or friends for financial assistance. The median has one family member and zero friends that they could go to for money. Only one percent of households report assistance (money, stipend, allowance, scholarships, food, or supplies) from a government entity and two percent of households report similar assistance from a non-governmental organization (NGO).

At baseline, we asked about four potential shocks at the household level:

1. *Has anyone in the household lost his/her main source of income in the last 12 months?*
2. *Has anyone in the household lost property or major assets because someone stole it, or because of fire or flooding in the last 12 months?*
3. *Did any members of our household have to stop working for a period of more than 2 weeks due to illness or injury in the past 12 months?*
4. *Did any members of your household die in the past 12 months?*

---

[7]Figure A.1.3 shows that the PMT score is weakly correlated with per capita expenditures, as estimated through the simple expenditure module.

[8]We used the same wording for FRIENDS.

Table 1.4: Baseline Summary Statistics, by PCE Quintile

	Population			**Predicted PCE Quintile**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	N=2656	N=532	N=531	N=531	N=531	N=531	p-value
Welfare:							
Household expenditure (LD)	914.02 (886.7)	594.38 (524.8)	789.15 (704.2)	860.43 (737.7)	1115.8 (1062.1)	1149.1 (1054.7)	<0.01
Per capita expenditure (LD)	259.39 (284.26)	106.39 (100.83)	158.77 (153.26)	188.80 (150.87)	289.00 (271.29)	492.64 (382.19)	<0.01
Predicted PCE (LD)	169.71 (89.76)	91.30 (18.88)	123.70 (5.28)	143.12 (7.34)	180.60 (15.33)	309.96 (104.22)	0.00
Household rents dwelling	0.71 (0.45)	0.63 (0.48)	0.66 (0.47)	0.70 (0.46)	0.77 (0.42)	0.77 (0.42)	<0.01
Number of rooms in dwelling	1.53 (1.12)	1.59 (1.16)	1.69 (1.36)	1.59 (1.16)	1.51 (1.08)	1.36 (0.87)	<0.01
Proxy means score	53.85 (14.57)	43.81 (12.41)	49.55 (12.12)	52.42 (12.82)	56.94 (12.63)	64.35 (12.80)	<0.01
Cantril ladder (1–5)	2.32 (1.10)	2.00 (0.98)	2.26 (1.11)	2.33 (1.11)	2.46 (1.17)	2.52 (1.08)	<0.01
Meals per day	1.79 (0.70)	1.46 (0.55)	1.69 (0.67)	1.83 (0.65)	1.90 (0.71)	2.03 (0.75)	<0.01
Any shock in past 12 months	0.68 (0.47)	0.63 (0.48)	0.80 (0.40)	0.71 (0.46)	0.63 (0.48)	0.64 (0.48)	<0.01
Wealth shock in past 12 months	0.50 (0.50)	0.49 (0.50)	0.52 (0.50)	0.50 (0.50)	0.48 (0.50)	0.52 (0.50)	0.60
Health shock in past 12 months	0.32 (0.47)	0.35 (0.48)	0.35 (0.48)	0.29 (0.46)	0.32 (0.47)	0.30 (0.46)	0.16
Demographic/Social:							
Household size	4.41 (2.63)	6.05 (3.22)	5.34 (2.42)	4.70 (2.10)	3.86 (1.95)	2.51 (1.47)	<0.01
Household head is female	0.25 (0.44)	0.33 (0.47)	0.24 (0.43)	0.22 (0.42)	0.21 (0.41)	0.25 (0.44)	<0.01
Years in community	10.08 (10.83)	13.12 (12.11)	11.91 (11.50)	10.22 (10.93)	7.84 (8.83)	8.06 (9.83)	<0.01
Household includes community leader	0.14 (0.35)	0.15 (0.36)	0.14 (0.35)	0.16 (0.37)	0.13 (0.34)	0.13 (0.33)	0.51
Religion is Christian	0.71 (0.45)	0.73 (0.45)	0.72 (0.45)	0.69 (0.46)	0.70 (0.46)	0.72 (0.45)	0.77
In-degree centrality	2.84 (3.04)	3.45 (3.29)	2.32 (2.86)	2.75 (3.27)	3.02 (2.94)	2.68 (2.71)	<0.01
Betweenness centrality	271.83 (410.02)	342.13 (450.99)	206.10 (362.66)	270.20 (408.88)	287.49 (418.20)	253.09 (393.56)	<0.01
Eigenvector centrality	0.12 (0.18)	0.16 (0.20)	0.09 (0.15)	0.12 (0.18)	0.14 (0.18)	0.12 (0.16)	<0.01
# Family assistance	0.86 (1.18)	0.70 (0.99)	0.80 (1.05)	0.80 (1.15)	0.87 (1.16)	1.06 (1.41)	<0.01
# Friends assistance	0.73 (1.19)	0.61 (1.05)	0.67 (1.12)	0.67 (1.07)	0.73 (1.16)	0.91 (1.45)	<0.01
Government assistance	0.01 (0.10)	<0.01 (0.06)	0.01 (0.10)	0.01 (0.10)	0.01 (0.10)	0.02 (0.13)	0.34
NGO assistance	0.02 (0.15)	0.03 (0.16)	0.03 (0.16)	0.02 (0.12)	0.02 (0.14)	0.02 (0.15)	0.85

Tables A.1.3, A.1.4, and A.1.5 provide baseline descriptive statistics by neighborhood. P-value is derived from the t-test for equality of all quintiles.

We categorized the first two shocks (loss of income source and loss of property as wealth shocks. The other two shocks (injury and death) are categorized as health shocks. Loss of income source and stop working due to illness or injury are uncorrelated ($\rho = 0.15$). As such, we believe that the two questions successfully differentiate economic and health shocks.

Half of all households report an economic shock in the preceding 12 months. 28 percent report a loss of income source and 36 percent report property loss. About one-third report a health shock. 24 percent stated that a member of the household stopped work due to illness or injury. 11 percent of households report a death within the household. 62 percent of households experienced at least one form of a shock in the 12 months preceding the baseline survey.

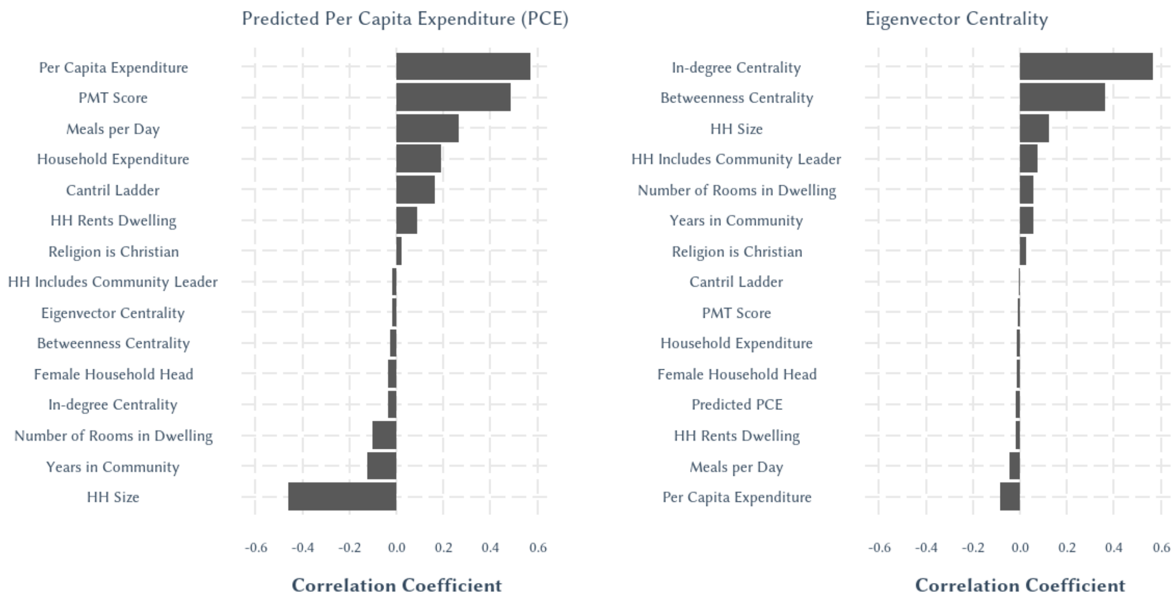

Figure 1.2: Correlation between Welfare and Social Network Measures

Social network centrality

Households, on average, were named by 2.84 other households in their block. This measure is termed as in-degree centrality, or how many incoming social ties a household has. Betweenness centrality provides a measure of how well individual households are positioned to connect socially-disconnected clusters of households in the social network. eigenvector accounts for the centrality of households that a given node is connected to and ranges from zero to one. On average, households have an eigenvector centrality of 0.12.

Table 1.4 presents summary statistics from households surveyed at baseline. Column 1 provides aggregate statistics for all households surveyed. On average, households comprised of 4.41 people and had lived in the community for ten years. The median household has been living in their community of Monrovia for six years; 16.8 percent of households have lived in the community for less than 2 years. As such residency in the study areas is largely permanent with recent in-migration composing less than one-fifth of the population.

Panel A of Figure 1.2 displays the correlation coefficient between Predicted PCE and survey-based welfare measures. We see that the absolute value of the correlation coefficient is greater than 0.4 for three measures: per capita expenditures (simple module), PMT score, and household size.

Columns 2–6 of Table 1.4 show summary statistics broken down by quintile of Predicted PCE.[9] We see that poorer households have lived in their community longer, state that they are lower on the Cantril Ladder, and consume fewer meals per day. We see no difference in the likelihood that a household experienced a wealth or health shock in the preceding 12 months.

The poorest quintile of households are no more likely than other households to include a community leader. Women constitute one-quarter of household heads. The poorest households are more likely to be female-headed.

The population of the communities is diverse. Christianity is the main religion within the study area with 70 percent of households reporting affiliation with a Christian denomination. However, Muslim households constitute 40 to 55 percent of households in five neighborhoods and 39 percent of the total sample. The religion of households is similar across all quintiles of Predicted PCE. Households come from more than 21 different tribes with Kru (14.5%), Kpelle (13.7%), Grebo (11.36%), and Vai (9.89%) being the most common tribes household heads.

In Panel B of Figure 1.2, we show that eigenvector centrality is correlated with in-degree centrality and betweenness centrality, however, there is no evident *linear* correlation with welfare or demographic characteristics. When we compare means across quintiles of Predicted PCE, as shown in Table 1.4, we do see that households in the poorest quintile have *higher* social network centrality.

Who are considered knowledgeable community members?

In accordance with the design presented in Section 1.3, we elicited nominations from each of the households surveyed at the time of the baseline. In Table 1.5, we present summary statistics for nominated households as well as households that include a community leader.

Thirty-five percent of households received at least one nomination. On average, households received 0.63 nominations. Conditional on receiving at least one nomination, households received an average of two nominations. One household received 49 nominations; however, the majority of nominated households received only one nomination. Among nominated households, 84 percent had only one household member nominated. Only 12 households had more than two household members nominated.

Nominated households were distinct along several key welfare and social dimensions. Nominated households were larger and had lived in the community longer. Nominated households reported lower per capita expenditures and higher PMT score. However, using our preferred

[9]Quintiles are calculated for each neighborhood.

Table 1.5: Summary Statistics for Leaders and Nominated Households

	(1) Population N=2656	(2) All Leaders N=345	(3) Nominated N=820	(4) Leaders vs. Non-Leaders p-value	(5) Nominated vs. Un-Nominated p-value
Welfare:					
Household expenditures (LD)	914 (887)	1005 (918)	984 (880)	0.052	0.006
Per capita expenditures (LD)	259 (284)	263 (325)	242 (268)	0.805	0.027
Predicted PCE (LD)	170 (89.8)	170 (95.5)	169 (95.2)	0.427	0.780
Household rents dwelling	0.71 (0.45)	0.61 (0.49)	0.65 (0.48)	$<$0.001	$<$0.001
Number of rooms in dwelling	1.53 (1.12)	1.81 (1.27)	1.74 (1.27)	$<$0.001	$<$0.001
Proxy means score	53.9 (14.6)	56.4 (15.3)	55.3 (13.7)	0.001	$<$0.001
Wealth index (PCA)	-0.60 (1.17)	-0.22 (1.20)	-0.32 (1.14)	$<$0.001	$<$0.001
Cantril ladder (1–5)	2.32 (1.10)	2.45 (1.14)	2.44 (1.10)	0.021	$<$0.001
Meals per day	1.79 (0.70)	1.80 (0.68)	1.81 (0.69)	0.794	0.327
Any shock in past 12 months	0.68 (0.47)	0.69 (0.46)	0.67 (0.47)	0.090	0.515
Wealth shock in past 12 months	0.50 (0.50)	0.57 (0.50)	0.52 (0.50)	0.006	0.166
Health shock in past 12 months	0.32 (0.47)	0.34 (0.47)	0.34 (0.47)	0.571	0.192
Demographic/Social:					
Household size	4.41 (2.63)	5.16 (3.00)	5.18 (2.91)	$<$0.001	$<$0.001
Household head is female	0.25 (0.44)	0.20 (0.40)	0.24 (0.43)	0.005	0.225
Years in community	10.1 (10.8)	12.3 (11.5)	11.7 (11.1)	$<$0.001	$<$0.001
Household includes leader	0.14 (0.35)	1.00 (0.00)	0.20 (0.40)	-	$<$0.001
Religion is Christian	0.71 (0.45)	0.85 (0.36)	0.75 (0.44)	$<$0.001	0.009
In-degree centrality	2.84 (3.04)	4.09 (3.69)	5.23 (3.52)	$<$0.001	$<$0.001
Betweenness centrality	272 (410)	438 (646)	479 (557)	$<$0.001	$<$0.001
Eigenvector centrality	0.12 (0.18)	0.17 (0.22)	0.19 (0.23)	0.002	$<$0.001
Nominations:					
Count of nominations	0.63 (1.78)	1.59 (3.94)	2.06 (2.71)	$<$0.001	$<$0.001
Unique HH Members nominated	0.36 (0.59)	0.59 (0.72)	1.17 (0.43)	$<$0.001	0.000

Notes: Standard deviations in parentheses.

measure of welfare, Predicted PCE, we do not see any statistically significant difference between nominated and un-nominated households.

Nominated households were more central within the social network. Across all three of our centrality measures, nominated households displayed much higher centrality scores than un-nominated households. Nominated households were more likely to include a leader; 20 percent of nominated households included a leader. When we look exclusively at leaders within the community (Column 2 of Table 1.5), we see that leaders were more central within the social network. The direction of differences for reported per capita expenditures and asset-based welfare measures mirrors that of nominated households. Again, on our preferred measure of welfare that includes non-linear relationships between expenditures and assets, Predicted PCE, we fail to reject the null hypothesis that leaders were different, on average, from non-leaders.

Table 1.6 presents the main findings for the first stage of the targeting study. The table presents coefficients from the following model for a host of dependent variables as derived through the process of eliciting nominations for knowledgeable community members. We use the following regression specification for the models in Table 1.6:

$$Y_i = \rho_2 Prompt_2 + \rho_3 Prompt_3 + \boldsymbol{X_i}\boldsymbol{\beta} + \nu_c + \epsilon_i \tag{1.1}$$

As outlined in Section 1.3, we elicited nominations through one of three randomly selected prompts. $Prompt_2$ and $Prompt_3$ are dummy variables indicating whether the nominated through prompt 2 or prompt 3. Prompt 1 is the held-out reference group. $\boldsymbol{X_i}$ is a 1×6 vector of the following six nominating household characteristics: (1) Predicted PCE Rank, (2)gender of respondent, (3) years in the community, (4) leader households, (5) religion is Christian, and (6) eigenvector centrality. We also adjust for community fixed effects, ν_c.

In Panel A of Table 1.6 shows the results for regression Equation 1.1 where the dependent variable is a dummy for whether or not the household provided a nomination (Column 1) and, conditional on providing a nomination, whether or not the nominee was found within the census listing database for the neighborhood (Column 2). We see no differential nomination rates across prompt type. We observe that Christian households were five percentage points more likely to provide a nomination. And higher social network centrality is correlated with higher nominations rates — a 1 standard deviation increase in eigenvector centrality is correlated with a 3 percentage point increase in providing a nomination and a 6 percentage point jump in the probability that we were able to identify a nomination in the census listing.

Panel B presents our homophily specifications, where the dependent variable is characteristic k of the nominee, j, that the baseline household, i, proposes. We slightly modify the regression equation to the following:

$$X_{j,k} = \rho_2 Prompt_2 + \rho_3 Prompt_3 + \boldsymbol{X_i}\boldsymbol{\beta} + \nu_c + \epsilon_i \tag{1.2}$$

Table 1.6: Homophily in Nominating Knowledgeable Community Members

	(A) Nomination Outcome		(B) Nominee (j) Characteristics						(C) Dyadic Distance	
	(1) Any Nomination	(2) Nomination Found	(3) Pred. PCE Rank	(4) Female	(5) Years in Community	(6) Leader	(7) Christian	(8) Eigenvector Centrality	(9) Social Distance	(10) Geographic Distance
Nomination Prompt:										
Prompt 2 (ρ_2)	-0.01	-0.00	-41.33	0.09^{**}	-0.53	0.03	-0.01	-0.02	0.15^{*}	3.99
	(0.01)	(0.02)	(45.00)	(0.03)	(0.73)	(0.03)	(0.02)	(0.01)	(0.06)	(3.16)
Prompt 3 (ρ_3)	-0.01	0.01	-101.52^{*}	0.05	-0.70	-0.00	0.01	0.00	0.18^{**}	-0.23
	(0.01)	(0.02)	(45.25)	(0.03)	(0.70)	(0.03)	(0.02)	(0.01)	(0.06)	(2.84)
Nominating (i) Characteristics:										
Predicted PCE Rank (β_1)	0.00	0.00	$\mathbf{0.07}^{**}$	0.00	-0.00	-0.00	0.00	-0.00^{*}	-0.00^{*}	0.00
	(0.00)	(0.00)	$(\mathbf{0.02})$	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (β_2)	0.01	-0.02	6.06	$\mathbf{0.15}^{***}$	0.24	-0.03	-0.01	0.01	-0.02	2.06
	(0.01)	(0.02)	(40.24)	$(\mathbf{0.02})$	(0.63)	(0.02)	(0.02)	(0.01)	(0.06)	(2.48)
Years in Comm. (SD) (β_3)	0.01	0.00	-15.71	0.04^{***}	$\mathbf{1.03}^{**}$	0.01	-0.01	-0.01^{*}	0.04	1.59
	(0.01)	(0.01)	(18.60)	(0.01)	$(\mathbf{0.33})$	(0.01)	(0.01)	(0.00)	(0.03)	(1.12)
Leader HH (β_4)	0.03^{*}	-0.03	-20.33	0.05	-0.30	$\mathbf{0.15}^{***}$	-0.02	-0.02	-0.07	13.14^{***}
	(0.01)	(0.03)	(52.47)	(0.03)	(0.81)	$(\mathbf{0.03})$	(0.02)	(0.01)	(0.06)	(3.82)
Christian (β_5)	0.05^{***}	-0.03	118.60^{**}	-0.00	1.15	0.05^{*}	$\mathbf{0.36}^{***}$	-0.00	0.02	-3.43
	(0.01)	(0.02)	(43.13)	(0.03)	(0.64)	(0.03)	$(\mathbf{0.03})$	(0.01)	(0.06)	(2.82)
Eigenvector (SD) (β_6)	0.03^{***}	0.06^{***}	-57.12^{**}	-0.01	-0.13	0.04^{**}	0.01	$\mathbf{0.15}^{***}$	-0.13^{***}	-3.41^{**}
	(0.00)	(0.01)	(18.24)	(0.01)	(0.30)	(0.01)	(0.01)	$(\mathbf{0.01})$	(0.02)	(1.06)
Adj. R^2	0.05	0.03	0.05	0.08	0.03	0.04	0.23	0.40	0.04	0.10
Num. obs.	2231	2059	1675	1675	1614	1675	1623	1675	1651	1675
$\rho_2 = \rho_3$ (p-value)	0.39	0.91	0.07	0.005	0.32	0.58	0.95	0.57	0.001	0.46
Mean of Dep. Var.	0.78	0.81	0.00	0.50	10.07	0.14	0.71	0.12	3.88	105.85

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. All regressions use ordinary least squares with Eicker-Huber-White robust standard errors in parentheses. Neighborhood fixed effects included. One of three nomination prompts was randomly assigned to each baseline respondent, see table A.1.6 for balance statistics by prompt type.

Nomination Prompt 1 (Reference): If we want to spread information about a social assistance program, to whom do you suggest we speak? (N = 811);

Nomination Prompt 2: If we want to identify which people would be best to help us identify which people in this block would benefit most from a social assistance program, to whom do you suggest we speak? (N = 698);

Nomination Prompt 3: If we want to identify which people would be best to help us identify which people in this block would benefit most from a cash gift for social assistance, to whom do you suggest we speak? (N = 744)

"Any Nomination" equals 1 if the baseline respondent provided a TA nomination, 0 if not. "Nomination Found" is conditional on the respondent providing a nomination and equals 1 if the nominee was found within the community census database, 0 if not. Predicted PCE Rank is the household ranking of the per capita expenditures as predicted by our random forest algorithm, centered at zero. Social Distance is the shortest path between the nominating and nominated household as calculated using social network data from the baseline. Geographic distance is the number of meters between the nominating and nominated household's dwelling as measured by the spherical geodesic distance between the latitude-longitude coordinates of the dwellings.

We observe a great deal of homophily in nominations. As evidenced along the diagonal in Panel B of Table1.6, we observe a high degree of positive correlation across nominator-nominee characteristics. A one-step increase in the Predicted PCE rank of the nominator is associated with a 0.07 step increase of the nominee's Predicted PCE rank. Women were 15 percentage points more likely to be nominated if the nominator is a woman. One standard deviation increase in the nominator's tenure in the community is associated with a 1.03 years increase in the nominee's tenure. Leaders were 15 percentage points more likely to nominate another leader. Christian households nominated other Christian households at a much higher rate. And high-eigenvector centrality nominators proposed central nominees.

In Panel C, we examine the dyadic distance between nominator i and nominee i. In Column 9, our dependent variable is the shortest distance along un-directed social network edges from i to j. We see that prompts 2 and 3 elicited nominees that were 0.15 to 0.18 degrees more socially distant from the nominator. This result suggests that when we mention targeting of a social assistance program, individuals conduct a more extended search of their network. The prompts do not show a clear pattern in correlation with nominee characteristics. Two exceptions are worth noting. Prompt 3 resulted in nominees that had lower Predicted PCE. Moreover, Prompt 2 led to more female nominees. Column 10 uses geographic distance (in meters) to assess the dyadic distance in nominations. We fail to reject the null hypothesis that the prompts led to different geographic distance.

Do nominated community members possess more accurate information?

As discussed in Section 1.3, we conducted one-on-one interviews with 394 community members through what we refer to as Targeting Assistant (TA) surveys. 254 of the TAs were drawn from those nominated by other community members during the baseline survey. Table 1.7 provides aggregate summary statistics from the TA surveys. In total, we queried 11,315 dyadic relationships between the TAs and randomly selected households from the TA's neighborhood. When asked if they know a randomly selected neighbor, nominated TAs knew 36 percent of their neighbors.

In comparison, un-nominated TAs knew 32 percent of neighbors. This can be explained by the fact that nominated TAs were more socially connected. Confirming the friendship paradox (Jackson 2019), we see that the social distance between nominated TAs and randomly selected neighbors is less than similar measure for un-nominated TAs. Nominated TAs were more likely to talk to known neighbors on a daily basis. We also note that the geographic distance between TAs and random neighbors is greater for the nominated TAs, suggesting that those nominated were seen to be knowledgeable across more area.

Among the neighbors that TAs know, approximately 37 percent were considered friends.[10]

[10]Because TAs were asked about a large number of neighbors, we may be concerned about survey fatigue influencing the truthfulness of reporting. For this reason, we randomized the order of the neighbors that each TA was asked about. Appendix Figure A.1.4 shows that there is a downward trend in TAs reporting that they know a neighbor. In our regression analysis, we control for the order in which neighbors appear in the TA survey. We do not find evidence of survey fatigue that effected nominated and un-nominated TAs differentially.

Nominated TAs were no more likely to consider known neighbors as friends or family. TAs state that they have borrowed from or lent money to thirteen percent of neighbors that were known (4.5 percent of all neighbors when we assume that they have not borrowed from or lent to unknown neighbors).

Table 1.7: TA Assessment of Randomly Selected Neighbors

	All TAs	Un-Nominated	Nominated	p-value
TA knows the neighbor	0.35 (0.48)	0.32 (0.47)	0.36 (0.48)	<0.001
Social distance to neighbor	3.60 (1.21)	3.69 (1.21)	3.56 (1.20)	<0.001
Distance to neighbor (meters)	96.6 (84.9)	90.7 (81.4)	99.7 (86.5)	<0.001
Conditional on TA knowing neighbor:				
Relationship to neighbor:				0.380
Family	223 (5.29%)	77 (5.66%)	146 (5.11%)	
Friend	1582 (37.5%)	525 (38.6%)	1057 (37.0%)	
Other Neighbor	2413 (57.2%)	758 (55.7%)	1655 (57.9%)	
Frequency of communication with neighbor: (*)				<0.001
Daily	2206 (52.3%)	649 (47.8%)	1557 (54.5%)	
Weekly	1358 (32.2%)	487 (35.8%)	871 (30.5%)	
Less than weekly	652 (15.5%)	223 (16.4%)	429 (15.0%)	
TA has borrowed from or lent to neighbor (**)	0.13 (0.34)	0.12 (0.33)	0.14 (0.34)	0.244
TA's view of neighbor's welfare: (***)				0.370
1st step	1249 (29.9%)	404 (29.9%)	845 (29.9%)	
2nd step	1695 (40.6%)	566 (41.9%)	1129 (40.0%)	
3rd-5th step	1229 (29.5%)	380 (28.1%)	849 (30.1%)	
Prop. would say neighbor is in poorest 20%? (****)	0.52 (0.31)	0.52 (0.29)	0.53 (0.32)	0.433

Notes: Chi-squared test used for categorical variables.
Social Distance is the shortest path between the TA and the randomly selected neighbor as calculated using social network data from the baseline survey. Geographic distance is the number of meters between the TA's and neighbor's dwelling as measured by the geodesic distance between the latitude-longitude coordinates of the dwellings.
(*) "How often would you say that you communicate with anyone in [NAME]'s household?"
(**) "Have you ever borrowed money from, lent money, or given or received a gift from anyone in [NAME]'s household?"
(***) "Please look at this 5-step ladder. On the bottom, the first step stand the poorest 20% of households in [NEIGHBORHOOD]. On the highest step, the fifth, stands the richest 20% of households in [NEIGHBORHOOD]. Where do you think [NAME]'s household stands when you think of how poor or wealthy [NAME]'s household is compared to others in [NEIGHBORHOOD]?"
(****) "If we asked 10 people from your community in [NEIGHBORHOOD], how many would say that [NAME]'s household is on the first step of the ladder?" (Response divided by 10 to show proportion in table.)

In Table A.1.7, we show TA knowledge of neighbors by three types of TAs: Randomly Selected, Nominated, and Leaders. Leaders were more likely to state that they know randomly selected neighbors. However, among known neighbors, leaders were no more likely to classify neighbors as poor.

When presented with the opportunity to vote on the welfare of a randomly selected neighbor (refer to question text in Section 1.3), TAs reported that 30 percent of neighbors were on the bottom step of the Cantril Ladder. This proportion is the same for nominated and un-nominated TAs. Eliciting this information from leaders, nominated TAs, or random TAs yielded the same result among known neighbors. TAs appeared to think that other community members are more likely to consider their neighbors as poor. When asked what proportion of ten community members would say that the random neighbor is on the bottom step of the Cantril Ladder, TAs estimated that 5.3 community members would respond in the affirmative. This suggests that the TAs possess some unique knowledge of their neighbors.

In Table 1.8, we display aggregate knowledge and votes of households from the TA survey. We assigned households to appear in five TA surveys, we see that there was near complete compliance with this goal; households appeared on average in 4.98 TA surveys. Consistent with evidence in Table 1.4, we see that the bottom two quintiles are more likely to be known. However, we cannot reject equivalence across quintiles. We do, though, see clear evidence that TAs provided meaningful votes in that households in the poorest two quintiles are more likely to be seen as poor. The difference is minimal, however, due to the fact that 88 percent of TA votes for households in the bottom quintile of Predicted PCE *do not* classify the household on the bottom step of the Cantril Ladder.

Table 1.8: TA Knowledge of Neighbors, by Predicted PCE Quintile

		Predicted PCE Quintile of Neighbor					
	All	1	2	3	4	5	p-value
# times mentioned in TA Survey	4.98 (0.13)	4.98 (0.16)	4.98 (0.12)	4.98 (0.15)	4.99 (0.11)	4.99 (0.10)	0.225
Prop. of TAs that know HH	0.35 (0.26)	0.36 (0.26)	0.36 (0.25)	0.35 (0.27)	0.35 (0.26)	0.33 (0.25)	0.218
Prop. of TAs votes as poor HH (when known) (*)	0.31 (0.37)	0.34 (0.38)	0.34 (0.38)	0.29 (0.37)	0.29 (0.36)	0.30 (0.37)	0.072
Prop. of TAs votes as poor HH (All TAs)	0.10 (0.15)	0.12 (0.16)	0.12 (0.16)	0.10 (0.14)	0.10 (0.14)	0.09 (0.13)	0.004
Prop. of community that would say HH is poor (**)	0.53 (0.26)	0.53 (0.26)	0.53 (0.27)	0.55 (0.25)	0.53 (0.25)	0.51 (0.27)	0.222
Errors:							
Prop. of votes as poor household	0.10 (0.14)	-	0.12 (0.16)	0.10 (0.14)	0.10 (0.14)	0.09 (0.13)	0.017
Prop. of votes as non-poor household	0.88 (0.16)	0.88 (0.16)	-	-	-	-	-

Notes: Standard deviations in parentheses. P-value is from the test for equivalence of all quintiles.
(*) "Please look at this 5-step ladder. On the bottom, the first step stand the poorest 20% of households in [NEIGHBORHOOD]. On the highest step, the fifth, stands the richest 20% of households in [NEIGHBORHOOD]. Where do you think [NAME]'s household stands when you think of how poor or wealthy [NAME]'s household is compared to others in [NEIGHBORHOOD]?"
(**) "If we asked 10 people from your community in [NEIGHBORHOOD], how many would say that [NAME]'s household is on the first step of the ladder?" (Response divided by 10 and geometric mean of TA beliefs is displayed in the table.)

We find that social distance — the steps in the social network graph from the TA to the random neighbor — was highly correlated with the quality of TA knowledge. For each additional step away from a household in social distance, the TA was 8 percentage points less likely to know the other household. Table 1.9 shows this relationship between the social distance between a TA and their neighbor and the accuracy of the TAs reporting.[11] Here accuracy, the dependent variable, is equal to one if the TA correctly classified the household either above or below the 20th percentile

[11] In the Appendix, we provide similar tables where the dependent variable is whether or not the TA knows the neighbor (Table A.1.8).

of the welfare distribution. Moving from one-degree social distance to two degrees of separation reduces the likelihood of accurate reporting by 13 to 17 percentage points. After two degrees of separation, reporting accuracy drops again but stays rather level all the way to unconnected neighbors (Social Distance = ∞).

We find little evidence that nominated TAs provided more accurate information. Across most social distances, nominated TAs show no statistically significant increase in accuracy over un-nominated TAs. This is explained by the dueling tendencies of nominated TAs to provide *lower* exclusion errors and *higher* inclusion errors for first-degree social contacts. We define an exclusion error as a vote, V_{ij}, from TA i that incorrectly classifies a poor neighbor j as non-poor (i.e. not in the bottom 20% of neighbors, $V_{ij} \neq 1$). W_j is a dummy variable equal to one if the Predicted PCE places household j in the bottom quintile of the welfare distribution for the neighborhood. Similarly, we identify an error of inclusion where the TA's vote incorrectly labeled a non-poor household as poor.

$$AccurateVote_{ij} = \sum \mathbf{1}_{[V_{ij} \neq 1 | W_j \neq 1]} + \mathbf{1}_{[V_{ij} = 1 | W_j = 1]} \tag{1.3}$$

$$ExclusionError_{ij} = \mathbf{1}_{[V_{ij} \neq 1 | W_j = 1]} \tag{1.4}$$

$$InclusionError_{ij} = \mathbf{1}_{[V_{ij} = 1 | W_j \neq 1]} \tag{1.5}$$

Table 1.10 presents regression results where exclusion error is the dependent variable. Table 1.11 shows our model results for inclusion errors. This suggests TAs were prone to provide inaccurate votes for adjacent social network ties.

When examined in the aggregate, TAs with more nominations tended to be more accurate. We sum exclusion errors and inclusion errors for each TA to calculate an aggregate score.

$$AccuracyRate_i = \frac{\sum_{j=1}^{J} AccurateVote_{ij}}{J} \tag{1.6}$$

$$ExclusionRate_i = \frac{\sum_{j=1}^{J} ExclusionError_{ij}}{J} \tag{1.7}$$

$$InclusionRate_i = \frac{\sum_{j=1}^{J} InclusionError_{ij}}{J} \tag{1.8}$$

Figure 1.3 shows the scatterplot and bivariate linear regression of TA information value on the number of nominations that a TA received. We see that the accuracy of TAs is correlated with the number of nominations that at TA received, even when excluding the outlier TA that received 49 nominations. The lift gained from additional nominations, however, is rather small — each additional nomination is associated with a one percentage point increase in the accuracy rate.

We do not observe a statistically significant correlation between the number of TA nominations and the inclusion or exclusion error rates of TAs (Panels B and C of Figure 1.3). The coefficient on number of nominations is negative for exclusion errors, -0.003, but not statistically significant. Working in the opposite direction, the coefficient on number of nominations is positive for inclusion errors, 0.003, but not statistically significant.

Table 1.9: Targeting Assistant Accuracy

	(1)	(2)	(3)	(4)
Social Distance = 2	-0.132^{***}	-0.172^{***}	-0.132^{***}	-0.174^{***}
	(0.025)	(0.046)	(0.026)	(0.046)
Social Distance = 3	-0.214^{***}	-0.220^{***}	-0.208^{***}	-0.221^{***}
	(0.023)	(0.043)	(0.024)	(0.043)
Social Distance = 4	-0.251^{***}	-0.245^{***}	-0.245^{***}	-0.246^{***}
	(0.023)	(0.043)	(0.023)	(0.043)
Social Distance = 5	-0.276^{***}	-0.252^{***}	-0.267^{***}	-0.245^{***}
	(0.024)	(0.044)	(0.024)	(0.044)
Social Distance = 6	-0.252^{***}	-0.239^{***}	-0.246^{***}	-0.229^{***}
	(0.029)	(0.050)	(0.030)	(0.052)
Social Distance = 7	-0.285^{***}	-0.186^{*}	-0.250^{***}	-0.164
	(0.044)	(0.076)	(0.047)	(0.084)
Social Distance = 8	-0.201	-0.290^{*}	-0.248^{**}	-0.367^{***}
	(0.104)	(0.127)	(0.091)	(0.047)
Social Distance = ∞	-0.280^{***}	-0.285^{***}	-0.302^{***}	-0.306^{***}
	(0.026)	(0.045)	(0.028)	(0.046)
Nominated TA		0.028		0.021
		(0.049)		(0.049)
Nominated TA * Social Distance = 2		0.059		0.062
		(0.055)		(0.055)
Nominated TA * Social Distance = 3		0.011		0.022
		(0.052)		(0.052)
Nominated TA * Social Distance = 4		-0.006		0.005
		(0.051)		(0.051)
Nominated TA * Social Distance = 5		-0.035		-0.030
		(0.052)		(0.053)
Nominated TA * Social Distance = 6		-0.017		-0.023
		(0.062)		(0.063)
Nominated TA * Social Distance = 7		-0.178^{*}		-0.147
		(0.090)		(0.099)
Nominated TA * Social Distance = 8		0.176		0.216
		(0.200)		(0.153)
Nominated TA * Social Distance = ∞		0.019		0.030
		(0.055)		(0.060)
Adj. R^2	0.023	0.025	0.043	0.044
Num. obs.	12130	12130	11140	11140
Neighbor Covariates	NO	NO	YES	YES
Neighbor Order F.E.	YES	YES	YES	YES

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variable is a binary variable equal to one if the respondent (TA) correctly classifies a household as poor or non-poor based on the household being in the bottom quintile of the predicted per-capita expenditure distribution of the neighborhood. SD indicates that a variables has been mean-centered and scaled to represent one standard standard deviation.
Eicker-Huber-White robust standard errors in parentheses. Neighbor-Order fixed effects included. Social Distance is the shortest path between the nominating and nominated household as calculated using social network data from the baseline.

Table 1.10: Targeting Assistant Exclusion Error

	(1)	(2)	(3)	(4)
Social Distance = 2	0.100^{*}	-0.099	0.095^{*}	-0.104
	(0.046)	(0.071)	(0.046)	(0.070)
Social Distance = 3	0.201^{***}	-0.020	0.191^{***}	-0.023
	(0.042)	(0.062)	(0.042)	(0.061)
Social Distance = 4	0.219^{***}	0.019	0.204^{***}	0.010
	(0.041)	(0.060)	(0.042)	(0.059)
Social Distance = 5	0.236^{***}	0.067	0.226^{***}	0.052
	(0.042)	(0.060)	(0.043)	(0.060)
Social Distance = 6	0.199^{***}	-0.045	0.204^{***}	-0.047
	(0.051)	(0.085)	(0.050)	(0.084)
Social Distance = 7	0.252^{***}	-0.027	0.221^{***}	-0.087
	(0.061)	(0.120)	(0.064)	(0.143)
Social Distance = 8	0.301^{***}	0.093	0.268^{***}	0.056
	(0.044)	(0.061)	(0.047)	(0.062)
Social Distance = ∞	0.236^{***}	0.032	0.213^{***}	0.016
	(0.046)	(0.065)	(0.048)	(0.066)
Nominated TA		-0.260^{***}		-0.257^{***}
		(0.075)		(0.075)
Nominated TA * Social Distance = 2		0.257^{**}		0.260^{**}
		(0.090)		(0.089)
Nominated TA * Social Distance = 3		0.290^{***}		0.281^{***}
		(0.080)		(0.079)
Nominated TA * Social Distance = 4		0.258^{***}		0.252^{**}
		(0.078)		(0.078)
Nominated TA * Social Distance = 5		0.206^{**}		0.217^{**}
		(0.080)		(0.079)
Nominated TA * Social Distance = 6		0.331^{**}		0.345^{***}
		(0.104)		(0.101)
Nominated TA * Social Distance = 7		0.391^{**}		0.427^{**}
		(0.132)		(0.153)
Nominated TA * Social Distance = 8		0.281^{**}		0.313^{***}
		(0.087)		(0.089)
Nominated TA * Social Distance = ∞		0.265^{**}		0.253^{**}
		(0.089)		(0.093)
Adj. R^2	0.039	0.044	0.041	0.045
Num. obs.	2627	2627	2542	2542
Neighbor Covariates	NO	NO	YES	YES
Neighbor Order F.E.	YES	YES	YES	YES

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variable is a binary variable equal to one if the respondent (TA) incorrectly classifies a poor household as non-poor. SD indicates that a variables has been mean-centered and scaled to represent one standard standard deviation. Eicker-Huber-White robust standard errors in parentheses. Neighbor-Order fixed effects included. Social Distance is the shortest path between the nominating and nominated household as calculated using social network data from the baseline.

Table 1.11: Targeting Assistant Inclusion Error

	(1)	(2)	(3)	(4)
Social Distance = 2	-0.087^{***}	0.006	-0.088^{***}	0.019
	(0.026)	(0.041)	(0.026)	(0.041)
Social Distance = 3	-0.148^{***}	-0.083^{*}	-0.145^{***}	-0.070
	(0.024)	(0.037)	(0.024)	(0.037)
Social Distance = 4	-0.176^{***}	-0.083^{*}	-0.167^{***}	-0.062
	(0.024)	(0.037)	(0.024)	(0.037)
Social Distance = 5	-0.194^{***}	-0.106^{**}	-0.183^{***}	-0.085^{*}
	(0.024)	(0.038)	(0.024)	(0.038)
Social Distance = 6	-0.218^{***}	-0.140^{***}	-0.205^{***}	-0.119^{**}
	(0.025)	(0.039)	(0.026)	(0.040)
Social Distance = 7	-0.197^{***}	-0.143^{**}	-0.171^{***}	-0.111^{*}
	(0.036)	(0.046)	(0.041)	(0.051)
Social Distance = 8	-0.173^{*}	0.029	-0.131	0.191
	(0.076)	(0.180)	(0.096)	(0.267)
Social Distance = ∞	-0.198^{***}	-0.133^{***}	-0.206^{***}	-0.121^{**}
	(0.025)	(0.038)	(0.026)	(0.038)
Nominated TA		0.120^{**}		0.135^{**}
		(0.046)		(0.046)
Nominated TA * Social Distance = 2		-0.135^{**}		-0.152^{**}
		(0.052)		(0.053)
Nominated TA * Social Distance = 3		-0.093		-0.107^{*}
		(0.048)		(0.048)
Nominated TA * Social Distance = 4		-0.135^{**}		-0.152^{**}
		(0.047)		(0.047)
Nominated TA * Social Distance = 5		-0.128^{**}		-0.141^{**}
		(0.048)		(0.048)
Nominated TA * Social Distance = 6		-0.111^{*}		-0.122^{*}
		(0.050)		(0.051)
Nominated TA * Social Distance = 7		-0.065		-0.069
		(0.071)		(0.080)
Nominated TA * Social Distance = 8		-0.314		-0.460
		(0.182)		(0.269)
Nominated TA * Social Distance = ∞		-0.087		-0.111^{*}
		(0.049)		(0.052)
Adj. R^2	0.023	0.025	0.024	0.026
Num. obs.	9503	9503	8598	8598
Neighbor Covariates	NO	NO	YES	YES
Neighbor Order F.E.	YES	YES	YES	YES

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variable is a binary variable equal to one if the respondent (TA) incorrectly classifies a non-poor household as poor. SD indicates that a variables has been mean-centered and scaled to represent one standard standard deviation. Eicker-Huber-White robust standard errors in parentheses. Neighbor-Order fixed effects included. Social Distance is the shortest path between the nominating and nominated household as calculated using social network data from the baseline.

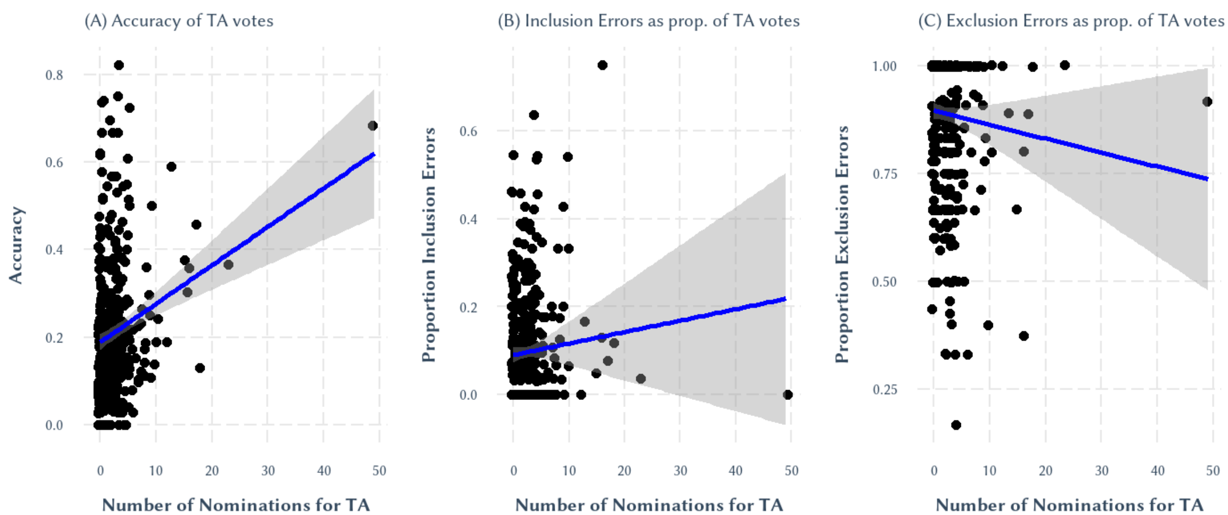

Figure 1.3: **Aggregate performance of TAs.** Points jittered for visibility. Panel A: Slope point estimate equals 0.009 with 95% CI for coefficient is [0.006, 0.012]. Excluding the outlier with 49 nominations, the point estimate is 0.008 with 95% CI of [0.003, 0.013]. Panel B: Slope point estimate equals 0.003 with 95% CI for coefficient is [-0.004, 0.009]. Excluding the household with 49 nominations, the point estimate is 0.008 with 95% CI of [-0.001, 0.012]. Panel C: Slope point estimate equals -0.003 with 95% CI for coefficient is [-0.003, 0.003]. Excluding the household with 49 nominations, the point estimate is 0.008 with 95% CI of [-0.013, 0.001].

Who receives votes for a cash grant social program?

TAs were allowed to nominate up to two members of their neighborhood to receive a cash grant. We elicited the cash grant nominations through two questions, as discussed in Section 1.3. We find that households nominated for the cash grant were more likely to be socially connected. Female-headed households, households that include a community leader, and Christian households were more likely to receive cash grant nominations. We see in Table 1.12 that the method of cash grant elicitation did not lead to differences in demographic and social characteristics of cash grant nominees. However, the 'Hard Times' prompt led to nominees with lower subjective welfare (Cantril Ladder) and more likely to have experienced any shock in the preceding 12 months. We take this to suggest that framing the nomination around shocks instead of best use leads to increased targeting of households that are experiencing some recent hardships.

We also see that nominated households are, on average, poorer across multiple dimensions. When we examine the relationship between household characteristics and cash grant nominations in a multivariate regression specification, as in Table A.1.9, we see that the coefficients on eigenvector centrality and Predicted PCE Rank are statistically significant. Thus, controlling for eigenvector centrality, poorer households are slightly more likely to receive a cash grant nomination.

Cash grant nominations are characterized by homophily. Similar to what we observed with targeting assistant nominations, Table 1.13

Similar to our specification in Equation 1.2, we use the following regression specification to test for homophily in cash grant nominations, where TA i has a 1×6 vector of k characteristics. The dependent variable is the kth characteristic, $X_{j,k}$, of the nominated household j. We include a term, $NominatedTA_i$, to indicate if the targeting assistant was nominated at baseline. Also, we interact the nominated TA indicator with an indicator, $Elicitation_2$, for the "Hard Times" question used to elicit the cash grant nominee.[12] We adjust for ν_c neighborhood fixed effects.

$$\begin{aligned} X_{j,k} =& \eta NominatedTA_i + \rho Elicitation_2 + \theta(NominatedTA_i \times Elicitation_2) \\ &+ \boldsymbol{X_i \beta} + \nu_c + \epsilon_i \end{aligned} \tag{1.9}$$

Panel A of Table 1.13 shows regression results from the Equation 1.9 specification. Across all household characteristics, we fail to reject the null hypothesis that cash grant nominees from nominated and un-nominated TAs were equivalent. We also fail to reject the null hypothesis that the two elicitation prompts, "Best Use" and "Hard Times", yield equivalent cash grant nominees.

We see that TAs demonstrated homophily across all six dimensions. The Predicted PCE Rank of the TA is positively correlated with the predicted expenditures of the cash grant nominee. Women were twelve percentage points more likely to nominate another woman to receive a cash grant. Christian TAs were 36 percentage points more likely to nominate another Christian household to receive a cash grant. One standard deviation increase in the eigenvector centrality of the TA

[12] In Table A.1.10 we restrict the sample of cash grant nominations to the first nominee elicited from each Targeting Assistant. The order of the elicitation question was randomly determined.

is associated with a 0.12 unit increase in the eigenvector centrality of the cash grant nominee. We also observe positive correlations between years in the community of the TA and the cash grant nominee and whether or not both households include a community leader. However, we have weaker evidence to reject the null hypothesis; these homophily coefficients for years in the community and leadership have p-values of 0.06 and 0.08, respectively.

We do not observe any clear relationship between the TA type and whether the nominated cash grantee experienced a shock. Likewise, the "Hard Times" prompt did not result in nominees that were more likely to have experienced a shock. The "Hard Times" elicitation prompt increased the social distance between the TA and the cash grant nominee among nominated and un-nominated TAs. However, this result only holds when we include all cash grant nominees (Table A.1.10 shows regression results for the first cash grant nominated by the TA.)

Table 1.12: Welfare and Social Network Summary Statistics, by Cash Transfer Nomination

	Any Nomination			Any 'Best Use' Nomination			Any 'Hard Times' Nomination		
	No N=2163	Yes N=493	p-value	No N=2348	Yes N=308	p-value	No N=2372	Yes N=284	p-value
Welfare:									
Household expenditures (LD)	899 (863)	969 (968)	0.156	897 (865)	1023 (1011)	0.042	917 (887)	891 (882)	0.650
Per capita expenditures (LD)	268 (298)	229 (222)	0.002	263 (291)	234 (235)	0.050	265 (293)	218 (204)	0.001
Predicted PCE (LD)	172 (90.2)	160 (87.1)	0.006	171 (90.7)	160 (81.6)	0.033	171 (89.4)	157 (92.0)	0.016
Household rents dwelling	0.73 (0.44)	0.62 (0.49)	$<$0.001	0.72 (0.45)	0.61 (0.49)	$<$0.001	0.72 (0.45)	0.60 (0.49)	$<$0.001
Number of rooms in dwelling	1.51 (1.12)	1.61 (1.11)	0.068	1.51 (1.11)	1.68 (1.16)	0.018	1.53 (1.13)	1.55 (1.04)	0.777
Proxy means score	54.4 (14.5)	51.8 (14.6)	0.001	54.1 (14.5)	52.1 (14.9)	0.034	54.4 (14.6)	50.2 (14.1)	$<$0.001
Cantril ladder (1–5)	2.33 (1.08)	2.27 (1.18)	0.337	2.31 (1.09)	2.34 (1.19)	0.713	2.34 (1.10)	2.17 (1.14)	0.019
Meals per day	1.81 (0.71)	1.69 (0.67)	0.001	1.79 (0.70)	1.75 (0.68)	0.292	1.81 (0.70)	1.62 (0.65)	$<$0.001
Any shock in past 12 months	0.69 (0.46)	0.64 (0.48)	0.037	0.69 (0.46)	0.62 (0.49)	0.022	0.68 (0.47)	0.67 (0.47)	0.789
Wealth shock in past 12 months	0.50 (0.50)	0.49 (0.50)	0.531	0.51 (0.50)	0.46 (0.50)	0.129	0.50 (0.50)	0.51 (0.50)	0.631
Health shock in past 12 months	0.32 (0.46)	0.35 (0.48)	0.208	0.32 (0.46)	0.36 (0.48)	0.098	0.31 (0.46)	0.37 (0.48)	0.071
Demographic/Social:									
Household Size	4.24 (2.58)	5.04 (2.73)	$<$0.001	4.29 (2.58)	5.21 (2.82)	$<$0.001	4.32 (2.60)	5.04 (2.78)	$<$0.001
Household head is female	0.24 (0.42)	0.32 (0.47)	$<$0.001	0.25 (0.43)	0.31 (0.46)	0.026	0.24 (0.43)	0.38 (0.49)	$<$0.001
Years in community	9.31 (10.3)	13.0 (12.2)	$<$0.001	9.54 (10.4)	13.6 (12.9)	$<$0.001	9.61 (10.5)	13.5 (12.4)	$<$0.001
Household includes leader	0.13 (0.34)	0.18 (0.39)	0.005	0.13 (0.34)	0.19 (0.39)	0.024	0.14 (0.34)	0.18 (0.38)	0.075
Religion is Christian	0.69 (0.46)	0.80 (0.40)	$<$0.001	0.69 (0.46)	0.83 (0.38)	$<$0.001	0.70 (0.46)	0.78 (0.41)	0.004
In-degree centrality	2.45 (2.71)	4.55 (3.76)	$<$0.001	2.55 (2.78)	5.05 (3.93)	$<$0.001	2.68 (2.94)	4.23 (3.55)	$<$0.001
Betweenness centrality	240 (379)	411 (501)	$<$0.001	250 (395)	437 (482)	$<$0.001	258 (397)	391 (491)	$<$0.001
Eigenvector centrality	0.11 (0.16)	0.20 (0.23)	$<$0.001	0.11 (0.16)	0.21 (0.24)	$<$0.001	0.12 (0.17)	0.19 (0.23)	$<$0.001
Cash Grant nominations:									
'Best use' cash nominations	0.00 (0.00)	0.71 (0.64)	$<$0.001	0.00 (0.00)	1.13 (0.42)	$<$0.001	0.09 (0.31)	0.44 (0.71)	$<$0.001
'Hard times' cash nominations	0.00 (0.00)	0.69 (0.70)	$<$0.001	0.09 (0.34)	0.41 (0.67)	$<$0.001	0.00 (0.00)	1.20 (0.50)	$<$0.001

'Best Use' elicitation question: "Think of households within this neighborhood, which household would make the most out of a USD$100 cash grant?"
'Hard Times' elicitation question: "Within this neighborhood, is there a household who recently fell on hard times and would benefit most from a USD$100 cash grant?"

Table 1.13: Homophily in Cash Grant Nominations

	(A) Nominee Characteristics						(B) Nominee Shocks		(C) Dyadic Distance	
	Pred. PCE Rank	Female	Years in Community	Leader	Christian	Eigenvector Centrality	Wealth Shock	Health Shock	Social Distance	Geographic Distance
TA Type and Prompt:										
Nominated TA (ν)	-21.52	-0.07	0.80	-0.05	-0.01	-0.03	-0.00	0.08	-0.12	6.82
	(93.80)	(0.06)	(1.53)	(0.05)	(0.04)	(0.02)	(0.06)	(0.06)	(0.14)	(5.11)
Hard Times Elicitation (ρ)	-126.97	-0.00	1.24	-0.05	-0.03	-0.02	0.04	0.00	0.37^{*}	9.58
	(104.88)	(0.06)	(1.62)	(0.05)	(0.05)	(0.03)	(0.07)	(0.07)	(0.17)	(7.45)
Nominated TA * Hard Times (θ)	115.14	0.06	-2.64	0.06	0.00	0.01	0.02	0.01	-0.11	-3.31
	(131.85)	(0.08)	(2.08)	(0.06)	(0.06)	(0.03)	(0.08)	(0.08)	(0.20)	(9.45)
Nominating TA Characteristics:										
Predicted PCE Rank (β_1)	$\mathbf{0.13}^{**}$	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00
	$(\mathbf{0.04})$	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (β_2)	-34.68	$\mathbf{0.12}^{**}$	1.93	0.01	0.04	-0.01	0.01	0.11^{**}	0.21^{*}	2.28
	(67.15)	$(\mathbf{0.04})$	(1.10)	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)	(0.10)	(4.63)
Years in comm. (SD) (β_3)	6.45	-0.01	$\mathbf{1.18}$	0.01	0.03^{*}	-0.01	-0.01	-0.01	-0.06	-1.18
	(32.24)	(0.02)	$(\mathbf{0.62})$	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(2.23)
Leader (β_4)	1.41	-0.01	-0.41	$\mathbf{0.07}$	-0.03	-0.02	0.02	0.02	0.20	-2.51
	(77.04)	(0.04)	(1.25)	$(\mathbf{0.04})$	(0.04)	(0.02)	(0.05)	(0.05)	(0.10)	(4.28)
Christian (β_5)	-80.09	-0.06	0.70	-0.00	$\mathbf{0.36}^{***}$	-0.02	0.07	0.03	-0.19	-4.56
	(82.68)	(0.05)	(1.28)	(0.04)	$(\mathbf{0.05})$	(0.02)	(0.05)	(0.05)	(0.11)	(5.69)
Eigenvector (SD) (β_6)	-6.58	-0.03	-0.12	0.00	0.01	$\mathbf{0.12}^{***}$	0.01	0.01	-0.10^{*}	-4.44^{*}
	(31.78)	(0.02)	(0.56)	(0.02)	(0.01)	$(\mathbf{0.01})$	(0.02)	(0.02)	(0.05)	(1.94)
Adj. R^2	0.10	0.04	0.06	0.01	0.24	0.31	0.02	0.03	0.08	0.08
Num. obs.	613	613	588	613	597	613	597	597	599	613
$\rho + \theta = 0$ (p-value)	0.88	0.21	0.27	0.89	0.36	0.36	0.25	0.86	0.015	0.27

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Eicker-Huber-White robust standard errors in parentheses. Neighborhood fixed effects included. Nominated TA is a binary variable equal to one if the Targeting Assistant was nominated by another community member. Targeting assistants were asked to provide two nominations for a cash grant. We elicited the cash grant nominations through two independent questions. 'Best Use' elicitation question (reference group in table): "Think of households within this neighborhood, which household would make the most out of a USD$100 cash grant?" 'Hard Times' elicitation question: "Within this neighborhood, is there a household who recently fell on hard times and would benefit most from a USD$100 cash grant?" The order of the two elicitation questions was randomized for each Targeting Assistant. Table A.1.10 shows regression results restrict to the first nominee elicited from each Targeting Assistant. Predicted PCE Rank is the household ranking of the per capita expenditures as predicted by our random forest algorithm, centered at zero. Social Distance is the shortest path between the nominating and nominated household as calculated using social network data from the baseline. Geographic distance is the number of meters between the nominating and nominated household's dwelling as measured by the spherical geodesic distance between the latitude-longitude coordinates of the dwellings.

Do cash grants impact welfare and entrepreneurial outcomes?

Utilizing the matched-pairs randomization design, we sought to measure treatment effects of the cash transfer on outcomes related to household expenditure, welfare, and business activity. Table 1.14 shows that the treatment and control groups were well balanced across baseline characteristics.[13] We find that this is evidence of a proper randomized assignment.

Our preferred method of estimating treatment effects is a differences-in-means estimand stratified by matched pairs. The difference-in-means estimator provides an unbiased estimate of the average treatment effect. With matched pairs, the average treatment effect, $\widehat{\tau}$, is simply the within-pair difference between outcomes for the treatment and control averaged over all pairs. Equation 1.10 lays out the calculation for the difference-in-means estimator, where J is the number of matched pairs, N is the total number of households, and $\widehat{\tau_j}$ is the estimated difference between the mean of the treatment and control household in the jth matched pair. The standard error, $\widehat{SE}(\widehat{\tau})$, is estimated using Equation 1.11. (Gerber and Green 2012; Imbens and Wooldridge 2009). We do not control for any baseline characteristics in our estimation strategy as baseline characteristics are indirectly controlled for through the matched-pairs assignment.

$$\widehat{\tau} = \sum_{j=1}^{J} \frac{2}{N} \widehat{\tau_j} \tag{1.10}$$

$$\widehat{SE}(\widehat{\tau}) = \frac{1}{J(J-1)} \sum_{j=1}^{J} (\widehat{\tau_j} - \widehat{\tau})^2 \tag{1.11}$$

In the endline survey, we sought to confirm the delivery of the cash transfer. Nearly all — 96 percent — of cash beneficiaries recalled receiving the cash transfer. The treatment group was 79 percentage points more likely to report that they received financial assistance from an NGO. There is no equivalent effect on the household reporting assistance from the government. Thus, we note that the receipt of the cash grant was a salient event for the treatment group and they understood that the cash grant was from a non-governmental organization.

Among the cash grant beneficiaries, respondents reported a diverse set of ways that they used the cash (Figure 1.4). The most common reason was for food expenses (41%). 31.8 percent stated that they used the cash transfer to start a business or to improve an existing business. 23 percent said they used the cash to pay for school fees. 19 percent said they spent the money on clothing. Also, 16 percent said that they saved part or all of the USD$80 provided through the cash transfer from IPA.

While it is confirming that the treatment households report that they received the cash transfer through an NGO, these treatment effects are mechanical and expected. What we are interested in are treatment effects on household behaviors. Table 1.15 displays difference-in-means estimates for the four main outcomes of interest. For each outcome, we display the pooled difference-in-means estimates using all matched pairs. The "No Nomination" row of the table displays the

[13]The one characteristic where we reject the null hypothesis for equality of means is whether or not a household rents its dwelling. The treatment group is more likely to be a renter.

Table 1.14: Baseline Summary Statistics, by Treatment Group

	(1) Control N=280	(2) Treatment N=280	(3) p-value
Welfare:			
Household expenditure (LD)	858.30 (730.39)	901.22 (1048.76)	0.58
Per capita expenditure (LD)	202.81 (174.11)	210.78 (215.20)	0.63
Predicted PCE (LD)	149.32 (71.02)	147.04 (67.11)	0.70
Household rents dwelling	0.63 (0.48)	0.71 (0.46)	0.06
Number of rooms in dwelling	1.59 (1.20)	1.46 (1.01)	0.16
Proxy means score	46.68 (14.92)	46.54 (15.12)	0.91
Cantril ladder (1−5)	2.18 (1.12)	2.19 (1.19)	0.96
Meals per day	1.68 (0.67)	1.65 (0.67)	0.62
Any shock in past 12 months	0.67 (0.47)	0.66 (0.47)	0.93
Wealth shock in past 12 months	0.49 (0.50)	0.54 (0.50)	0.25
Health shock in past 12 months	0.37 (0.48)	0.33 (0.47)	0.37
Demographic/Social:			
Household size	4.82 (2.30)	4.80 (2.23)	0.92
Household head is female	0.31 (0.47)	0.31 (0.46)	0.93
Years in community	10.44 (10.01)	10.15 (9.76)	0.74
Household includes community leader	0.14 (0.35)	0.13 (0.34)	0.71
Religion is Christian	0.75 (0.43)	0.76 (0.43)	0.79
In-degree centrality	3.45 (2.95)	3.70 (3.16)	0.33
Betweenness centrality	275.19 (360.17)	313.28 (378.48)	0.22
Eigenvector centrality	0.16 (0.20)	0.17 (0.20)	0.62
# Family assistance	0.79 (1.07)	0.86 (1.23)	0.48
# Friends assistance	0.70 (1.09)	0.63 (1.05)	0.42
Government assistance	0.01 (0.10)	0.01 (0.09)	0.65
NGO assistance	0.03 (0.16)	0.02 (0.13)	0.55

Notes: Standard deviations in parentheses. P-value is derived from the t-test for equality of treatment and control means.

difference-in-means estimates for the 69 matched pairs that did not receive a nomination, i.e. were randomly sampled from the poorest quintile of the proxy-means score distribution. The "Best Use" row uses the 78 matched pairs that were nominated using the "Best Use" elicitation prompt in the TA survey. The "Hard Times" row uses the 75 matched pairs that were nominated using the "Hard Times" elicitation prompt in the TA survey.

Panel A of Table 1.15 shows the effects of the cash grant on per capita expenditures. Across all sub-samples—as well as the pooled estimate—we fail to reject the null hypothesis that the cash grant had zero impact on per capita expenditures. Panel B shows that, likewise, we do not observe any statistically significant impacts on household savings.

In Panel C of Table 1.15, we observe that the cash grant led to a *decrease* in financial satisfaction.[14] The estimate, however, does not survive Bonferroni for the four hypotheses that we test under this outcome.

During the endline survey, we asked households about business activities. Panel D of Table 1.15 shows our difference-in-means estimates for whether or not a household has an active business in operation. We see that those assigned to receive a cash transfer were more likely to be operating a

[14]Financial satisfaction is a dummy variable equal to one for the 'Satisfied' and 'Very Satisfied' responses to the question, "How satisfied are you personally with the financial situation of your household?"

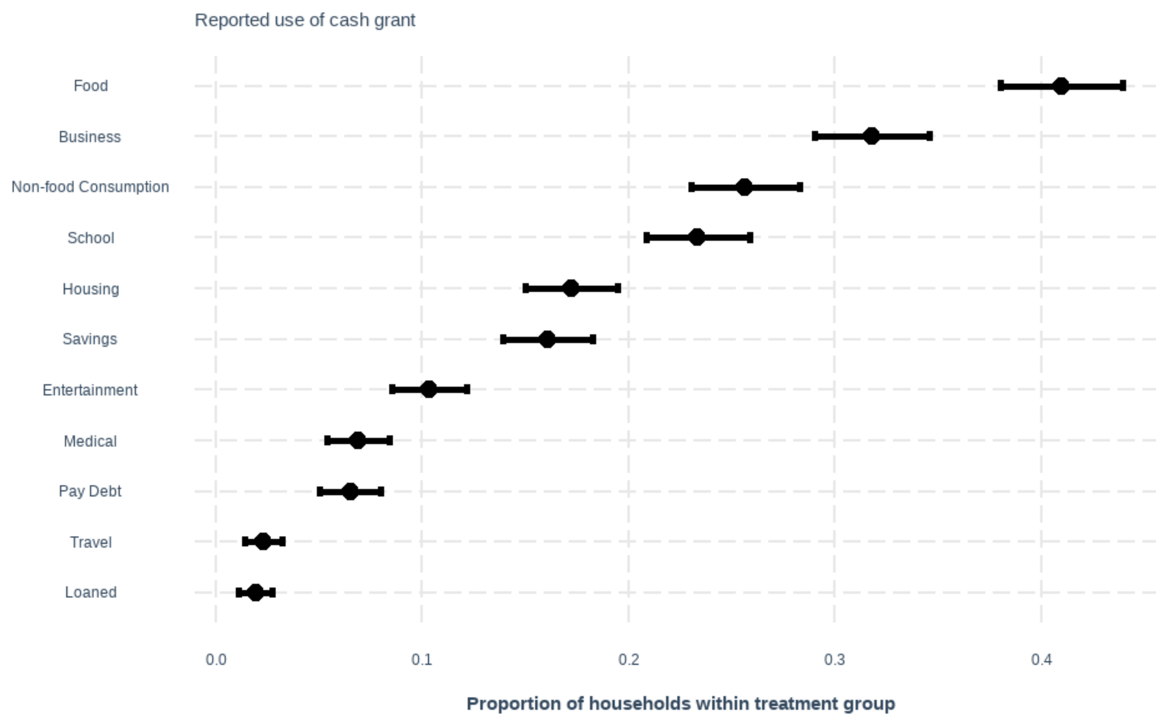

Figure 1.4: **Reported Use of Cash Grant.** Multiple responses permitted, thus values do not sum to one.

business (p-value = 0.08). This effect is highest among cash beneficiaries that were not nominated through the TA survey, the coefficient on the "No Nomination" group survives a Bonferroni adjustment for four hypotheses at the 0.05 significance level.

Taken together, our results do not suggest that the cash grant impacted welfare and business activities in a clear and systematic manner. Moreover, there is no evidence that leveraging nominations directed cash grants towards households that would have a higher impact.

1.5 Conclusion

This paper analyses data from a multi-stage experimental approach to targeting a social program. This paper provides novel data to understand individual preferences for targeting social programs. Whereas previous work has focused on algorithmic approaches to identifying poor households in comparison to community targeting, we present evidence to help decipher individual perceptions of knowledge within a neighborhood social network.

Using data collected from over 2000 households in 13 densely-populated urban neighborhoods in Monrovia, Liberia, we developed a novel approach to estimating household welfare by training a random forest model with expenditure and asset data from households. We show that predicted

Table 1.15: Cash Grant Treatment Effects

	$\widehat{\tau}$ ($\widehat{SE}(\widehat{\tau})$)	p-value	95% CI
Panel A: Per capita Expenditures			
Pooled	5.75 (11.09)	0.60	[-16.11, 27.60]
No Nomination	9.13 (19.99)	0.65	[-30.76, 49.01]
Best Use	-21.04 (17.29)	0.23	[-55.47, 13.39]
Hard Times	30.49 (20.18)	0.14	[-9.73, 70.71]
Panel B: Household Savings			
Pooled	-69.01 (60.67)	0.26	[-188.56, 50.55]
No Nomination	-201.07 (189.17)	0.29	[-578.56, 176.42]
Best Use	-26.11 (37.60)	0.49	[-100.97, 48.75]
Hard Times	7.88 (20.70)	0.70	[-33.36, 49.12]
Panel C: Satisfied with HH Financial Situation			
Pooled	-0.05 (0.05)	0.33	[-0.14, 0.05]
No Nomination	-0.19 (0.08)	0.02	[-0.35, -0.03]
Best Use	0.10 (0.08)	0.21	[-0.06, 0.26]
Hard Times	-0.07 (0.07)	0.37	[-0.21, 0.08]
Panel D: Has Active Business			
Pooled	0.08 (0.05)	0.08	[-0.01, 0.17]
No Nomination	0.22 (0.08)	0.01	[0.05, 0.39]
Best Use	0.09 (0.07)	0.21	[-0.05, 0.23]
Hard Times	-0.05 (0.08)	0.51	[-0.21, 0.11]

Notes: Difference-in-means estimator with matched pairs shown in the column 1. Eicker-Huber-White robust standard errors in parentheses. Degrees of freedom: Pooled = 221, No Nomination = 68, Best Use = 77, Hard Times = 74.

expenditures through this measure yield a meaningful indicator of household welfare across a host of measures.

We found that targeting through social network leads to a high proportion of nominations defined by homophily. Principals nominate agents that are similar to them across a vector of six characteristics. This holds at two stages of our analysis, when nominating knowledgeable individuals within the neighborhood and when nominating cash grant recipients.

We found that social distance is important in the accuracy of information provided by agents (targeting assistants). Agents provide more accurate information in targeting poor neighbors if the neighbor is a one-degree distance from the agent on the social network. However, agents provide inaccurate information when their first-degree neighbor is not poor. Agents that were nominated by others (Nominated TAs) are particularly good at minimizing exclusion error and particularly bad at minimizing inclusion error.

Cash grants targeted through a decentralized system do indeed go to poorer households. On average, those nominated for a cash grant are poorer than those households that are not nominated. We do not observe any evidence that agents direct the cash grants to households that have experienced a shock.

We believe that the evidence presented in this paper can help to understand the role that social networks in setting preferences for targeting of social programs. This is particularly true in urban neighborhoods. As targeting of social programs in urban areas gains importance, a better understanding of social networks is important to best design mechanisms for targeting social programs that leverage social connections.

Chapter Two

Demographics, Social Networks, and Mobile Phone Usage

Introduction and Take up of Community Cellular Networks in the Philippines

Abstract

We investigate the determinants of cellular network adoption in the context of remote locations of the Philippines. We leverage unique circumstances where all households in seven localities were interviewed before the installation of cellular network towers. We link rich socio-economic as well as social network data to call detail records from the first five months after community cellular networks were introduced. First, we show that 68 percent of households owned a cellphone before the cellular network installation, although other channels were much more commonly used as sources of information. Second, we examine and attempt to explain variance in network usage across the seven localities. Finally, we provide a unique descriptive analysis of social and economic factors that correlate with cellular network adoption. We find that network adoption and volume are correlated with the wealth of the household. A one-standard-deviation increase in the wealth of a household is associated with a 3 percentage point increase in cellular network adoption. Farming and Fishing households are particularly likely to join the cellular network. Wealthy households also record more transactions (calls and texts) on the network. Also, households with higher levels of education make and receive more text messages. Our results contribute to the literature on cellular networks by revealing household-level characteristics that should be taken into consideration as "last-mile" ICT interventions are considered.[†]

[†]The material in this chapter is based on joint work with Joshua Blumenstock, Arman Rezaee, and Erin Troland. Innovations for Poverty Action conducted the data collection and project management in the Philippines. A team from the University of the Philippines installed and managed the community cellular networks.

2.1 Introduction

More than 5 billion people have mobile phone subscriptions, yet the expansion of subscribers has slowed in recent years. The reduced pace of growth presents a considerable challenge in the pursuit of ubiquitous communication systems. One of the Sustainable Development Goals is to, "Significantly increase access to information and communications technology and strive to provide universal and affordable access to the internet in least developed countries by 2020."[1] The World Economic Forum pinpoints a similar goal of "Internet for All" (World Economic Forum 2018). The GSMA points to mobile subscriptions as a predicate to expanding access to phone and internet communications but cites demographic and geographic challenges to expanding mobile connections (GSMA 2019). Namely, the GSMA notes that individuals with limited formal education, low employment potential, and the elderly are less likely to be connected. Rural households are also disadvantaged due to infrastructural constraints.

To better understand the constraints as well as the potential to expanding ICT connections, it is worthwhile to examine the constraints and potential for expanding mobile phone access. Much of the literature on communications networks in developing countries relies on administrative data and lacks demographic and socio-economic data. Sarker and Wells (2003) provides an early analysis of cellphone adoption and usage. Shah et al. (2017) discuss characteristics of the phone used by community cellular network subscribers in the Philippines. Ahmad et al. (2016) also examine device characteristics of mobile network subscribers on a major network in Pakistan. Heimerl, Menon, et al. (2015), using call detail records from one community cellular network in Indonesia, explore the uptake of the network and expansion of smartphones. The studies mentioned lack detailed socio-economic data to unpack the determinants of mobile phone usage.

We examine access to information prior to the installation of cellular networks in seven remote locations of the Philippines and mobile phone use during the first five months of network service. Resembling the work of Blumenstock and Eagle (2012), Blumenstock (2014), Blumenstock, Cadamuro, and On (2015), and Blumenstock (2018), we leverage detailed socio-economic survey data that can be paired with individual-level call detail records (CDR). Unlike previous work in which survey data and CDR are analyzed, we utilize a detailed social network census to measure social connections as reported by households with their observed communication activity through call detail records. As such, the contribution of this paper is to examine demographic, economic, and social characteristics that correlate with mobile phone usage prior to and after the installation of cellular networks.

We structure our presentation as follows. First, we discuss the broad communications landscape in the Philippines. Next, we describe the installation of an innovative approach to expanding phone coverage, the Community Cellular Network. Between September 2017 and January 2019, seven new networks were installed in remote areas of the Philippines. We discuss the selection of sites as well as demographic characteristics of households in the sites. In Section 2.2, we describe our primary data sources, namely baseline survey data from all households and call detail records from all phone transactions in the first five months of the cellular networks. In Section 2.3, we

[1] sustainabledevelopment.un.org

provide a detailed descriptive analysis of household demographics, economic welfare, and social network position at the time of the baseline survey. We show that phone owners differ from non-phone owners before the launch of cellular networks. In Section 2.4, we discuss details of the network installations and registration of mobile subscribers. In Section 2.5 we provide a descriptive analysis of network take up for each of the seven sites. We also examine the determinants of network usage by households. We provide a discussion of the findings and broader relevance in Section 2.6.

Communications in the Philippines

The Philippines provides an ideal setting to study Community Cellular Networks. The country is an archipelago composed of about 7,641 mountainous islands. The topology of the Philippines results in thousands of localities that are isolated from other parts of the country. Much of this country faces the "last mile" connectivity gap. This gap is caused by the fact that telecommunications companies have not found it commercially viable to bring cellular towers to many of the country's small, remote islands nor to the mountainous, coastal regions of its larger islands. Approximately 63 percent of the Philippines population subscribe to a mobile network operating system, leaving over 25 million people disconnected (GSMA 2018).

Community Cellular Networks

This paper contributes to a burgeoning literature on low-cost mobile phone networks in rural areas of the world. Originally developed by Heimerl and Brewer (2010) and named the Village Base Station (VBTS), the network installation provides four main benefits:

- flexible, low-power deployment requirements that leverage local power generation via solar or wind;
- support for local services within the locality with the potential to run autonomously;
- power/coverage optimization to minimize loss of coverage; and
- a portfolio of data and voice services.

For the purposes of this study, a VBTS provides a practical, open-source technology (both hardware and software) that can be implemented to create Community Cellular Networks (CCN) in remote, off-the-grid localities in the Philippines. This technology was adapted for the Philippines by a team of researchers from the University of the Philippines (UP), University of Washington, and University of California Berkeley (Barela et al. 2016), with the regulatory support of a national mobile network operator, Globe. This process included hardware procurement and fabrication, software design and integration, site selection, engineering and construction of towers and solar grids for VBTS boxes in these sites, working with the major telecommunications company and satellite connection providers, and mobilizing local communities to maintain a CCN. The literature on the CCN technology focuses on the hardware and software systems (Barela et al. 2016; Hasan et al. 2019; Heimerl and Brewer 2010; Heimerl, Hasan, et al. 2013; Jang et al. 2018). Our concentration in this paper is on the descriptive analysis of social and economic factors that correlate with the adoption and usage of CCNs.

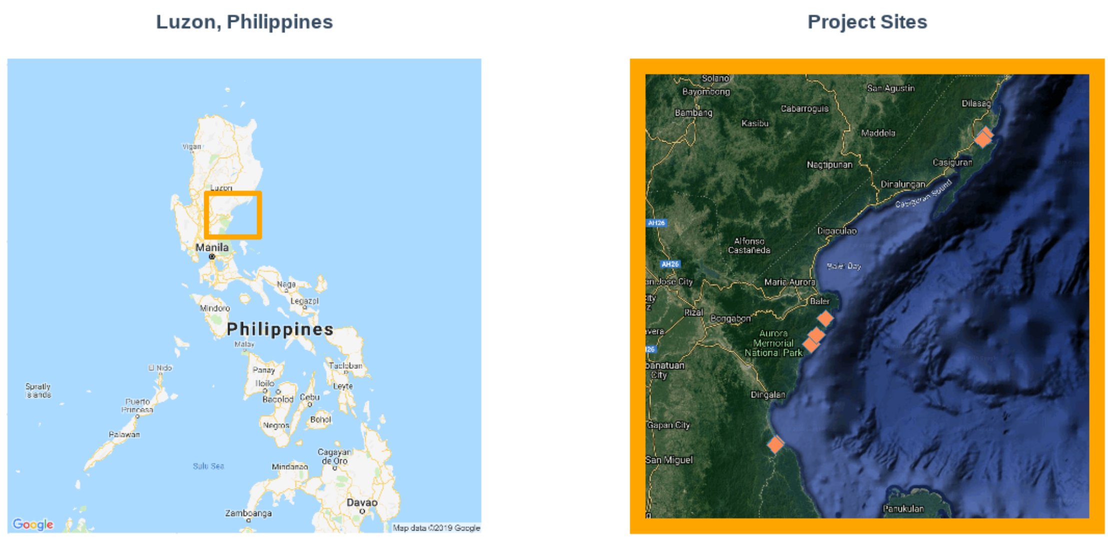

Figure 2.1: Project Sites

Site Identification

As described in Barela et al. (2016), we identified candidate sites for CCNs in the Philippines—areas that were both remote enough to lack cellular network coverage but not so remote as to make the logistics of research infeasible. In each potential site, the team from UPD conducted spectrum analysis to assess the existence and quality of cellular network signal as well as any potential topographical features that could pose a challenge to implementing a community cellular network. We identified fourteen candidate sites along the west coast of Luzon, the largest island in the Philippines. All sites are located in Aurora Province. Sites are villages, or "sitios" using the Philippines term for the lowest-level administrative unit, located near or along the coast. The localities where this study was conducted has some of the highest poverty rates in the country. The coastal and mountainous terrain make mobile connectivity difficult. Many of the potential sites are located in sea coves only accessible by boat. Aurora is located within a path frequently hit by typhoons. During typhoon season, coves are often inaccessible for multiple days.

From the original fourteen candidate sites, seven were randomly selected to receive a CCN tower. In this paper, we focus our analysis on the seven project sites that received a CCN tower.[2]

[2]For more details on the randomized controlled trial, see Blumenstock, Keleher, et al. (2019).

Figure 2.1 shows the location of the seven selected sites as well as the location of the region within the Philippines. The main regional town in the area, Baler, is a multiple hour commute by bus or boat from the project sites. Each candidate site was deemed able to support a CCN. CCN base stations transmit to a 500-meter radius, though terrain often cuts this distance. Field teams visited all potential sites to verify eligibility (no current cellular connection), determine possible logistics, and meet with local government units (LGUs). We then randomly selected seven sites that would receive an initial installation of a CCN tower. Figure 2.1 shows the location of the seven sites that were selected to receive a Community Cellular Network tower.

2.2 Data

Survey Data

Prior to the installation of CCN towers, we interviewed adults from all households in the selected rural localities of Aurora Province. Among the seven CCN sites, 1,131 households were interviewed at baseline.[3] The household survey consisted of modules about household demographic composition, asset ownership, and economic activity. To participate in the study, we asked for voluntary consent from all survey respondents.

Table 2.1 provides site-level population counts. Sites range from 50 to 382 households. On average, there is one child under the age of 15 for every two adults living within the study population. In total, we listed 3,057 adults out of which we conducted 1,617 one-on-one surveys. The core modules contained within the adult survey were a travel diary and a social network module. Women comprise 62 percent of the adult survey respondents. In sum, before CCN tower installation, we surveyed all household living in the seven CCN project sites and 53 percent of all adults.

Table 2.1: Site Information

	Population				**Social Network**				
VBTS Site	Households	Adults	Children	HH Owns Phone	Edges	Largest Component	Density	Avg. Clustering	Avg. Distance
Site 1	88	215	146	0.80	390	87	0.10	0.33	3.09
Site 2	382	1103	787	0.69	2354	381	0.03	0.37	2.34
Site 3	176	495	312	0.47	919	175	0.06	0.31	3.20
Site 4	100	251	159	0.57	288	95	0.06	0.25	3.50
Site 5	255	646	371	0.85	1777	255	0.05	0.41	2.22
Site 6	50	128	85	0.58	163	50	0.13	0.27	2.63
Site 7	80	219	139	0.62	282	78	0.09	0.18	2.58

Phone ownership varies across sites. We show in Table 2.1 that even in the absence of cellphone service in the site many households own a cellphone. In two sites, more than 80 percent of

[3]In total, we collected baseline data from 2,370 households across the fourteen sites included in the randomized controlled trial.

households owned a cellphone at the time of the baseline survey. In only one site, Site 3, was phone ownership below 50 percent of households.

We took painstaking efforts to identify local and non-local social network ties at the time of the baseline survey. For local social networks, we asked respondents that participated in the adult survey to name their closest friends and family that lived in the same barangay. We then matched the names of their contacts with names from our household listing. Using these data, we were able to construct a social network graph for each site to identify social linkages, or "edges", between households. In total, we were able to identify 6,173 edges among the 1,131 households. As shown in Table 2.1, most households are included in the largest component of the social network graph. Between 95 and 100 percent of households can be reached via connections in the social network. The density of the social network graph—that is the fraction potential links that are present—is low but on par with social network data from other settings (See Alatas et al. (2016) and Bloch and Olckers (2018)). Average social distance, that is the shortest path between any two households, in each site ranges between 2.2 and 3.5 steps.

Call Detail Records

Once a cellular network launched, all cellular transactions were logged for the CCN. In this paper, we work with the raw Call Detail Records (CDR) from the seven CCN sites. SIM cards were assigned to all adults living in the site. We provide details of the SIM assignments in Section 2.4. We were able to link each adult to a unique record from the baseline survey data. CDR include an identifier for the initiating and receiving parties, the type of transaction, the date-time, the tower used, the cost of the transaction, and the duration of calls. We limit our analysis to call and text message transactions in the first 144 days of the CCN in each site. We limit our analysis to this period for two main reasons. First, we want to focus on the adoption of the mobile network, thus we concentrate on the earliest period of the network. Second, the dates of CCN launches were staggered for logistical reasons. The last CCN tower was installed in January 2019. Restricting our analysis to the first 144 days of each site maximizes our window of time for this last site while also creating a uniform period of time to examine each CCN site.

We also limit our analysis to phone calls and text messages. We drop invalid calls and texts as well as text messages sent to special codes (i.e. to check account balance). In total, 952,876 phone calls and text messages were sent or received by CCN subscribers in the first 144 days of tower activity across all seven CCN sites.

2.3 Information Networks prior to installation

Panel A of Table 2.2 shows demographic, welfare, asset ownership, and social network characteristics of households before the launch of the CCN. Households comprise, on average, 2.7 adults and 1.8 children under the age of 15. One-third of household heads are women. One-quarter of household heads have a secondary school degree.

Economic activity and sources of income were mixed. One-quarter of adults did not work in the week preceding the baseline interview. The majority of residents lived and worked within the sitio. Only 25 percent of adults traveled outside of their sitio for work in the twelve months preceding the baseline interview. Individuals do, however, travel for non-work reasons. Half of all adults expected to travel outside the sitio in the 12 months following the baseline interview.

Electricity coverage was fairly widespread, with 63 percent of households having access to some form of electricity. Communication technologies were, however, observed in fewer households. 32 percent of households owned a radio, 52 percent owned a television, and 31 percent owned a satellite dish.

To estimate household welfare, we constructed two commonly-used metrics for assessing the relative wealth of households. First, we included questions from the Poverty Probability Index (PPI) Scorecard.[4] The PPI scorecard is a set of ten questions that, when considered together, are most predictive of per capita expenditures (see Kshirsagar et al. (2017) for discussion of the PPI methodology). The ten questions from the PPI scorecard can be found in Appendix Section B. The PPI score indicates the probability of a household being below the poverty line. Lower scores indicate that a household is more likely to fall below the poverty line. Among the seven CCN sites, the mean household has a PPI score of 42.17, which translates to a 56.4 percent probability that an individual in the mean household lives on less than USD$2.50 per day.

As our second measure of household welfare, we calculate an asset wealth index using the first component of a principal component analysis of 14 asset questions in the baseline household survey.[5] The first component has been shown to be a reliable predictor of household socio-economic status (Filmer and Pritchett 2001). Several of our asset variables are categorical factor variables; thus, we follow the suggestion of Kolenikov and Angeles (2009) by using the polychoric correlation matrix in the principal component analysis. The Pearson correlation coefficient between the PPI score and the PCA wealth index is 0.55.

Despite the lack of mobile network access prior to the CCN installation, the majority (68%) of households owned a cellphone and a SIM card (64%). On average, households owned 1.2 phones and 1.3 SIM cards. There are two major mobile network operators in the Philippines. One-quarter of households owned a SIM card from Globe while 55 percent had a SIM card from the Smart network[6]. To use the SIM card, people need to travel outside of the sitio. The CCN provided, for the first time, reliable local cellular network service within the localities.

Using the social network data that was collected in the adult survey, we construct two measures of social importance for each household. First, we calculate the in-degree centrality of a household by summing the number of times members of a household were named as a close friend or family member by others during the baseline adult survey. Second, we compute the eigenvector centrality of each household, which is a measure of the position of a social network node that accounts for

[4]https://www.povertyindex.org/country/philippines

[5]The 14 assets used in the principal component analysis were: land ownership (0/1), number of rooms in the dwelling, access to electricity in the dwelling (0/1), wall type, roof type, floor type, and ownership (0/1) of sala sets, refrigerator, television set, video player, radio, satellite dish, vehicle, and gas stove.

[6]The two largest mobile network operators have dual-branded cellular network offerings. Globe includes the Globe and Touch Mobile (TM) SIM cards. Smart includes Smart and TNT SIM cards.

the centrality of nodes that are connected to it. Thus, households with high eigenvector centrality are connected to central households, which are connected to central households, and so on.

Columns 2 and 3 of Table 2.2 provide comparative statistics for households that reported phone ownership versus those that said they did not own a phone at the time of the baseline survey. As seen in these columns, households that did not own a phone differ significantly from those that do. In addition to not owning a phone, asset ownership was lower for all of the main assets that we asked about. Households without phones ranked lower on both welfare scores and are less likely to have a bank account. However, these households were no less central to social networks in the sites. Households that lacked a cellphone were equally likely to have a family member in political office and were equally, if not more, central to the social network — we fail to reject the null hypothesis that in-degree centrality is equivalent but reject the null that eigenvector centrality is equivalent. Adults in households without phones were no less likely to report social ties inside the barangay.

Phone ownership correlates with non-local communication networks. Adults from households that owned a phone reported that they traveled outside their barangay more than adults from households lacking a phone. Phone owners were more likely to have close friends outside of the barangay. Also, phone owners stated that they were more likely to communicate with family outside of the barangay in case of emergency.

Before the CCN launch, in-person communications and television were dominant sources of information about daily events in the Philippines (see Figure 2.2). More than 60 percent of adult survey respondents stated that they spoke with friends and family and watched television on a daily basis to learn about events and developments in the Philippines and around the world. Nearly 90 percent of households receive information from these sources on a weekly or more frequent basis. As shown in Figure 2.2, the majority of households reported that they never received information via mobile phone prior to the baseline survey.

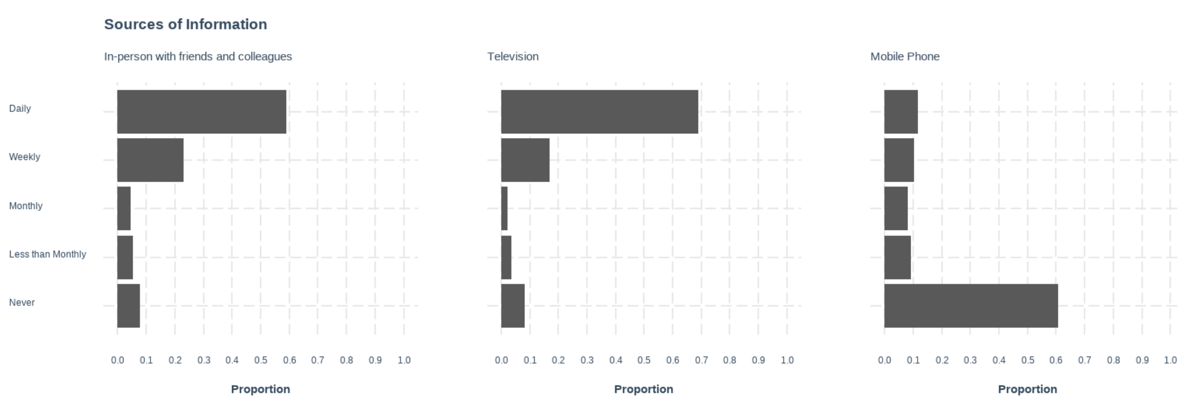

Figure 2.2: Sources of Information

2.4 Installation of Community Cellular Networks

CCN Installations and Launch

After the conclusion of the baseline survey, the team from the University of the Philippines Diliman (UP) initiated activities to install the CCN towers. The schedule of installation was staggard for logistical reasons and the order of tower installation was randomized.

The UP team met with local community cooperatives to identify key contacts for managing the community cellular networks. They also received the appropriate permissions from administrative leaders, regulatory bodies, and the leading telecommunications company through which the cellular network would receive spectrum. A team of engineers installed towers and solar panels to provide power to the CCN tower in central locations of each site to maximize coverage. The UP team ran network tests to confirm that the tower worked adequately. Also, they trained a select group of local community members to use and maintain the tower. Finally, they designated local vendors to sell phone credit, "e-load." After the construction of the CCN tower, we organized launch and registration events in each site. For the project, the CCNs were referred to as the VBTS Konekt Network.[7]

The VBTS Konekt launch was a significant event for each locality. Several days in advance of the registration event, we informed each community of the scheduled launch of the new cellular network. Prior to doing so, we obtained local authorization and agreement on the date and time of the community registration event. With assistance from local officials such as kagawads, barangay tanods, sitio/purok, leaders, or tribal council members, we established the date, time, and location for the community registration event. The local leaders were responsible for identifying barangay representatives to assist in spreading the word about the community registration event. Community members were informed that any residents, 15 years of age or older, would be able to collect one free SIM card at the registration.

At the community registration event, we explained the purpose and motivation for the community cellular network. Community members were informed that the towers were part of a research project and the towers would not necessarily remain beyond the duration of the research study. We described how to use VBTS Konekt and details for utilizing the SIM cards.

Residents that attended the community registration event were then asked to register and activate their SIM cards. A team of registration enumerators from Innovations for Poverty Action verified the identity of individuals interested in activating a VBTS Konekt SIM card.

For each customer, a registration staff member explained the steps to activate a SIM card (see Appendix Section B for the script that was followed). VBTS Konekt SIM cards were provided for free to customers. To be eligible for a SIM card, a customer needed to have a functioning GSM 900 or multi-band cellphone. No phones were provided to customers by the research team. Customers were read the user agreement (see Appendix Section B) and required to accept the terms of the agreement before receiving their SIM card. Registration staff assisted customers to

[7] VBTS stands for Village Base Station.

activate their SIM card and inform them of their unique phone number (MSISDN). SIM cards could be replaced if they were lost or malfunctioned in which case the customer would retain the same phone number in the event of a SIM replacement. Additional SIM cards could be bought for 15 Pesos.

Using the baseline survey data, we assigned a SIM card to every adult in CCN sites. The unique phone number was assigned to an IMSI subscriber identifier which was in turn linked to a unique identifier from the baseline household survey. When an interested customer was unidentified in the baseline survey database, we required a household visit to confirm that the individual was a resident of the treatment site. SIM cards were provided at no charge to the subscriber.

The VBTS Konekt Network allowed for calls to and from other mobile and landline phones within the Philippines. Text messages were limited to the local Konekt network and long-distance on-network phone numbers in the Philippines. Table 2.3 provides the schedule of tariffs for all network interaction types. Local calls and texts were the lowest cost, on-network long-distance calls and texts were billed at a higher rate than local interactions, and off-network interactions were the most costly. All incoming calls and texts were free of charge to the customer; however, the calling party for incoming calls and texts were charged at standard long-distance rates. Due to regulatory restrictions, texts from off-network numbers could not be received. Similarly, international transactions were prohibited on the VBTS Konekt Network.

Customers were informed that they could purchase phone credit through retailers based within the site. Each site had between one and three retailers. To promote the take-up of the network and encourage customers to try the network, all customers that activated their SIM card received five free text messages. Customers were also informed that promotions might be offered to them at a later date.

Following the launch event, UP network administrators enabled the cellular network for all activated SIM cards. The network could only work through VBTS Konekt SIM cards. Customers could purchase phone credit directly through local retailers.

2.5 Mobile Network Adoption

Site-level Network Usage

Figure 2.3 shows daily calls and text messages for each of the 7 CCN sites. We have shaded the period of analysis — the first 144 days after the network launch — in gray. The actual dates when the periods begin range from September 2017 for Site 1 to January 2019 for Site 7. There is a clear and marked difference in site-level usage. Sites 2, 3, and 4 consistently had more than 1,000 transactions per day. Site 7 also shows high activity with over 100 transactions on most days. In contrast, we observed approximately 100 transactions per day in the first two weeks of the network in Sites 1, 5, and 6 that rather quickly decreased to below ten by the third month of the network.

Table 2.4 provides aggregate statistics of usage in each site during the first 144 days of service. Several findings stand out. Mean usage per household ranges from 10.6 transactions in site 5 to

178 transactions per household in Site 2. In sites 2, 3, 4, and 7, more than 3 transactions were logged per household on an average day in the period covered by our analysis. Incoming calls were much more common than outgoing calls. Incoming calls were also longer in duration as compared to outgoing calls. However, outgoing texts tended to be more common than incoming text messages. This pattern of communication comports with practices observed elsewhere in that low-income households concentrated their network usage on text messages to avoid more costly phone calls.

In Figure 2.4, we show that sites also differed in their use of local and long-distance communications. By long distance, we mean any transaction that is outside of the CCN, i.e. to a phone number that is registered with another mobile network operator. Across all sites, long-distance communications were more common. In two sites, Sites 1 and 5, local calls and texts were avoided almost entirely. The solid lines in Figure 2.4 represent incoming transactions while the dashed line represents outgoing transactions. Local calls and SMS messages are paired; thus, the outgoing and incoming lines are indistinguishable.

Household-level Network Usage

Nearly two-thirds of households initiated or received at least one call or text message during the first 144 days of the cellular network. As shown in Table 2.5, ownership of a phone prior to the network launch does not explain network adoption across all sites. However, as shown in Table A.2.2, network adoption ranges from 17 percent of households in site 5 to 94 percent in site 2. Households that reported owning a phone at the time of the baseline survey had, on average, 150 greater transactions than households that did not report phone ownership during the baseline survey.

Approximately 60 percent of households had at least one outgoing call, one incoming call, one outgoing text and one incoming text. Most of the difference in phone transactions by baseline phone ownership is explained by higher use of text messages by households that owned a phone prior to the CCN launch. This could be due to a number of factors, including literacy levels that are correlated with asset wealth as well as digital literacy that might increase comfort in using a phone to type and read text messages. We do not see any differences, on average, in outgoing calls based on previous phone ownership. Households with phones at baseline received more phone calls, but we are unable to reject the null hypothesis that incoming call volume is equivalent to that of households that did not own a phone at the time of the baseline.

We now turn to multivariate regression analysis of the determinants of household usage of the CCN. Table 2.6 shows results from running the regression specification in Equation 2.1. Our dependent variable, Y_i, is one of six quantitative measures of cellular network usage. Each measure is identified by a column in Table 2.6. Any Transaction is a binary variable that takes the value of one if a household has at least one transaction (call or text) in the CCN call detail records. Total Transactions is the count of all incoming and outgoing calls and texts associated with phone numbers registered to a given household. Out Calls, Out SMS, In Calls, and In SMS correspond to a similar count at the household level for outgoing calls, outgoing text messages, incoming calls, and incoming text messages, respectively.

$$
\begin{aligned}
Y_i =&\beta_1 OwnedPhone_i + \beta_2 WealthIndex_i + \beta_3 Contacts_i^{LD} \\
&+ \beta_4 NetCentrality_i^{Local} + \beta_5 HHSize_i + \beta_6 FemaleHOH_i \\
&+ \beta_7 SecSchoolHOH_i + \beta_8 FarmingIncome_i + \beta_9 FishingIncome_i \\
&+ \beta_1 0 WorkOutsideBgy_i + \beta_1 1 CogAbility_i \\
&+ \nu_s + \epsilon_i
\end{aligned} \tag{2.1}
$$

We include several household characteristics from the baseline survey as covariates in our regression specification. $OwnedPhone_i$ is a dummy variable indicating whether household i reported owning a phone at the time of the baseline survey. $WealthIndex_i$ is the first component of the principal component analysis described in Section 2.3. For our regressions, we transform the wealth index to a standardized value. $Contacts_i^{LD}$ is the number of contacts outside the household's barangay reported in the adult survey. For households with more than one adult survey, we use the highest number of contacts reported by an adult in household i. $NetCentrality_i^{Local}$ is a measure of social network centrality using social ties within the site. For our regressions, we use in-degree centrality; results are similar using eigenvector centrality. $HHSize_i$ is the number of adults and children living in the household. We include a dummy variable, $FemaleHOH_i$, to indicate if the head of the household is a woman and a dummy variable, $SecSchoolHOH_i$, that equals one if the head completed secondary school. We include dummy variables for the primary source of income at baseline. $FarmingIncome_i$ and $FishingIncome_i$ are equal to one if the household reports farming or wage labor, respectively, as the primary income source for the household. We include a dummy variable, $WorkOutsideBgy_i$, equal to one if any member of the household worked outside of the barangay in the 12 months preceding the baseline survey. Finally, we include a measure of cognitive ability, $CogAbility_i$, which is the maximum score from the first component of a PCA using three ravens test questions and four numeracy questions. We also include site fixed effects, ν_s, in all regressions.

In Column 1 of Table 2.6, we see that CCN adoption was correlated with household wealth. Controlling for other household characteristics and site fixed effects, we see that a one standard deviation increase in the wealth index is correlated with a 3 percentage point increase in network adoption. Households that reported farming or fishing as their main source of income were 10–11 percentage points more likely to adopt the network than other households.

Although we fail to reject the null hypothesis at the 5 percent significance level, we also see that households that owned a phone at baseline were five percentage points more likely to have at least one transaction on the CCN ($p = 0.062$). We find that larger households and female-headed households were more likely to have used the CCN. The result for female-headed households is encouraging as it suggests that female headed households were 5 percentage points more likely to join the cellular network compared to male-headed households, controlling for other covariates. This indicates a demand that could help to reduce gender barriers to ICT technologies.

When we look at the intensive margin of CCN usage, we observe that wealth and education are primary determinants of cellular network activity. The household wealth index is positively correlated with all types of transactions. A one-standard-deviation increase in the wealth index is associated with 44 additional cellular network transactions. The majority of increased transactions come from outgoing and incoming text messages and incoming calls. Put another way a standard deviation increase in household wealth is associated with one additional outgoing text message every two weeks and one additional outgoing call every month.

We do not observe any meaningful correlations between non-local and local social network measures and CCN usage. However, we do observe that in households where the head has a secondary school degree were more active on the cellular network. These household were notably more likely to send more text messages. Controlling for the head of household's education, we fail to reject the null hypothesis that cognitive ability un-associated with network adoption and usage.

In summary, we observe several characteristics influenced whether or not a household participated in the community cellular network. Foremost among these were the wealth of the household, the number of people living in the household, the primary income source, and whether the household head was a woman. However, the volume of network usage was primarily driven by the wealth of the household. Households where the head has a secondary education displayed higher text message volumes.

2.6 Discussion and Conclusions

The descriptive analysis presented in this paper provides insight into social networks and economic well-being before the installation of cellular towers in seven remote locations of the Philippines. We leverage a unique scenario where we are able to examine economic and social characteristics that correlate with the adoption of the cellular network.

Our findings point to evidence that most households in the study actively sought out information even in the absence of a cellular network, primarily through in-person communications and television. Moreover, nearly two-thirds of households owned a cellphone despite the absence of a cellular network signal where they live. Phone ownership is highly correlated with indicators of wealth and the presence of long-distance social networks.

Take-up of the community cellular networks varied greatly across installation sites. Three sites displayed anemic usage statistics. Two of the three sites with low usage had phone ownership rates in excess of 80 percent prior to the CCN installation. We believe that this is a signal that these sites had easier access to cellular networks prior to the project. In future research, we will investigate the influence of network quality on take-up.

Four sites demonstrated rapid take-up and sustained activity. Because this is the largest installation of CCNs in a research project to date, we believe that the four most successful sites serve as good examples of how a low-cost cellular network can provide lasting revenue to sustain operability and involve a wide swath of subscribers.

Furthermore, we highlight several demographic characteristics that deserve attention in the pursuit of universal access to mobile and internet services. We document critical barriers to the adoption of the cellular network. We also show that wealth and secondary education are positively correlated with cellular network use. Female headed households revealed a high demand for the cellular network, as did households involved in fishing and farming. We also find, though evidence is weaker, that the lack of phone ownership was associated with lower probability of transacting on the cellular network.

We believe that this analysis is an important step towards deciphering the promises and challenges of expanding cellular networks to the remaining 10% of the world's population that currently lacks phone service. Firstly, our analysis shows the promise of expanding cellular network coverage through low-cost community cellular networks. Secondly, we highlight key considerations and potential barriers that may limit the ability for households to partake in the mobile communications revolution.

Table 2.2: Baseline Summary Statistics, by Phone Ownership

	(1) All	(2) No Phone	(3) Owns Phone	(4)
Panel A: Household Summary Statistics	N=1131	N=364	N=767	p-value
Adults in household	2.70 (1.29)	2.26 (0.98)	2.91 (1.37)	<0.01
Children (0-14) in household	1.77 (1.49)	1.75 (1.61)	1.78 (1.43)	0.76
Female head of household	0.36 (0.48)	0.39 (0.49)	0.35 (0.48)	0.19
HOH - Secondary school	0.27 (0.44)	0.12 (0.33)	0.34 (0.47)	<0.01
Rooms in dwelling	1.79 (0.81)	1.60 (0.73)	1.88 (0.83)	<0.01
Household owns land	0.45 (0.50)	0.44 (0.50)	0.45 (0.50)	0.70
Family member in political/govt office	0.21 (0.41)	0.20 (0.40)	0.21 (0.41)	0.57
Household has bank account	0.16 (0.36)	0.07 (0.26)	0.19 (0.40)	<0.01
Income Source - Farming	0.34 (0.47)	0.37 (0.48)	0.33 (0.47)	0.14
Income Source - Fishing	0.24 (0.43)	0.28 (0.45)	0.22 (0.41)	0.03
Income Source - Wage Labor	0.20 (0.40)	0.14 (0.35)	0.22 (0.42)	<0.01
Poverty Score	42.17 (11.97)	37.25 (10.48)	44.51 (11.93)	<0.01
Wealth Index - Polychoric PCA	-0.11 (1.33)	-0.70 (1.12)	0.17 (1.33)	<0.01
Electricity in dwelling	0.63 (3.30)	0.42 (4.10)	0.74 (2.83)	0.18
Owns sala/sofa set	0.17 (0.38)	0.08 (0.28)	0.22 (0.41)	<0.01
Owns refrigerator	0.10 (0.30)	0.03 (0.17)	0.13 (0.34)	<0.01
Owns television	0.52 (0.50)	0.37 (0.48)	0.59 (0.49)	<0.01
Owns VHS/DVD player	0.30 (0.46)	0.21 (0.41)	0.35 (0.48)	<0.01
Owns radio	0.32 (0.47)	0.27 (0.45)	0.34 (0.47)	0.03
Owns satellite TV dish	0.31 (0.46)	0.18 (0.39)	0.37 (0.48)	<0.01
Owns motor vehicle or boat	0.32 (0.47)	0.22 (0.41)	0.37 (0.48)	<0.01
Owns gas stove	0.19 (0.39)	0.07 (0.26)	0.25 (0.43)	<0.01
Owns cellphone	0.68 (0.47)	0.00 (0.00)	1.00 (0.00)	-
Number of cellphones	1.20 (1.19)	0.00 (0.00)	1.77 (1.04)	<0.01
Owns SIM card	0.64 (0.48)	0.01 (0.07)	0.95 (0.23)	0.00
Number of SIM cards	1.33 (1.52)	0.01 (0.07)	1.95 (1.47)	<0.01
Owns Globe SIM card	0.25 (0.43)	0.00 (0.00)	0.36 (0.48)	<0.01
Owns SMART SIM card	0.55 (0.50)	0.01 (0.07)	0.81 (0.39)	<0.01
In-degree centrality	5.40 (29.58)	5.64 (34.69)	5.29 (26.83)	0.87
Eigenvector centrality	0.09 (0.15)	0.11 (0.18)	0.09 (0.13)	0.01
Panel B: Adult Survey Module	N=1617	N=516	N=1101	p-value
Do you see yourself as part of your community?	0.57 (0.49)	0.54 (0.50)	0.59 (0.49)	0.10
Do you feel isolated from the rest of your country?	0.30 (0.46)	0.31 (0.46)	0.30 (0.46)	0.88
Could you communicate with family in case of emergency?	0.46 (0.50)	0.38 (0.49)	0.50 (0.50)	<0.01
Traveled to Neighboring Bgy. (12 mo.)	0.44 (0.50)	0.38 (0.48)	0.48 (0.50)	<0.01
Traveled to Manila (3 yrs)	0.14 (0.35)	0.07 (0.25)	0.18 (0.38)	<0.01
Total contacts within barangay	6.10 (6.34)	5.91 (4.46)	6.19 (7.05)	0.35
Total contacts outside barangay	4.02 (7.54)	3.33 (4.60)	4.34 (8.56)	<0.01
Close friends/family within barangay	4.60 (1.69)	4.59 (1.58)	4.60 (1.75)	0.86
Close friends/family outside barangay	2.35 (1.67)	1.98 (1.49)	2.52 (1.72)	<0.01

Table 2.3: VBTS Konekt Tariff Schedule

Network Interaction Type	Tariff (PHP)
Call from a Konekt number to another Konekt number	1.00/minute
Call from a Konekt number to a long-distance on-network number	3.00/minute
Call from a Konekt number to an long-distance off-network number	5.50/minute
Text from Konekt number to Konekt number	0.25/message
Text from Konekt number to long-distance on-network number	0.50/message
Text from Konekt number to long-distance off-network number	1.00/message
All incoming calls	FREE
Incoming text messages (on-network local and long-distance)	FREE
Incoming text messages (off-network)	NOT ALLOWED

Table 2.4: Summary of Call Detail Record, By VBTS Site

VBTS Site	All	Outgoing Calls	Mean Outgoing Call Time	Incoming Calls	Mean Incoming Call Time	Outgoing SMS	Incoming SMS
Site 1	2,073	287	39.44	1,024	144.93	309	453
Site 2	680,556	76,133	29.00	238,265	84.06	191,584	174,574
Site 3	110,782	19,516	33.69	50,555	129.71	26,843	13,868
Site 4	110,284	22,768	31.26	59,253	118.58	15,980	12,283
Site 5	2,711	895	33.07	1,310	78.48	207	299
Site 6	4,645	2,060	22.77	1,671	89.98	542	372
Site 7	41,825	7,876	35.36	20,320	98.18	8,427	5,202

Table 2.5: Call Usage Summary Statistics, by Baseline Phone Ownership

	(1) All N=1131	(2) No Phone N=364	(3) Owned Phone N=767	(4) p-value
Any Transaction (prop.)	0.65 (0.48)	0.66 (0.47)	0.64 (0.48)	0.47
Count of Calls & SMS	773.75 (1462.56)	671.33 (1349.35)	822.36 (1511.74)	0.09
Any Outgoing Call (prop.)	0.56 (0.50)	0.60 (0.49)	0.55 (0.50)	0.14
Count of Outgoing Calls	100.78 (203.43)	99.71 (187.46)	101.29 (210.71)	0.90
Count of Long-distance Outgoing Calls	70.82 (140.75)	71.30 (136.18)	70.59 (142.95)	0.94
Any Outgoing SMS (prop.)	0.53 (0.50)	0.56 (0.50)	0.52 (0.50)	0.29
Count of Outgoing SMS	201.99 (481.14)	165.87 (418.10)	219.13 (507.69)	0.06
Count of Long-distance Outgoing SMS	120.08 (274.63)	102.24 (263.40)	128.55 (279.57)	0.12
Any Incoming Call (prop.)	0.61 (0.49)	0.63 (0.48)	0.60 (0.49)	0.32
Count of Incoming Calls	299.37 (576.33)	263.65 (504.48)	316.32 (607.04)	0.13
Count of Long-distance Incoming Calls	268.70 (542.41)	234.23 (465.22)	285.05 (574.99)	0.11
Any Incoming SMS (prop.)	0.60 (0.49)	0.61 (0.49)	0.60 (0.49)	0.81
Count of Incoming SMS	171.61 (413.95)	142.11 (379.36)	185.62 (428.92)	0.08
Count of Long-distance Incoming SMS	89.79 (217.88)	75.60 (214.38)	96.52 (219.34)	0.13

Standard deviations in parentheses. P-value for t-test of equality of means for No Phone and Owned Phone groups in column 4. Summary statistics by pre-existing phone ownership shown in Table A.2.2.

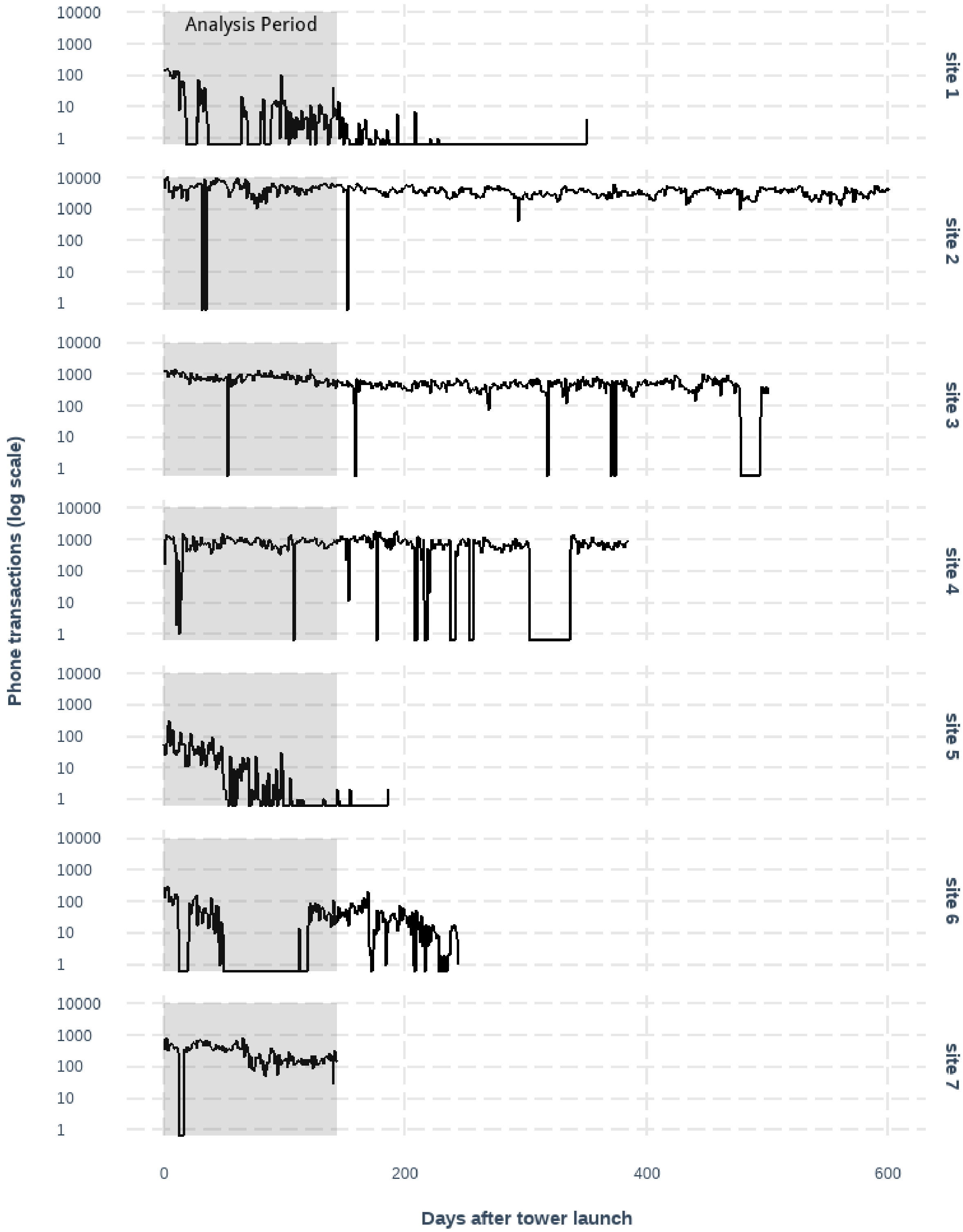

Figure 2.3: Site Activity

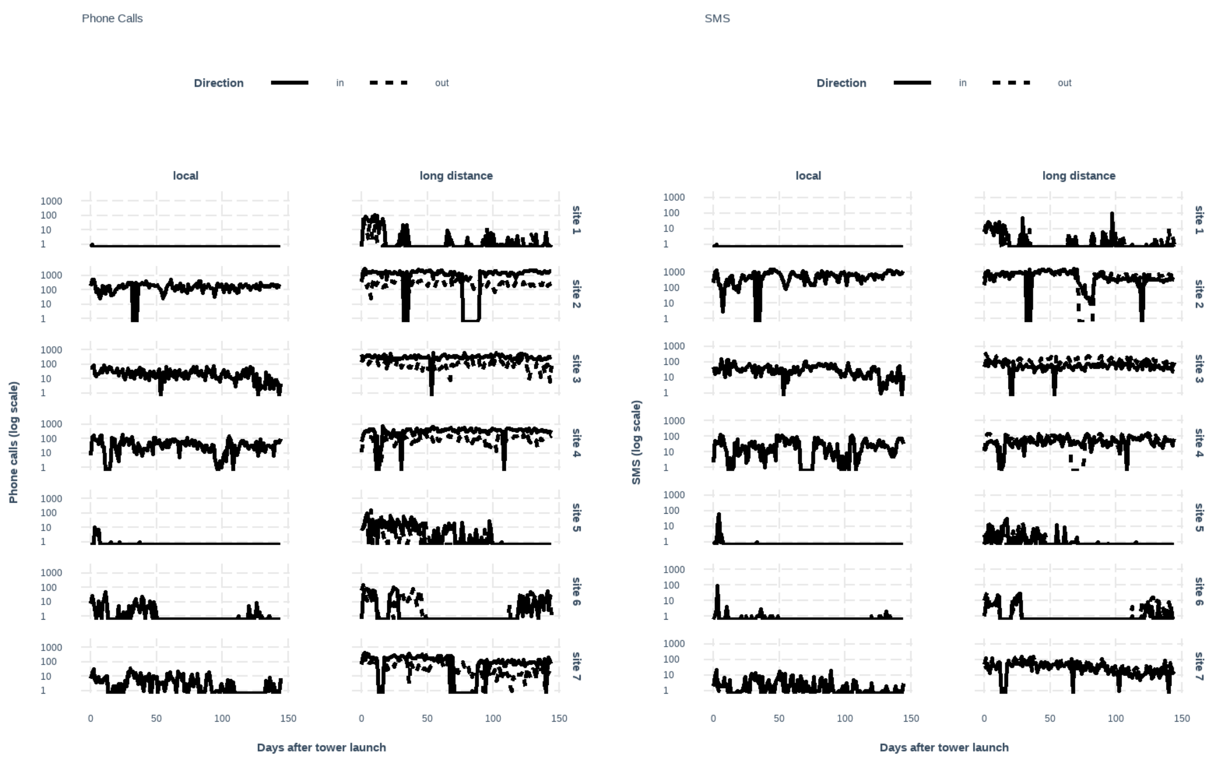

Figure 2.4: Network Usage by VBTS Site

Table 2.6: Determinants of Mobile Network Usage

	(1) Any Transaction	(2) Total Transactions	(3) Out Calls	(4) Out SMS	(5) In Calls	(6) In SMS
Owned Phone at Baseline	0.05	−19.70	−5.72	−3.32	−2.40	−8.26
	(0.03)	(40.05)	(5.81)	(13.32)	(15.51)	(11.45)
Wealth Index (SD)	0.03*	42.19**	5.96*	12.78*	11.77*	11.68**
	(0.01)	(14.30)	(2.74)	(5.06)	(5.71)	(4.23)
Contacts Outside Barangay	−0.00	−1.17	−0.18	−0.20	−0.59	−0.20
	(0.00)	(1.46)	(0.20)	(0.44)	(0.57)	(0.44)
In-degree Centrality (SD)	0.00	−4.96	−0.75	−2.47	0.58	−2.32
	(0.01)	(9.81)	(1.33)	(3.69)	(2.09)	(2.95)
Household Size	0.02**	2.07	−0.90	0.40	1.53	1.04
	(0.01)	(7.52)	(1.08)	(2.43)	(3.02)	(2.18)
HOH - Female	0.05*	−3.03	4.35	−1.52	−4.16	−1.70
	(0.02)	(30.28)	(5.13)	(10.47)	(12.08)	(8.88)
HOH - Secondary	−0.01	87.44*	7.25	28.18*	27.14	24.86*
	(0.03)	(39.15)	(5.53)	(14.09)	(15.01)	(11.73)
Income - Farming	0.11***	−10.40	−2.86	−3.12	1.33	−5.75
	(0.03)	(36.57)	(5.88)	(12.84)	(14.05)	(10.90)
Income - Fishing	0.10**	−67.69	−3.89	−20.01	−22.86	−20.94
	(0.03)	(39.53)	(6.59)	(13.78)	(15.71)	(11.10)
Work Outside Barangay	0.00	−30.15	−1.96	−8.99	−13.91	−5.30
	(0.02)	(29.94)	(4.79)	(10.33)	(11.55)	(8.72)
Cog. Ability Score (SD)	0.02	6.52	1.82	1.57	1.61	1.52
	(0.01)	(13.00)	(2.19)	(4.21)	(5.84)	(3.37)
R^2	0.39	0.27	0.16	0.21	0.22	0.23
Adj. R^2	0.38	0.26	0.15	0.20	0.21	0.22
Num. obs.	1131	1131	1131	1131	1131	1131
RMSE	0.38	456.93	75.10	155.02	183.44	131.84
Mean of Outcome	0.65	773.75	100.78	201.99	299.37	171.61
S.D. of Outcome	(0.48)	(1462.65)	(203.43)	(481.14)	(576.33)	(413.95)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Eicker-Huber-White robust standard errors in parentheses. Dependent variables with (SD) are in standardized units. Site fixed effects included.

Chapter Three

The Price is Right?

Statistical evaluation of a crowd-sourced market information system in Liberia

Abstract

Many critical policy decisions depend upon reliable and up-to-date information on market prices. Such data are used to construct consumer price indices, measure inflation, detect food insecurity, and influence macroeconomic policy. In developing countries, where many of these problems are most acute, reliable market price information can be hard to come by. Here, we evaluate data from Premise, a new technology for measuring price information using crowd-sourced data contributed by local citizens. Our evaluation focuses on Liberia, a country with a history of economic and political instability. Using data from Premise, which recently began data collection in Liberia, we analyze tens of thousands of individual price observations collected at hundreds of different locations in Monrovia. We illustrate how these data can be used to construct composite market price indices, and compare these constructed indices and prices for individual products to "ground truth" data from the Central Bank of Liberia and the United Nations World Food Programme. Our results indicate that the crowd-sourced price data correlates well with traditional price indices. However, we find statistically and economically significant deviations from traditional measures that require more in-depth investigation. We conclude by discussing how indices based on Premise data can be further improved with simple supervised learning methods that use traditional low-frequency data to calibrate and cross-validate the high-frequency Premise-based indices.[†]

[†]The material in this chapter is based on material originally published with Joshua Blumenstock in 2015, see Blumenstock and Keleher (2015).

3.1 Introduction and Motivation

The Consumer Price Index (CPI) is one of the most important economic statistics used by policy-makers to determine macroeconomic policy and to evaluate the health of an economy (Mankiw 2014). It is the primary barometer for measuring inflation, which in turn impacts both fiscal and monetary policy; it is used to calculate purchasing power and determine exchange rates; it also directly impacts wages and welfare payments in both the private and public sector.

The CPI is intended to capture the overall cost of goods and services paid for by a typical individual at local markets. It requires two primary inputs: the "basket of goods" that is intended to be representative of a typical consumer; and the price for each of the goods in the basket. Prices of goods in the common basket are the focus of this paper.

The lack of reliable and up-to-date price information is particularly problematic in fragile economies, where dependence on subsistence agriculture and the lack of insurance and other social safety nets can exacerbate the welfare impacts of weather shocks, political instability, and food insecurity as mediated through price changes. In Liberia and neighboring countries, for instance, food prices have been one of the primary instruments used in assessing the economic impacts of the recent Ebola outbreak (Glennerster and Suri 2015; Himelein and Kastelic 2015; World Food Programme 2015b).

Here, we investigate a new technology for collecting price information in developing countries, which relies on "crowd-sourced" observations collected by local citizens with mobile phones. We focus on Premise, a technology platform that allows for mobile-equipped citizens to capture and upload price information to a central service. This technology, described in greater detail in Section 3.3, is similar to a small number of related platforms that enable crowd-sourced data collection in developing countries. The mClerk (Gupta et al. 2012) and txtEagle (Eagle 2009) platforms are the most directly comparable systems of which we are aware; both use mobile-based platforms to gather data from low-end mobile phones. More broadly, several examples of researchers sourcing data from the crowd exist. For instance, Starbird and Palen (2011) study microblogging in response to a large earthquake in Haiti in 2010, and Ashley et al. (2009) describe several other emerging technological systems that facilitate participatory contribution in development areas including agriculture, rural development and natural resource management. In the closest study to our own, Hamadeh, Rissanen, and Yamanaka (2013) conduct a feasibility study to explore the use of the Jana platform for collecting price data. While Hamadeh, Rissanen, and Yamanaka (2013) demonstrate the potential of the platform, they focus on a description of the technological platform, rather than on evaluating the accuracy of the collected data in comparison to external sources of validation data.[1]

Different from prior work, our focus is not on the technological artifact or interface used to collect data; rather, we study the data generated by this platform, and statistically evaluate its potential for use as an index of inflation and related economic activity. The empirical analysis

[1] A much more extensive literature, which we do not review here, explores the potential for data collection using mobile phones in developed economies (Lane et al. 2010). A separate literature discusses how mobile data can be used for population-based inferences of social and economic indicators (Blumenstock 2014; Eagle, Macy, and Claxton 2010; Frias-Martinez et al. 2013).

relies on data generated from the Premise network of contributors in Liberia, which has been collecting price information on 38 market goods in Monrovia since late 2014. We present the data set and document several prominent features related to the stability and noise present in the raw data (Section 3.3), discuss corrections that improve the reliability of derivative metrics (Section 3.3), and describe a basic method for computing a consumer price index from the crowd-sourced data (Section 3.3). Section 3.4 compares these indices to alternative measures of inflation in Liberia.

This paper thus makes three primary contributions. First, we illustrate how crowd-sourced market data can be used to construct price indices, and characterize the statistical properties of these indices. Second, we carefully evaluate these indices by comparing them to traditional methods for measuring food insecurity. Finally, we discuss a simple supervised learning framework that can be used to improve the accuracy of high-resolution estimates through calibration and cross-validation with low-resolution sources of data.

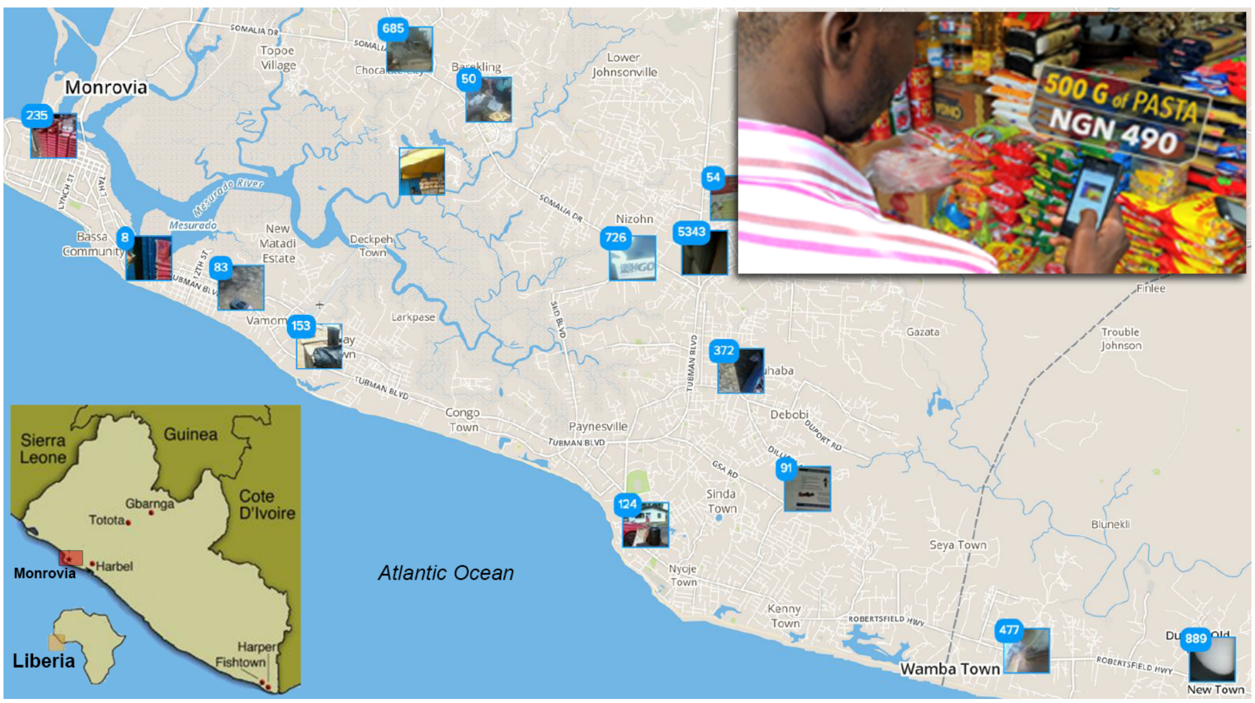

Figure 3.1: **Premise Data collection methodology.** Premise indexes and analyzes data captured by a global network of contributors. Bottom-left: Location of Liberia and its capital Monrovia. Main figure: Locations from which contributors have captured data. Numbers indicate the number of data points collected from each location over the past three months; square icons are actual images uploaded by contributors. Top-right: Schematic of data capture process in which a contributor uses a cameraphone to photograph the prices of pasta at a local market. These photos are sent to Premise and form the basis for the data we analyze.

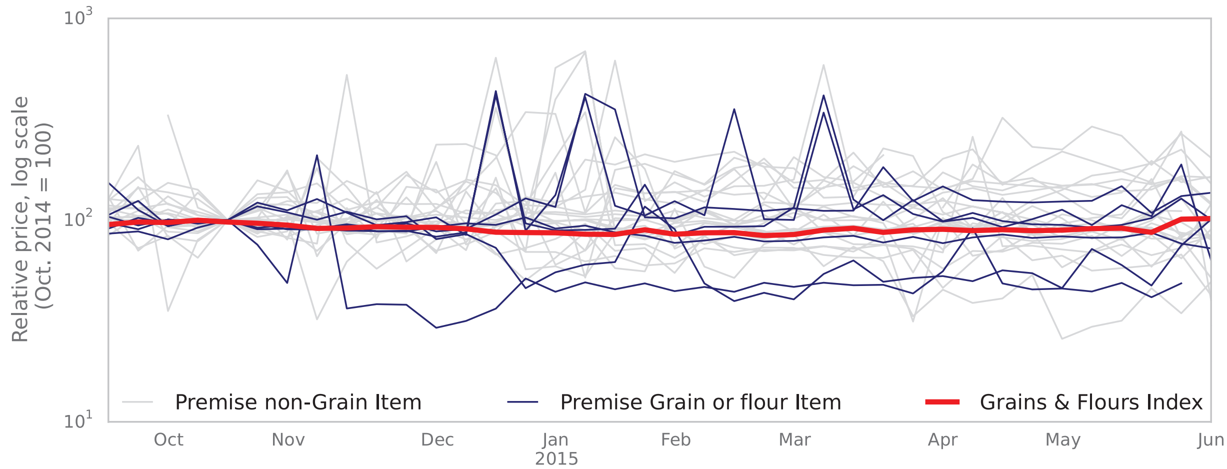

Figure 3.2: **Raw data captured by Premise contributors.** Data for 38 different products is captured by Premise contributors (light grey lines). Of these goods, 6 are in the "Grains and Flours" product group (dark blue lines), corresponding to bread, bulgur wheat, butter rice, cassava flour, fan-fan rice, and USA parboiled rice. Using the methods described in Section 3.3, these six product price-series are aggregated into a single sub-index for the product group (red line). Each series is normalized to a value of 100 on October 15, 2015.

3.2 Consumer Price Indices in Liberia

Following Liberia's long-entrenched civil war and post-conflict recovery, which ended in 2003, the country took several years to achieve political and economic stability. Between 2011 and 2014, consumer prices in Liberia followed a roughly linear trend with mean year-to-year inflation of approximately 9%. In the period following the Ebola Virus Disease epidemic, the country saw an uptick in the price index, with year-to-year inflation rising to a peak of 16.3% in September 2014. With the rise in consumer prices, concern over access to food was heightened. As of writing, the government and international agencies remained vigilant to the threat to food security that increased prices posed for Liberian households (Himelein and Kastelic 2015).

Price data is historically accessible to a limited set of actors and on an infrequent basis. Official price data is collected by the Liberian government and analyzed by the Central Bank of Liberia in order to produce aggregate price indices on a monthly basis. Data collected by the United Nations World Food Programme (WFP) is typically collected for the purpose of monitoring food insecurity. However, data collected for food security monitoring covers a more limited set of consumption items and locations relative to the government statistics.

Price data plays an integral roll in the WFP's forecasts of food shortages, as food prices influence household spending decisions. As such, the WFP uses food prices as an indicator of the impact of economic shocks on households. Access to timely price data has been especially important during the 2014–15 Ebola epidemic in West Africa, where the combination of restricted economic activity, loss of life, and loss of sources of income have culminated in a threat to food security (World Food Programme 2015a).

3.3 Premise Data

Technology platform

Premise (`www.premise.com`) is a technology company based in San Francisco that offers a platform for capturing data from a distributed network of individual contributors.[2] The intent of the platform is to enable rapid and adaptive measurement of local economic and social infrastructure, using data collected by local citizens. Premise recruits individuals in urban and rural regions of developing countries to perform simple, structured tasks that capture information about their local community (Figure 3.1). Premise currently operates in 32 countries worldwide. A major focus of Premise's efforts to date has been on collecting price data from developing countries.

Premise contributors use photo-enabled phones to capture prices in local markets. Contributors are compensated in the local currency, and are trained in person prior to submitting data. Each day, contributors receive a list of *tasks* which detail the items for which price observations are needed. Contributors are typically paid a piece rate for each successfully completed task, for an amount "on the order of the price of an egg" for each data point captured.[3] Photos and price information submitted by the contributor go through a quality control screening process, described in more detail below.

Beginning in September 2014, Premise initiated data collection in Liberia in an effort to produce a Food Staples Price Index (FSPI) as well as other measures of economic and business activity. Premise data collection in Liberia launched in response to the 2014 Ebola Virus Disease epidemic in West Africa. The initial goal was to provide timely information about prices of key consumer goods. The initial focus has been in Monrovia, where Premise collects daily price observations for staple foods and non-food consumer items in all of Monrovia's major market areas (Figure 3.1). In the summer of 2015, Premise expanded data collection in rural markets of Liberia, beginning with the towns of Voinjama and Fish Town.

Premise currently receives approximately 600 price observations per week for a basket of 38 unique products, from a small network of independent contributors in Liberia.[4] As can be observed in Appendix Table A1, which presents summary statistics for the raw contributor data, there is a great deal of variation in the frequency at which each product is observed, the number of unique locations at which a product is captured, and the price level and variance for each product over time. As we describe in the following section, the number of errant observations also varies considerably by product.

[2]Contributors are analogous to enumerators or surveyors in traditional terminology for primary data collection.

[3]Based on private correspondence with Premise staff, July 2015.

[4]Data volumes in Liberia are low compared to other countries. For example, Premise contributors submit upwards of 15,000 observations per week for 150 items in Nigeria.

Detecting outliers in crowd-sourced data

The raw data captured by Premise contributors in Liberia are illustrated in Figure 3.2. Here, we plot a separate time series for each of the 38 products as a semi-transparent grey line. Each of these lines represents the daily average price for that product, averaged across all observations taken on that day by all contributors in all locations. Highlighted in blue are the six time series corresponding to products in the "Grains and flours" category. In red is the composite sub-index, calculated using a procedure we will shortly describe.

As is evident in the product-level time series in Figure 3.2, the raw data collected by Premise contributors is subject to several sources of error. On some occasions there appear to be idiosyncratic spikes where a single product's price will change by as much as 200%; at others, these spikes appear to be correlated across products. Of primary concern is disentangling from actual changes in prices from measurement error. As has been documented in related work, there are many possible sources of such error, both accidental and deliberate (Birnbaum, Borriello, et al. 2013; Birnbaum, DeRenzi, et al. 2012; Hamadeh, Rissanen, and Yamanaka 2013). These include input errors (for instance, a misplaced decimal point or a photograph of the floor), as well as outright fraud where a contributor intentionally falsifies data. Premise implements several measures to detect and prevent such deliberate fraud (Premise 2014), but many of these are not publicly disclosed, and in practice affect a relatively small number of total captured data.

In follow-up work, we are developing more refined techniques for identifying and removing erroneous data, which may constitute as much as 20% of the total data captured on the Premise platform. Here, we describe a simple procedure that, based on manual verification, appears to catch a large share of these errors. Formally, we denote by P_{itlk} an observation recorded by individual i at time t in location l for item k. We define price outliers as those observations that deviate significantly from historical prices for a given product, i.e.,

$$|\log(P_{itlk}) - \log(\mu_{tlk})| > \lambda_k \sigma_{tlk} \tag{3.1}$$

where $\mu_{tlk} = \frac{1}{Nn} \sum_i \sum_{s<t} P_{islk}$ is the average historical value for k at location l (assuming N individuals and n observations where $s < t$), and σ_{tlk} is the corresponding standard deviation. When insufficient observations exist from which to derive reliable estimates of μ_{tlk} and σ_{tlk}, a "bootstrap" process is used to manually curate and reject anomalous observations. In this framework, λ_k is the key parameter which determines the stringency with which outliers will be identified.

Currently, Premise employs a common threshold across all products of $\lambda_k = \lambda \approx 3$. In ongoing work, we are exploring a supervised learning approach to determining a product-specific λ_k, which will allow for some products with greater expected variation over time to exhibit more intertemporal variability. We are also testing density-based techniques (Ester et al. 1996) and other alternative methods for outlier detection.

Computing CPI from crowd-sourced data

A primary objective of the Premise application is to convert the disaggregated price data collected by the network of contributors into more meaningful price indices, similar to the CPI, which can then be used to measure inflation and inform food policy decision-making. Here, we describe the process used to construct the Food Staples Price Index (FSPI), the Premise equivalent of the CPI, from the disaggregated data. Formally, our goal is to compute an aggregate CPI_{rm} for a region r in month m, from a large number of disaggregated price observations P_{itlk}.

Given the set of P_{itlk} with outliers removed, the FSPI CPI_{rm} is constructed using a process based loosely on the methods employed by the US Bureau of Labor Statistics BLS (2008). Initially, the average daily price P_{Tlk} for item k at location l is constructed by taking the average of all $|T|$ observations collected on day T,

$$P_{Tlk} = \frac{1}{N|T|} \sum_{i \in l} \sum_{t \in T} P_{itlk} \tag{3.2}$$

This value is still quite specific, as l can be as precise as a single storefront location, and k can be unique to an item SKU, such as the price of one bottle of Club Beer (a local beer), or the cost of a bucket of low-grade gari (a local flour). These item-day averages are next aggregated into product-day averages P_{TlK} by standardizing units of measurement (e.g., pounds to grams) and taking the geometric mean of related products (e.g., beef briskets), i.e.,

$$P_{TlK} = \left(\prod_{k=1}^{K} P_{Tlk} \right)^{1/k} \tag{3.3}$$

Product-day averages are similarly aggregated across all locations in a region and across all days in a time window to produce monthly estimates of the regional cost of a specific product. These product averages are further aggregated into product sub-groups (e.g., standard cut beef), product groups (e.g., beef), and sub-indices (e.g., meat). This aggregation uses weights that are determined based on the estimated expenditures of consumers on the various items. In Liberia, these weights are determined by a recent consumer expenditure survey (Group 2013).

In Liberia there are 12 such *sub-indices*, which indicate the prices of the most common items in the country. The final step in constructing the FSPI is to combine the sub-indices into a single consumer price index that reflects the price level of a typical market basket of food staples. The weights w_k for each of the sub-indices used in constructing the Liberian FSPI are given in Table 3.1. Thus, the composite FSPI can be expressed as a weighted aggregate of the original contributor observations:

$$CPI_{rm} = \frac{1}{|r||m|} \sum_{l \in r} \sum_{T \in m} w_k P_{TlK} \tag{3.4}$$

The FSPI for Liberia, computing using the above methodology, as well as the 12 sub-indices, is shown in Figure 3.3. To emphasize the fluctuations within a product category over time, each index is normalized so that the value on October 15, 2014 is set to 100.

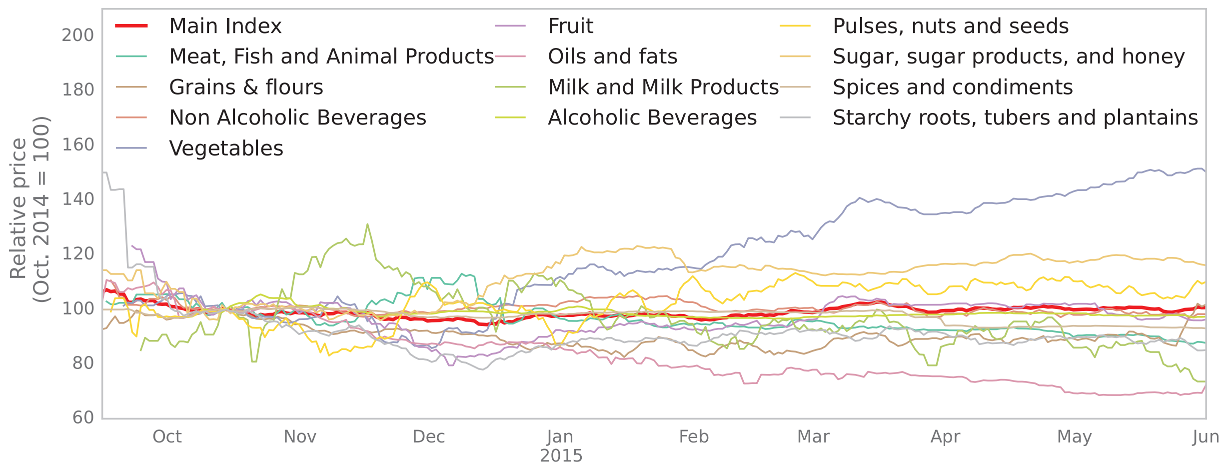

Figure 3.3: **Premise FSPI and sub-indices, as calculated from contributor data.** Following the procedure described in Section 3.3, sub-indices are computed for each of the 12 product groups listed in Table 3.1. Using the weights listed in the same table, the composite Food Staples Price Index, the Premise equivalent of a CPI, is calculated and shown as a thick red line.

3.4 Evaluation Results

The methods described above make it possible to construct a CPI-like metric, the FSPI, as well as several sub-indices of product-group prices, from the data captured by Premise contributors (Figure 3.3). In order to validate the relevance of these constructs, we take two approaches — one at the aggregate index level and another at the item level. We draw from two sources of data for the validation. The purpose of this comparison is to judge the consistency and reliability of Premise data vis-a-vis an economic indicator for the Liberian economy as well as a best-available option for policymakers interested in food security as proxied by the price of individual goods. The data for the former comparison comes from the Central Bank of Liberia, while the item-level comparison is conducted with a primary source of data collected for the United Nations World Food Programme with the purpose of tracking threats to food security in Liberia.

Comparison data

Central Bank of Liberia

The Central Bank of Liberia releases headline consumer price index data from the 15th day of each month. National price level are based on monthly price surveys conducted by the Liberia Institute of Statistics and Geo-information Services (LISGIS). Price indices are provided for the overall price level, food and non-alcoholic beverages (split by domestic and imported food), transportation, and imported fuel. The inset of Figure 3.4 displays the time series of the Food and Non-Alcoholic

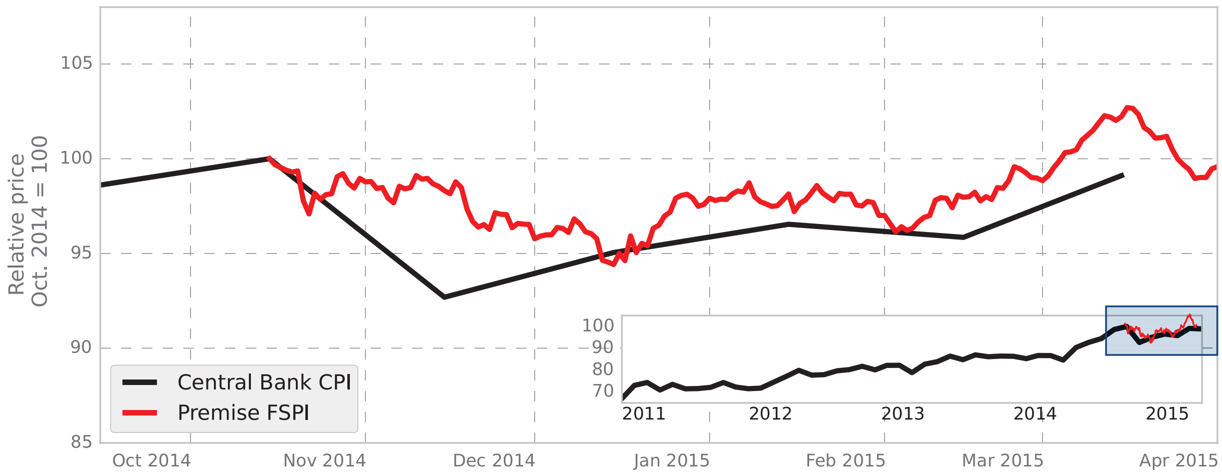

Figure 3.4: **Comparison of FSPI to Central Bank CPI.** Main figure shows the CPI calculated by the Central Bank of Liberia and the FSPI based on Premise data. Inset figure displays the Food and Non-Alcoholic Beverages relative price from May 2011 to April 2015.

Beverages Index for Liberia between May 2011 and April 2015. To emphasize relative changes in prices, the index is normalized so that the value of the index in October 2014 is equal 100. The impact of the Ebola epidemic, which caused year-to-year price inflation to peak at 16% in September 2014, is evident in the figure. Note that the Central Bank index is intended to be nationally representative, but Premise data for this period is restricted to the capital city, Monrovia.[5]

World Food Programme

We additionally compare the product-level data captured by Premise contributors to data acquired from United Nations World Food Programme (WFP), which conducts regular data collection for key food prices in order to assess food security and identify price shocks that disproportionately affect poor households. We utilize data collected for the WFP on a key set of consumption goods in Liberia, which were initiated in part to monitor the impact of Ebola on price inflation. For the purpose of comparability with the Premise data, we restrict the WFP data to a subset of four goods: imported rice, cassava, palm oil and charcoal.[6] Figure 3.6 shows the time series of mean weekly price of both rice and cassava, as independently observed in Premise and WFP data.

[5]In July 2015, Premise initiated data collection in two rural markets, but these data were not available at the time this research conducted.

[6]The WFP data captured six products in total, including brown cowpeas and gari flour; however, the Premise data collection does not include these items.

Index component	Weight (%)
Alcoholic Beverages	1.2%
Fruit	9.2%
Grains and Flours	13.7%
Meat, Fish and Animal Products	16.8%
Milk and Milk Products	6.1%
Non Alcoholic Beverages	12.0%
Oils and Fats	3.0%
Pulses, Nuts and Seeds	2.9%
Spices and Condiments	8.0%
Starchy Roots and Tubers	11.4%
Sugar and Sugar Products	2.2%
Vegetables	13.4%

Table 3.1: **Weights used in constructing the FSPI.** The composite Food Staple Price Index is constructed as the weighted sum of 12 primary sub-indices, where the above weights are determined based on a recent household expenditure survey.

Statistical comparison

Comparison to baselines

We compare the Premise data to both the Central Bank food and non-alcoholic beverages index and the WFP individual item prices for the period from September 2014 to April 2015. Effectively, we seek to quantify the differences in Figures 3.4 and 3.6. These results are presented in Table 3.2, which indicates the average per-month error (RMSE) as well as the correlation between datasets over time. Since the CPI data is collected at the monthly level while the Premise data exists in daily averages, we compare the two series by making monthly comparison of the CPI to the average Premise data over the preceding 30-day period (row 1), and also by linearly interpolating the CPI data between monthly observations and comparing at the daily level (row 2).

As is evident in Table 3.2, we find suggestive evidence of a correlation between the aggregate indices. Importantly, these correlations do not account for potential delays and offsets in the different series. For instance, if the Premise FSPI is a leading indicator of the CPI, or vice versa, such patterns would not be reflected in the results in Table 3.2. However, correlation of individual food item prices is weaker. Our analysis thus suggests two areas for further investigation. First, through a longer time series of indices, we will be able to test for a more robust lagged structure to the relationship between the Central Bank indicators and Premise indicators. Second, the variation in food item prices within and across sources suggest that methods for screening outliers and validating price data with multiple source may improve stability of official price data for individual products.

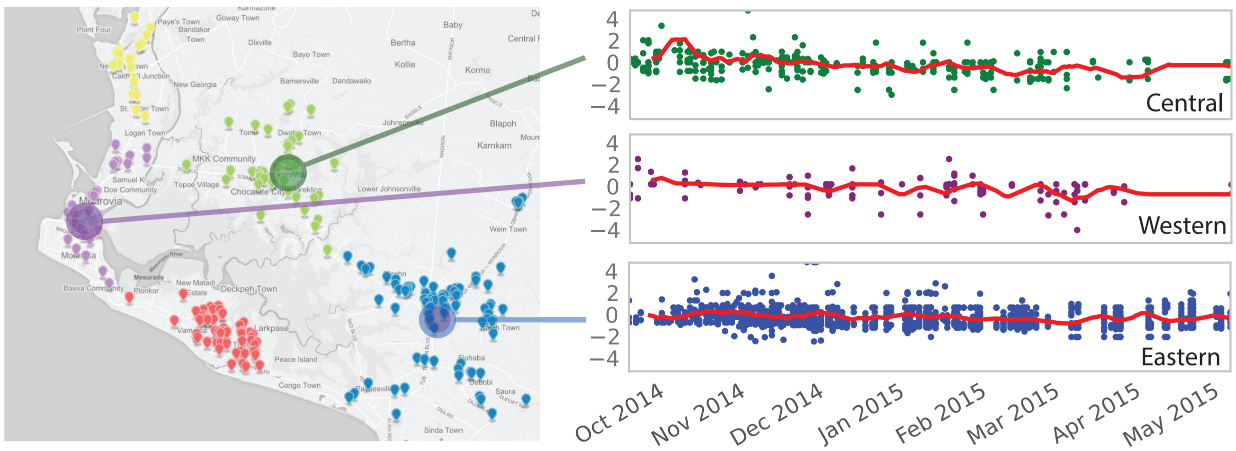

Figure 3.5: **Geospatially disaggregated time series.** Premise data include individual price data for 38 consumer goods from 310 unique locations in Monrovia, Liberia. This figure shows normalized prices for grain and flour products in the three geographic locations, with a red line to indicate the 10-day moving average across all observations from each sub-region. Sub-regions are identified using k-means clustering with five clusters specified using latitude and longitude of the observations submitted by the Premise contributor.

	Corr	RMSE	RMSE (% of mean)
A. Central Bank Food Price Index			
FSPI (Monthly Average)	0.64	2.97	3.58%
FSPI (Daily vs. Central Bank linear interpolation)	0.73	2.39	2.88%
B. Food Items (WFP, relative prices)			
Rice	0.54	8.94	9.77%
Cassava	0.03	16.27	18.71%
Charcoal	0.35	6.29	6.64%

Table 3.2: **Model performance.** Measures of model accuracy and error, comparing Premise data to data collected by the Liberian Central Bank and the World Food Programme.

Modeling improvements

Our analysis thus far indicates a suggestive correlation between Premise data and the indices collected by the Liberian Central Bank. Moving forward, we believe a promising area for research lies in using different sources of "ground truth" data to better calibrate the model used to construct the FSPI. To take a simple example, one might imagine using cross-validation to select the optimal parameters for outlier detection (in our case, the λ_k described in Section 3.3). This is a complicated task, however, as there are known issues with existing sources of price as CPI data, so it is important that any supervised learning approach not overfit the ground truth data. However, by calibrating with multiple sources of external data, collected through different processes and at different spatial and temporal resolution, we believe it should be possible to build more robust and accurate price indices.

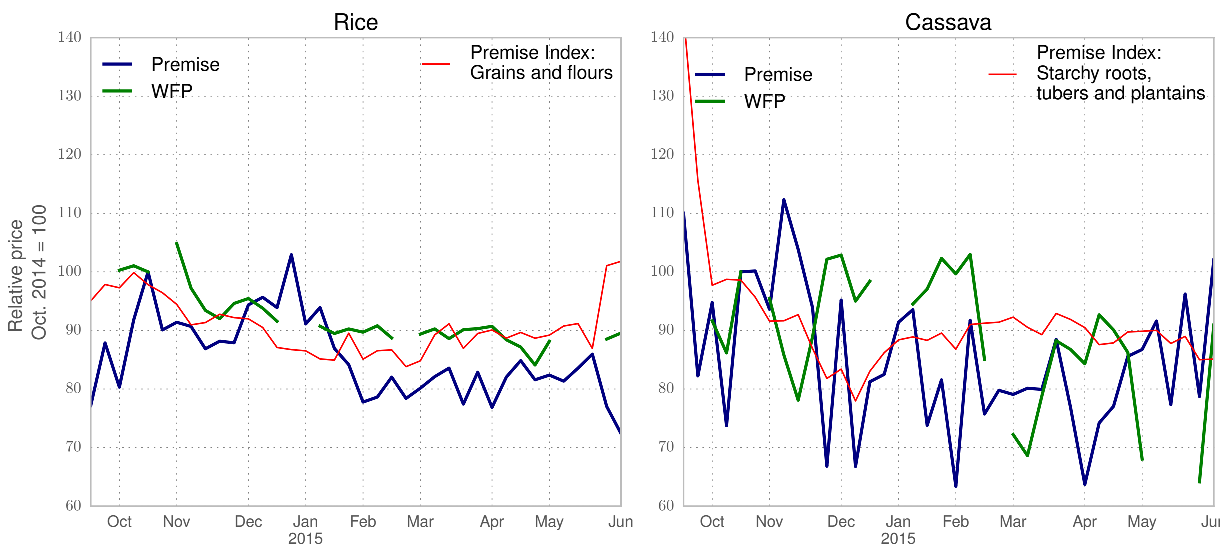

Figure 3.6: **Comparison of Premise and WFP Item Prices.** Average price data for two food items in Liberia. The blue lines represent the average weekly price data collected for rice (left) and cassava (right). The green lines indicate the corresponding price data for the same item, collected by the World Food Program's Building Markets initiative. The red line indicates the composite sub-index from Premise data, constructed from all goods in the grains (left) and starchy roots (right) product categories. Each series is normalized to a value of 100 on the week of October 15, 2015.

3.5 Discussion

While Premise data is correlated with the two traditional price measures we were able to obtain, there are statistical discrepancies that are not easily resolved.[7] While we largely treat the Central Bank of Liberia's CPI data as "ground truth" and assume that deviations observed in the Premise FSPI are errors, it is also conceivable that the Premise data might at times be accurate where Central Bank data is not. Indeed, the two sources of data, while comparable, have distinct advantages and disadvantages that we discuss briefly before concluding. Similarly, the WFP data

[7]One obvious source of these difference may be the differences in sampling frames used in data collection, for instance the fact that the Central Bank surveys the entire country's prices while Premise is thus far focused in Monrovia; as Premise expands to additional markets it will be possible to test this hypothesis.

relies upon the accuracy of reports from on-the-ground surveyors that are no less prone to errors than Premise contributors. In fact, real-time data quality controls can be put in place through Premise's technology that are often lacking or slow to implement through traditional price data collection methods.

Advantages of traditional price data

A primary advantage of traditional sources of price information is that they are familiar to most consumers of price data. Governments and international organizations have well-established mechanisms for collecting and processing price data, and a standard set of best practices exist for determining sample frames, deciding price frequencies, and integrating the resulting measures into macroeconomic decision-making. Centralized administration of these efforts further ensures that governments and international organizations play an integral role in the collection of critical economic data. Private sector efforts may be more subject to a different optimization problem, i.e. profit maximization.

There are also economies of scale in centralized data collection. Governments and international organizations conduct data collection, such as censuses, expenditure surveys and firm surveys, that are complimentary to price data. For example, expenditure surveys are integral to updating consumption basket estimates that feed into consumer price indices.

Advantages of Premise CPI

Relative to traditional models of price data collection, the Premise platform offers several distinct advantages. In particular, Premise data is highly granular, and can be sourced continuously in time and space, increasing the ability to track prices in real-time, in sub-regions of the country. In Figure 3.5, for instance, we show how the original data, collected from over 300 unique locations, can be aggregated to form sub-regional price indicators. For the figure, we cluster all of these locations geographically (using k-means on the market geo-coordinates), then aggregate all data by cluster. The three time-series graphs on the right show the original price data in the grains and flour category, from each of these market clusters (the points on the graphs). The red line indicates the 10-day moving average, similar to the sub-index described earlier. While generally correlated, there is within-market volatility that is not clearly reflected in the Monrovia aggregate. Such patterns are not visible in most traditional sources of price data.

While our focus has been on price data collection, the crowd-sourcing framework can also be used to capture a much wider array of data types. Near-term possibilities include street mapping, collecting information on the availability of public utilities, and mapping financial inclusion. The Premise platform for recruiting and compensating contributors makes it possible to quickly scale data collection efforts, and to target specific regions with less reliable data.

3.6 Conclusions

We describe and evaluate data from Premise, a platform for collecting crowd-sourced price information from networks of local contributors in developing countries. Our focus on the statistical properties of the Premise data, and on comparing Premise-based indices to more traditional measure of price inflation, reveals several promising areas for future work.

First, further quantitative work would benefit greatly from a longer panel of price data. While the Premise data contains tens of thousands of observations and is collected at extremely high frequency, the authoritative central bank data is collected only monthly. With only 9 months of overlapping data, it is difficult to make robust statistical comparisons.

Second, there is considerable scope for improvement in the techniques used to identify erroneous data, caused both by innocent error and intentional fraud. The outlier removal system we describe and implement appears to be reasonably effective, but is rather coarse and relies heavily on (possibly unjustified) intuition. Given a labeled training set, where the source of erroneous data points is known, it would be possible to develop more sophisticated methods tailored to specific products, locations, and contributors.

Finally, and perhaps most promising, we believe significant progress can still be made in developing methods for supervised learning that use authoritative data to improve the accuracy of estimates based on high-frequency data from Premise and related sources. Here, we made the simple point that inflation forecasts appear to improve when historical CPI data is supplemented with data from Premise. Analogous approaches could be used to increase the granularity of official CPI estimates (beyond the country-month), or to construct monthly Premise estimates that correspond more closely to official benchmarks. As before, the absence of a long panel of training data makes these exercises difficult in the immediate term, but as data from Premise's global network continues to stream in, it will open many opportunities for research that can impact how prices and inflation are measured in developing economies.

Bibliography

Chapter 2 References

Adamic, Lada A. and Eytan Adar (2005). "How to Search a Social Network". *Social Networks*, 27(3): 187–203.

Alatas, Vivi, Abhijit Banerjee, Arun G. Chandrasekhar, Rema Hanna, and Benjamin A. Olken (2016). "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia". *American Economic Review*, 106(7): 1663–1704.

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): 1206–1240.

Aral, Sinan and Dylan Walker (2012). "Identifying Influential and Susceptible Members of Social Networks". *Science*, 337(6092): 337–341.

Athey, Susan and Guido W. Imbens (2017). "The Econometrics of Randomized Experiments". In: *Handbook of Economic Field Experiments*. Ed. by Abhijit Vinayak Banerjee and Esther Duflo. Vol. 1. Handbook of Field Experiments. North-Holland, 73–140.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou (2006). "Who's Who in Networks. Wanted: The Key Player". *Econometrica*, 74(5): 1403–1417.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson (2014). *Gossip: Identifying Central Individuals in a Social Network*. Working Paper 20422. National Bureau of Economic Research.

Basurto, Pia M., Pascaline Dupas, and Jonathan Robinson (2017). *Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi*. Working Paper 23383. National Bureau of Economic Research.

Baumann, Leonie (2016). "Identifying the Best Agent in a Network". *SSRN Electronic Journal*.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak (2018). *Can Network Theory-Based Targeting Increase Technology Adoption?* Working Paper 24912. National Bureau of Economic Research.

Bloch, Francis and Matthew Olckers (2018). "Friend-Based Ranking". arXiv: `1807.05093 [physics]`.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet (2015). "Social Networks and the Decision to Insure". *American Economic Journal: Applied Economics*, 7(2): 81–108.

Cantril, Hadley (1966). *The Pattern of Human Concerns*. OCLC: 2082808. New Brunswick, N.J.: Rutgers University Press.

Conley, Timothy G. and Christopher R. Udry (2010). "Learning about a New Technology: Pineapple in Ghana". *The American Economic Review*, 100(1): 35–69. JSTOR: 27804921.

Conley, Timothy and Christopher Udry (2001). "Social Learning Through Networks: The Adoption of New Agricultural Technologies in Ghana". *American Journal of Agricultural Economics*, 83(3): 668–673.

Deaton, Angus (2008). "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll". *Journal of Economic Perspectives*, 22(2): 53–72.

Fafchamps, Marcel and Julien Labonne (2019). "Family Networks and Distributive Politics". *Journal of the European Economic Association*.

Fernald, Lia CH, Paul J Gertler, and Lynnette M Neufeld (2008). "Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's Oportunidades". *The Lancet*, 371(9615): 828–837.

Filmer, Deon and Lant Pritchett (1999). "The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries". *Population and Development Review*, 25(1): 85–120.

Fiszbein, Ariel and Norbert R. Schady (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty.* World Bank Publications. 384 pp.

Galeotti, Andrea, Benjamin Golub, and Sanjeev Goyal (2017). "Targeting Interventions in Networks". arXiv: 1710.06026 [cs].

Gerber, Alan S. and Donald P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation.* 1st edition. New York: W. W. Norton & Company. 480 pp.

Golub, Benjamin and Matthew O. Jackson (2012). "How Homophily Affects the Speed of Learning and Best-Response Dynamics". *The Quarterly Journal of Economics*, 127(3): 1287–1338.

Hussam, Reshmaan, Natalia Rigol, and Benjamin Roth (2018). *Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design In The Field.* Working Paper.

Imbens, Guido W. and Jeffrey M. Wooldridge (2009). "Recent Developments in the Econometrics of Program Evaluation". *Journal of Economic Literature*, 47(1): 5–86.

Jackson, Matthew O. (2019). "The Friendship Paradox and Systematic Biases in Perceptions and Social Norms". *Journal of Political Economy*, 127(2): 777–818.

Kim, David A, Alison R Hwong, Derek Stafford, D Alex Hughes, A James O'Malley, James H Fowler, and Nicholas A Christakis (2015). "Social Network Targeting to Maximise Population Behaviour Change: A Cluster Randomised Controlled Trial". *The Lancet*, 386(9989): 145–153.

Kitsak, Maksim, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse (2010). "Identification of Influential Spreaders in Complex Networks". *Nature Physics*, 6(11): 888–893.

Kleinberg, Jon (2000). "The Small-World Phenomenon: An Algorithmic Perspective". *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing* (Portland, Oregon, USA). STOC '00. New York, NY, USA: ACM, 163–170.

Kolenikov, Stanislav and Gustavo Angeles (2009). "Socioeconomic Status Measurement with Discrete Proxy Variables: Is Principal Component Analysis a Reliable Answer?" *Review of Income and Wealth*, 55(1): 128–165.

Kshirsagar, Varun, Jerzy Wieczorek, Sharada Ramanathan, and Rachel Wells (2017). "Household Poverty Classification in Data-Scarce Environments: A Machine Learning Approach". arXiv: `1711.06813 [stat]`.

McBride, Linden and Austin Nichols (2016). "Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning". *The World Bank Economic Review*.

Milgram, Stanley (1967). "The Small World Problem". *Psychology today*, 2(1): 60–67.

Niehaus, Paul, Antonia Atanassova, Marianne Bertrand, and Sendhil Mullainathan (2013). "Targeting with Agents". *American Economic Journal: Economic Policy*, 5(1): 206–238.

Premand, Patrick and Pascale Schnitzer (2018). *Efficiency, Legitimacy and Impacts of Targeting Methods: Evidence from an Experiment in Niger.* SSRN Scholarly Paper ID 3165227. Rochester, NY: Social Science Research Network.

Rai, Ashok S. (2002). "Targeting the Poor Using Community Information". *Journal of Development Economics*, 69(1): 71–83.

Schüring, Esther (2014). "Preferences for Community-Based Targeting - Field Experimental Evidence from Zambia". *World Development*, 54: 360–373.

Shakya, Holly B., Derek Stafford, D. Alex Hughes, Thomas Keegan, Rennie Negron, Jai Broome, Mark McKnight, Liza Nicoll, Jennifer Nelson, Emma Iriarte, Maria Ordonez, Edo Airoldi, James H. Fowler, and Nicholas A. Christakis (2017). "Exploiting Social Influence to Magnify Population Level Behaviour Change in Maternal and Child Health: Study Protocol for a Randomised Controlled Trial of Network Targeting Algorithms in Rural Honduras". *BMJ Open*, 7(3).

Vera-Cossio, Diego A. (2017). *Targeting Credit through Community Members.* Working Paper 07/2017. Institute for Advanced Development Studies.

Watts, Duncan J., Peter Sheridan Dodds, and M. E. J. Newman (2002). "Identity and Search in Social Networks". *Science*, 296(5571): 1302–1305.

Chapter 3 References

Ahmad, Sohaib, Abdul Lateef Haamid, Zafar Ayyub Qazi, Zhenyu Zhou, Theophilus Benson, and Ihsan Ayyub Qazi (2016). "A View from the Other Side: Understanding Mobile Phone Characteristics in the Developing World". *Proceedings of the 2016 ACM on Internet Measurement Conference*. ACM, 319–325.

Alatas, Vivi, Abhijit Banerjee, Arun G. Chandrasekhar, Rema Hanna, and Benjamin A. Olken (2016). "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia". *American Economic Review*, 106(7): 1663–1704.

Barela, Mary Claire, Mae Sincere Blanco, Philip Martinez, Miguel Carlo Purisima, Kurtis Heimer, Matthew Podolsky, Eric Brewer, and Cedric Angelo Festin (2016). "Towards Building a Community Cellular Network in the Philippines: Initial Site Survey Observations". *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*. ICTD '16. New York, NY, USA: ACM, 55:1–55:4.

Bloch, Francis and Matthew Olckers (2018). "Friend-Based Ranking". arXiv: `1807.05093 [physics]`.

Blumenstock, Joshua E. (2014). "Calling for Better Measurement: Estimating an Individual's Wealth and Well-Being from Mobile Phone Transaction Records". *Proc. 20th ACM Conference on Knowledge Discovery and Mining (KDD '14), Workshop on Data Science for Social Good.*

Blumenstock, Joshua E. (2018). "Estimating Economic Characteristics with Phone Data". *AEA Papers and Proceedings*, 108: 72–76.

Blumenstock, Joshua E., Gabriel Cadamuro, and Robert On (2015). "Predicting Poverty and Wealth from Mobile Phone Metadata". *Science*, 350(6264): 1073–1076. pmid: `26612950`.

Blumenstock, Joshua E. and Nathan Eagle (2012). "Divided We Call: Disparities in Access and Use of Mobile Phones in Rwanda". *Information Technologies & International Development*, 8(2): pp. 1–16.

Blumenstock, Joshua E., Niall Keleher, Arman Rezaee, and Erin Troland (2019). "Community Cellular Networks: An Experimental Evaluation in the Philippines (Registered Trial)". *AEA RCT Registry*.

Filmer, Deon and Lant H. Pritchett (2001). "Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application To Educational Enrollments In States Of India*". *Demography*, 38(1): 115–132.

GSMA (2018). *The Mobile Economy: Asia Pacific 2018*. GSMA Intelligence.

GSMA (2019). *The Mobile Gender Gap Report 2019*. GSMA Intelligence.

Hasan, Shaddi, Mary Claire Barela, Matthew Johnson, Eric Brewer, and Kurtis Heimerl (2019). "Scaling Community Cellular Networks with CommunityCellularManager". 16th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 19), 735–750.

Heimerl, Kurtis and Eric Brewer (2010). "The Village Base Station". *Proceedings of the 4th ACM Workshop on Networked Systems for Developing Regions.* NSDR '10. San Francisco, CA: ACM, 14:1–14:2.

Heimerl, Kurtis, Shaddi Hasan, Kashif Ali, Eric Brewer, and Tapan Parikh (2013). "Local, Sustainable, Small-Scale Cellular Networks". *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers - Volume 1.* ICTD '13. New York, NY, USA: ACM, 2–12.

Heimerl, Kurtis, Anuvind Menon, Shaddi Hasan, Kashif Ali, Eric Brewer, and Tapan Parikh (2015). "Analysis of Smartphone Adoption and Usage in a Rural Community Cellular Network". *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development.* ACM, 40.

Jang, Esther, Mary Claire Barela, Matt Johnson, Philip Martinez, Cedric Festin, Margaret Lynn, Josephine Dionisio, and Kurtis Heimerl (2018). "Crowdsourcing Rural Network Maintenance and Repair via Network Messaging". *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18.* The 2018 CHI Conference. Montreal QC, Canada: ACM Press, 1–12.

Kolenikov, Stanislav and Gustavo Angeles (2009). "Socioeconomic Status Measurement with Discrete Proxy Variables: Is Principal Component Analysis a Reliable Answer?" *Review of Income and Wealth*, 55(1): 128–165.

Kshirsagar, Varun, Jerzy Wieczorek, Sharada Ramanathan, and Rachel Wells (2017). "Household Poverty Classification in Data-Scarce Environments: A Machine Learning Approach". arXiv: `1711.06813 [stat]`.

Sarker, Suprateek and John D. Wells (2003). "Understanding Mobile Handheld Device Use and Adoption". *Commun. ACM*, 46(12): 35–40.

Shah, Kushal, Philip Martinez, Emre Tepedelenlioglu, Shaddi Hasan, Cedric Festin, Joshua E. Blumenstock, Josephine Dionisio, and Kurtis Heimerl (2017). "An Investigation of Phone Upgrades in Remote Community Cellular Networks".

World Economic Forum (2018). *Financing Forward-Looking Internet for All.* World Economic Forum.

Appendices

Appendix A

Chapter 1 Additional Materials

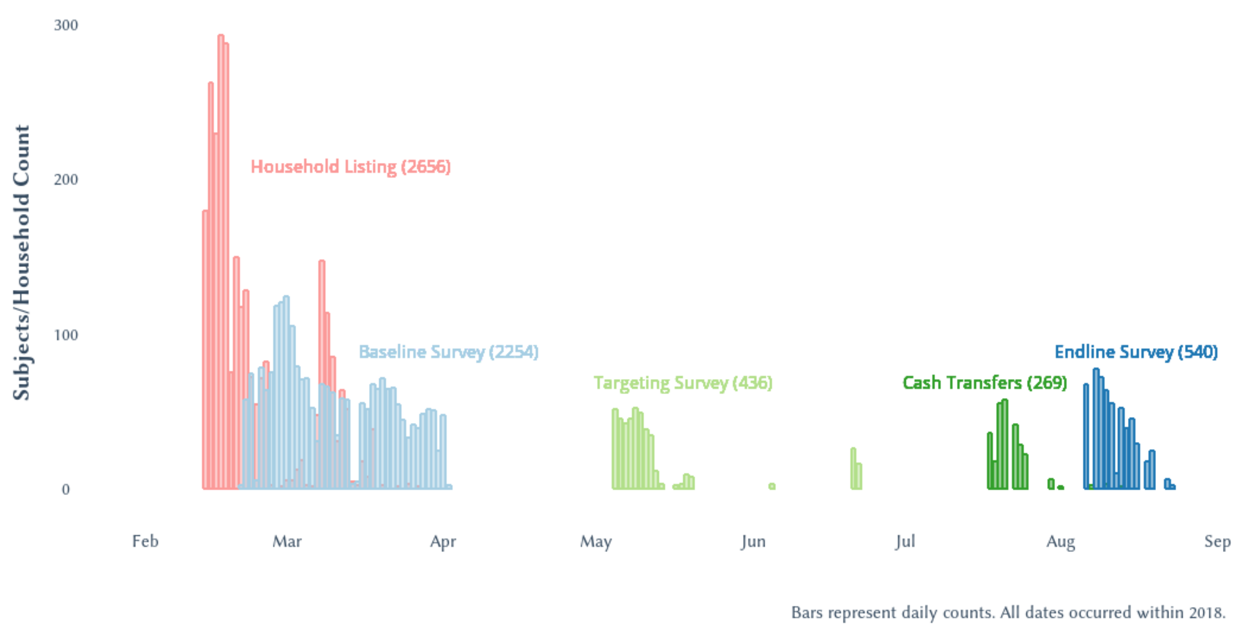

Figure A.1.1: Project Timeline

Table A.1.1: Baseline Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Median	Max
Household Size	2,253	4.441	2.639	1	4	25
Household Head is female	2,253	0.257	0.437	0	0	1
Years in community	2,236	10.075	10.828	0.000	6.000	72.000
Household includes community leader	2,253	0.144	0.351	0	0	1
Household rents dwelling	2,253	0.709	0.454	0	1	1
Rooms in dwelling	2,252	1.531	1.118	1.000	1.000	12.000
Religion is Christian	2,253	0.712	0.453	0	1	1
Household expenditures (LD)	2,251	914.019	886.646	0.000	640.251	12,453.330
Per capita expenditures (LD)	2,251	259.392	284.261	0.000	168.888	4,151.111
Predicted per capita expenditures (LD)	2,253	176.171	95.773	23.079	151.379	952.211
Proxy means score	2,253	53.851	14.567	10	55	100
Wealth index (PCA)	2,253	−0.596	1.173	−3.600	−0.638	2.925
Cantril ladder (1–5)	2,253	2.318	1.104	1	2	5
Meals per day	2,248	1.788	0.700	1.000	2.000	4.000
Shock in past 12 months	2,253	0.623	0.485	0	1	1
Wealth Shock in past 12 months	2,253	0.499	0.500	0	0	1
Health Shock in past 12 months	2,253	0.322	0.467	0	0	1
In-degree centrality	2,253	3.230	3.080	0	2	22
Betweenness centrality	2,253	316.627	428.388	0.000	170.971	5,332.499
Eigenvector centrality	2,253	0.146	0.183	0.000	0.079	1.000

Notes: Household summary statistics for all 13 community blocks.

Table A.1.2: Baseline Summary Statistics, by Community

	C N=722	L N=958	W N=976	p-value
Household size	4.12 (2.33)	4.58 (2.96)	4.44 (2.47)	<0.01
Household head is female	0.27 (0.44)	0.27 (0.44)	0.23 (0.42)	0.14
Years in community	9.18 (10.32)	9.26 (10.42)	11.54 (11.42)	<0.01
Household includes community leader	0.13 (0.34)	0.14 (0.35)	0.15 (0.36)	0.55
Household rents dwelling	0.75 (0.43)	0.63 (0.48)	0.75 (0.43)	<0.01
Number of rooms in dwelling	1.43 (0.92)	1.82 (1.43)	1.31 (0.77)	<0.01
Religion is Christian	0.58 (0.49)	0.74 (0.44)	0.78 (0.41)	<0.01
Household expenditures (LD)	689.45 (508.38)	658.44 (473.27)	1334.02 (1207.18)	<0.01
Per capita expenditures (LD)	208.93 (174.73)	189.62 (167.98)	366.14 (390.15)	<0.01
Predicted per capita expenditures (LD)	185.14 (92.71)	161.57 (90.20)	166.28 (85.70)	<0.01
Proxy means score	56.04 (13.97)	53.13 (14.75)	52.97 (14.66)	<0.01
Wealth index (PCA)	-0.34 (1.09)	-0.50 (1.21)	-0.88 (1.14)	<0.01
Cantril ladder (1–5)	2.32 (1.18)	2.39 (1.02)	2.25 (1.12)	0.04
Meals per day	1.77 (0.71)	1.69 (0.69)	1.90 (0.69)	<0.01
Any shock in past 12 months	0.63 (0.48)	0.66 (0.47)	0.73 (0.44)	<0.01
Wealth shock in past 12 months	0.45 (0.50)	0.49 (0.50)	0.55 (0.50)	<0.01
Health shock in past 12 months	0.28 (0.45)	0.31 (0.46)	0.37 (0.48)	<0.01
In-degree centrality	2.82 (2.94)	2.94 (3.26)	2.76 (2.90)	0.40
Betweenness centrality	266.24 (359.96)	337.50 (514.57)	211.50 (306.88)	<0.01
Eigenvector centrality	0.13 (0.17)	0.13 (0.17)	0.12 (0.19)	0.74

Table A.1.3: Baseline Summary Statistics (Clara Town)

	C1 N=321	C2 N=251	C3 N=150	p-value
Household size	3.99 (2.19)	4.10 (2.43)	4.41 (2.41)	0.24
Household head is female	0.26 (0.44)	0.31 (0.46)	0.21 (0.41)	0.08
Years in community	8.19 (9.94)	10.64 (11.27)	8.59 (9.02)	0.03
Household includes community leader	0.16 (0.36)	0.11 (0.31)	0.11 (0.32)	0.25
Household rents dwelling	0.70 (0.46)	0.77 (0.42)	0.83 (0.37)	0.01
Number of rooms in dwelling	1.41 (0.87)	1.35 (0.87)	1.62 (1.08)	0.02
Religion is Christian	0.57 (0.50)	0.60 (0.49)	0.55 (0.50)	0.59
Household expenditures (LD)	645.61 (428.14)	727.93 (639.53)	709.82 (378.31)	0.19
Per capita expenditures (LD)	204.24 (193.06)	222.78 (174.62)	194.18 (130.35)	0.29
Predicted per capita expenditures (LD)	178.64 (98.06)	197.39 (94.70)	178.55 (74.22)	0.03
Proxy means score	57.26 (13.68)	54.36 (14.29)	56.56 (13.77)	0.07
Wealth index (PCA)	-0.30 (1.08)	-0.41 (1.06)	-0.26 (1.14)	0.38
Cantril ladder (1–5)	2.26 (1.23)	2.33 (1.16)	2.41 (1.12)	0.48
Meals per day	1.69 (0.67)	1.86 (0.78)	1.77 (0.63)	0.03
Any shock in past 12 months	0.64 (0.48)	0.72 (0.45)	0.47 (0.50)	<0.01
Wealth shock in past 12 months	0.42 (0.49)	0.55 (0.50)	0.31 (0.47)	<0.01
Health shock in past 12 months	0.28 (0.45)	0.36 (0.48)	0.13 (0.34)	<0.01
In-degree centrality	2.45 (2.75)	3.03 (3.23)	3.27 (2.72)	0.01
Betweenness centrality	291.99 (396.68)	316.66 (383.91)	126.75 (128.22)	<0.01
Eigenvector centrality	0.07 (0.11)	0.13 (0.18)	0.24 (0.19)	<0.01

Table A.1.4: Baseline Summary Statistics (Logan Town)

	L1 N=190	L2 N=133	L3 N=227	L4 N=408	p-value
Household size	3.48 (1.93)	4.75 (3.07)	4.44 (2.79)	5.07 (3.24)	$<$0.01
Household head is female	0.20 (0.40)	0.32 (0.47)	0.30 (0.46)	0.27 (0.44)	0.08
Years in community	8.06 (9.63)	10.09 (12.16)	8.18 (10.70)	10.07 (10.00)	0.07
Household includes community leader	0.17 (0.38)	0.22 (0.41)	0.14 (0.35)	0.11 (0.32)	0.03
Household rents dwelling	0.61 (0.49)	0.60 (0.49)	0.61 (0.49)	0.66 (0.47)	0.47
Number of rooms in dwelling	1.41 (0.86)	2.35 (1.90)	1.81 (1.34)	1.85 (1.47)	$<$0.01
Religion is Christian	0.86 (0.35)	0.88 (0.33)	0.85 (0.35)	0.59 (0.49)	$<$0.01
Household expenditures (LD)	555.87 (361.98)	716.57 (660.44)	698.48 (465.16)	665.51 (448.57)	0.01
Per capita expenditures (LD)	207.78 (173.38)	196.42 (196.15)	207.79 (180.06)	170.68 (147.91)	0.03
Predicted per capita expenditures (LD)	160.84 (77.73)	155.13 (84.35)	174.20 (109.62)	156.98 (84.97)	0.10
Proxy means score	49.92 (14.13)	52.51 (16.68)	56.16 (16.26)	53.18 (13.29)	$<$0.01
Wealth index (PCA)	-0.94 (1.09)	-0.52 (1.30)	-0.22 (1.25)	-0.45 (1.16)	$<$0.01
Cantril ladder (1–5)	2.47 (1.14)	2.53 (0.92)	2.51 (1.01)	2.25 (0.98)	0.01
Meals per day	1.77 (0.68)	1.55 (0.67)	1.86 (0.72)	1.62 (0.66)	$<$0.01
Any shock in past 12 months	0.65 (0.48)	0.75 (0.43)	0.66 (0.47)	0.64 (0.48)	0.14
Wealth shock in past 12 months	0.48 (0.50)	0.56 (0.50)	0.40 (0.49)	0.52 (0.50)	0.03
Health shock in past 12 months	0.24 (0.43)	0.35 (0.48)	0.34 (0.47)	0.31 (0.46)	0.14
In-degree centrality	3.22 (3.61)	2.47 (3.33)	2.75 (2.64)	3.08 (3.37)	0.13
Betweenness centrality	194.51 (306.56)	107.23 (194.30)	285.40 (381.56)	508.14 (654.44)	$<$0.01
Eigenvector centrality	0.14 (0.20)	0.13 (0.18)	0.04 (0.12)	0.17 (0.16)	$<$0.01

Table A.1.5: Baseline Summary Statistics (West Point)

	W1 N=149	W2 N=73	W3 N=159	W4 N=102	W5 N=190	W6 N=303	p-value
Household size	4.54 (2.44)	4.88 (2.80)	4.58 (2.31)	4.53 (2.29)	4.34 (2.76)	4.26 (2.38)	0.49
Household head is female	0.21 (0.41)	0.22 (0.42)	0.23 (0.42)	0.31 (0.47)	0.30 (0.46)	0.17 (0.38)	0.02
Years in community	10.01 (9.14)	12.60 (10.59)	12.38 (13.26)	12.97 (14.11)	10.68 (10.25)	11.69 (11.20)	0.31
Household includes community leader	0.09 (0.29)	0.17 (0.38)	0.16 (0.37)	0.19 (0.40)	0.15 (0.36)	0.15 (0.35)	0.38
Household rents dwelling	0.72 (0.45)	0.75 (0.43)	0.72 (0.45)	0.70 (0.46)	0.87 (0.34)	0.74 (0.44)	0.01
Number of rooms in dwelling	1.46 (0.91)	1.46 (0.95)	1.23 (0.61)	1.34 (0.54)	1.19 (0.52)	1.31 (0.90)	0.02
Religion is Christian	0.90 (0.30)	0.95 (0.23)	0.83 (0.37)	0.78 (0.42)	0.78 (0.41)	0.65 (0.48)	<0.01
Household expenditures (LD)	1567.22 (1476.87)	1150.76 (870.41)	1633.11 (1473.54)	1698.04 (1203.57)	1195.51 (927.92)	1044.30 (1010.51)	<0.01
Per capita expenditures (LD)	433.28 (469.51)	276.46 (230.45)	382.84 (302.44)	443.26 (389.97)	383.36 (454.89)	302.93 (359.63)	<0.01
Predicted per capita expenditures (LD)	162.64 (83.33)	156.85 (84.57)	172.87 (80.85)	168.27 (78.12)	169.82 (83.69)	163.99 (93.25)	0.75
Proxy means score	50.93 (13.98)	52.18 (16.04)	50.69 (14.18)	52.44 (15.13)	53.33 (14.14)	55.47 (14.87)	0.02
Wealth index (PCA)	-0.98 (1.16)	-0.70 (1.18)	-0.93 (1.07)	-0.93 (1.08)	-0.92 (1.05)	-0.80 (1.22)	0.50
Cantril ladder (1–5)	2.13 (0.96)	2.67 (1.39)	2.59 (1.34)	2.13 (0.89)	2.24 (1.44)	2.08 (0.73)	<0.01
Meals per day	1.92 (0.72)	1.89 (0.70)	1.89 (0.68)	1.91 (0.67)	1.82 (0.69)	1.93 (0.70)	0.74
Any shock in past 12 months	0.74 (0.44)	0.73 (0.45)	0.73 (0.45)	0.81 (0.39)	0.69 (0.46)	0.72 (0.45)	0.38
Wealth shock in past 12 months	0.57 (0.50)	0.44 (0.50)	0.54 (0.50)	0.62 (0.49)	0.53 (0.50)	0.54 (0.50)	0.43
Health shock in past 12 months	0.32 (0.47)	0.40 (0.49)	0.40 (0.49)	0.39 (0.49)	0.29 (0.45)	0.41 (0.49)	0.13
In-degree centrality	2.36 (2.66)	2.22 (2.53)	2.75 (3.02)	2.86 (2.65)	2.91 (2.88)	2.96 (3.10)	0.18
Betweenness centrality	173.24 (233.93)	54.03 (78.83)	181.81 (266.85)	93.91 (126.19)	211.08 (253.58)	323.69 (410.42)	<0.01
Eigenvector centrality	0.16 (0.19)	0.12 (0.24)	0.09 (0.18)	0.17 (0.21)	0.05 (0.13)	0.15 (0.19)	<0.01

Table A.1.6: Baseline Summary Statistics, by TA Nomination Prompt

	Nomination Prompt			
	1	2	3	F-Test
	N=811	N=698	N=744	p-value
Household Size	4.48 (2.77)	4.46 (2.55)	4.39 (2.58)	0.80
Household head is female	0.27 (0.44)	0.25 (0.44)	0.25 (0.43)	0.58
Years in Community	10.66 (11.27)	9.47 (10.03)	10.01 (11.04)	0.11
Household includes a community leader	0.17 (0.38)	0.11 (0.32)	0.14 (0.35)	0.01
Household rents dwelling	0.71 (0.46)	0.70 (0.46)	0.72 (0.45)	0.62
Number of rooms in dwelling	1.62 (1.27)	1.49 (0.97)	1.47 (1.06)	0.02
Christian	0.74 (0.44)	0.68 (0.47)	0.72 (0.45)	0.07
Household expenditure (LD)	900.60 (913.50)	941.70 (942.26)	902.65 (799.19)	0.61
Per capita expenditure (LD)	255.92 (313.57)	257.06 (247.74)	265.38 (283.01)	0.78
Predicted PCE (LD)	173.87 (100.69)	175.34 (88.11)	179.46 (97.20)	0.50
Proxy means score	53.54 (14.56)	53.97 (14.60)	54.08 (14.56)	0.74
Cantril Ladder	2.29 (1.07)	2.36 (1.13)	2.31 (1.11)	0.44
Meals per day	1.77 (0.71)	1.79 (0.69)	1.81 (0.70)	0.65
In-degree Centrality	3.20 (2.99)	3.17 (3.00)	3.32 (3.25)	0.61
Betweenness Centrality	318.46 (423.27)	312.03 (394.05)	318.94 (463.92)	0.94
Eigenvector Centrality	0.15 (0.19)	0.14 (0.18)	0.15 (0.19)	0.58

Notes: One of three nomination prompts was randomly assigned to each baseline respondent.
Nomination Prompt 1: If we want to spread information about a social assistance program, to whom do you suggest we speak?;
Nomination Prompt 2: If we want to identify which people would be best to help us identify which people in this block would benefit most from a social assistance program, to whom do you suggest we speak?;
Nomination Prompt 3: If we want to identify which people would be best to help us identify which people in this block would benefit most from a cash gift for social assistance, to whom do you suggest we speak?

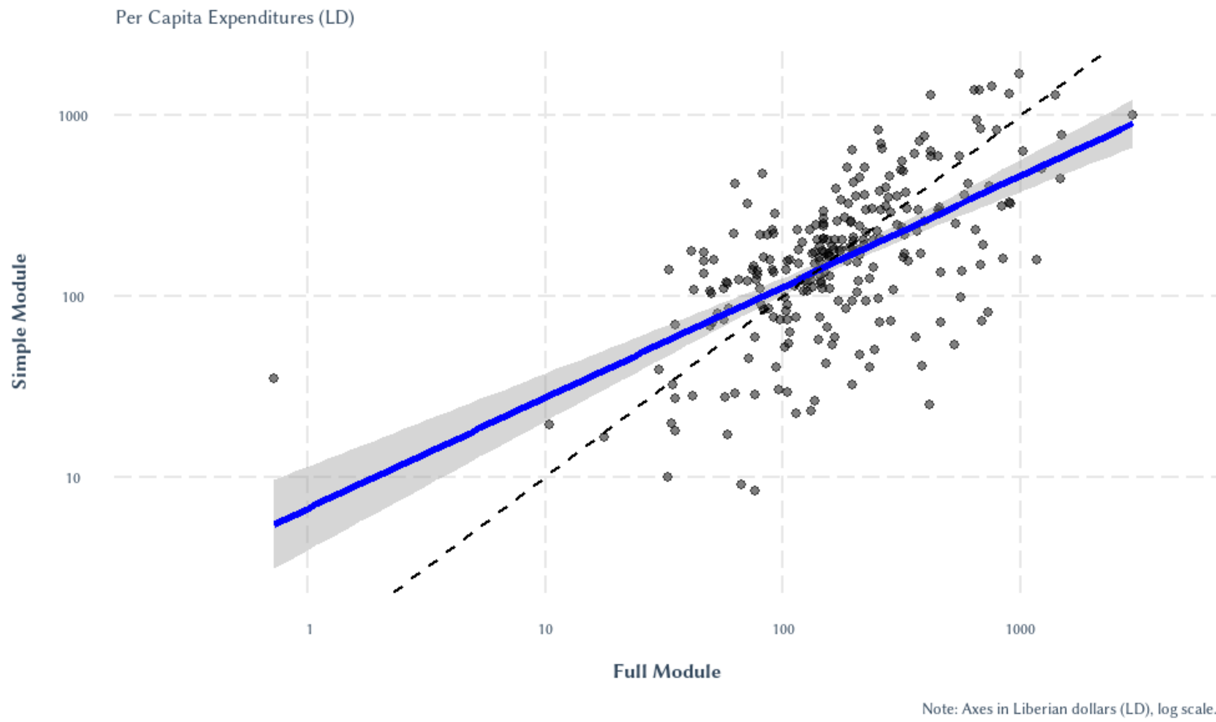

Figure A.1.2: Per capita expenditure, simple module vs. full module

Table A.1.7: Knowledge of Neighbors, by TA Type

	All TAs	Random TAs	Nominated TAs	Leader TAs
	N=393	N=132	N=254	N=94
# of neighbors	30.9 (6.05)	31.4 (5.46)	30.9 (6.19)	32.0 (6.11)
Prop. known	0.34 (0.22)	0.32 (0.22)	0.36 (0.22)	0.41 (0.24)
Prop. classified as poor	0.30 (0.28)	0.32 (0.30)	0.30 (0.27)	0.29 (0.26)
Accuracy	0.62 (0.22)	0.60 (0.23)	0.63 (0.22)	0.61 (0.21)
Inclusion Error Rate	0.10 (0.12)	0.09 (0.11)	0.10 (0.13)	0.12 (0.15)
Exclusion Error Rate	0.89 (0.16)	0.89 (0.15)	0.88 (0.17)	0.88 (0.16)

Notes: Standard deviations in parentheses. TA types are not mutually exclusive.

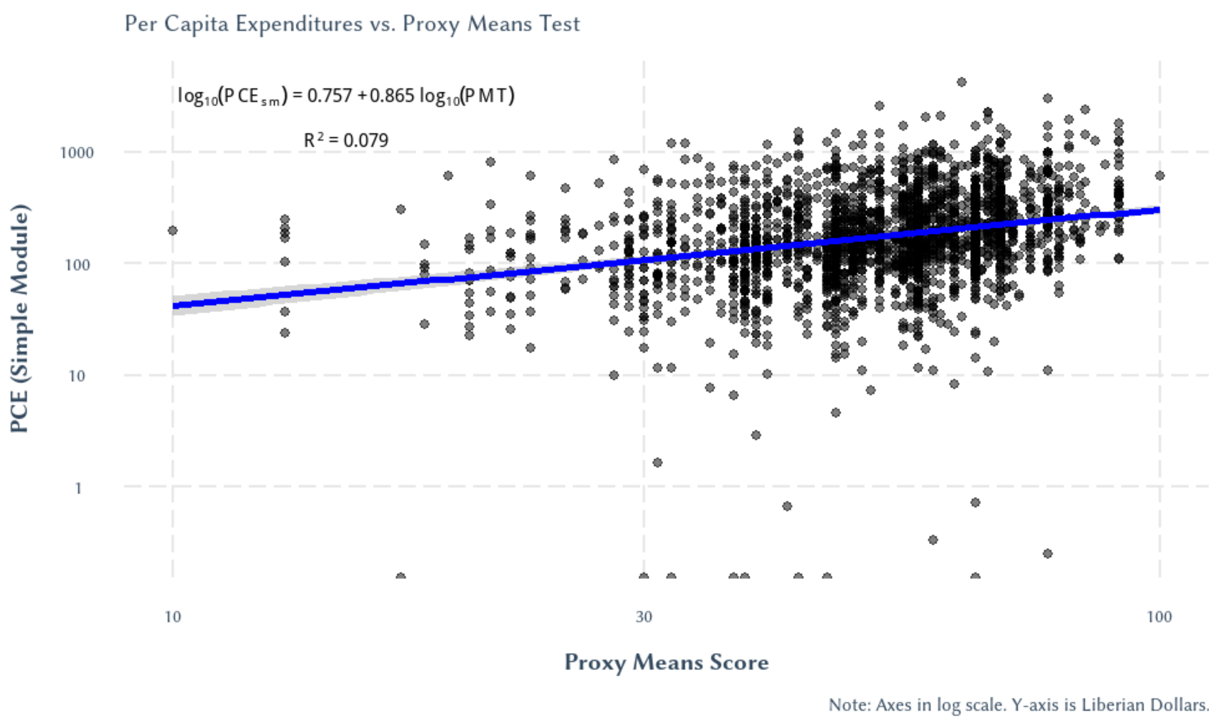

Figure A.1.3: Per capita expenditure (simple module) vs. PMT Score

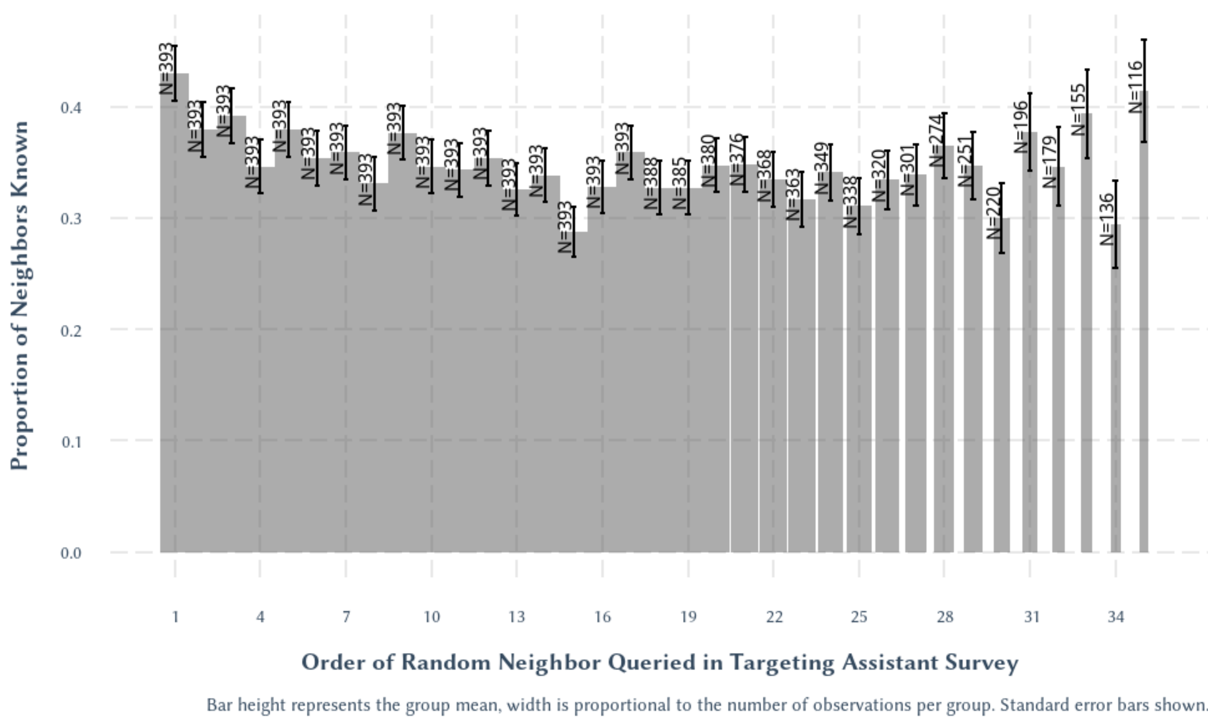

Figure A.1.4: Order of Random Neighbor and TA knowledge

Table A.1.8: Targeting Assistant Knows Random Neighbor

	(1)	(2)	(3)	(4)
Social Distance = 2	-0.236^{***}	-0.249^{***}	-0.236^{***}	-0.249^{***}
	(0.024)	(0.046)	(0.024)	(0.046)
Social Distance = 3	-0.383^{***}	-0.371^{***}	-0.373^{***}	-0.368^{***}
	(0.021)	(0.041)	(0.022)	(0.041)
Social Distance = 4	-0.446^{***}	-0.400^{***}	-0.428^{***}	-0.388^{***}
	(0.021)	(0.041)	(0.021)	(0.041)
Social Distance = 5	-0.491^{***}	-0.423^{***}	-0.462^{***}	-0.400^{***}
	(0.022)	(0.043)	(0.023)	(0.043)
Social Distance = 6	-0.486^{***}	-0.424^{***}	-0.454^{***}	-0.385^{***}
	(0.028)	(0.051)	(0.030)	(0.053)
Social Distance = 7	-0.541^{***}	-0.417^{***}	-0.490^{***}	-0.390^{***}
	(0.046)	(0.078)	(0.051)	(0.085)
Social Distance = 8	-0.448^{***}	-0.441^{**}	-0.463^{***}	-0.475^{**}
	(0.112)	(0.162)	(0.113)	(0.163)
Social Distance = ∞	-0.520^{***}	-0.515^{***}	-0.545^{***}	-0.530^{***}
	(0.025)	(0.044)	(0.028)	(0.045)
Nominated TA		0.073		0.066
		(0.045)		(0.045)
Nominated TA * Social Distance = 2		0.021		0.023
		(0.053)		(0.053)
Nominated TA * Social Distance = 3		-0.014		-0.001
		(0.048)		(0.048)
Nominated TA * Social Distance = 4		-0.064		-0.053
		(0.048)		(0.047)
Nominated TA * Social Distance = 5		-0.099^{*}		-0.087
		(0.050)		(0.050)
Nominated TA * Social Distance = 6		-0.090		-0.100
		(0.062)		(0.063)
Nominated TA * Social Distance = 7		-0.209^{*}		-0.157
		(0.095)		(0.105)
Nominated TA * Social Distance = 8		0.011		0.040
		(0.224)		(0.224)
Nominated TA * Social Distance = ∞		0.014		0.018
		(0.054)		(0.059)
Adj. R^2	0.057	0.060	0.071	0.073
Num. obs.	12130	12130	11140	11140
Neighbor Covariates	NO	NO	YES	YES
Neighbor Order F.E.	YES	YES	YES	YES

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variable is a binary variable equal to one if the respondent responds in the affirmative to the question, "Have you heard of [NAME] in your community?", where [NAME] is the name of a randomly selected adult from the TA's neighborhood. SD indicates that a variables has been mean-centered and scaled to represent one standard standard deviation. Eicker-Huber-White robust standard errors in parentheses. Neighbor-Order fixed effects included. Social Distance is the shortest path between the nominating and nominated household as calculated using social network data from the baseline.

Table A.1.9: Cash Grant Nominations

	(1) Any Nomination	(2) Count of Nominations	(3) 'Best Use' Nomination	(4) 'Hard Times' Nomination
Female HOH	0.041^{*}	0.105^{**}	0.015	0.055^{**}
	(0.020)	(0.036)	(0.017)	(0.017)
Years in community (SD)	0.041^{***}	0.075^{***}	0.033^{***}	0.027^{***}
	(0.010)	(0.017)	(0.008)	(0.008)
Leader	0.044	0.040	0.025	0.020
	(0.026)	(0.041)	(0.022)	(0.021)
Christian	0.070^{***}	0.123^{***}	0.059^{***}	0.034^{*}
	(0.018)	(0.029)	(0.015)	(0.015)
PCE Rank	-0.000^{***}	-0.000^{**}	-0.000^{*}	-0.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
Wealth Shock	-0.020	-0.037	-0.032^{*}	-0.001
	(0.017)	(0.028)	(0.014)	(0.014)
Health Shock	0.017	0.075^{*}	0.022	0.024
	(0.019)	(0.033)	(0.016)	(0.015)
Eigenvector Centrality (SD)	0.064^{***}	0.103^{***}	0.049^{***}	0.034^{***}
	(0.009)	(0.017)	(0.009)	(0.008)
R^2	0.068	0.071	0.054	0.045
Adj. R^2	0.060	0.063	0.046	0.036
Num. obs.	2236	2236	2236	2236

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. SD indicates that a variables has been mean-centered and scaled to represent one standard standard deviation. Eicker-Huber-White robust standard errors in parentheses. Neighbor fixed effects included.

Table A.1.10: Cash Grant Nominations, First TA Nomination Only

	(A) **Nominee Characteristics**						(B) **Nominee Shocks**		(C) **Dyadic Distance**	
	Pred. PCE Rank	Female	Years in Community	Leader	Christian	Eigenvector Centrality	Wealth Shock	Health Shock	Social Distance	Geographic Distance
TA Type and Prompt:										
Nominated TA	-57.01	-0.10	1.47	-0.05	-0.05	-0.02	0.05	0.07	0.10	8.22
	(133.37)	(0.08)	(2.15)	(0.07)	(0.06)	(0.03)	(0.08)	(0.08)	(0.18)	(7.65)
Hard Times Elicitation	-118.37	-0.00	0.00	-0.03	-0.08	-0.01	0.14	0.07	0.26	8.64
	(137.89)	(0.09)	(2.07)	(0.07)	(0.06)	(0.03)	(0.09)	(0.09)	(0.20)	(11.12)
Nominated TA * Hard Times	110.33	0.07	-2.56	0.01	0.03	0.00	-0.01	-0.06	-0.15	-8.02
	(177.24)	(0.10)	(2.70)	(0.09)	(0.08)	(0.04)	(0.11)	(0.11)	(0.25)	(13.08)
Nominating TA Characteristics:										
Predicted PCE Rank	0.12^{*}	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00
	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female	-69.10	0.10	2.44	-0.03	0.03	-0.02	-0.04	0.11^{*}	0.21	6.16
	(94.41)	(0.05)	(1.42)	(0.04)	(0.04)	(0.02)	(0.06)	(0.06)	(0.13)	(6.42)
Years in comm. (SD)	-15.20	-0.01	1.18	-0.01	0.06^{**}	0.00	0.02	0.02	-0.06	-2.79
	(47.71)	(0.03)	(0.94)	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.07)	(2.88)
Leader	-8.44	0.07	-0.82	0.10	0.03	-0.01	-0.07	0.01	0.05	-3.82
	(102.18)	(0.06)	(1.65)	(0.05)	(0.05)	(0.03)	(0.07)	(0.07)	(0.15)	(6.09)
Christian	-39.28	-0.07	0.80	0.05	0.39^{***}	0.00	0.07	-0.04	-0.13	-13.17
	(116.56)	(0.06)	(1.78)	(0.05)	(0.06)	(0.03)	(0.07)	(0.07)	(0.16)	(8.75)
Eigenvector (SD)	14.49	-0.00	-1.06	0.02	0.01	0.11^{***}	0.02	0.01	-0.11	-4.30
	(48.96)	(0.03)	(0.72)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.07)	(2.37)
Adj. R^2	0.07	0.07	0.04	0.02	0.26	0.26	0.03	0.02	0.04	0.11
Num. obs.	332	332	314	332	322	332	322	322	323	332
$\rho + \theta = 0$ (p-value)	0.94	0.23	0.14	0.80	0.27	0.79	0.08	0.97	0.50	0.93

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Eicker-Huber-White robust standard errors in parentheses. Neighborhood fixed effects included.
Nominated TA is a binary variable equal to one if the Targeting Assistant was nominated by another community member. Targeting assistants were asked to provide two nominations for a cash grant. We elicited the cash grant nominations through two independent questions. 'Best Use' elicitation question (reference group in table): "Think of households within this neighborhood, which household would make the most out of a USD$100 cash grant?" 'Hard Times' elicitation question: "Within this neighborhood, is there a household who recently fell on hard times and would benefit most from a USD$100 cash grant?" Table 1.13 shows regression results including all cash grant nominees elicited from targeting assistants. Predicted PCE Rank is the household ranking of the per capita expenditures as predicted by our random forest algorithm, centered at zero. Social Distance is the shortest path between the nominating and nominated household as calculated using social network data from the baseline. Geographic distance is the number of meters between the nominating and nominated household's dwelling as measured by the spherical geodesic distance between the latitude-longitude coordinates of the dwellings.

Appendix B

Chapter 2 Additional Materials

Table A.2.1: Baseline Summary Statistics, by VBTS Site

	(1) site 1	(2) site 2	(3) site 3	(4) site 4	(5) site 5	(6) site 6	(7) site 7	
Panel A: Household Summary Statistics	N=88	N=382	N=176	N=100	N=255	N=50	N=80	p-value
Adults in household	2.44 (1.18)	2.89 (1.36)	2.81 (1.31)	2.51 (1.15)	2.53 (1.24)	2.56 (1.15)	2.74 (1.38)	<0.01
Children (0-14) in household	1.66 (1.29)	2.06 (1.62)	1.77 (1.51)	1.59 (1.40)	1.45 (1.28)	1.70 (1.52)	1.74 (1.56)	<0.01
Female head of household	0.43 (0.50)	0.45 (0.50)	0.32 (0.47)	0.33 (0.47)	0.35 (0.48)	0.13 (0.34)	0.12 (0.32)	<0.01
HOH - Secondary school	0.24 (0.43)	0.33 (0.47)	0.12 (0.32)	0.21 (0.41)	0.36 (0.48)	0.17 (0.38)	0.18 (0.39)	<0.01
Rooms in dwelling	1.75 (0.75)	1.96 (0.87)	1.72 (0.78)	1.54 (0.77)	1.72 (0.76)	1.62 (0.67)	1.77 (0.78)	<0.01
Household owns land	0.39 (0.49)	0.50 (0.50)	0.62 (0.49)	0.28 (0.45)	0.38 (0.49)	0.50 (0.51)	0.29 (0.46)	<0.01
Family member in political/govt office	0.05 (0.21)	0.25 (0.44)	0.32 (0.47)	0.32 (0.47)	0.11 (0.31)	0.16 (0.37)	0.12 (0.33)	<0.01
Household has bank account	0.05 (0.21)	0.23 (0.42)	0.14 (0.35)	0.07 (0.26)	0.15 (0.35)	0.06 (0.24)	0.15 (0.36)	<0.01
Income Source - Farming	0.14 (0.35)	0.40 (0.49)	0.48 (0.50)	0.40 (0.49)	0.29 (0.45)	0.46 (0.50)	0.04 (0.19)	<0.01
Income Source - Fishing	0.42 (0.50)	0.21 (0.41)	0.35 (0.48)	0.30 (0.46)	0.04 (0.19)	0.02 (0.14)	0.62 (0.49)	<0.01
Income Source - Wage Labor	0.16 (0.37)	0.22 (0.41)	0.09 (0.29)	0.13 (0.34)	0.31 (0.46)	0.26 (0.44)	0.06 (0.24)	<0.01
Poverty Score	42.0 (9.7)	41.6 (12.6)	38.2 (12.1)	42.3 (10.6)	46.8 (11.6)	38.2 (8.1)	41.2 (11.5)	<0.01
Wealth Index - Polychoric PCA	-0.37 (1.21)	0.19 (1.32)	-0.67 (1.36)	-0.25 (1.25)	0.20 (1.28)	-0.93 (1.05)	-0.32 (1.14)	<0.01
Electricity in dwelling	0.89 (0.32)	0.92 (0.28)	0.08 (5.86)	0.82 (0.39)	0.64 (4.89)	0.10 (0.30)	0.31 (0.47)	0.10
Owns sala/sofa set	0.12 (0.33)	0.22 (0.41)	0.12 (0.33)	0.11 (0.31)	0.23 (0.42)	0.08 (0.27)	0.10 (0.30)	<0.01
Owns refrigerator	0.05 (0.21)	0.13 (0.33)	0.05 (0.22)	0.11 (0.31)	0.15 (0.36)	0.00 (0.00)	0.01 (0.11)	<0.01
Owns television	0.49 (0.50)	0.60 (0.49)	0.40 (0.49)	0.47 (0.50)	0.59 (0.49)	0.22 (0.42)	0.51 (0.50)	<0.01
Owns VHS/DVD player	0.32 (0.47)	0.37 (0.48)	0.27 (0.44)	0.30 (0.46)	0.30 (0.46)	0.10 (0.30)	0.19 (0.39)	<0.01
Owns radio	0.23 (0.42)	0.36 (0.48)	0.30 (0.46)	0.29 (0.46)	0.20 (0.40)	0.56 (0.50)	0.50 (0.50)	<0.01
Owns satellite TV dish	0.22 (0.41)	0.37 (0.48)	0.13 (0.34)	0.27 (0.45)	0.43 (0.50)	0.12 (0.33)	0.31 (0.47)	<0.01
Owns motor vehicle or boat	0.39 (0.49)	0.23 (0.42)	0.31 (0.46)	0.32 (0.47)	0.35 (0.48)	0.56 (0.50)	0.50 (0.50)	<0.01
Owns gas stove	0.15 (0.36)	0.20 (0.40)	0.08 (0.27)	0.12 (0.33)	0.28 (0.45)	0.12 (0.33)	0.29 (0.46)	<0.01
Owns cellphone	0.80 (0.41)	0.69 (0.46)	0.47 (0.50)	0.57 (0.50)	0.85 (0.36)	0.58 (0.50)	0.62 (0.49)	<0.01
Number of cellphones	1.20 (0.91)	1.23 (1.22)	0.75 (1.07)	1.13 (1.38)	1.55 (1.11)	0.94 (1.17)	1.21 (1.24)	<0.01
Owns SIM card	0.78 (0.41)	0.65 (0.48)	0.43 (0.50)	0.50 (0.50)	0.82 (0.38)	0.56 (0.50)	0.56 (0.50)	<0.01
Number of SIM cards	1.34 (1.14)	1.34 (1.51)	0.84 (1.53)	1.21 (1.90)	1.73 (1.30)	1.04 (1.32)	1.38 (1.73)	<0.01
Owns Globe SIM card	0.33 (0.47)	0.15 (0.36)	0.15 (0.36)	0.14 (0.35)	0.33 (0.47)	0.54 (0.50)	0.51 (0.50)	<0.01
Owns SMART SIM card	0.70 (0.46)	0.61 (0.49)	0.39 (0.49)	0.48 (0.50)	0.73 (0.44)	0.12 (0.33)	0.26 (0.44)	<0.01
In-degree centrality	4.38 (3.56)	6.13 (37.83)	5.15 (4.53)	2.85 (2.89)	6.88 (41.32)	3.18 (3.30)	3.52 (4.06)	0.89
Eigenvector centrality	0.07 (0.16)	0.04 (0.06)	0.17 (0.20)	0.14 (0.21)	0.06 (0.07)	0.21 (0.22)	0.20 (0.18)	<0.01
Panel B: Adult Survey Module	N=118	N=584	N=264	N=134	N=333	N=64	N=120	p-value
Do you see yourself as part of your community?	0.31 (0.47)	0.60 (0.49)	0.48 (0.50)	0.61 (0.49)	0.60 (0.49)	0.68 (0.47)	0.74 (0.44)	<0.01
Do you feel isolated from the rest of your country?	0.19 (0.39)	0.28 (0.45)	0.18 (0.39)	0.22 (0.41)	0.45 (0.50)	0.51 (0.50)	0.35 (0.48)	<0.01
Could you communicate with family in case of emergency?	0.53 (0.50)	0.41 (0.49)	0.23 (0.42)	0.31 (0.46)	0.73 (0.45)	0.40 (0.49)	0.65 (0.48)	<0.01
Traveled to Neighboring Bgy. (12 mo.)	0.34 (0.48)	0.39 (0.49)	0.38 (0.49)	0.60 (0.49)	0.45 (0.50)	0.66 (0.48)	0.65 (0.48)	<0.01
Traveled to Manila (3 yrs)	0.30 (0.46)	0.12 (0.32)	0.07 (0.26)	0.18 (0.39)	0.16 (0.37)	0.11 (0.31)	0.19 (0.40)	<0.01
Total contacts within barangay	4.95 (2.20)	6.34 (6.47)	5.34 (2.44)	5.44 (3.17)	5.28 (2.86)	6.73 (6.08)	10.41 (15.7)	<0.01
Total contacts outside barangay	2.86 (2.09)	4.96 (10.92)	2.65 (2.85)	3.99 (5.93)	2.77 (3.45)	4.62 (4.90)	6.76 (7.67)	<0.01
Close friends/family within barangay	5.01 (2.17)	4.37 (1.22)	5.12 (1.92)	4.51 (1.70)	4.83 (2.01)	3.78 (1.00)	4.03 (1.38)	<0.01
Close friends/family outside barangay	2.55 (1.93)	2.33 (1.59)	2.22 (1.67)	2.46 (1.64)	2.30 (1.93)	2.11 (1.20)	2.64 (1.19)	0.17

Table A.2.2: Call Usage Summary Statistics, by VBTS Site

	(1) All sites N=1131	(2) site 1 N=88	(3) site 2 N=382	(4) site 3 N=176	(5) site 4 N=100	(6) site 5 N=255	(7) site 6 N=50	(8) site 7 N=80	(9) p-value
Any Transaction (prop.)	0.65	0.56	0.94	0.73	0.65	0.17	0.70	0.70	$<$0.01
	(0.48)	(0.50)	(0.24)	(0.45)	(0.48)	(0.38)	(0.46)	(0.46)	
Calls & SMS	773.75	22.47	1793.83	479.85	619.42	3.47	80.26	457.51	$<$0.01
	(1462.56)	(64.74)	(2001.78)	(760.68)	(1023.76)	(21.72)	(99.42)	(678.30)	
Any Outgoing Call (prop.)	0.56	0.20	0.90	0.66	0.63	0.05	0.64	0.66	$<$0.01
	(0.50)	(0.41)	(0.30)	(0.48)	(0.49)	(0.22)	(0.48)	(0.48)	
Outgoing Calls	100.78	3.03	206.99	79.81	118.55	0.51	32.78	87.22	$<$0.01
	(203.43)	(11.72)	(284.68)	(114.87)	(198.18)	(3.59)	(55.88)	(135.66)	
Long-distance Outgoing Calls	70.82	3.02	137.80	65.47	77.38	0.45	29.70	79.17	$<$0.01
	(140.75)	(11.71)	(193.90)	(93.89)	(131.40)	(3.04)	(51.09)	(122.49)	
Any Outgoing SMS (prop.)	0.53	0.15	0.90	0.64	0.59	0.02	0.48	0.62	$<$0.01
	(0.50)	(0.36)	(0.31)	(0.48)	(0.49)	(0.14)	(0.50)	(0.49)	
Outgoing SMS	201.99	3.19	499.99	115.93	87.39	0.22	8.40	94.42	$<$0.01
	(481.14)	(13.36)	(709.54)	(251.60)	(165.78)	(2.36)	(16.54)	(186.49)	
Long-distance Outgoing SMS	120.08	3.18	277.37	94.37	53.22	0.11	8.08	90.16	$<$0.01
	(274.63)	(13.37)	(391.76)	(209.86)	(97.45)	(1.28)	(16.42)	(177.56)	
Any Incoming Call (prop.)	0.61	0.35	0.92	0.69	0.65	0.12	0.62	0.66	$<$0.01
	(0.49)	(0.48)	(0.27)	(0.46)	(0.48)	(0.33)	(0.49)	(0.48)	
Incoming Calls	299.37	11.33	640.99	223.09	339.56	2.15	31.80	217.19	$<$0.01
	(576.33)	(41.44)	(754.34)	(370.70)	(687.46)	(17.82)	(49.22)	(311.09)	
Long-distance Incoming Calls	268.70	10.70	570.43	205.30	303.27	1.97	27.00	209.22	$<$0.01
	(542.41)	(41.49)	(717.91)	(357.45)	(660.49)	(17.75)	(48.00)	(305.75)	
Any Incoming SMS (prop.)	0.60	0.51	0.93	0.68	0.64	0.10	0.52	0.60	$<$0.01
	(0.49)	(0.50)	(0.25)	(0.47)	(0.48)	(0.30)	(0.50)	(0.49)	
Incoming SMS	171.61	4.91	445.87	61.02	73.92	0.60	7.28	58.67	$<$0.01
	(413.95)	(10.98)	(612.04)	(128.98)	(170.58)	(3.52)	(17.07)	(111.36)	
Long-distance Incoming SMS	89.79	4.62	225.70	37.10	37.07	0.31	5.06	54.44	$<$0.01
	(217.88)	(10.85)	(323.68)	(84.95)	(77.05)	(1.91)	(12.25)	(106.39)	

Standard deviations in parentheses. P-value for t-test of equality of means across all sites shown in column 9. Summary statistics by pre-existing phone ownership shown in Table 2.5.

Poverty Probability Index Score Card for the Philippines

Source: https://www.povertyindex.org/country/philippines

1. How many members does the household have?
2. Are all household members ages 6 to 17 currently attending school?
3. How many household members did any work for at least one hour in the past week
4. In their primary occupation or business in the past week, how many household members were farmers, forestry workers fishers, laborers, or unskilled workers?
5. What is the highest grade completed by the female head/spouse?
6. What type of construction materials are the outer walls made of?
7. Does the family own any sala sets?
8. Does the family own a refrigerator/freezer or a washing machine?
9. Does the family own a television set or a VTR/VHS/VCD/DVD player?
10. How many telephones/cellphones does the family own?

Script for Registration

Hello! We're going to give you your new SIM and we'll help you activate it as well. First, as part of the NTC regulatory requirements, we are informing you that this is a test and an experimental network and the grade of service may NOT be the same as those of conventional networks. Here's a copy of the waiver/subscriber conforme, kindly affix your name and signature below.

(Customer signs. There should be two copies. Give one copy to the customer)

Here is your new SIM card! Please insert it into your phone.

(Staff would then assist the customer to follow the step-by-step instructions stated in this SIM provisioning guide)

You have now activated your SIM! As stated in the text message, your number is 09XXXXXXX.

(Staff will input the number on the form)

Finally, to celebrate the launch of the new network, we will be offering several different promos in the coming months. Subscribers will be selected through a lottery system for the opportunity to avail of one specific promo each. Please refer to this flyer for more information on the promo opportunities. Some of subscribers will receive promos for a certain number of free text messages, some will receive promos to use for a select group of friends and family members, some will receive promos to contact people living outside the network. In the coming months, look for a text message on your phone notifying you of your promo.

Thank you. Please proceed to the next table for the final step.

Customer Agreement

Rules and Guidelines on using the KONEKT Barangay Promo SIM (English Translation)

1. To be able to use the services offered by the Konekt Barangay Promo SIM, you must have the following:
   (a) A working GSM 900 or multiband cellphone
   (b) A working Konekt Barangay Promo SIM
   (c) Konekt Barangay Promo SIM prepaid load credit
2. The Konekt Barangay Promo SIM is a promo of GK Mabuhay and may not provide the same quality of service as the conventional Globe or TM Prepaid SIM. The Konekt Barangay Promo SIM will only work within a 300–500m radius from the location of the Konekt Barangay base station site. The Konekt Barangay Promo SIM allows you to:
   (a) Call to other mobile and landline numbers within the Philippines
   (b) Receive calls from other mobile numbers within the Philippines
   (c) Send texts (SMS) from any Globe or TM mobile number within the Philippines. Texts from other operators (non-Globe/TM) will not be received.
   (d) The Konekt Barangay Promo SIM will not work if you are outside the 500m radius from the base station
   (e) You cannot make or receive calls from outside the Philippines
   (f) You cannot avail any Globe/TM value added services and promotions such as ring-back tones, etc.
3. The Konekt Barangay Promo SIM is a promo that may start anytime from [START DATE] until [END DATE]. Globe reserves the right to terminate the promo and selling of the Konekt Barangay Promo SIM and load at any time and date after the promo period. Because this is a promo service, Globe may terminate the service at any time and Globe is not obliged to finish the promo period.
4. The Konekt Barangay Promo SIM can be bought at its suggested retail price (SRP) of Php 15.00 from your local retailer.
5. The use of the services of the Konekt Barangay Promo SIM has a corresponding cost. Globe may change the service rates, with the approval of the National Telecommunications Commission (NTC), without any notifications to its subscribers.
6. The Konekt Barangay Promo SIM has its own retails and load sellers. Prepaid load for the Konekt Barangay Promo SIM is not available from the Globe/TM loading outlets. You cannot also buy prepaid load using Globe Autoload Max.

Network Interaction Type	Tariff (PHP)
Call from a Konekt number to another Konekt number	1.00/minute
Call from a Konekt number to a long-distance on-network number	3.00/minute
Call from a Konekt number to an long-distance off-network number	5.50/minute
Text from Konekt number to Konekt number	0.25/message
Text from Konekt number to long-distance on-network number	0.50/message
Text from Konekt number to long-distance off-network number	1.00/message
All incoming calls	FREE
Incoming text messages (on-network local and long-distance)	FREE
Incoming text messages (off-network)	NOT ALLOWED

7. It is the responsibility of the SIM and phone owner to protect his/her own mobile phone, SIM and its corresponding PIN/PUK from load theft and unauthorized load usage. All calls, SMS and other transactions will be charged accordingly.

8. You shall not use the Konekt Barangay SIM or Load for any unlawful, fraudulent, elicit or abusive purpose. The Konekt distributor or partner may terminate the service of any subscriber who shall exhibit abusive usage.

9. The Konekt Barangay load of any subscriber who deliberately violates the terms and conditions stated in this document will be voided. The Konekt partner will not return any value or load to the offending subscriber.

10. Globe Telecom reserves the right to amend these terms and conditions at any time, with or without prior notice.

11. The Konekt Barangay Promo SIM is owned by Globe and uses new technology to provide cellular access to previously unserved sites.

12. By using the Konekt Barangay Promo SIM, the following information will be stored in a cloud server owned by a third-party:

    (a) Konekt barangay number and all called or texted numbers

    (b) Details of all calls, SMS, and data made/used through the Konekt Barangay Promo SIM

    (c) Details of sending/receiving load

13. You will continue to enjoy your Konekt SIM card by ensuring that your SIM card is loaded. The Konekt SIM card will expire if you fail to reload within 60 days from the initial activation and use. The Konekt SIM card will also expire if you fail to load within 120 days after the date you had a zero account balance. Once a SIM card expires, it is automatically deactivated and the phone number cannot be used anymore.

14. Since the network and services of the Konekt Promo service is powered by experimental technology, Globe cannot ensure 24/7 operation and fast network restoration in the event of typhoons, rains, earthquakes and other acts of God.

15. Globe will not be responsible for any service disruptions caused by any untoward incidents like typhoons, earthquakes, fire, terrorism, or force majeure.

Subscriber:
Name, Signature, Date

Appendix C

Chapter 3 Additional Materials

Market good	# observations	# locations	Units	Min	Max	Mean	SD
Banana	819	201	piece	0.33	25	10.47	3.73
Beef Brisket (fresh, raw)	317	102	piece	5	200	42.42	37.30
Beer	718	177	cl	0	1833.33	11.56	74.65
Bitter Ball	476	130	kg	0	475	74.28	53.98
Bitter Kola	823	192	piece	0.91	35	13.96	4.7
Boiled Eggs	891	206	piece	1	25	13.24	2.54
Bread	1002	202	piece	5	500	72.35	48.49
Bulgur Wheat	830	168	kg	0.03	3187.5	49.91	116.3
Butter Rice	886	183	kg	0.68	7437.5	85.2	353.04
Cassava	749	165	piece	0.11	100	15.12	9.12
Cassava Flour	979	203	kg	0	4250	57.45	228.58
Cassava Leaf	377	109	piece	8.33	100	18.90	14.07
Charcoal	807	193	kg	2.4	1190	17.25	59.93
Chloride	842	199	liter	0	2000	152	88.76
Fan-fan Rice	839	170	kg	1	4250	68.77	147.18
Fuel (Diesel)	569	164	liter	10.57	1733.33	100.95	111.06
Instant Coffee	518	154	gram	0.01	10	2.72	2.63
Kidney Beans (dry)	805	175	kg	0.18	15937.5	203.45	869.62
Live Chicken (medium size)	361	81	piece	0	1300	615.95	244.02
Mayonnaise	613	168	ml	0.01	739.34	35.02	96.95
Onion	1098	214	piece	0.4	125	21.47	20.59
Orange	852	193	piece	2.5	125	16.57	11.29
Palm Butter	711	158	kg	3	2500	35.37	149.35
Palm Oil	997	198	liter	9.51	1800	110.71	80.40
Petrol (gas)	682	175	gallon	0.31	710	252.43	111.52
Plantain (Cooking Banana)	900	198	piece	0	250	20.37	11.26
Potato Greens	608	141	piece	0.5	35	14.2	4.74
Powdered Milk	846	197	gram	0	70	0.86	2.78
Salt	918	190	gram	0	4.25	0.09	0.23
Sardines (canned)	332	99	gram	0.04	4.25	1.55	1.53
Seasoning cube	434	108	gram	0	127.5	0.72	6.11
Smoked Fish	592	162	piece	0.52	550	70.24	81.93
Sugar	967	198	gram	0	8.75	0.12	0.54
Tomato	564	132	piece	0.08	500	22.07	33.62
USA Parboiled Rice	981	205	kg	5	6250	92.21	279.91
Vegetable Oil	1012	197	liter	1.35	580	130.4	41.09
Water (Bag)	1035	217	ml	0	5.71	0.02	0.18

Table A1: **Summary statistics of Premise contributor data.** Data collection in Liberia began in October, 2014, with products and market locations being gradually added since then. Significant variation between products in data quality and quantity.