Title
SMS Surveys of Selected Expenditures

Authors
Lang, Megan
Ligon, Ethan

Publication Date
2022-05-09

Evaluating the efficacy of SMS surveys for measuring household welfare and firm performance

Megan Lang* and Ethan Ligon†

May 10, 2022

Abstract

High-frequency measures of economic well-being can allow policymakers and researchers to understand and quickly respond to dynamic problems, but collecting such data is expensive. Can short message service (SMS) surveys enable researchers and policymakers to measure household welfare and firm performance at a high frequency in low-income countries? We detail the implementation of two SMS surveys and evaluate their efficacy for gathering high-frequency data. One measures consumption expenditures in Rwanda and the other measures microenterprise revenues in Uganda. We successfully calculate a measure of household welfare for households that respond to the SMS survey in Rwanda and track changes in revenues over time for microenterprises in Uganda. Our SMS surveys are substantially less costly than equivalent in-person surveys; however, nonresponse is a significant problem. We propose combining SMS surveys with in-person data collection to compute weights that correct for nonresponse bias, then evaluate the performance of our method using the revenues data from Uganda.
Keywords: Panel data, text message survey, nonresponse bias, weighting

*King Climate Action Initiative and University of California Santa Barbara. Correspondence may be sent to meganlang@ucsb.edu.

†University of California Berkeley.
This study has approval from the Rwandan National Ethics Committee (No. 0018/RNEC/2019), the Ugandan National Ethics Committee (REF 0207-2018), and IRB approval in the U.S. (U.C. Berkeley CPHS 2018-02-10770 and U.C. Berkeley CPHS 2018-04-10959).

1 Introduction

Household consumption expenditures are generally the preferred measure of household welfare in low income countries (Deaton 1997). Similarly, data on firm performance provides a valuable snapshot of economic conditions. However, surveying households and firms is costly (Scott, Steele, and Temesgen 2005). These measures are important for a range of academic and policy purposes: setting and adjusting the consumer price index (Atkin, Faber, and Gonzalez-Navarro 2018), evaluating impacts of poverty alleviation programs Banerjee1260799,givedirectly, and tracking economic growth (Deaton 2005; Chen and Ravallion 2010; Young 2012) to name just a few. For this reason, we typically only observe them when governments conduct nationally representative surveys or when researchers conduct independent data collection. Such infrequent measures limit our ability to observe dynamics over time or to precisely identify changes caused by new policies that do not vary across a population.

New technologies proliferating at a relatively low cost in low-income countries raise the exciting possibility of collecting data at a high frequency, as they allow for surveys to be implemented at a fraction of the cost of in-person surveys. However, there are important trade-offs between surveys implemented in-person versus over a phone or a computer. Instruments need to be short to ensure completion and response rates are often lower than they are for in-person surveys, raising concern about selection bias.

We use two simple short message service (SMS) surveys to assess the feasibility of using SMS-based tools to measure household welfare and firm performance. The first survey asks about consumption expenditures and basic household demographics among low-income households in rural Rwanda. The second survey asks about firm revenues among women-owned microenterprises in central Uganda. Both surveys take place alongside randomized control trials (RCTs), allowing us to describe selection in terms of respondent characteristics that we observe in either administrative data, phone surveys, or in-person surveys.

We find promising results in terms of individual responses. (Ligon 2019) shows that we can estimate a measure of household utility using household expenditures on a subset of goods, meaning that we do not require the complete demand system to estimate welfare. Our first survey shows that we can estimate the (Ligon 2019) measure of welfare using data on disaggregated consumption expenditures on eighteen goods and a four-question survey about household composition. Our second survey of microenterprise revenues demonstrates that we can successfully capture high-frequency variation in revenues over time.

While our results are promising from the perspective of the data collected on individual households and businesses, we also document significant limitations stemming from low response rates and selection bias that make data collected through SMS surveys an unreliable method for measuring population characteristics. Response rates for the household consumption expenditures survey are only around 25%. Response rates for the simpler, single question survey on microenterprise revenues are higher but still fall far short of response rates for in-person surveys at 40%–56%, depending on the month. In Rwanda, households who appear to be wealthier respond at higher rates. In Uganda, response patterns are less clear: older women with larger social networks tend to respond more often, but single women appear to respond more often than those who have ever been married.

We propose a method that leverages in-person surveys to correct for selection bias, thereby allowing for high-frequency measures of population-level outcomes. We use in-person data collected around the same time as SMS data to compute weights for each SMS responder. Applying the weights to all SMS data collected after the in-person survey allows us to adjust for selection bias. We use two rounds of in-person data collection in Uganda to test the performance of our method. The weights we calculate using the first round of in-person data collection do not allow us to perfectly recover the distribution of revenues in the second round of in-person data collection, but do substantially improve upon the unweighted SMS data. We conclude by highlighting areas that could benefit from additional research.

2 Methodology

In Rwanda, we construct the welfare measure derived by Ligon (2019) using data elicited using a series of SMS surveys. To construct the welfare measure, we require disaggregated consumption expenditures for a subset of goods and basic information on household composition: the number of men, women, boys, and girls who regularly eat and sleep in the household. In Uganda, we simply want to elicit sales revenues. In both settings we send all surveys using only SMS technology, so respondents do not need a smart phone to participate.

2.1 Household consumption expenditures survey - Rwanda

In Rwanda we implement two surveys: one on consumption expenditures over the past week, and one on household composition that we only administer once. The questions on household composition ask how many men, women, girls, and boys regularly eat and sleep in the respondent's household. The questions on consumption expenditures ask how much, in Rwandan francs (RWF), the respondent spent on eighteen different goods between the previous Monday and Sunday (the past week in terms of calendar days, not a moving window of 7 days). We randomly assign respondents to receive one of four different surveys, with the order of the goods randomized between surveys.

Individuals receive a notice the day before the survey starts informing them when the survey will begin. We send this notice on Sunday evenings to minimize the recall time required for questions about the previous week's expenditures. The message informs respondents that they will be receiving a survey the following evening, that data are private and will only be used for research purposes, that responses are free, and that they will receive RWF 200 in airtime for completing the survey.

The next day, we send a notice that the survey is about to start and that responses can earn RWF 200 in airtime. Shortly after, we send the first question. Conditional on responding to the first question, the respondent receives a second question. If a respondent does not answer the first evening they receive the survey, they get a reminder the following morning. Individuals who respond to less than six questions do not receive RWF 200 in airtime.

Surveys expire after 24 hours to allow the next set of questions to start. We send another introductory SMS message the following day before starting the next set of six questions. We repeat the same procedure on the third day. A respondent who answered all three surveys in a given week could earn up to RWF 600 in airtime. We only send the household composition survey once, at the end of the survey period, and respondents can earn RWF 200 for answering four questions about the individuals living in their household.

We select the goods to include in our survey using data from the most recent Integrated Household Living Survey (EICV) in Rwanda, a nationally representative survey (National Institute of Statistics of Rwanda, December 2017). Using the cross-sectional data enables us to identify the goods that vary most with income, which is in part a function of the expenditure shares going to each good. While not strictly necessary, choosing goods in this way has two benefits. First, it increases the likelihood that households are spending money on the subset of goods that we select. Second, choosing the goods that are the most income elastic puts us in the best position to measure changes in welfare.

We conduct the SMS surveys among individuals in Rwanda who were part of a randomized control trial conducted in partnership with a solar company. To ensure that responses are free, we limit to individuals with phone numbers on MTN, the largest network provider in Rwanda. This brings our total sample size to 1,915 individuals.

In designing the survey, we have control over a range of relevant attributes: whether to require respondents to opt in to begin a survey, the initial invitation message, the number of questions included in each survey, the size of the airtime incentive for survey completion, the response window, and initial sensitization about the survey. We performed a number of informal A/B tests with a subset of respondents before rolling out the survey to everyone and found that not requiring opt in and reminding individuals that responses were free of charge were the only changes that substantially increased response rates. Interestingly, even offering very high airtime incentives did little to change response rates.

Individuals learned about the SMS surveys when they got a sensitization call about the RCT in June and surveys started in September. We collected data on the following dates, starting in 2018 and continuing to 2019: September 3, November 26, December 4, December 10, January 15, January 21, and January 28.

2.2 Microenterprise revenues survey - Uganda

In Uganda our survey is much simpler: we ask respondents to report their total sales revenues the day before the survey. If respondents answer, they receive UGX 1000 in airtime. We randomize respondents to receive the survey on one day of the week falling from Tuesday through Saturday, with the day of the week remaining constant for each respondent over time to limit confusion. Respondents receive the survey question directly, with no introductory message, although the survey message does remind respondents of the completion incentive. Since we only ask one question in Uganda, we leave the survey open for 72 hours and send a reminder if a woman does not respond within the first 48 hours. At the end of each month, we compile a list of all women who have not responded at least once over in the past month. An enumerator calls all non-responders and records their sales the previous day directly in the SMS platform.[1]

Our sample of respondents in Uganda is 940 women participating in a randomized control trial designed to evaluate a business training program. Women self select into study participation, so the population of interest is women who are either aspiring business owners or business owners eager to improve or grow their business. Unlike in Rwanda, responses are not free for respondents: they pay the normal price of sending a SMS message (roughly $0.01). Individuals first learn about the SMS surveys when they sign up for the study. Enumerators remind them about the SMS surveys during the an in-person baseline survey, and again during an initial meeting where all women learn their treatment status in the study. We always emphasize that even if a woman does not have a business or did not have any sales the day before she receives the survey, she can reply with zero and still receive the incentive.

We recruit the women in the study in Uganda in five different locations, with implementation of the study and the business training staggered between locations. For this reason, we present all results in event time, where event month zero corresponds to the month when we conducted baseline and began sending SMS surveys in a given location. We conduct in-person midline surveys six months after baseline and in-person endline surveys 18–24 months after baseline.[2] The weekly SMS surveys begin just after baseline and continue through the endline survey.

[1] We include the cost of these follow-up calls in our cost calculations for the revenue surveys in Uganda.

[2] The variable timing of endline surveys is due to COVID lockdowns that delayed data collection in some locations.

3 Results

3.1 Survey Response Patterns

In Rwanda, a total of 477 individuals ever respond to one of our surveys: 24.9%. Conditional on ever responding, respondents reply to an average of 4.2 (of 6) surveys with a median of five surveys. Of the eighteen goods we include, we receive responses about an average of 11.8. Conditional on starting the survey, completing the entire survey is the modal response.[3] There are spikes in survey completion at six and twelve, as expected given the six-question implementation structure. Therefore, there are two important features of response patterns. First, respondents who reply to at least one survey are highly likely to reply in most survey rounds. Second, item nonresponse is low relative to selection into not responding at all. Response rates are very similar across all goods, including luxury or temptation goods such as banana beer, ubushera, snacks, and sweets.[4]

Next we consider how many questions respondents answer on average across our survey rounds. We again see small spikes at six and twelve questions, but we also see that many consumers average three or fewer questions.[5] This indicates that we potentially lose a number of responses as a result of the six question implementation structure: respondents complete only one or two of the three sets of six questions, leading to incomplete information for a number of respondents in a given round of surveys. It could also indicate that some respondents start surveys and fail to complete them, which appears particularly common

[3]See Appendix Figure A1 for the distribution of responses across all consumption expenditure surveys.

[4]Appendix Figure A3 plots response rates by good.

[5]See Appendix Figure A2 for the average number of goods each respondent provide data on, conditional on responding to at least one survey.

during the first set of six questions.

Turning to the results of the household composition survey, we observe that the median household size is six and the mean is 6.6.[6] The median household contains one man, one woman, 1–2 boys, and 1–2 girls. Unfortunately, only 57% of respondents who responded to at least one consumption expenditure survey also responded to the household composition survey, and 13% of respondents only responded to the household composition survey but never responded to a consumption expenditure survey.

The top panel of Figure 1 shows median weekly expenditures on each good. Fish and meat make up the largest expenditure category by far, with median weekly expenditures on fish and meat at RWF 2,000. These expenditure patterns match what we would predict, providing some reassurance that the quality of data we receive from individual respondents is not substantially worse than it would be for an equivalent in-person survey. For the subset of consumers for whom we have data on household composition, we can estimate the distribution of the measure of welfare derived in Ligon (2019), denoted $\log \lambda$. The bottom panel of Figure 1 shows that there is variation in the distribution across survey rounds, suggesting that even eliciting expenditures on a small set of goods is sufficient to capture short-term changes in welfare, albeit for a self-selected sample of consumers. This is useful validation of the survey design, as it shows that asking about a small number of goods is sufficient to construct meaningful welfare measures, at least for individual consumers.

Response rates for our firm revenues survey in Uganda are significantly higher. Of the

[6] Appendix Figure A4 shows the distribution of household size for respondents who completed the household composition survey.

940 women in our sample, 85% respond to at least one survey. However, as the top panel of Figure 2 shows, monthly response rates range from 40%–56% after the first month of surveys. Higher response rates in our Uganda revenue survey are unsurprising given key implementation differences: respondents only need to answer a single question, we allow for a longer survey response window, and we explain how to respond to the SMS survey multiple times in-person. The bottom panel of Figure 2 shows corresponding mean daily sales revenues each month, taken from first computing individual means for each month and then computing overall means among the women who responded at least once in a given month. As expected given the experiment running alongside the revenues SMS survey and negative shocks from COVID lockdowns, revenues fluctuate over the course of the experiment. We clearly capture this dynamic information with our SMS data, again providing some reassurance that the individual data points we collect via SMS surveys contain useful information.

3.2 Selection Bias

Next, we examine respondent characteristics to identify correlates with SMS survey response. In Rwanda, our SMS survey took place as part of an RCT with solar customers and so we have a considerable amount of administrative data on these households as well as an independently conducted phone survey. In Uganda, we lack administrative data, but use data collected from an in-person baseline survey.

In Rwanda, the treatment in the RCT involved many SMS reminders, raising the concern that treated respondents may be less attentive to SMS messages overall. In Table 1, we check for selection based on treatment status. We also create bins of pre-experimental solar demand based on the proportion of time respondents purchase access to solar to flexibly test for selection based on respondents' level of engagement with the solar company. These bins are also useful as rough proxies for the proportion of time a respondent's phone may be charged. We include the price each respondent pays for a day of solar access because higher prices correspond to higher appliance ownership, providing one proxy for wealth. Finally, we use data from a phone survey conducted after the SMS surveys ended in March, 2019 to test for the effect of high quality housing materials: clay tile roofs, cement floors, and brick walls (providinng a second wealth proxy).

We find no evidence that treatment status is significantly correlated with the probability of responding to at least one SMS consumption expenditure survey in Rwanda. As Table 1 shows, we also see no significant differences in response rates based on pre-experimental demand for solar. However, we do find evidence that wealthier households are more likely to respond to the survey. Respondents with higher solar prices, meaning those who have chosen more expensive systems with more appliances, are significantly more likely to respond to the survey than those with lower prices. Similarly, we find that respondents with cement floors and brick walls are nearly 6pp more likely to respond to the survey than those whose houses are built from lower quality materials. Taken together, the correlates of selection in Table 1 raise the concern that our SMS surveys in Rwanda may be biased toward wealthier respondents.

In Uganda we have a richer set of covariates that we can use to study selection patterns.

Given the larger number of survey rounds in Uganda, we test for correlates with the number of months respondents reply to at least one survey. As in Rwanda, we test for correlations between treatment status and selection. We also test for correlations between selection and a number of baseline covariates: age, marital status, respondent's education level, parental education levels, employment status, household size, number of dependent children, business ownership, size of the social network, and business profits.

Table 2 shows that unlike in Rwanda, our measures of economic well-being in Uganda are not strongly correlated with response rates. Business ownership is positively correlated with response rates but profits are negatively correlated, and both are only significant at the 10% level. We do not find any significant correlation with treatment status. We do find that age is significantly positive correlated with response rates and that single women respond more than ever married women. Women with larger social networks also tend to respond more, perhaps due to a greater familiarity with SMS messaging or a higher value for the airtime incentive. Taken together, the results in Table 2 do not raise the same concerns about bias towards wealthier respondents that we see in Rwanda, but there is still clear evidence that the sample of women who selects into responding is not representative of the sample of our in-person survey.

So far we have shown that it is possible to estimate household welfare and to measure firm revenues from a series of incentivized SMS surveys. This raises the possibility of obtaining substantially higher-frequency measures of poverty than is currently feasible using in-person surveys. In the next two subsections, we consider the two most significant trade-offs between our SMS surveys and equivalent in-person surveys: their relative costs, and differences in selection bias induced by the significantly higher nonresponse rates we found when conducting SMS surveys.

3.3 Cost Analysis

While we have detailed data on the costs of in-person surveys in Uganda, we did not conduct our own in-person surveys in Rwanda. The Rwandan government periodically conducts a nationally representative survey of all households to measure poverty and living conditions, in partnership with international organizations like the World Bank and the African Development Fund (National Institute of Statistics of Rwanda, December 2017). We use data on their reported survey costs to estimate the cost of in-person data collection in Rwanda and our own budget data to calculate the cost of equivalent in-person data collection in Uganda.[7]

In Rwanda, we consider the costs of a single in-person survey that elicits the same amount of information that ours does: one round of consumption expenditures plus basic household composition. At best, if nearly all survey costs vary with survey time, it would cost an average of USD 0.70 per respondent. More realistically, if only 75% of survey costs are a function of time spent actually surveying respondents, it would cost an average of USD 16.58 per respondent.[8]

How do these costs compare to our SMS surveys? Based on our work, we conservatively assume a 20% response rate. Setting aside concerns about selection bias, if we simply wanted

[7]See the appendix for a detailed discussion of the assumptions we make in our cost analysis.
[8]Appendix Table A1 shows cost comparisons assuming different proportions of costs that vary with survey time.

to achieve the same sample size as the EICV4, this would necessitate that we survey five times as many households: 72,095. Taken together, this implies an average cost per household that actually responds of $3.89, which is less than the cost of in-person surveys across nearly our entire range of cost estimates. Furthermore, nearly all costs associated with the SMS surveys scale with the number of people surveyed, or the number of survey rounds. The cost of collecting household expenditure data each month for a year using SMS surveys is $46.61 per respondent, or $3.89 per respondent per month. In-person surveys have to repeatedly incur many fixed costs every time enumerators conduct household visits, so average costs are unlikely to substantially decrease with frequent survey rounds.

Our SMS surveys in Uganda do not have reverse billing, so the cost to send unconditional SMS messages is only $0.01. Even if we assume that we have to send a reminder to every respondent every week and that we have to over-sample to compensate for low response rates, the average costs for each respondent are only $0.78 per month. Assuming that we do weekly surveys every month for a year, the average cost per respondent is only $7.97 compared to an estimated $123.60 for the equivalent in-person data collection.

Cost estimates from both Rwanda and Uganda make it clear that SMS surveys present a much less costly option for high frequency data collection than in-person surveys. As such, the critical question becomes whether we can collect sufficiently useful data using SMS surveys given concerns about selection bias. In the next section, we propose a specific use for high-frequency SMS data along with a method to correct for selection bias.

3.4 Correcting for Selection Bias in SMS Surveys

There is a large literature on selection bias in mail and internet surveys (see, for instance, Whitehead, Groothuis, and Blomquist 1993; Schonlau et al. 2009; Bethlehem 2010). The conclusions from this literature are generally discouraging: it is difficult to correct for selection bias even when the researcher can observe covariates of respondents and non-respondents, and selection bias typically has detrimental effects on the reliability of survey results. Our proposed use of SMS surveys is distinct from much of this literature: we are not suggesting that high-frequency SMS surveys replace more comprehensive in-person surveys. Instead, we propose that researchers and policymakers use SMS surveys to obtain high-frequency measures between costly in-person surveys. Doing so can allow us to better understand economic dynamics and, where possible, enable more nimble policy responses to changes in economic conditions for households or firms.

Our proposed strategy for combining in-person and SMS surveys leverages the panel nature of the data to generate weights that researchers can use to obtain high-frequency approximations of the distribution of economic measures in spite of the bias caused by nonresponse. Let y_{it} be the outcome of interest for household or firm i at time t.

Suppose that the researcher conducts (or can observe the results of) a representative in-person survey at baseline ($t = 0$) as well as a SMS survey at time $t = 0$. Then researcher can estimate the true density of y in the full sample at time zero, say $f_0(y)$, from the in-person survey. We also observe the selected density of y in the SMS survey at time zero, say $g_0(y)$. Then the researcher can assign weights to each SMS response using the ratio of these densities.[9] Then we can use data from SMS responders in period t to construct the density of y in the population at that period by calculating $f_t(y) = g_t(y)\frac{f_0(y)}{g_0(y)}$. This re-weighting allows us to estimate the population distribution of y at any period from the distribution of y just among SMS responders. The success of this procedure hinges on an assumption that the factors governing selection into the SMS sample are invariant over time.

Thus, if we assume that the outcome of interest evolves similarly for responders and non-responders, then we can simply re-implement the weights from baseline in subsequent rounds of SMS surveys and recover the true distribution of y_t. This enables us to leverage the panel nature of the data to adjust for selection bias in every survey round. It also exploits the features of the data we observed in our data collection in Rwanda: the modal response among those individuals who responded at least once was to respond to every survey we sent, indicating that we are unlikely to have a large degree of attrition among those individuals who first respond.[10]

The primary limitation of our approach is that any attrition that does occur may be a function of the outcome of interest. For instance, households that initially respond but then experience a large increase in welfare may become non-responders because the value of their time is higher than it was in $t = 0$. Conversely, households that experience a substantial decline in welfare may not regularly charge their mobile phones, responding less. Any changes in nonresponse that are a function of changes in the outcome of interest will bias estimates

[9]Note that this is the simplest case, which makes no assumption about other covariates that the researcher can observe from the in-person survey at baseline. However, the research could in principle use a range of covariates to predict selection into the SMS survey and to compute weights.

[10]In fact, researchers could further reduce survey costs by only sending SMS surveys in $t = 1, 2, \ldots, T$ to those individuals who respond in $t = 0$.

obtained using our proposed methodology.

We lack the data to test our proposed weighting strategy for the consumption expenditure surveys in Rwanda, but we have precisely the data required to test it using firm revenues in Uganda. As we show in Figure 2, response rates for the firm revenues SMS survey take a few months to stabilize. For this reason, we define weights using the in-person midline survey and the SMS surveys conducted around the same time. We then apply those weights to SMS data collected around the same time as the endline survey and compare the resulting distribution to the actual distribution of revenues at the endline survey.

The top panel of Figure 3 illustrates the initial part of the weighting process at midline. We take the distribution of revenues at the in-person midline survey as given, then use it to compute weights for the SMS surveys conducted around the same time. As the blue dots show, these weights allow us to recover the distribution of sales revenues from the in-person midline survey. The bottom panel of Figure 3 tests how well the weights we define using the midline survey perform at endline. The weights do not perform as well as they did when defined at midline. In particular, they lead us to underestimate the proportion of women in the lowest revenue bin by 6.3pp and to overestimate women in the second and third to lowest bins by 4pp–4.5pp. However, using the unweighted SMS data would nearly double the size of the error across the distribution of revenues, so the weights still offer an improvement. Whether the remaining errors are acceptable will depend on the specific application.

4 Conclusion

Our findings suggest that SMS surveys can elicit meaningful data from respondents but suffer from selection bias. Here, we summarize three key findings and propose steps for future work.

First, we document higher response rates in our survey on microenterprise revenues in Uganda relative to our survey of household consumption expenditures in Rwanda. Given that the size of the incentive was similar, there are a number of plausible explanations for these differences. The sales revenue survey was a single question rather than a series of questions. In addition to a single question being easier for respondents to answer, it also allowed us to extend the survey response period from 24 to 72 hours. We also had multiple in-person interactions with respondents in Uganda to sensitize them to the SMS survey, which we did not have the opportunity to do in Rwanda. In the future, it would be useful to undertake a more formal study of the technical survey design features for SMS surveys eliciting different types of information.

Second, we compare the costs of conducting equivalent SMS and in-person surveys. Unsurprisingly, SMS surveys are at least an order of magnitude less costly to perform than in-person surveys. Given that individual responses to SMS surveys appear to contain useful information, this suggests that SMS surveys are, at a minimum, a low cost tool in settings where researchers and policymakers need individual-level information rather than estimating population-level parameters.

Finally, we document differential selection patterns in different contexts and show that our proposed weighting method can only partially correct for selection bias. In Rwanda, it appears that respondents who (according to rough proxy measures) may be more wealthy are more likely to respond to our survey, but we find little evidence for a similar selection mechanism in Uganda. When we apply our proposed weights to data in Uganda, we come closer to recovering the true distribution of revenues at endline than we would using unweighted SMS data, but still document substantial errors. The deteriorating performance of the weights over time suggests that selection is a partially dynamic process, highlighting a key limitation to SMS surveys.

Beyond exploring techniques for increasing response rates, it would be useful to consider which types of outcomes are best suited to SMS surveys. For instance, an outcome like firm revenues varies significantly more than consumption expenditures, potentially leading to larger differences in selection over time. If a stable set of respondents answer consumption expenditure surveys, our weighting procedure may prove efficacious at consistently recovering the true distribution of expenditures even with low response rates. As more researchers gather SMS data alongside periodic in-person surveys, we will be able to better understand when we can correct for selection bias well enough to benefit from the low-cost data collection available through SMS surveys.

References

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro. 2018. Retail globalization and household welfare: evidence from mexico. *Journal of Political Economy* 126:1–73.

Bethlehem, Jelke. 2010. Selection bias in web surveys. *International Statistical Review* 78:161–188.

Chen, Shaohua, and Martin Ravallion. 2010. The developing world is poorer than we thought, but no less successful in the fight against povery. *Quarterly Journal of Economics* 125:1577–1625.

Deaton, Angus. 1997. *The analysis of household surveys.* The World Bank. https://doi.org/10.1596/0-8018-5254-4. eprint: https://elibrary.worldbank.org/doi/pdf/10.1596/0-8018-5254-4. https://elibrary.worldbank.org/doi/abs/10.1596/0-8018-5254-4.

———. 2005. Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics* 87:1–19.

Ligon, Ethan. 2019. Estimating household welfare from disaggregate expenditures. Unpublished manuscript. http://escholarship.org/uc/item/5gc4h1fm.

National Institute of Statistics of Rwanda. December 2017. Integrated household living conditions survey(2016-2017), version 0.1 of the public use dataset. provided by the National Data Archive. http://microdata.statistics.gov.rw.

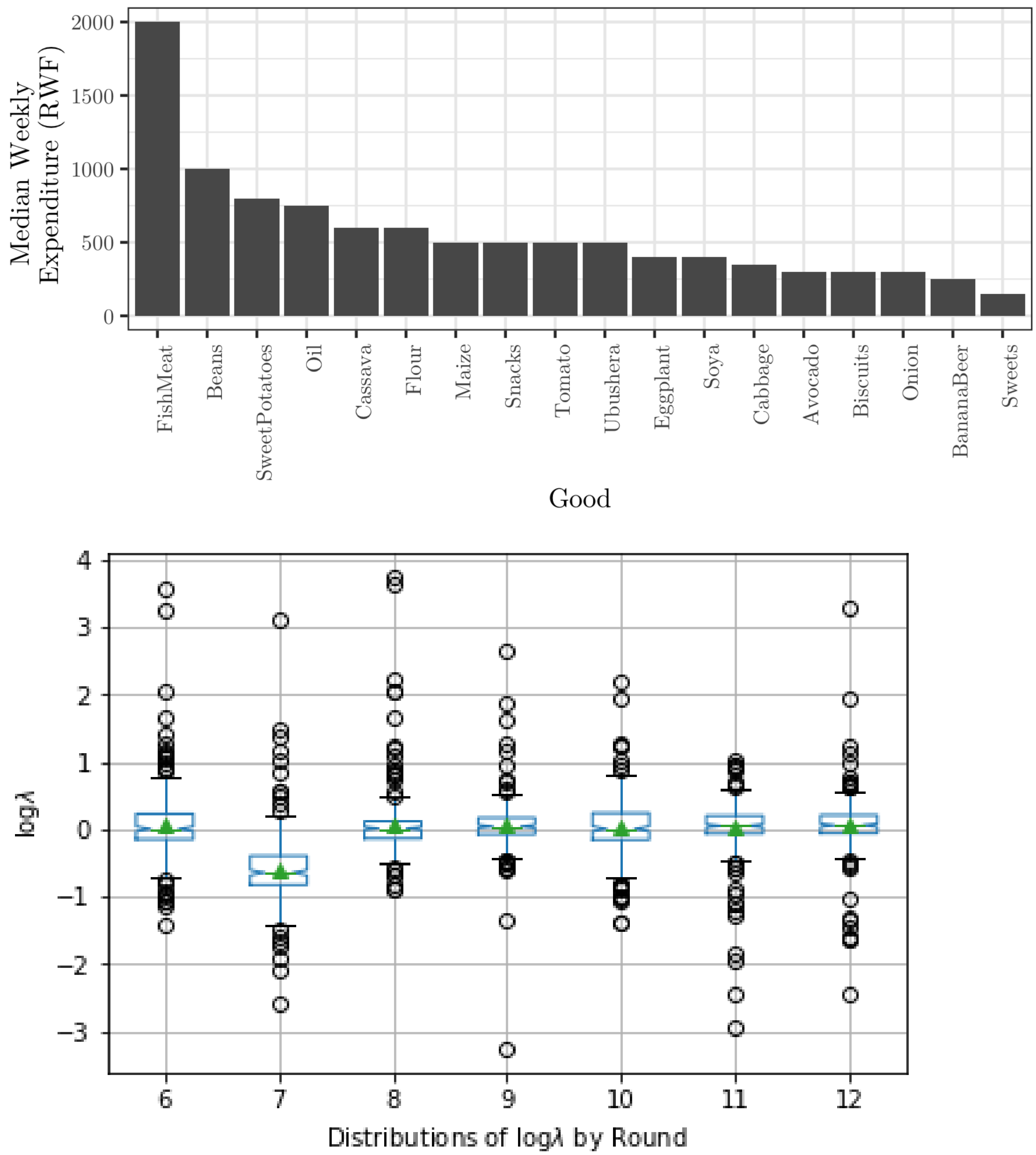
Schonlau, Matthias, Arthur van Soest, Arie Kapteyn, and Mick Couper. 2009. Selection bias in web surveys and the use of propensity scores. *Sociological Methods and Research* 37:291–318.

Scott, Kinnon, Diane Steele, and Tilahun Temesgen. 2005. Living standards measurement study surveys. In *Household sample surveys in developing and transition countries.*

Whitehead, John C., Peter A. Groothuis, and Glenn C. Blomquist. 1993. Testing for non-response and sample selection bias in contingent valuation: analysis of a combined phone/mail survey. *Economics Letters* 41:215–220.

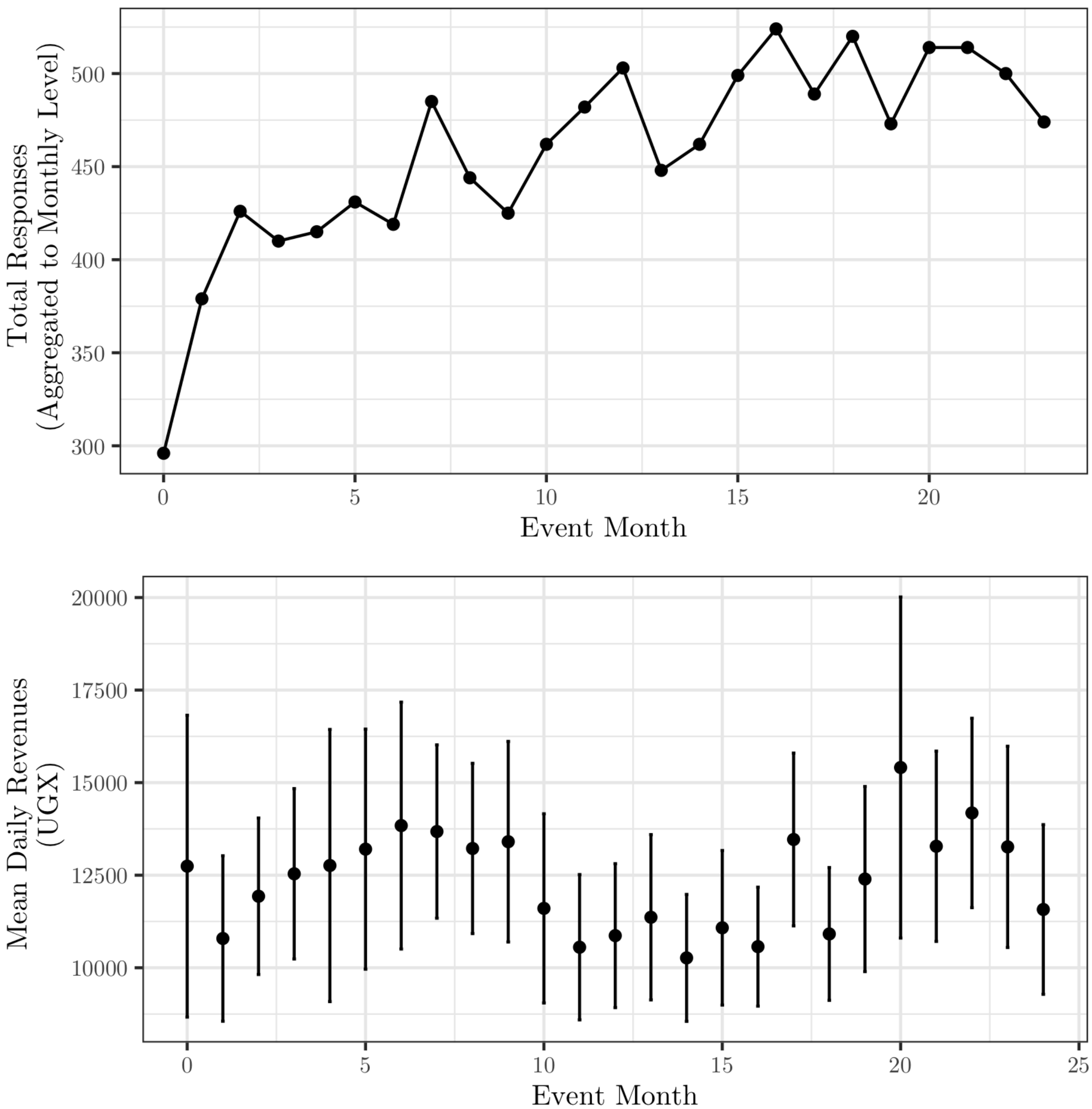
Young, Alwyn. 2012. The african growth miracle. *Journal of Political Economy* 120:696–739.

Figures and Tables



Note: Top panel shows median weekly expenditures for each of the 18 goods we ask about in Rwandan francs, across all rounds of the survey. Bottom panel shows the distributions of $\log(\lambda)$ for each survey round. Green triangles show means for each round. Blue notched boxes show medians with 95% confidence intervals. Blue bars show 1.5 times the interquartile range. Black circles show all observations that lie outside 1.5 times the interquartile range.

Figure 1: Median Expenditures and Distribution of $\log \lambda$: Rwanda Consumption Expenditure Surveys



Note: Top panel shows total number of responses for each month of the firm revenue surveys in Uganda. Note that response rates reflect phone calls that we make at the end of each month to all respondents who did not reply to at least one survey over the course of the previous month. Bottom panel shows mean daily revenues in Ugandan shillings for each month of the firm revenue surveys. We send SMS surveys on firm revenues weekly, average each respondent's answers within each month, then compute the mean among all respondents within a month. Bars are 95% confidence intervals.

Figure 2: Response Rates and Mean Revenues: Uganda Firm Revenue Surveys

Table 1: Selection into SMS Response: Rwanda Consumption Surveys

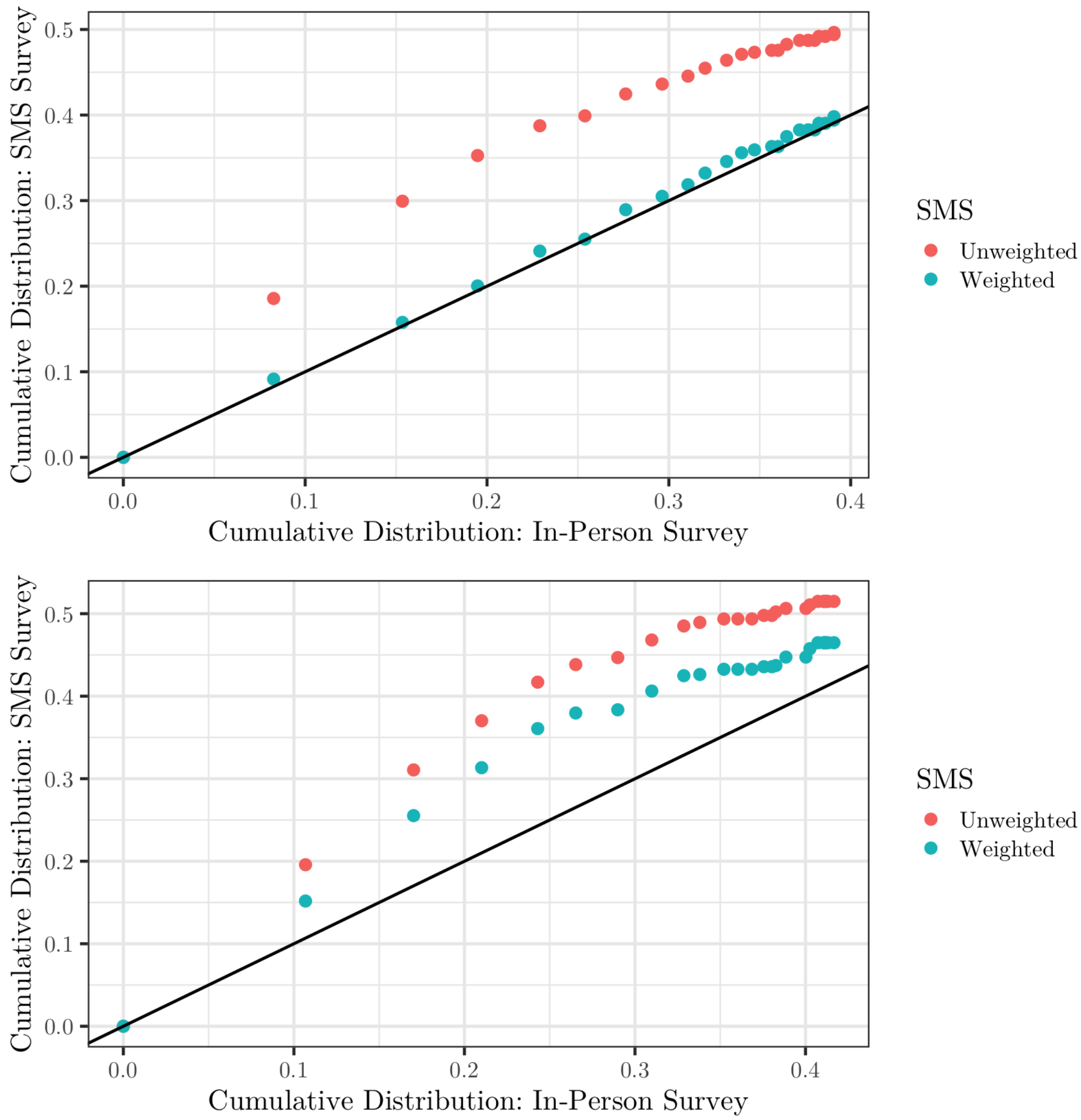
	Pr(Ever Respond)
Treated	-0.038
	(0.025)
Solar access: 95% +	-0.002
	(0.037)
Solar access: 85% - 95%	0.069
	(0.038)
Solar access: 75%-85%	0.025
	(0.039)
Solar access: 65% - 75%	-0.029
	(0.037)
Solar price	0.00033
	(0.00014)
Tile Roof	-0.042
	(0.026)
Cement Floor	0.06
	(0.026)
Brick Walls	0.058
	(0.028)
N	1433

Notes: Correlates of responding to a consumption expenditures survey at least once in Rwanda. Treated is a binary variable equal to one if the respondent was part of the treatment group in the solar RCT that the survey ran alongside. Solar access ranges refer to the proportion of days the respondent purchased solar access prior to the start of the study (the reference group are consumers who buy access less than 65% of the time). Solar price is the price the consumer pays for a day of solar access (higher prices correspond to owning more appliances). Tile roof is a binary variable equal to one if the respondent's roof is constructed with clay tiles rather than metal sheets. Cement floor is a binary variable equal to one if the floor is made of cement rather than earth. Brick walls is a binary variable equal to one if the walls are made of brick rather than mud or timber. We report White robust standard errors in parentheses.

Table 2: Selection into SMS Response: Uganda Revenue Surveys

	No. Months Responding
Treated	-0.004
	(0.582)
Age	0.049
	(0.025)
Married	-4.212
	(1.023)
Divorced	-2.466
	(1.166)
Widowed	-4.334
	(1.325)
Primary Ed	-0.04093
	(0.916)
Secondary Ed	1.08
	(0.93)
Father Primary Ed	-0.346
	(0.686)
Father - Secondary Ed	-0.176
	(0.799)
Mother Primary Ed	0.329
	(0.644)
Mother Secondary Ed	-0.231
	(0.865)
Employed	-0.713
	(0.632)
HH Size	-0.067
	(0.232)
No. Children	0.396
	(0.278)
Own Business	1.09
	(0.665)
Network Size	0.18
	(0.071)
Business Profits	-2e-06
	(1e-06)
N	767

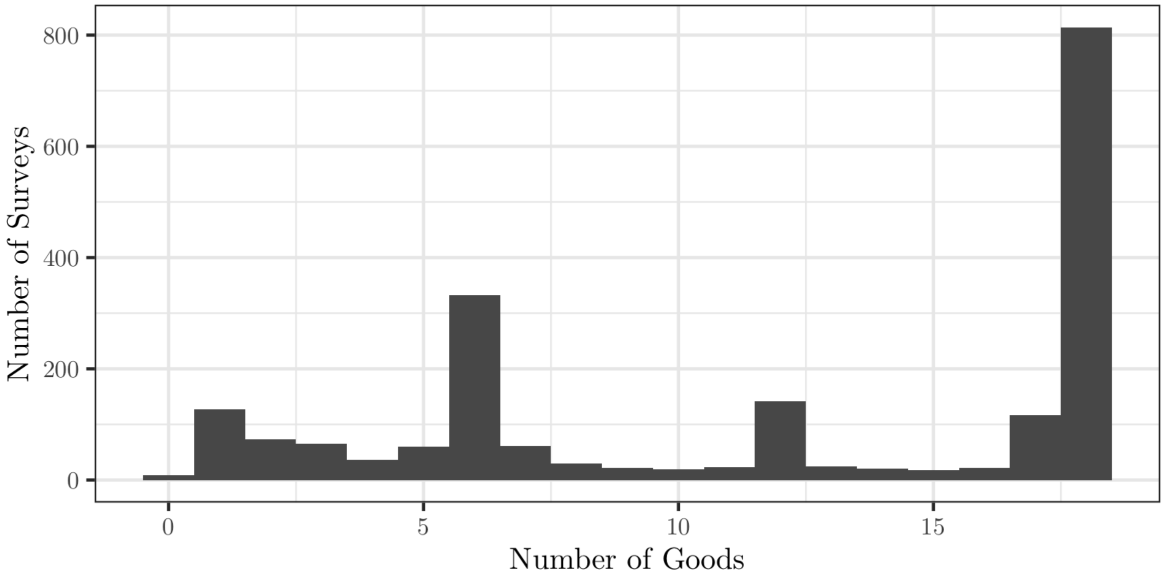
Notes: Correlates of the number of responses to SMS firm revenue surveys in Uganda. Treated is a binary variable equal to one if the respondent was part of the treatment group in the business training RCT that the survey ran alongside. Married, divorced, and widowed are binary variables equal to one if the respondent has the corresponding marital status at midline (the survey round we treat as "baseline" for the purpose of our analysis in this paper), with single women acting as the reference group. Primary Ed and Secondary Ed are binary variables equal to one if the respondent's highest level of education is primary or secondary, respectively, with less than primary education acting as the reference group. Father Primary Ed, Father Secondary Ed, Mother Primary Ed, and Mother Secondary Ed are analogous variables for the respondent's parents. Employed is a binary variable equal to one if the respondent is emplyed at midline. Household size is the total number of people who regularly



Note: The top figure compares the cumulative distribution of daily revenues in the in-person midline survey to the SMS daily revenues survey in Uganda. For the in-person survey, we compute daily revenues as the average of reported sales each day over the seven days prior to the survey. For the SMS survey, we take the average of all respondent's weekly answers for the month of the midline survey. Red dots show the comparison with the raw, unweighted SMS data and blue does compare to the weighted SMS data. The black line is a 45 degree line. The bottom panel shows the analogous comparison for endline data using weights computed using the midline data.

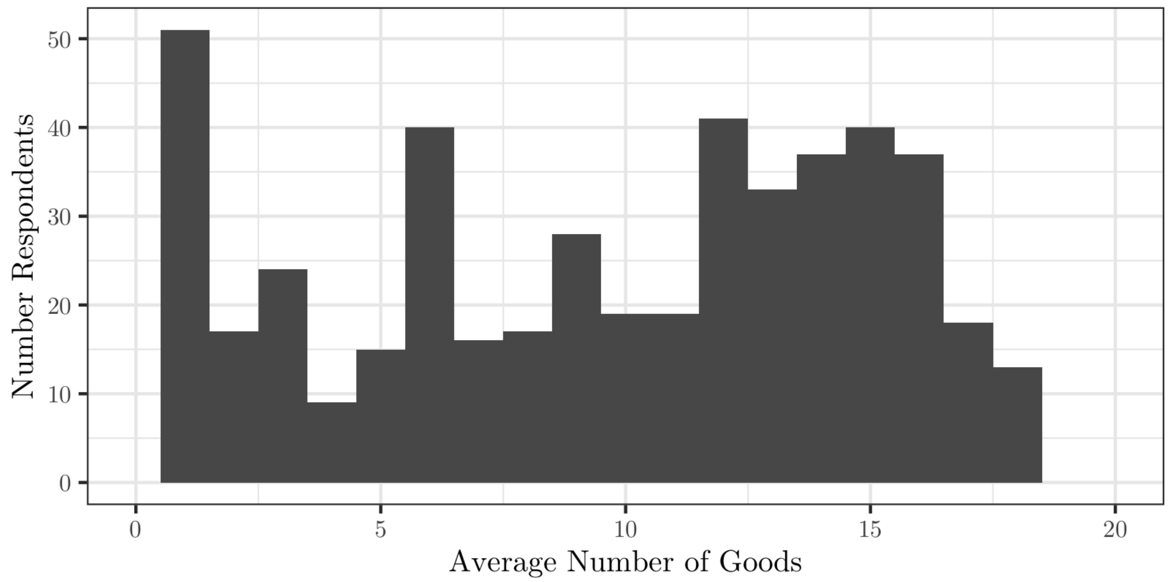
Figure 3: Comparing Distributions from In-person, SMS, and Weighted SMS surveys

A Appendix



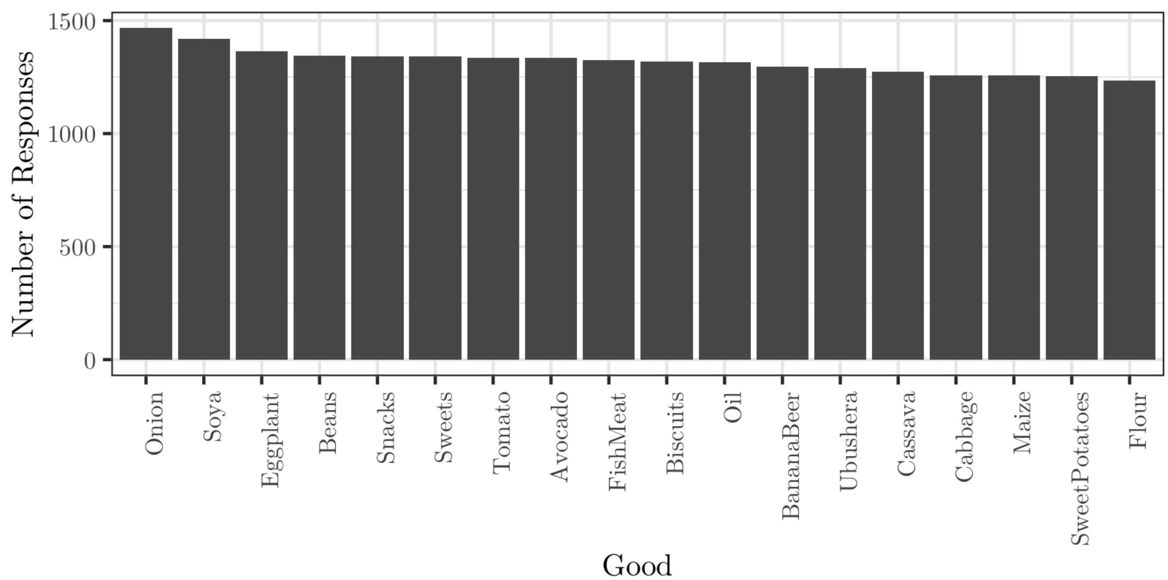
Note: Distribution of the number of consumption expenditure questions completed in each SMS survey. Each question asks about weekly expenditures on a different good. The full survey asks about a total of 18 goods. We send questions in sets of six and consumers receive a completion incentive for each set, which is why we observe spikes at 6, 12, and 18.

Figure A1: Distribution of Survey Completeness: Rwanda Consumption Expenditure Surveys



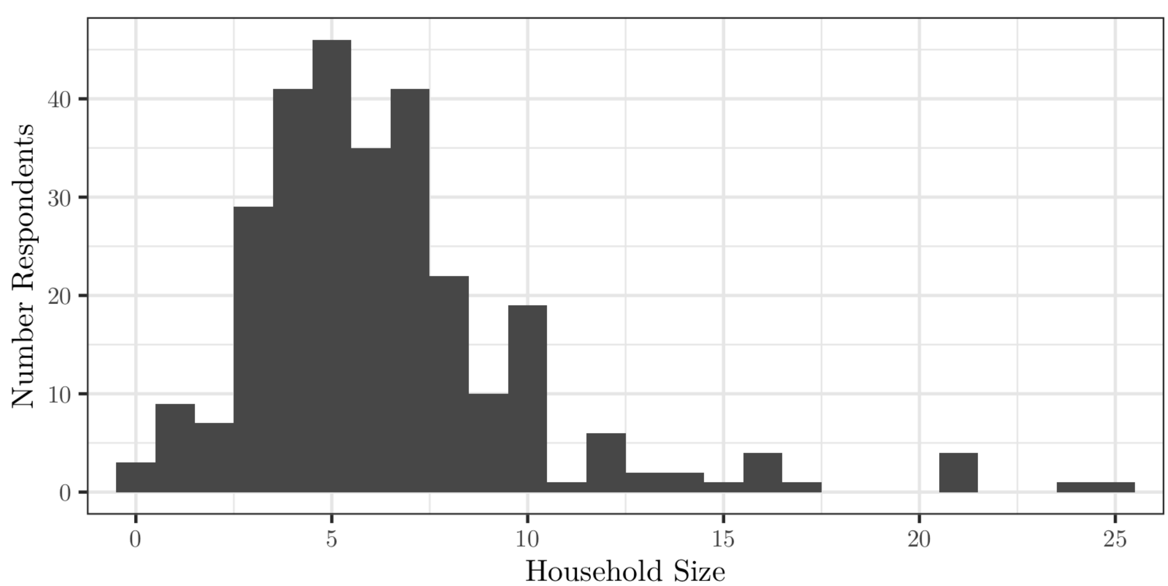
Note: Distribution of the average number of goods each respondent provides data on across all surveys they responded to over the study period. Each question asks about weekly expenditures on a different good. The full survey asks about a total of 18 goods. We send questions in sets of six and consumers receive a completion incentive for each set.

Figure A2: Average Survey Completion by Respondents: Rwanda Consumption Expenditure Surveys



Note: Total number of responses received for each of the 18 goods that we ask about in the survey. Note that we randomize the order of goods that we ask about in the survey.

Figure A3: Responses for Each Good: Rwanda Consumption Expenditure Surveys



Note: Distribution of household size, where a member of the household is defined as someone who regularly eats and sleeps in the household. We only send the survey on household composition once over the course of the study.

Figure A4: Distribution of Household Size: Rwanda Consumption Expenditure Surveys

A.1 Cost Comparison Assumptions

The African Development Fund's report on the fourth Integrated Household Living Conditions Survey (EICV4) in Rwanda estimated data collection costs at approximately USD 948,600.[11] EICV4 sampled 14,419 households, indicating an average survey cost of USD 65.79 per household. While it is difficult to correctly estimate the total survey time dedicated to the consumption expenditures module, we will conservatively estimate that it is only 4.3% of the survey.[12] The consumption expenditures module in the EICV contains 147 items, but households are visited 11 times to ask about all items. Therefore, the equivalent of our 18 consumption expenditure questions and our four household composition questions would amount to just 0.06% of the total EICV4 survey.

Finally, we need to make assumptions about the proportion of total data collection costs for the EICV4 that vary as a function of the actual time spent surveying respondents versus costs that are not dependent on respondent time (e.g., transport to the field site, accommodation for enumerators, supervisor salaries, etc.). Lacking a strong basis for such assumptions, we present a range of cost estimates for a survey equivalent to our SMS survey based on the proportion of costs that vary as a function of survey time with the respondent.

[11] The report states that data collection costs were UA 620,000 and that the currency conversion is 1UA = 1.53 USD

[12] We arrive at 4.3% in the following manner. There are a total of 77 pages in the EICV4. Of those, 8 pages do not require any data entry. 9 pages are dedicated to the consumption expenditure module. However, the EICV asks three questions about each good: how many months the household purchased the good out of the last 12, how much the household spent on the good since the previous visit, and where the household buys the good most often. Therefore, we estimate that the proportion of the survey dedicated to the equivalent questions that we ask is only 3 pages.

Assuming a single round of SMS household consumption expenditure and household composition surveys in Rwanda, we face a fixed cost of USD 65 to set up a system to allow for reverse billed SMS messages as well as a USD 60 subscription fee for the platform used to send the SMS messages. We send a total of 11 SMS messages that are not conditioned on receiving a response and an additional 41 conditional messages. Respondents who complete each sub-survey receive an incentive of USD 0.25, and there are a total of four sub-surveys. Each SMS message costs USD 0.03.

We use our detailed budget data from Uganda to calculate the cost of eliciting the same amount of data in an in-person survey as we elicit through our SMS revenue surveys. We estimate that the average cost to complete the full in-person survey is $29.48 per respondent and that (conservatively) eliciting sales revenues only comprises 0.8% of the survey. We calculate that only around 35% of survey costs vary with survey time, so the average cost per household to perform the in-person equivalent of our SMS surveys is around $10.30 for a single round of data collection. Even if we were to replace weekly SMS surveys with monthly in-person surveys, the total cost for a year would be $123.60.

Table A1: Cost for Equivalent In-Person Survey: Rwanda

Proportion of Costs that Vary with Survey Time	Average Cost Per Sampled Respondent (USD)
1%	\$65.13
5%	\$62.50
10%	\$59.21
25%	\$49.35
50%	\$32.91
75%	\$16.47
90%	\$6.61
95%	\$3.33
99%	\$0.70

Note: The average cost per SMS respondent is \$3.89 in Rwanda. We derive estimates of costs for in-person surveys from The African Development Fund's report on the fourth Integrated Household Living Conditions Survey (EICV4).