

Report No: ACS14146

Somalia Poverty TA Program

Informing the Somali High Frequency Survey

June 2015

GPVDR
AFRICA



Standard Disclaimer:

This volume is a product of the staff of the International Bank for Reconstruction and Development/ The World Bank. The findings, interpretations, and conclusions expressed in this paper do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Copyright Statement:

The material in this publication is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The International Bank for Reconstruction and Development/ The World Bank encourages dissemination of its work and will normally grant permission to reproduce portions of the work promptly.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA, telephone 978-750-8400, fax 978-750-4470, <http://www.copyright.com/>.

All other queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA, fax 202-522-2422, e-mail pubrights@worldbank.org.

This report contains two working papers and one note. The working paper “Second Stage Sampling for Conflict Areas: Methods and Implications” was prepared by Kristen Himelein, Stephanie Eckman, Siobhan Murray and Johannes Bauer. The working paper “Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu High Frequency Survey” as well as the note “Utilizing Mobile Technology to Innovate CAPI Data Collection in Fragile Contexts” are authored by Utz Pape and Johan Mistiaen. The activity was led by Johan Mistiaen (GPVDR) and Utz Pape (GPVDR). The team is grateful for inputs and comments from Kathleen Beegle, Andrew Dabalen, Matthieu Dillais, Geoff Handley, Peter Lanjouw, Alvin Etang Ndip, Roy Van der Weide, Nobou Yoshida and Paolo Zacchia.

Vice President
Country Director
Senior Director
Practice Manager
Task Team Leaders

Makhtar Diop
Bella Bird
Ana Revenga
Pablo Fajnzylber
Johan Mistiaen, Utz Pape

Informing the Somali High Frequency Survey

16th June 2015

Overview

The historical civil war and political insecurity in Somalia has resulted in a lack of socioeconomic, perception and other key data in Somalia. The Somalia Socioeconomic Survey 2002 was the last Somalia-wide representative survey. This lack of data makes it difficult for the government and its development partners to plan and implement appropriate policies and programs that are needed to support economic growth and stability. Especially the lack of poverty numbers undermines the development of an interim poverty reduction strategy paper, which is required to apply for HIPC debt relief.

World Bank aims to support Somalia by implementing a Somalia High Frequency Survey with a focus on consumption for the first wave to estimate poverty numbers. However, data collection in Somalia is challenging due to insecurity in some areas. First, traditional sampling methodologies require a full listing of enumeration areas, which is impossible in insecure areas. Second, face-to-face time is limited to about 60 minutes while a full consumption questionnaire takes 90 to 120 minutes. Third, limited field access makes monitoring of data quality difficult. The poverty team developed solutions to overcome these challenges and allow household consumption data collection in Somalia.

The first challenge is resolved by employing a segmentation approach instead of requiring a full listing in insecure areas. Each enumeration area is segmented into smaller parts. Instead of listing all households in the enumeration areas, one segment is randomly chosen and all structures within the segment are listed. After choosing a structure randomly, only the households within the structure are listed and used to randomly select a household. The first poverty note shows that this methodologies delivers unbiased estimates with reasonable precision. In addition, the methodology is a good compromise between preparation time, ease of implementation, and the time and complexity necessary for weight calculations.

The second challenge is overcome by a newly developed methodology to collect consumption data in 60 minutes described in the second poverty note. First, the large number of consumption items is distributed across different households. Second, the 'missing' consumption of items for a particular household are imputed from the consumption of the remaining households within the survey. An ex-post simulation using data from Hergeiza shows that the methodology is able to provide accurate poverty estimates. Implementation of the methodology in the Mogadishu pilot shows that the methodology is essential for unbiased poverty estimates and practical to implement.

The third challenge of monitoring limitations is tackled in the third poverty note. Data quality can be improved – especially for consumption data – by implementing a dynamic on-the-fly data validation system. This system flags unusually high or low entries and asks enumerators to confirm the correctness of the entry. In addition, an innovative solution utilizing available software is proposed to monitor and manage tablets remotely. Using GPS tracking software, the location and trajectory of tablets can be determined – even retrospectively when the tablet re-enters 3G/WiFi areas. This helps to determine whether interviews were conducted at the correct location. Using remote management software, errors in the tablet configuration can be detected and solved while tablets can be updated remotely. Thirdly, a real-time monitoring system can identify challenges in the field work as well as weak enumerators early on and mitigate their impact on data quality. In addition, the analysis system calculates outcome indicators of the survey, e.g. consumption, education attainment or unemployment, to check incoming data while field work is still ongoing. The availability of the analysis code at the end of the data collection considerably accelerates the process from data collection to publication of results. Implementing these innovations in the Somalia High Frequency Survey will ensure high data quality despite limitations for field monitoring.

Based on the three poverty notes, it will be possible to implement a Somalia High Frequency Survey to collect consumption data and estimate poverty. The sampling and the new rapid consumption methodologies are critical choices and their validity is of utmost importance. The two first notes provide sufficient evidence for the validity of the methodologies. Both notes were peer-reviewed and presented at the Annual Bank Conference for Africa (ABCA), 2015.

Executive Summaries

Second Stage Sampling for Conflict Areas: Methods and Implications

The collection of survey data from war zones or other unstable security situations provides important insights into the socio-economic implications of conflict. Data collected during these periods, however, is vulnerable to error, because conflict often limits the options for survey implementation. For example, the traditional two-stage sample design for face-to-face surveys in many countries is to first select census enumeration areas (EAs) proportional to size and then to conduct a household listing operation. However, such an approach is not always feasible in areas of conflict. At the first stage, updated counts are often not available, making probability proportional to size selection inefficient. At the second stage, the listing visit, in which survey staff canvass the entire selected area, is seen as too dangerous. To collect high-quality data in areas of conflict, new methods of selection are needed, particularly at the second stage.

Security concerns limit the amount of time that the interviewers can spend in the field: more time in the field, particularly outside of households, increases interviewers' exposure to robbery, kidnapping, and assault, or simply that the local militias would object to their presence. Limited field access for international supervisors demands a sample design that can be implemented by interviewers without extensive training and which is verifiable.

Several alternative sampling approaches can be considered for face-to-face household surveys in areas with limited security. Overall, each of the methods can prove to be the best option for second stage sampling in a conflict zone depending on the context of the survey. Satellite mapping, segmentation, and the Mecca methods with full area weights are all probability methods for which it is possible (though perhaps not easy) to calculate weights, and thus all should produce unbiased estimates of the population mean. In addition, the Mecca method with simple distance weights is a close approximation of an unbiased sample.

The choice between the different methods is really one of cost and variance, as well as issues specific to the survey area. If there are no restrictions on time and back office resources, a full listing yields the most consistently unbiased and efficient design, provided recent maps are available and potential issues with out-of-sample buildings can be adequately addressed. The Mecca method provides promising results in simulations but has yet to be tried in the field. The simple distance variation of the Mecca method shows particular promise as it removes the requirement of updated satellite maps and greatly reduces the calculation burden for the weights. The non-probability methods, random walk and the unweighted Mecca method, do not produce unbiased results. Random walk, in particular, did not perform well in the simulations despite being common practice for many surveys. Given the expanding availability of satellite maps and decreasing costs of GPS technology, much of which is integrated into the phones and tablets used by interviewers, alternative methods based on probability sampling may reduce bias with little impact on cost or complexity of implementation.

For the Somalia High Frequency Survey, the segmentation method is recommended for areas where a full listing is not possible. The segmentation method is a compromise between preparation time, ease of implementation, and the time and complexity necessary for the weight calculations. Special focus should be put on verifying the correct location of enumerators since one in four interviews had to be discarded in the pilot for being conducted outside the selected enumeration area. Regardless of this challenge, however, the pilot prove that it is possible to implement a complex and yet rapid, high-quality survey in one of the most challenging urban contexts known to date.

Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu High Frequency Survey

Consumption aggregates are estimated traditionally by time-consuming household consumption surveys. A household consumption questionnaire records consumption and expenditures for a comprehensive list of food and non-food items. With around 300 to 400 items, the administering time of the questionnaire often exceeds 90 – 120 minutes. In addition to higher costs due to longer administering time, response fatigue can increase measurement error especially for items at the end of the questionnaire. In a fragile country context, security concerns can restrict the duration of a visit to less than 60 minutes.

The extensive nature of household consumption surveys make it difficult to obtain updated poverty estimates especially when they are needed the most: after a shock and in fragile countries. Therefore, approaches were developed to reduce administering time to allow collection of consumption data with significantly lower administering time. The most straight-forward approach to minimize administering time reduces the number of items either by asking for aggregates or by skipping less frequently consumed items, which we call reduced consumption methodology. However, both approaches have been shown to under-estimate consumption, which in turn over-estimates poverty. Splitting up the questionnaire for multiple visits is another solution but attrition issues – especially in fragile country contexts – increase required sample size and also have a high cost implication. In addition, multiple visits to the same household can increase security concerns. A second class of approaches utilizes a full consumption baseline survey and updates poverty estimates based on a small subset of collected indicators. This class of approaches is not feasible for Somalia due to the lack of a baseline.

The proposed rapid consumption methodology combines an innovative questionnaire design with standard imputation techniques. Consumption items are partitioned into a core module containing items with large consumption shares and optional modules for the remaining items. Instead of administering all modules to each household, only the core and one optional module are administered. As different optional modules are assigned to different households, the ‘missing’ consumption of the modules not administered to a particular household can be estimated within the survey.

This methodology reduces the administering time of a consumption survey to less than 60 minutes while at the same time credible poverty estimates are obtained. Thus, the gain in administering time is bought by the need to impute missing consumption values. Due to the design of the

questionnaire, the method circumvents systematic biases as identified for alternative methodologies.

The results of an ex-post simulation and a pilot survey in Mogadishu suggest that the rapid consumption methodology is a promising approach to estimate consumption and poverty in a cost-efficient and fast manner in Somalia. At the same time, the simulation clearly indicates that a reduced consumption approach would considerably over-estimate poverty. Thus, it is recommended to use the rapid consumption methodology for the Somalia High Frequency Survey.

Utilizing Mobile Technology to Innovate CAPI Data Collection in Fragile Contexts

Data collection in fragile countries faces two major constraints. First, field access is often limited to monitor data collection. This puts data quality at risk if enumerators cannot be sufficiently supervised. Second, the context is highly volatile demanding timely data collection with swift results. The note presents three measures to overcome these barriers and deliver high quality household consumption data using tablet face-to-face interviews. The three innovations go beyond traditional tablet survey collection by an innovative online management of the data collection, dynamic on-the-fly data validation and real-time monitoring and analysis of the data.

Three applications can be used to improve online management of the tablets. It is assumed that encrypted tablets equipped with Android will be used. In case of a stolen tablet, the data will not be readable even directly from the memory due to encryption of the device. The Android Device Manager allows remotely locking and wiping tablets. A second app called AirDroid enables managing remotely the tablet. This includes checking the phone call log, changing the configuration of the device as well as changing the content of the file system. Using this app, tablets can be remotely managed and monitored without direct access to the field. A third app Android GPS Tracker tracks the position all enumerators including pace and indicating stops while traveling. Paired with coordinates for households to be interviewed, it is easy to check whether enumerators indeed visit the assigned households.

The second component for improved data collection implements dynamic on-the-fly data validation with focus on consumption data. Questions for quantities consumed / purchased have dynamic unit validation to ensure that enumerators are not entering the wrong unit (e.g. 100 kg). In addition, built-in unit value tables validate entries of the value of consumption by comparing the unit value of the entered data with the built-in table. In both cases, entries outside pre-defined intervals are flagged to the enumerator requiring his explicit confirmation that entries are correct. Conversion tables of non-standard units (like heaps) are collected directly in the field by asking enumerators upon each entry for his estimate. These measures effectively clean data on-the-fly utilizing the context-specific knowledge of enumerators while visiting the household.

The third component performs real-time monitoring and analysis. Taking advantage of the almost instant availability of collected data, descriptive statistics of the implementation process are compiled, e.g. number of interviews per day and enumerator. In addition, interviews can be flagged for validity, e.g. by checking whether the interview was conducted within the enumeration area based on GPS information. Thus, with day 1 data can be monitored and action been taken on irregularities in the field. The descriptive statistics from the data collection are accompanied by

monitoring indicators collected by the survey, e.g. employment rate. This helps to ensure data quality at a higher level. From there, it is a continuum towards real time analysis. Once these statistics are implemented for the field collection, they are immediately available at the end of the field work reducing gap time between field work and the publication of results.

Second Stage Sampling for Conflict Areas: Methods and Implications

Kristen Himelein, Stephanie Eckman, Siobhan Murray and Johannes Bauer¹

Abstract: The collection of survey data from war zones or other unstable security situations is vulnerable to error because conflict often limits the options for implementation. Although there are elevated risks throughout the process, we focus here on challenges to frame construction and sample selection. We explore several alternative sampling approaches considered for the second stage selection of households for a survey in Mogadishu, Somalia. The methods are evaluated on precision, the complexity of calculations, the amount of time necessary for preparatory office work and the field implementation, and ease of implementation and verification.

Unpublished manuscript prepared for the Annual Bank Conference on Africa on June 8 – 9, 2015 in Berkeley, California. Do not cite without authors' permission.

Acknowledgments: The authors would like to thank Utz Pape, from the World Bank, and Matthieu Dillais from Altai Consulting for their comments on this draft, and Hannah Mautner and Ruben Bach of IAB for their research assistance.

¹ Kristen Himelein is a senior economist / statistician in the Poverty Global Practice at the World Bank. Stephanie Eckman is a senior researcher at the Institute for Employment Research (IAB) in Nuremberg, Germany. Siobhan Murray is a technical specialist in the Development Economics Research Group in the World Bank. Johannes Bauer is research fellow at the Institute for Sociology, Ludwigs-Maximilians University Munich and at the Institute for Employment Research (IAB). All views are those of the authors and do not reflect the views of their employers including the World Bank or its member countries.

1. Introduction

The collection of survey data from war zones or other unstable security situations provides important insights into the socio-economic implications of conflict. Data collected during these periods, however, is vulnerable to error, because conflict often limits the options for survey implementation. For example, the traditional two-stage sample design for face-to-face surveys in most developing countries first selects census enumeration areas (EAs) with probability proportional to size and then conducts a listing operation to create a frame of households from which a sample is selected. Such an approach, however, is not always feasible in conflict areas. At the first stage, updated counts are often not available, making probability proportional to size selection inefficient. At the second stage, survey staff canvas the entire selected area, requiring interviewers to spend additional time in the field approaching all households (see Harter et al, 2010 for description of household listing procedures). This may be too dangerous. To collect high-quality data in conflict areas, new methods of selection are needed, particularly at the second stage.

This paper explores several alternative sampling approaches considered for the baseline of the Mogadishu High Frequency Survey (MHFS). The baseline was a face-to-face household survey in Mogadishu, Somalia, conducted from October to December 2014 by the World Bank team and Altai Consulting. A full listing was deemed unsafe in Mogadishu. The additional time in the field and predictable movements increased interviewers' exposure to robbery, kidnapping, and assault, and also increased the likelihood that the local militias would object to their presence. The survey needed a sample design that would minimize the time spent in the field outside of the households, but also could be implemented without expensive equipment or extensive technical training. In addition, international supervisors from the consulting firm could not go to the field, necessitating a sample design that could be verified afterwards.

The consulting firm originally proposed a random walk procedure. While this methodology had the benefits of fast implementation and unpredictability of movement, the procedure gives biased results, even if implemented under perfect conditions (Bauer, 2014), and the circumstances in Mogadishu were far from ideal. Therefore the team considered four alternatives for household selection. The first option considered was to use a satellite map (of which many high quality options exist, due the arid conditions and political importance of the region) to map each structure. Of these, 10 would be selected. The second option considered was to cut EAs into 8-10 household non-uniform segments and ask enumerators to list and choose households from the segments.² The third option considered was to lay a uniform grid over the EA and ask enumerators to list and choose households from selected grid boxes. The final option considered was to start at a random point in the cluster and walk in a set direction, in this case towards Mecca, until the interviewer encountered a structure.

The paper will make use of data from the Mogadishu survey and geo-referenced maps and three example EAs to explore the following questions: (1) What are the implementation concerns for each method, including the options for verification and the impact of non-household structures? (2) What are the implications in terms of precision and bias for each of the methods described above? (3) What information is needed to calculate sampling weights for each method, and is this information available?

² Interviewers had a selection application on their smart phones that they used whenever subsampling was needed.

The next section briefly describes the literature as it relates to the questions above. Section 3 describes the data, section 4 gives detail on the methods considered, section 5 presents the results, and section 6 offers some discussion and conclusions.

2. Literature Review

The most common method for collecting household data in sub-Saharan Africa is to use a stratified two-stage sample, with census enumeration areas selected proportional to size in the first stage and a set number of households selected with simple random sampling in the second stage (Grosh and Munoz, 1996). Since often administrative records are incomplete and most structures do not have postal addresses, as is the case in Mogadishu, a household listing operation is usually necessary prior to the second stage selection. Due, however, to the security concerns cited above, listing was not feasible in Mogadishu.

A number of alternatives for second stage selection can be used when household lists are not available. The most common alternative is a random-walk in which the probabilities of selection are considered equivalent to simple random sampling. Random walk methodologies are commonly used in Europe (see Bauer, 2014, for recent examples), but are also implemented in the developing world. Specifically, the Afrobarometer survey, which has been conducted in multiple rounds in 35 African countries since 1999, and the Gallup World Poll, which conducted surveys in 29 sub-Saharan African countries in 2012, use random walk methodologies. Bauer (2014) tests the assumption that a random walk is equivalent to simple random sampling by simulating all possible random routes using standard within a German city and calculating the probability of selection for each household. The results show substantial deviation from simple random sampling expectations which lead to systematic bias. The simulations also assume perfect implementation of the routing instructions, which is unlikely given the limited ability to conduct in-field supervision and strong (though understandable) incentives for interviewers to select respondents who are willing to participate (Alt et al 1991).

The alternative methodologies discussed here use a combination of satellite maps and area-based sampling. As satellite technology has improved in quality and become more readily available, it has been increasingly used for research in the developing world. Barry and R  ther (2001) and Turkstra and Raithelhuber (2004) use satellite imagery to study informal urban settlements in South Africa and Kenya, respectively. Aminipouri et al (2009) use samples from high resolution satellite imagery to estimate slum populations in Dar-es-Salaam, Tanzania. Examples from the United States and Europe are less common, as usually there are traditional, reliable alternatives, but Dreiling et al (2009) tested the use of satellite images for household selection in rural counties of South Dakota.

While area-based selection methods are more common in agricultural and livestock surveys, Himelein et al (2014) used circles generated around randomly generated points to survey pastoralist populations in eastern Ethiopia, with the stratification developed from satellite imagery. A variation of this method was considered in Mogadishu, but the methodology surveys all eligible respondents living within the selected circle. The resulting uncertainty over the number of total number of selected households and the time spent in the field caused it to be discarded.

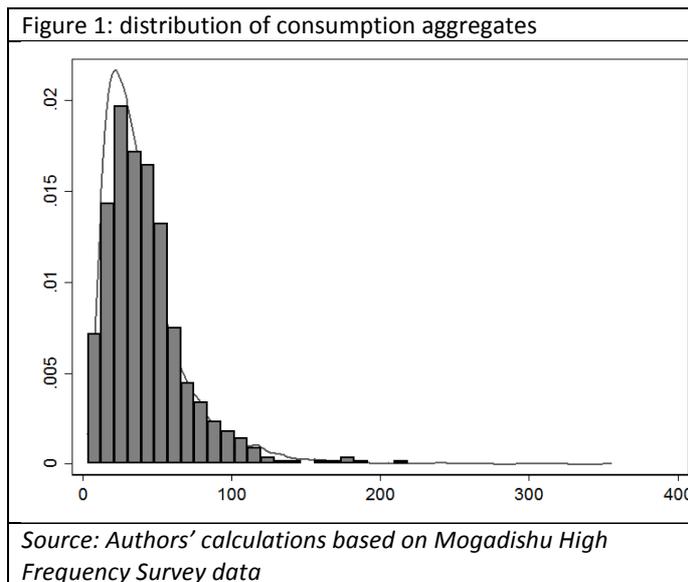
There are further examples of random point and satellite based selection in the public health literature. Grais et al (2007) also used a random point selection methodology in their study of vaccination rates in urban Niger, comparing the results to a random walk. They do not find statistically significant differences in the results from the three methods, though the sample size was limited, but conclude that interviewers found the random point selection methods most straightforward to implement than the random walk. Lowther et al (2009) uses satellite imagery to map more than 16,000 households in urban Zambia to select young children for a measles prevalence survey. They find the method straightforward to implement, but do not do a formal comparison with alternatives. Other public health studies such as the EPI studies use the “spin the pen” method to choose a starting household and then interview a tight cluster of households. This method is nonprobability (Bennett et al 1994) and was not considered for the Mogadishu study.

This paper brings together alternatives developed from this literature and applies them to a conflict environment. We take a rigorous approach using simulations and careful estimation of weights to compare the methods across a variety of potential field conditions. The results offer general guidelines for practitioners developing implementation plans for conflict settings.

3. Data and Methodology

To explore the challenges of the random walk and the four proposed alternatives, we simulated the use of each in three example PSUs from Mogadishu, Somalia. We purposefully chose three census enumeration areas as the PSUs for this exercise to illustrate the variation in physical layout present in Mogadishu. Maps of the three examples PSUs are shown in the appendix. The first is in Dharkinley district, a comparatively wealthy section of southwestern Mogadishu where the households are laid out relatively uniformly over gridded streets. The second is on the eastern edge of Heliwa district in the northeast of the city. This area is more irregular in layout with larger gaps between buildings. The third selected was in the more central Hodon district. It is densely populated with very irregularly laid out structures. (See maps in the appendix.)

To construct the dataset for the simulations, the values for consumption are taken from data collected by the MHFS. The survey covered both households in neighborhoods and those in internally displaced persons camps, but for the purposes of this simulation, we use only the neighborhood sample as the variation in the IDP sample is compressed due to reliance on food aid. Data was collected from the selected households on limited range of food and non-food items which we sum to calculate total consumption (see Mistiaen and Pape, forthcoming, for further



details on these calculations.) There were 624 cases outside of the IDP camps with non-missing values on the two consumption measures. The distribution of consumption across these cases has a strong right skew to the distribution (Figure 1) with mean 43.0 and standard deviation 27.5. In assigning values for our simulations, we drew consumption totals from this distribution.

To simulate the variety of situations that may be found in the field, we use three different mechanisms for assigning consumption values to households in the three example PSUs. In the first, values are randomly assigned across the households in each PSU. In the second, the same values are reassigned to households to create a moderate degree of spatial clustering. In the third assignment mechanism, the spatial clustering of consumption values is more extreme. We study the ability of each of the proposed methods to estimate consumption under these three conditions. While these distributions may not mimic actual conditions, they are illustrative of the different situations encountered in the field.

For each of the combinations of methods and assignment mechanisms discussed above, 10,000 simulated samples were drawn and relevant probability weights calculated. Ten structures were selected within each PSU. In the cases of segmentation and grid point selection, where the sample was selected in two stages, two clusters were selected and then five structures within each cluster.

4. Alternative Sampling Methodologies

4.1 Satellite mapping

A full mapping of the PSU entails using satellite maps to identify the outline of each structure (see appendix). In this case, we used maps publically available on Google Earth and PSU outlines provided by the Somali Directorate of Statistics. From these maps, the structures inside each PSU can be assigned numbers and selected easily in the office with simple random sampling. Once selected, interviewers can be provided with the GPS coordinates of the household to locate it in the field. Mapping is the closest of the proposed study methods to the gold standard of a well-implemented full household listing. The main differences are that in a field listing, enumerators can exclude ineligible structures, such as uninhabited and commercial buildings, and include information not available from satellite maps, such as new construction and the identification of individual units within multi-household structures. Selection from a satellite mapping therefore requires an additional set of field protocols for addressing and documenting the above issues.

Satellite mapping can also be time consuming in terms of preparation. In the simplest approach to this method, cluster outlines can be overlaid with maps from Google Earth, maps printed, and buildings outlined and numbered by hand. A fully digitalized approach would take longer. Based on the experience mapping the three PSUs used in the paper, it takes about one minute per household to construct an outline. If the PSUs contain approximately 250 structures (the ones used here contain 68, 309, and 353 structures, respectively), mapping the 106 PSUs selected for the full Mogadishu High Frequency Survey would have required more than 50 work days.

Once the mapping has been completed, however, the calculation of the probability of selection, and by extension the survey weight, is straight-forward. The probability would be $\frac{n}{N}$, where n is the number of

structures selected and N is the total number of structures mapped, plus any necessary adjustment for multi-household structures.

4.2 Segmenting

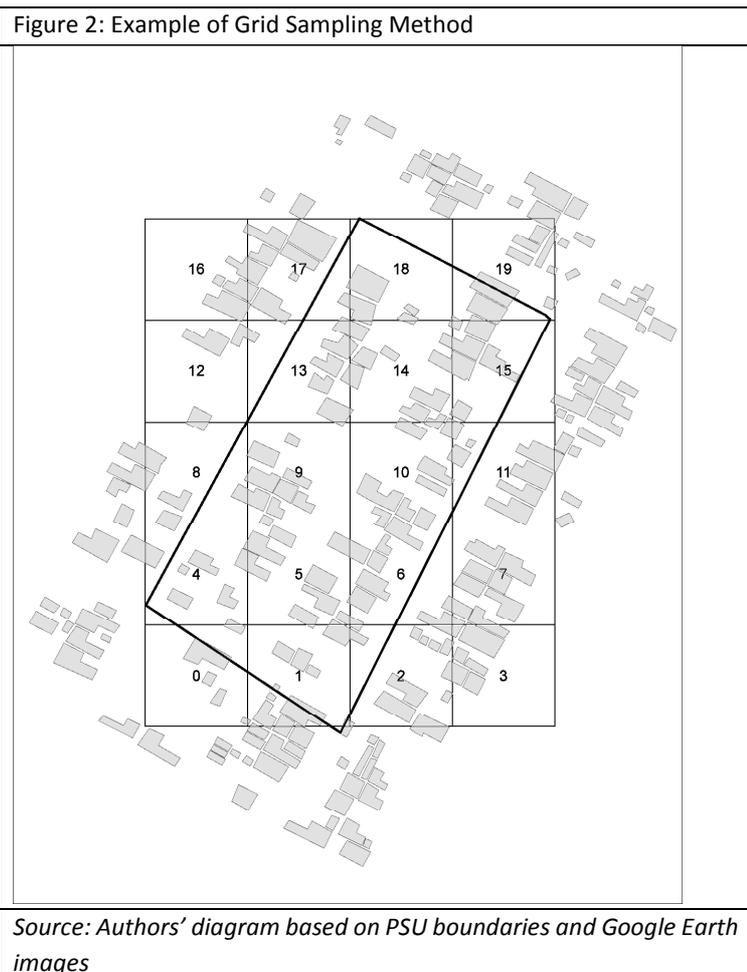
Segmenting is a standard field procedure of subdividing large PSUs into approximately equal sized smaller units for listing and selection purposes. The individual segments are then selected with simple random sampling, and listed by field enumerators, and households selected from these lists. Segmenting is less time consuming than a full mapping exercise in terms of office preparation, but still requires the manual demarcation of segment boundaries, which generally follow roads or other easily identifiable landmarks so that interviewers can identify the boundaries of the segments in the field. Also there would be additional time taken in the field to list the segment, which would have security implications.

The calculation of the probability of selection is straightforward: the product of the probability of selection of the segment, and the probability of selection of the household within the segment. There are, however, implications for precision as the additional clustering introduced by the selection of segments could increase the design effect (though the magnitude would depend on the number of segments selected, the number of households selected per segment, and the degree of homogeneity within clusters for the study variables).

4.3 Grid

To implement the grid method, a uniform grid of squares (or another shape) is overlaid on the PSU map. Figure 1 shows an example using 50 x 50 meter squares for the Dharkinley PSU. The area of a grid point includes all of the area that lies both within the grid point and within the PSU boundaries. For example, in grid point 17 in figure 1, the majority of the structures would not be eligible as they lay outside of the PSU boundaries. Only the structures which lie in the bottom left corner are both within the grid and PSU boundaries.

One or more squares are selected with simple random sampling from the set of all squares that overlap the selected PSU. Depending on the survey protocols, a structure may be defined as eligible if all or part of it lies



within the grid space. The more common protocol, including the structure if the majority lies within the grid point, has the benefit of simplifying the weight calculations, but the risk of subjective decisions made by interviewers in the field about where the majority of the building lies. Since the options for supervision and field re-verification were limited in this survey, it was decided to consider the structure as eligible if any portion of the structure lay within the grid boundaries.

To select a sample of households within the selected squares, a common approach would be for interviews to be conducted with all eligible respondents with the grid point. This could lead, however, to issues with verification as well as decreasing control over the final total sample size. Therefore, the protocol used in Mogadishu had interviewers list all households with the selected square and use the application to select households for the survey.

This variation of the grid method has the advantage that it requires less preparation time compared to mapping or segmenting. There are considerable drawbacks, however, in the ease of implementation and additional work to accurately calculate the selection probabilities. Since the squares do not follow landmarks on the ground, interviewers need to use GPS devices to find the squares' boundaries. This approach also still requires some listing work, which may have security implications depending on the size of the squares in the grid. The size can vary depending on the physical size of the PSU and the density of the population. Smaller squares require less listing work, but also mean that more buildings will lie on the boundaries between squares. Those selected structures which lie on boundary lines require additional time for field implementation, because the overlapping squares must also be listed, and also complicate the calculation of the probabilities of selection.

Let s_k be the number of squares selected in PSU_k and S_k be the number of squares that are partially or completely contained within PSU_k . For households that are entirely contained within square j , the probability of selection, given that PSU_k was selected, is:

$$\pi_{i|k} = \frac{n_j}{N_j} * \frac{s_k}{S_k} \quad (1)$$

where n_j is the number of structures selected from square j and N_j is the total number of eligible structures in the square. $\frac{s_k}{S_k}$ is the probability of selection of the square when a simple random sample of size s_k is selected from the S_k squares in PSU_k .

If household i lies in both squares j and j' , the probability of selection is:

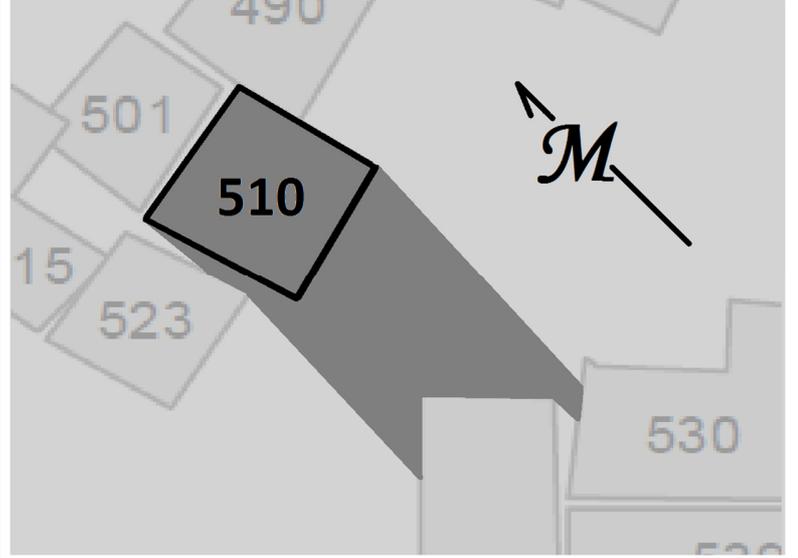
$$\pi_{i|k} = \left(\frac{n_j}{N_j} * \frac{s_k}{S_k} \right) + \left(\frac{n_{j'}}{N_{j'}} * \frac{s_k}{S_k} \right) - \left(\frac{n_j}{N_j} * \frac{n_{j'}}{N_{j'}} * \frac{s_k}{S_k} * \frac{s_k-1}{S_k-1} \right) \quad (2)$$

In an extreme case of a structure lying on a four way intersection, there would be additional terms in equation (2). Interviewers would also have to spend significant time on additional listing, which greatly increases exposure in the field and provides disincentives to interviewers to report such households.

4.4 Mecca Method

This sampling approach involves selecting multiple random locations within each PSU and travelling from those points in a given direction until a structure is found. In Somalia, the consulting firm suggested using the direction of Mecca, since it is common for interviewers to have an application on their cell phones which shows this direction. If that structure is a household, the interview is done with the household. We are not aware of any surveys that have used this approach, but it clearly has intuitive appeal, as it is straightforward and inexpensive to implement in the field.

Figure 3: Example of Mecca Method



Source: Authors' diagram based on PSU boundaries and Google Earth images

Figure 3 gives a stylized example of this method. Household 510 will be selected whenever any of the points in the shaded region are selected. This region includes the area of the dwelling itself (its roof) and all points in its “shadow” – that is, all land inside the PSU that lies in the direction opposite Mecca, excluding points that lead to the selection of other buildings.

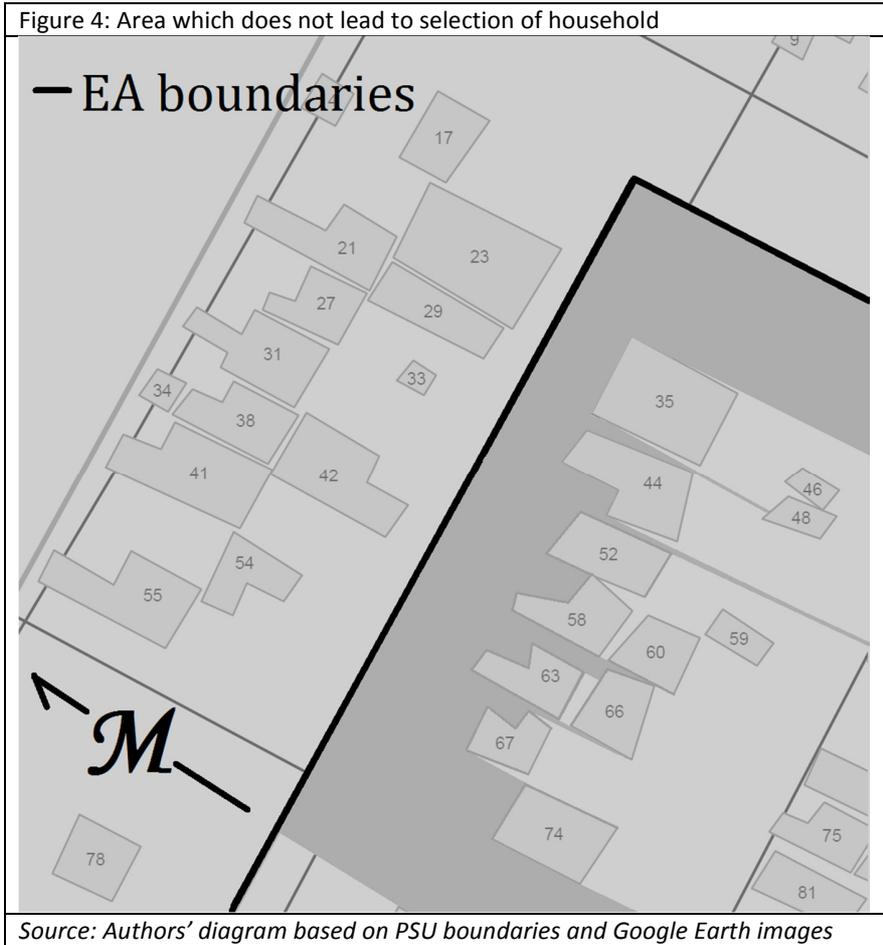
Despite its seeming ease-of-use, this approach contains many challenges. For one, it is not clear how non-residential structures should be handled. The interviewer could walk around business and vacant housing units, continuing in the direction of Mecca until she finds a residential unit. This approach would work in theory, but in addition to the difficulties in remote verification it would create, it would also complicate the calculation of probabilities of selection (discussed below). Therefore we do not suggest it. Instead, we suggest coding points that lead to non-household selections as out-of-scope, and selecting additional points to replace them.

Perhaps the biggest challenge with this method is the collection of the information needed to calculate probabilities of selection of the selected households. Figure 3 shows Household 510 and, in the shaded area, the set of all points that lead to the selection of this household. Each household, i , in the PSU has an associated selection region: call this region A_i . The probability of selection of household i (conditional on selection of PSU k), if c points in the PSU are selected, is one minus the probability that all c selected points are not in A_i :

$$\pi_{i|k} = 1 - \left(1 - \frac{\text{Area of } A_i}{\text{total area of PSU}_k}\right)^c \quad (3)$$

(based on Särndal et al. 1992, p.50). This approach is essentially probability proportional to size selection with replacement, where the measure of size is the area of A_i . The weight is then the inverse.

Looking at Equation 3, the hardest quantity to calculate is the area of A_i . For the purposes of this paper, we use relatively recent Google Earth maps (two of which were from March 2014 and one from December 2013) to calculate the A_i region for each selected household. If high quality and recent satellite photos of the survey region are not available, calculation of the area of the selection regions will be much harder and may not be possible at all. Any structures added since the imagery was captured would not be included and therefore it would be difficult to calculate the area. Similarly, if new structures were built in the shadow, these areas would be incorrectly still included in A_i .



As it is likely that there will be incidences where field maps are either too dated to be useful or completely unavailable, we also consider two alternatives: using weights based on the estimated distance to the next structure in the opposite direction of Mecca, and ignoring the weights completely. Neither method can provide unbiased results, but under certain conditions they may be a good alternative for those that find themselves in second or third best scenarios. The simple distance would be unbiased if the length of the line was exactly proportional to the area of the shadow. While this specific condition is unlikely, the variation does have the benefit of not requiring digitized maps and being more flexible in accounting for new construction. No weighting would be approximately unbiased if dwellings were identical in size and equidistant.

Figure 4 illustrates another potential issue with this group of methods. There are points in this PSU that would not lead to the selection of any households. Consider the points in the direction of Mecca from Household 35, for example. If any of these points were selected, the interviewer would not find any household before she left the boundaries of the PSU. This issue raises questions for the field protocols. Should interviewers stop at the PSU boundary, or should they continue and select housing units outside of the selected PSU? If the former, how would the interviewer know where the PSU boundaries are? If

the latter, the probabilities should be adjusted for the fact that the A_i region extends outside of the PSU, which is not straightforward. Additional structures outside of the boundaries of the PSU would need to be mapped, requiring additional preparation time. For the purposes of this paper, we mapped all households in a 50 meter buffer zone around the PSU boundaries. This increased the number of structures required from 309 to 408, 68 to 207, and 353 to 724, respectively, nearly doubling the required mapping time. A third option would be to allow interviewers to travel outside of the PSU in search of a selected household, but then remove these interviewed households outside the selected PSUs from the data set, because their probabilities of selection are too complex to calculate. This approach has preserves the probabilities of selection and is easy for the interviewer to implement, but deleting data is inefficient in terms of cost.

4.5 *Random Walk*

There are many different implementations of the random walk procedure. Each method invokes choosing a starting point within the selected area and then proceeding along a path, selecting every k^{th} household. The methods differ in how the path is defined. In this paper, we follow the method used by the Afrobarometer survey. The walking instructions are:

“Starting as near as possible to the SSP [Sampling Start Point], the FS [Field Supervisor] should choose any random point (like a street corner, a school, or a water source) being careful to randomly rotate the choice of such landmarks. From this point, the four Fieldworkers follow this Walk Pattern: Fieldworker 1 walks towards the sun, Fieldworker 2 away from the sun, Fieldworker 3 at right angles to Fieldworker 1, Fieldworker 4 in the opposite direction from Fieldworker 3.... Walking in their designated direction away from the SSP, they will select the fifth household for their first interview, counting houses on both the right and the left (and starting with those on the right if they are opposite each other). Once they leave their first interview, they will continue on in the same direction, and select the tenth household (i.e., counting off an interval of ten more households), again counting houses on both the right and the left. If the settlement comes to an end and there are no more houses, the Fieldworker should turn at right angles to the right and keep walking, continuing to count until finding the tenth dwelling” (Afrobarometer, pg. 35).

However, there are several documented problems with random walk methods. First, although the random walk methods do not necessarily produce equal probability samples, they do not collect any information with which to calculate probabilities of selection. For this reason, weights are not calculable for random walk samples; instead, the samples are analyzed as if they were equal probability. Bauer (2014) shows that this assumption is not correct (though the methodology differs from the one described above). Second, the method is difficult to verify. If a supervisor has GPS tracks from each interviewer, he or she can perhaps verify that the interviewers’ direction of travel was correct, but cannot be sure that the rules about which households to select were implemented correctly. Even if implemented according to protocols, two interviewers starting at the same point and traveling on the same path may select different samples depending on the distance that they consider close enough to be included or in what sequence they count the dwellings. Finally, interviewers using random walk tend to select people who are at home, rather than those who live in the households specified by the rules (Alt et al 1991).

To simulate the random walk in the Mogadishu context, we replicate the Afrobarometer protocols to the extent possible. First a random point is selected. Since it is not possible to identify landmarks with the level of detail available on the maps, the randomly selected point is taken as the starting point. To simulate the direction of the sun, a random angle is chosen and the direction of the interviewer's path assigned at 90 degree intervals. For example, if 13 degrees from due north was selected, then the four paths would be at 13 degrees, 103 degrees, 193 degrees, and 283 degrees. From these lines, it was assumed that every dwelling within 5 meters on either side of the direction of walking was within the interviewer's line of sight. These dwellings were sequentially numbered and every fifth dwelling selected. If the interviewer reached the PSU boundary before selecting the requisite number of households, the path made a 90 degree turn and continued. If each of the four interviewers selected three households, the total cluster size would be 12. In order to ensure comparability with the other methods, each of which aimed to select ten households, we dropped the last two selected households.³ (See illustrations in appendix for further detail.)

5. Results

5.1 Simulations

For each of the methods discussed above and the three different methods of allocating consumption values to households (random, some spatial clustering, extreme clustering), we simulated 10,000 samples and calculated a mean for each one. We report the mean and standard deviation⁴ of the distribution across all 10,000 samples and evaluate the different sampling approaches in terms of their bias and variance. If a sampling method is unbiased, the expected value of the sample means should be 40, the true mean consumption in each PSU. However, because our samples are quite small (only 10 cases at the most) and the underlying distribution is far from normal (see Figure 1), we should not expect that all methods will appear unbiased.

While generally it was possible to implement all of the methods in our simulations, there were notable challenges with two of the designs. In simulating the Mecca method, certain selected points did not lead to a selection within the EA. The impact was largely negligible in Heliwa or Hodon, where only 0.4 percent and 1.4 percent, respectively, of the total area led to no selection, but in Dharkinley, the smallest and most regular of the PSUs tested, 13 percent of the area led to no selection. In implementing the grid selection method, there was little control over the number of households in each grid point. In some cases, grid points were empty or did not have the minimum number of structures to achieve the expected sample size. In the most extreme case of the large and sparsely populated PSU of Heliwa, when 50 x 50 m grid points were used, 22 contained no structures. Of those remaining, a further 17 percent had less than the necessary five structures. Therefore the grid points were combined into 100 x 100 m squares. After combination, none of the larger grid points were empty, but 16 percent continued to contain less than the minimum. For the simulations, we dropped grid points without households, though this would likely not be possible in true field implementation, leading to cost inefficiencies.

³ The analogous action in the field would be for the supervisor to rotate the additional interviews between interviewers to assure an even workload, though most likely in the design stage the cluster size would have been set to be evenly divisible among interviewers.

⁴ The standard deviation of the distribution is the standard error of the estimate of the mean.

5.2 *Bias and Variance*

The mean, standard deviation, and coefficient of variation are shown in Table 1 for the seven methods and three PSUs under the three different consumption values. From this table we can evaluate how well each method worked in terms of bias and variance. From a true mean of 40, it was unsurprising that the full listing / satellite mapping method showed the most consistently unbiased results across the nine scenarios. Segmentation also showed consistently unbiased results. As expected, however, the segmentation method is more sensitive to clustering in the underlying distribution of the consumption values because the homogeneity within the segments.

The Mecca method also generally performed well when the results were weighted. The Mecca method with full weights calculated from the areas of the shadow showed an average difference of 1.3 from the true mean of 40 across the nine scenarios, in comparison to the 0.0 for full listing and 0.3 for segmentation. The method should be unbiased, as it is a probability method using appropriate weights, but the small sample size and skewed distribution of the underlying variable, as discussed in Section 4.6. The Mecca method with proxy weights, based on the distance from the selected point to the household, also performed well with an average difference of 1.5. As expected, the Mecca method without any weighting showed the most biased results of the methods considered, even when there was no clustering of wealth. It performed particularly poorly in Heliwa, where the sparse geography leads to long shadows and large differentials in the probabilities of selection.

The two remaining methods, the grid and random walk, showed more bias than the full listing, segmentation, and two of the three Mecca methods, but were better than the Mecca method with no weights. Both performed relatively well when the wealth values were randomly assigned, though the grid method was slightly better in these situations. The two methods also performed better in the more uniform Dharkinley and sparse Heliwa compared to the more chaotic Hodon PSU.

6.2 *Implementation Issues*

The final criteria on which the methods were evaluated was the ease of field implementation and remote supervision. The most straightforward to supervise remotely is the satellite mapping. Since the selection is done in the office, the interviewers can be sent to the field with their target locations loaded on a handheld GPS device, and supervisors can use tracks from the device to verify that interviewers visited the correct households. While there is still scope for cheating, such as going to the target location to record the GPS point and then interviewing another household, or claiming a refusal because the household appeared to be difficult or time-consuming, these behaviors are possible in all surveys, regardless of method, and can only be addressed through training. The Mecca method also provides the possibility for remote supervision by using GPS waypoints and tracks to let supervisors and central office staff verify that the interviewer travelled in the correct direction and interviewed households within the boundary of the select PSU (see Himelein et al, 2014, for an example of using interviewer tracks for supervision).

The grid and segmentation methods are more difficult to supervise because they are also more difficult for the interviewers to implement. When creating segments, best practice is to use clearly discernable landmarks used to draw boundaries, but these can change over time or not be correctly identified by the

interviewers. If the interviewer incorrectly identifies the segment, it may be necessary to exclude the resulting data as it cannot be properly weighted.

The grid method provides additional challenges as the boundaries between the grid squares do not follow existing landmarks. The boundaries must therefore be programmed into the GPS and identified by the interviewers. As it is unlikely that they will be able to walk straight along the boundary, additional training may be required to correctly identify eligible structures. In addition, as was the case in Heliwa, it may be necessary to have large grid points in sparsely populated areas. This increases the time necessary to do the listing, exposing the interviewer to increased danger that may result in the inability to complete interviews.

With both the grid and segmentation method, additional verification can be done for the listing portion by using satellite maps. If the interviewer enters fewer households into the selection tool than are expected based on the satellite map, it could be that they are purposely excluding or shirking. If the number is much higher without evidence of multi-household structures, it may be that the interviewers have gone to the wrong location, and this can be confirmed with the target household waypoint.

5.3 Replacements

One remaining potential issue not yet addressed relates to the use of non-response due to either refusals or are out of sample selections. Due to high transportation costs, most surveys in the developing world use replacements. This is done either through selecting additional households from PSU lists, as is recommended in the World Bank's Living Standards Measurement Study (Grosh and Munoz, 1996), or selecting a neighboring structure based on field protocols, such as selecting the dwelling immediately to the right (Lowther et al, 2009). While replacements for out of sample selections with new random points does not introduce bias, it is inefficient and increases costs. For non-response due to refusal, it is likely to be non-random, and therefore replacements will create at least some degree of bias in the data. The reason and method for the replacement may influence the degree. If refusals tend to come from the highest and lowest wealth households, as the opportunity cost of their time is high, and replacements come from the main part of the distribution, the use of replacements will attenuate the variation in the sample. This may cause the results to underestimate measures such as inequality that depend on accurately capturing the extremes of the distribution. When using a replacement method that uses near neighbors, if structures are abandoned or commercial buildings, those households living adjacent may be systematically different from the remainder of the PSU. In addition, those households near the boundary of the PSU would have a lower probability of selection since there are fewer households near them that would lead to them being selected as replacements.

Of the methods discussed above, segmenting and gridding require a short listing exercise at which time non-eligible structures can be excluded. Satellite mapping and the Mecca method rely on maps that cannot differentiate based on eligibility, and are therefore more vulnerable to issues with out of sample selections. In addition, regardless of method, the survey protocols should address procedures for the inevitable refusals, which may be more likely in conflict areas.

6. Discussion

Overall, each of the methods above could prove to be the best option for second stage sampling in a conflict zone depending on the context of the survey. Satellite mapping, segmentation, and the Mecca methods with full area weights are all probability methods for which it is possible (though perhaps not easy) to calculate weights, and thus all should produce unbiased estimates of the population mean.⁵ In addition, the Mecca method with simple distance weights is a close approximation of an unbiased sample. The choice between the different methods is really one of cost and variance, as well as issues specific to the survey area. If there are no restrictions on time and back office resources, a full listing yields the most consistently unbiased and efficient design, provided recent maps are available and potential issues with out-of-sample buildings can be adequately addressed. The Mecca method provides promising results in simulations but has yet to be tried in the field. The simple distance variation of the Mecca method shows particular promise as it removes the requirement of updated satellite maps and greatly reduces the calculation burden for the weights. The non-probability methods, random walk and the unweighted Mecca method, do not produce unbiased results. Random walk, in particular, did not perform well in the simulations despite being common practice for many surveys. Given the expanding availability of satellite maps and decreasing costs of GPS technology, much of which is integrated into the phones and tablets used by interviewers, alternative methods based on probability sampling may reduce bias with little impact on cost or complexity of implementation.

In the case specifically discussed here, the Mogadishu High Frequency Survey, the team opted to use segmentation as a compromise between preparation time, ease of implementation, and the time and complexity necessary for the weight calculations. The survey was generally successfully fielded, but the team encountered a number of difficulties in the field. Teams occasionally faced high-level security threats and exploitative rent-seeking from local leadership. The complexity of the survey protocols, including the sampling design, slowed the implementation of the survey. Also a substantial number of observations had to be discarded because the interviewed points did not fall within the boundaries of the selected segments. Regardless of these challenges, however, it was possible implement a complex and yet rapid, high-quality survey in one of the most challenging urban contexts known to date.

⁵ In this particular study, we saw small deviations from unbiasedness due to the small sample sizes used other issues discussed above.

References

- Afrobarometer Network, 2014. Afrobarometer Round 6 Survey Manual.
- Alt, C., Bien, W., Krebs, D., 1991. Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe. *ZUMA-Nachrichten* 28, 65–72.
- Aminipouri, M., Sliuzas, R., Kuffer, M., 2009. Object-oriented analysis of very high resolution orthophotos for estimating the population of slum areas, case of Dar-Es-Salaam, Tanzania, in: Proc. ISPRS XXXVIII Conf. pp. 1–6.
- Barry, M., Rütger, H., 2001. Data collection and management for informal settlement upgrades, in: Proc. International Conference on Spatial Information for Sustainable Development. Citeseer.
- Bauer, J.J., 2014. Selection Errors of Random Route Samples. *Sociological Methods & Research* 0049124114521150. doi:10.1177/0049124114521150
- Bennett, A., Radalowicz, A., Vella, V., Tomkins, A., 1994. A Computer Simulation of Household Sampling Schemes for Health Surveys in Developing Countries. *International Journal of Epidemiology* 23,6: 1282-1291.
- Dreiling, K., Trushenski, S., Kayongo-Male, D., & Specker, B. 2009. Comparing household listing techniques in a rural Midwestern vanguard center of the national children's study. *Public Health Nursing*, 26(2), 192-201.
- Eckman, S., Himelein, K., Dever, J., forthcoming. New Ideas in Sampling for Surveys in the Developing World, in: Johnson, T.P., Pennell, B.-E., Stoop, I., Dorer, B. (Eds.), *Advances in Comparative Survey Methodology*. 3MC.
- Gallup, 2014. Farm workers pessimistic about their lives [WWW Document]. URL <http://www.gallup.com/poll/169019/farm-workers-africa-pessimistic-lives.aspx> (accessed 1.27.15).
- Gallup, 2015. World Poll Methodology [WWW Document]. URL <http://www.gallup.com/poll/105226/world-poll-methodology.aspx> (accessed 1.27.15).
- Grais, R.F., Rose, A.M., Guthmann, J.-P., 2007. Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys. *Emerging Themes in Epidemiology* 4, 8. doi:10.1186/1742-7622-4-8
- Grosh, M.E., Munoz, J., 1996. A manual for planning and implementing the living standards measurement study survey (No. LSM126). The World Bank.
- Harter, R., Eckman, S., English, N., O'Muircheartaigh, C., 2010. Applied sampling for large-scale multi-stage area probability designs. *Handbook of survey research* 2, 169–199.
- Himelein, K., Eckman, S., Murray, S., 2014. Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations. *Journal of Official Statistics* 30.
- Lowther, S.A., Curriero, F.C., Shields, T., Ahmed, S., Monze, M., Moss, W.J., 2009. Feasibility of satellite image-based sampling for a health survey among urban townships of Lusaka, Zambia. *Tropical Medicine & International Health* 14, 70–78. doi:10.1111/j.1365-3156.2008.02185.x
- Mneimneh, Z.N., Axinn, W.G., Ghimire, D., Cibelli, K.L., Alkaisy, M.S., 2014. Conducting surveys in areas of armed conflict, in: *Hard-to-Survey Populations*. Cambridge University Press.
- Särndal, C.E., Swensson, B., Wretman, J.H., 1992. *Model assisted survey sampling*. Springer.
- Turkstra, J., Raithelhuber, M., 2004. *Urban Slum Monitoring*.

Table 1: Main Results

| Method/Clustering | Dharkinley (EA4) | | | Heliwa (EA5) | | | Hodon (EA6) | | |
|----------------------------------|------------------|-----------|------|--------------|-----------|------|-------------|-----------|------|
| | Mean | std error | cv | mean | std error | cv | mean | std error | cv |
| True Mean | 40.0 | | | 40.2 | | | 40.0 | | |
| Full Listing / Satellite Mapping | | | | | | | | | |
| Randomly assigned | 40.0 | 11.7 | 0.29 | 40.0 | 9.6 | 0.24 | 40.0 | 9.5 | 0.24 |
| Some spatial clustering | 40.0 | 11.6 | 0.29 | 40.0 | 9.6 | 0.24 | 40.0 | 9.6 | 0.24 |
| Extreme clustering | 40.1 | 11.7 | 0.29 | 40.0 | 9.6 | 0.24 | 40.0 | 9.5 | 0.24 |
| Mecca method (accurate weights) | | | | | | | | | |
| Randomly assigned | 40.9 | 10.4 | 0.26 | 39.7 | 13.7 | 0.34 | 39.3 | 14.1 | 0.36 |
| Some spatial clustering | 39.4 | 15.7 | 0.40 | 37.6 | 15.1 | 0.40 | 41.5 | 13.3 | 0.32 |
| Extreme clustering | 37.8 | 16.7 | 0.44 | 37.8 | 14.5 | 0.38 | 41.4 | 13.8 | 0.33 |
| Mecca method (proxy weights) | | | | | | | | | |
| Randomly assigned | 41.2 | 12.5 | 0.30 | 39.5 | 14.7 | 0.37 | 39.3 | 14.9 | 0.38 |
| Some spatial clustering | 39.4 | 17.8 | 0.45 | 37.3 | 15.7 | 0.42 | 41.7 | 14.5 | 0.35 |
| Extreme clustering | 37.7 | 17.2 | 0.46 | 37.6 | 15.2 | 0.40 | 41.7 | 15.0 | 0.36 |
| Mecca method (no weights) | | | | | | | | | |
| Randomly assigned | 49.4 | 14.1 | 0.29 | 37.7 | 8.8 | 0.23 | 38.0 | 8.1 | 0.21 |
| Some spatial clustering | 37.1 | 9.5 | 0.26 | 33.4 | 7.0 | 0.21 | 45.5 | 9.2 | 0.20 |
| Extreme clustering | 35.8 | 8.1 | 0.23 | 34.1 | 7.9 | 0.23 | 44.5 | 8.4 | 0.19 |
| Segmentation | | | | | | | | | |
| Randomly assigned | 40.2 | 11.5 | 0.29 | 40.0 | 10.2 | 0.25 | 40.1 | 10.8 | 0.27 |
| Some spatial clustering | 40.9 | 17.3 | 0.42 | 40.0 | 17.3 | 0.43 | 39.9 | 17.0 | 0.43 |
| Extreme clustering | 40.9 | 17.7 | 0.43 | 40.0 | 19.2 | 0.48 | 40.0 | 19.3 | 0.48 |
| Grid | | | | | | | | | |
| Randomly assigned | 40.2 | 13.2 | 0.33 | 40.7 | 11.5 | 0.28 | 39.7 | 11.0 | 0.28 |
| Some spatial clustering | 38.1 | 20.5 | 0.54 | 40.9 | 18.6 | 0.46 | 45.3 | 20.1 | 0.44 |
| Extreme clustering | 38.8 | 24.9 | 0.64 | 42.6 | 24.0 | 0.56 | 45.1 | 22.6 | 0.50 |
| Random walk | | | | | | | | | |
| Randomly assigned | 38.4 | 11.5 | 0.30 | 39.3 | 9.1 | 0.23 | 39.5 | 9.1 | 0.23 |
| Some spatial clustering | 39.2 | 11.8 | 0.30 | 40.8 | 12.0 | 0.29 | 45.8 | 17.7 | 0.39 |
| Extreme clustering | 38.8 | 10.5 | 0.27 | 39.2 | 10.3 | 0.26 | 46.2 | 20.9 | 0.45 |

Appendix

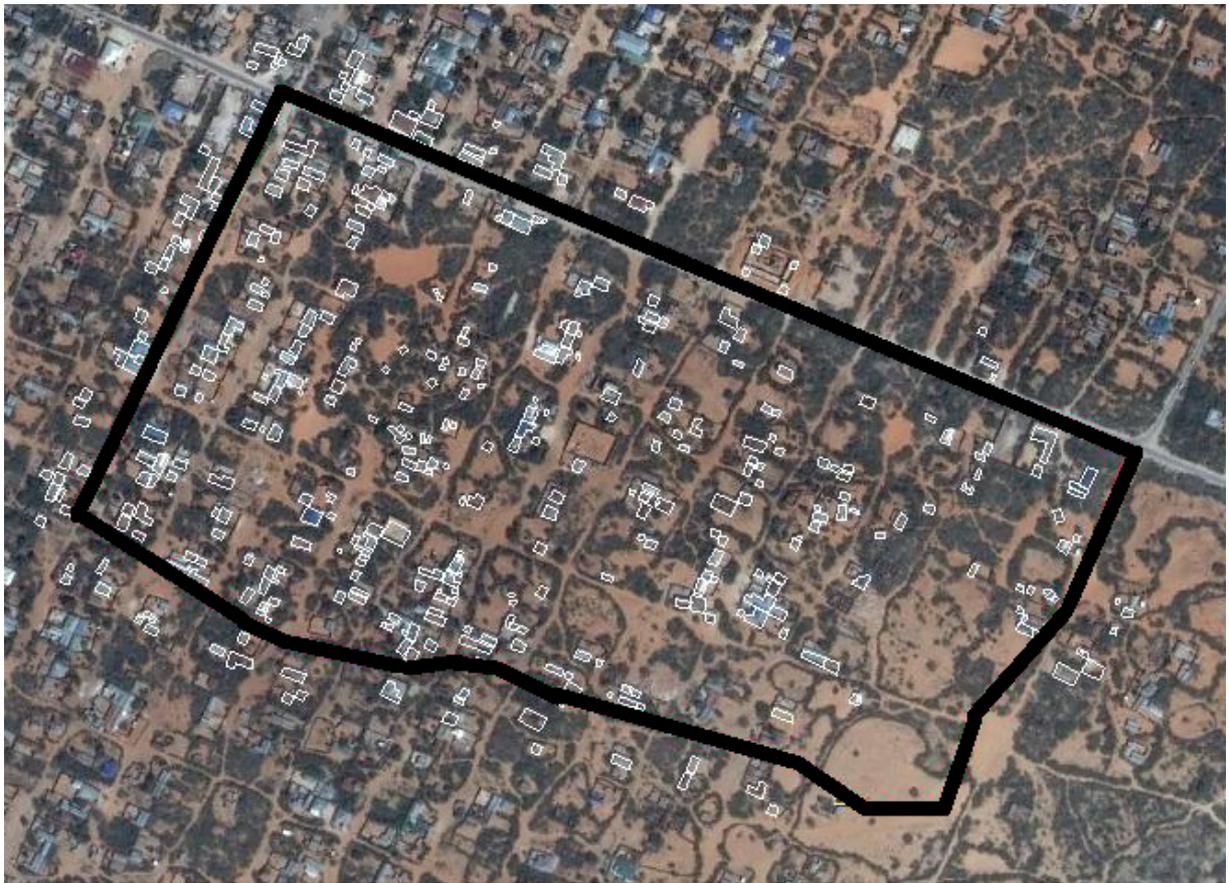
Table A1 : Description of sample PSUs

| Location | Total PSU Area (m ²) | Total PSU + Buffer Area (m ²) | Area in which no households would be selected with Mecca method (% of total) | Number of Structures | Number of Structures (including buffer) | Imagery date |
|------------|----------------------------------|---|--|----------------------|---|-------------------|
| Hodon | 42,615 | 95,707 | 1.4% | 309 | 408 | March 14, 2014 |
| Dharkinley | 24,390 | 65,447 | 13.0% | 68 | 207 | December 25, 2013 |
| Heliwa | 345,157 | 477,252 | 0.4% | 353 | 724 | March 14, 2014 |

Dharkinley



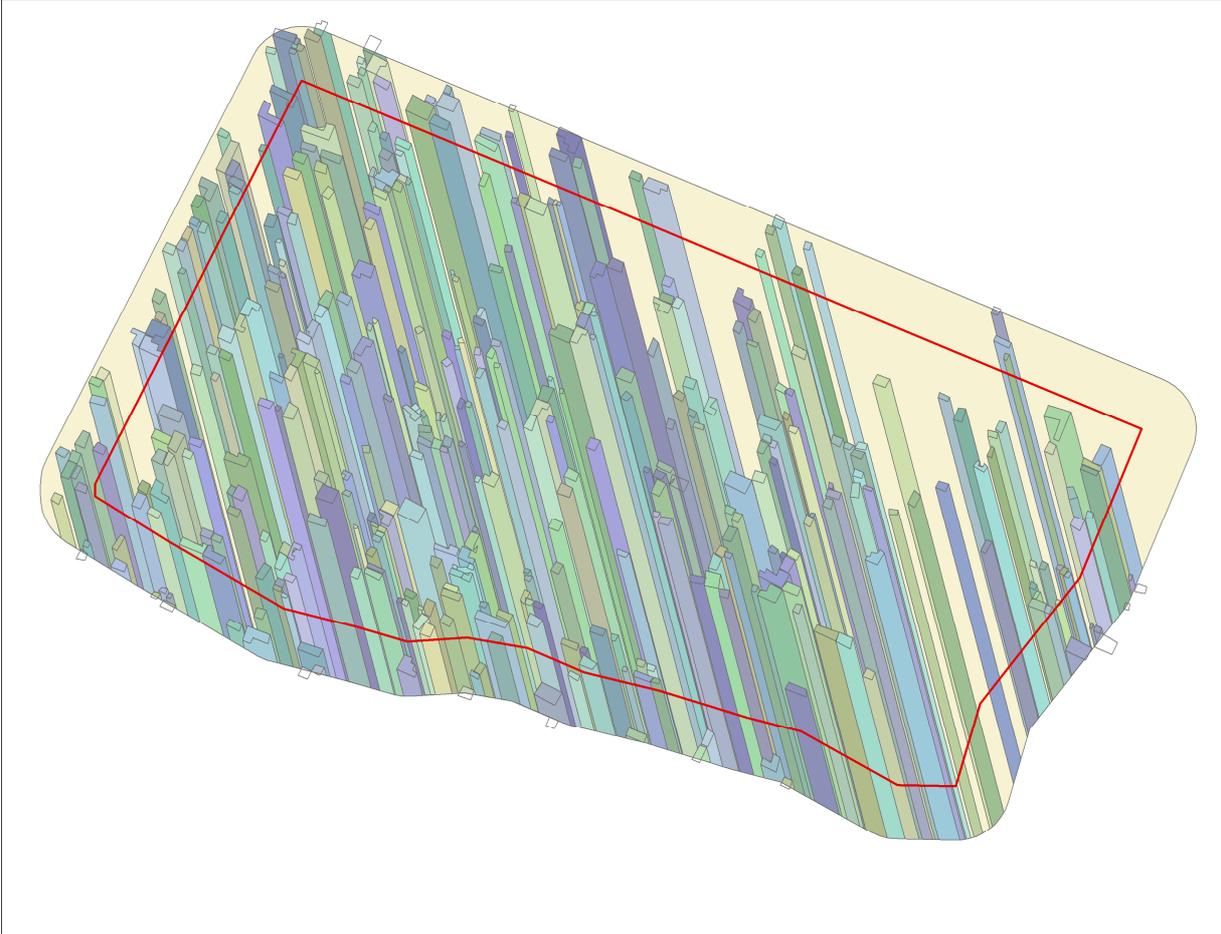
Heliwa



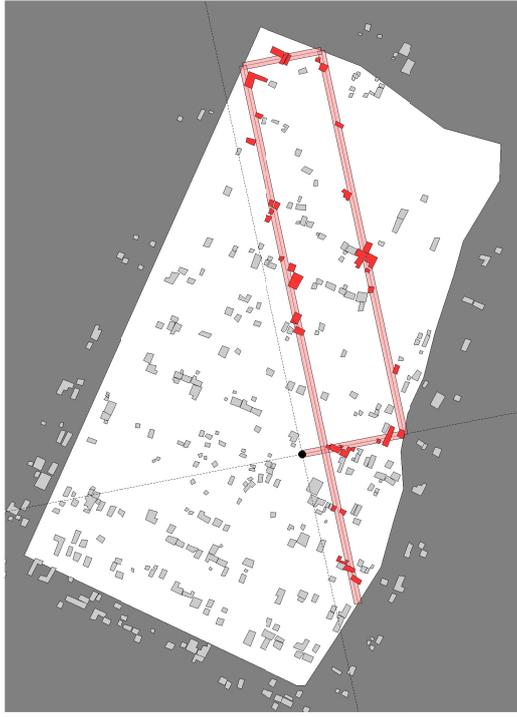
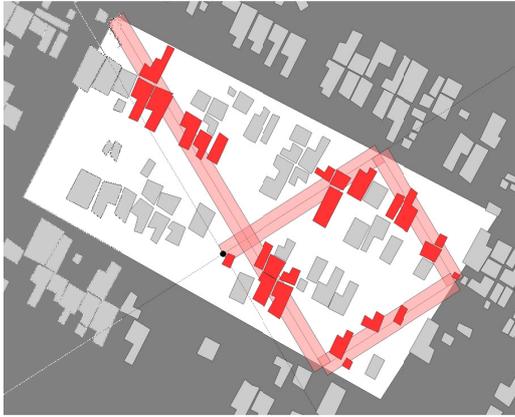
Hodon



Example of “shadows” for Mecca method



Example of path of random walk



Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu High Frequency Survey

Utz Pape and Johan Mistiaen

16th June 2015

Abstract

For Mogadishu, no poverty estimates exist because security risks constrain a face-to-face interview time to 60 minutes. This excludes traditional methods to estimate poverty based on time-consuming household consumption surveys. This paper presents the first approach to estimate total consumption reliably in such a context. Based on an innovative questionnaire design, the administering time is less than one hour for any households allowing fast and cost-efficient data collection even in areas with high security risks. The questionnaire design selects core consumption items administered to all households while remaining consumption items are algorithmically partitioned into optional modules assigned systematically to households. After data collection, multiple imputation techniques are used to estimate total household consumption. Based on *ex post* simulations, the approach is demonstrated to yield reliable estimates of poverty using household budget data from Hergeiza. The approach is then applied as part of the High Frequency Survey in Mogadishu to estimate consumption within 60 minutes of face-to-face interview time.

Introduction

Poverty is the paramount indicator to gauge socio-economic wellbeing of a population. Especially after a shock, poverty estimates can disentangle who in the population was affected how severely. As one of the main indicators for poverty, monetary poverty is measured by a welfare aggregate usually based on consumption in developing countries and a poverty line. The poverty line indicates the minimum level of welfare required for a healthy living.

Consumption aggregates are estimated traditionally by time-consuming household consumption surveys. A household consumption questionnaire records consumption and expenditures for a comprehensive list of food and non-food items. With around 300 to 400 items, the administering time of the questionnaire often exceeds 90 – 120 minutes. In addition to higher costs due to longer administering time, response fatigue can increase measurement error especially for items at the end of the questionnaire. In a fragile country context, a face-to-face time of 90 – 120 minutes can be

prohibitively high. For example, security concerns restricted the duration of a visit in Mogadishu to about 60 minutes.

The extensive nature of household consumption surveys make it difficult to obtain updated poverty estimates especially when they are needed the most: after a shock and in fragile countries. Therefore, approaches were developed to reduce administering time to allow collection of consumption data with significantly lower administering time. The most straight-forward approach to minimize administering time reduces the number of items either by asking for aggregates or by skipping less frequently consumed items, which we call reduced consumption methodology. However, both approaches have been shown to under-estimate consumption, which in turn over-estimates poverty.¹ Splitting up the questionnaire for multiple visits is another solution but attrition issues – especially in fragile country contexts – increase required sample size and also have a high cost implication. In addition, multiple visits to the same household can increase security concerns.

A second class of approaches utilizes a full consumption baseline survey and updates poverty estimates based on a small subset of collected indicators.² These approaches estimate a welfare model on the baseline survey using a small number of easy-to-collect indicators. This allows updating poverty estimates by collecting only the set of indicators instead of direct consumption data. While the approach is cost-efficient and easy to implement in normal circumstances, the approach has two major drawbacks in the context of fragility and shocks. First, the approach requires a baseline survey, which is sometimes – for example in Mogadishu – not existent. Second, the approach relies on a structural model estimated from the baseline survey.³ In the case of shocks, the structural assumptions, which cannot be tested, are often violated. Thus, poverty updates based on the violated assumption tend to under-estimate the impact of the shock on poverty. Therefore, cross-survey imputation methodologies are not applicable in the context of shocks and fragility.

We propose a new methodology combining an innovative questionnaire design with standard imputation techniques. This substantially reduces the administering time of a consumption survey to about 60 minutes while at the same time credible poverty estimates are obtained. Thus, the gain in administering time is bought by the need to impute missing consumption values. Due to the design of the questionnaire, the method circumvents systematic biases as identified for alternative methodologies.

After explaining the methodology in more detail in the next section, we will assess the performance of the methodology *ex post* using collected household budget data in Hergeiza, Somalia. Next, we apply the methodology to newly collected data in Mogadishu, Somalia, where full consumption data collection was impossible due to security constraints. We evaluate the consistency of the consumption estimates by performing validity checks. We conclude with a discussion of the limitations of the methodology, the benefits especially in combination of using CAPI technology and the need for further research.

¹ Beegle et al, 2012.

² Doudich et al, 2013; SWIFT

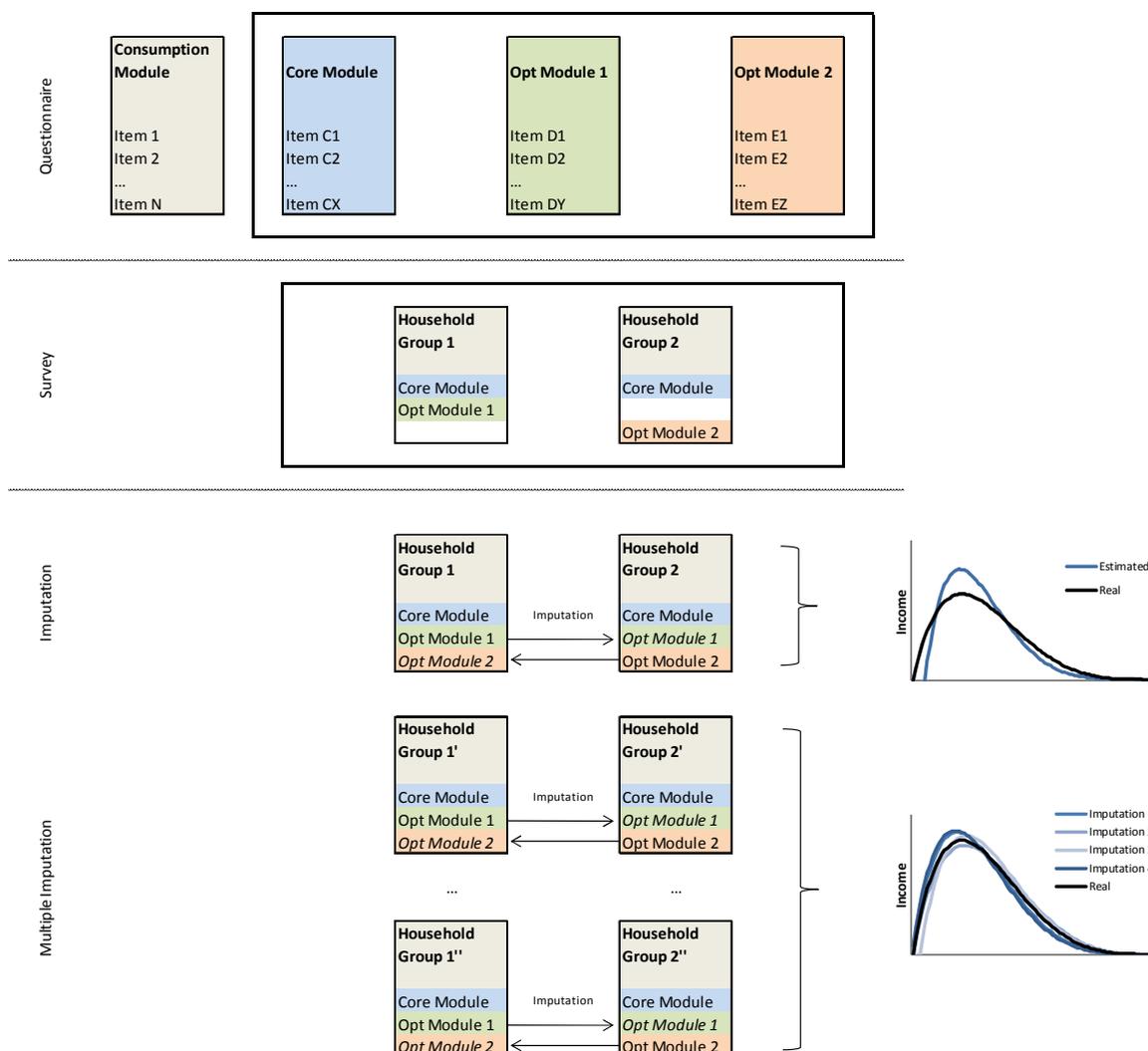
³ Christiaensen et al, 2010; Christiaensen et al, 2011.

Methodology

Overview

The rapid survey consumption methodology consists of five main steps (Figure 1). First, core items are selected based on their importance for consumption. Second, the remaining items are partitioned into optional modules. Third, optional modules are assigned to groups of households. After data collection, fourth, consumption of optional modules is imputed for all households. Fifth, the resulting consumption aggregate is used to estimate poverty indicators.

Figure 1: Illustration of the rapid consumption survey methodology (using illustrative data only). The consumption module is partitioned into core and optional modules, which in turn are assigned to households. Consumption is imputed utilizing the sub-sample information of the optional modules either by single or multiple imputation methods.



First, core consumption items are selected. Consumption in a country bears some variability but usually a small number of a few dozen items captures the majority of consumption. These items are assigned to

the core module, which will be administered to all households. Important items can be identified by its average food share per household or across households. Previous consumption surveys in the same country or consumption shares of neighboring / similar countries can be used to estimate food shares.⁴

Second, non-core items are partitioned into optional modules. Different methods can be used for the partitioning into optional modules. In the simplest case, the remaining items are ordered according to their food share and assigned one-by-one while iterating the optional module in each step. A more sophisticated method would take into account correlation between items and partition them into orthogonal sets per module. This would lead to high correlation between modules supporting the total consumption estimation.

Conceptual division into core and optional items should not be reflected in the layout of the questionnaire. More complicated partition patterns can result in a set of very different items in each module. However, the modular structure should not influence the layout of the questionnaire. Instead, all items per household will be grouped into categories of consumption items (like cereals) and different recall periods. Therefore, it is recommended to use CAPI technology, which allows hiding the modular structure of the consumption module from the enumerator.

Third, optional modules will be assigned to groups of households. Assignment of optional modules will be performed randomly stratified by enumeration areas to ensure appropriate representation of optional modules in each enumeration area. This step is followed by the actual data collection.

Fourth, household consumption will be estimated by imputation. The average consumption of each optional module can be estimated based on the sub-sample of households assigned to the optional module. In the simplest case, a simple average can be estimated. More sophisticated techniques can employ a welfare model based on household characteristics and consumption of the core items. We present six techniques in the next section and perform their performance on the dataset from Hergeiza.

Single imputation of the consumption aggregate under-estimates the variance of household consumption. Depending on the location of the poverty line relative to the consumption distribution, this can either consistently under- or over-estimate poverty. Multiple imputation based on bootstrapping can mitigate the problem but will render analysis more complicated. We use single as well as multiple imputation techniques for the evaluation of the methodology.

Module Construction

Consumption for a household is estimated by the sum of expenditures for a set of items

$$y_i = \sum_{j=1}^m y_{ij}$$

⁴ As shown later, the assignment of items to modules is very robust and, thus, even rough estimates of consumption shares are sufficient to inform the assignment without requiring a baseline survey.

where y_{ij} denotes the consumption of item j in household i . The list of items can be partitioned into $M+1$ modules each with m_k items:

$$y_i = \sum_{k=0}^M y_i^{(k)} \quad \text{with} \quad y_i^{(k)} = \sum_{j=1}^{m_k} y_{ikj}$$

For each household, only the core module $y_i^{(0)}$ and one additional optional module $y_i^{(k^*)}$ are collected.

The item assignment to the modules should be based on either a previous survey or a survey in a related country with similar consumption behavior. As the core module is administered to all households, it should include items covering the largest shares of consumption. Optional modules can be constructed in different ways. Currently, an algorithm is used to assign items iteratively to optional modules so that items are orthogonal within modules and correlated between modules. In each step, an unassigned item with highest consumption share is selected. For each module, total per capita consumption is regressed on household size, the consumption of all assigned items to this module as well as the new unassigned item. The item will be assigned to the module with the highest increase in the R2 relative to the regression excluding the new unassigned item. The sequenced assignment of items based on their consumption share can lead to considerable differences in the captured consumption share across optional modules. Therefore, a parameter is introduced ensuring that in each step of the assignment procedure the difference in the number of assigned items per module does not exceed d . Using $d=1$ assigns items to modules (almost) maximizing equal consumption share across modules.⁵ Increasing d puts increasing weight on orthogonality within and correlation between modules.

The assignment of optional modules must ensure that a sufficient number of households are assigned to each optional module. Household consumption can then be estimated using the core module, the assigned module and estimates for the remaining optional modules

$$\hat{y}_i = y_i^{(0)} + y_i^{(k^*)} + \sum_{k \in K^*} \hat{y}_i^{(k)}$$

where $K^* := \{1, \dots, k^* - 1, k^* + 1, \dots, M\}$ denotes the set of non-assigned optional modules.

Consumption Estimation

Consumption of non-assigned optional modules can be estimated by different techniques. Three classes each with two techniques are presented differing in their complexity and theoretical underpinnings. The first class of techniques simply uses summary statistics like the average to impute missing data. The second class is based on multiple univariate regression models. The third class uses multiple imputation techniques taking into account the variation absorbed in the residual term.

⁵ Even with $d=1$, equal consumption share across modules is not maximized because among the modules with the same number of assigned items, the new item will be assigned to the module it's most orthogonal to rather than to the module with lowest consumption share.

Summary Statistics (average and median)

This class of techniques applies a summary statistic on the collected module-specific consumption and applies the result to the missing modules. For each module k , the summary statistic f can be computed as

$$\hat{y}^{(k)} = f\left(\langle y_i^{(k)} \rangle_i\right).$$

For household i , household consumption is estimated as

$$\hat{y}_i = y_i^{(0)} + y_i^{(k^*)} + \sum_{k \in K^*} \hat{y}^{(k)}.$$

Thus, each household is assigned the same consumption per missing module. In the following, the average and the median are used as summary statistics. The median has the advantage of being more robust against outliers but cannot capture small module-specific consumption if more than half of the households have zero consumption for the module.

Module-wise Regression (OLS and tobit regression)

Module-wise estimation applies a regression model for each module. This allows capturing differences in core consumption as well as other household characteristics

$$\hat{y}_i^{(k)} = \beta_0^{(k)} y_i^{(0)} + x_i^T \beta^{(k)} + u_i^{(k)}$$

With x_i^T representing a vector of household characteristics and $u_i^{(k)}$ an error term assumed to be normally distributed with $N(0, \sigma^{(k)})$. Thus, module-wise estimation uses a regression separately for each module. Coefficients are estimated only based on the subsample assigned to module k . In general, a bootstrapping approach using the residual distribution could mimic multiple imputations but is not applied here. Given the impossibility of negative consumption, a tobit regression with a lower bound of 0 is used in addition to a standard OLS regression approach. For the OLS regression, negative imputed values are set to zero.

Multiple Imputation Chained Equations (MICE)

Multiple Imputation Chained Equations (MICE) uses a regression model for each variable and allows missing values in the dependent and independent variables. As missing values are allowed in the independent variables, the consumption of all optional modules can be used as explanatory variables:

$$\hat{y}_i^{(k)} = \beta_0^{(k)} y_i^{(0)} + \sum_{k' \in K^*} \beta_{k'}^{(k)} y_i^{(k')} + x_i^T \beta^{(k)} + u_i^{(k)}$$

Missing values in the explanatory variable ($y_i^{(k')}$) are drawn randomly in the first step. Iteratively, these values are substituted with imputed values drawn from the posterior distribution estimated from the regression for $\hat{y}_i^{(k')}$. While the technique of chained equations cannot be shown to converge in distribution theoretically, practical results are encouraging and the method is widely used.

Multi-Variate Normal Regression (Mimvn)

Multiple Imputation Multi-variate Normal Regression uses an EM-like algorithm to iteratively estimate model parameters and missing data. In contrast to chained equations, this technique is guaranteed to converge in distribution to the optimal values. An EM algorithm draws missing data from a prior (often non-informative) distribution and runs an OLS to estimate the coefficients. Iteratively, the coefficients are updated based on re-estimation using imputed values for missing data drawn from the posterior distribution of the model. Multiple Imputation Multi-variate Normal Regression employs a Data-Augmentation (DA) algorithm, which is similar to an EM algorithm but updates parameters in a non-deterministic fashion unlike the EM algorithm. Thus, coefficients are drawn from the parameter posterior distribution rather than chosen by likelihood maximization. Hence, the iterative process is a Monte-Carlo Markov –Chain (MCMC) in the parameter space with convergence to the stationary distribution that averages over the missing data. The distribution for the missing data stabilizes at the exact distribution to be drawn from to retrieve model estimates averaging over the missing value distribution. The DA algorithm usually converges considerably faster than using standard EM algorithms:

$$\hat{y}_i^{(k)} = \beta_0^{(k)} y_i^{(0)} + x_i^T \beta^{(k)} + u_i^{(k)}$$

Estimation Performance

The performance of the different estimation techniques is compared based on the relative bias (mean of the error distribution) and the relative standard error. We define the relative error as the percentage difference of the estimated consumption and the reference consumption (based on the full consumption module):

$$e_i = \frac{\hat{y}_i - y_i}{y_i}$$

The relative bias is the average of the relative error:

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i$$

The relative standard error is the standard deviation of the relative error:

$$se = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}$$

For estimation based on multiple imputations, e_i is averaged over all imputations.

Each proposed estimation procedure is run on random assignments of households to optional modules. A constraint ensures that each optional module is assigned equally often to a household per enumeration. The relative bias and the relative standard error are reported across all simulations.

The performance measures can be calculated at different levels. At the household level, the relative error is the relative difference in the household consumption. At the cluster level, the relative error is defined as the relative difference of the average reference household consumption and average estimated household consumption across the households in the cluster. Similarly, the global level compares total average consumption for all households.

Results

In this section, the rapid consumption methodology will first be applied to a dataset including a full consumption module from Hergeiza, Somalia. This will be used to assess the performance of the rapid consumption methodology compared to the traditional full consumption. Subsequently, we present results from the High Frequency Survey in Mogadishu. Security risks restrict face-to-face interview time to less than one hour. Therefore, we employed the rapid consumption methodology to derive the first ever consumption estimates for Mogadishu. We present the resulting consumption aggregate and perform consistency checks for its validation.

Ex-post Simulation

The rapid consumption methodology is applied *ex post* to household budget data collected in Hergeiza, Somalia. Hergeiza was chosen as it is the most similar city to Mogadishu. Using the full consumption dataset from Hergeiza allows a full-fledged assessment of the new methodology. Based on selected indicators, we compare the results after estimating consumption based on the rapid consumption methodology with the results from using the traditional full consumption module. We add a comparison with the results for a reduced consumption module.

The simulation assigns each household to one optional module. The consumption data for the modules not assigned to the household is deleted. Multiple simulations are performed with varying assignment of modules to households. Across the simulations, we calculate three consumption and four poverty and inequality indicators. The consumption indicators capture the accuracy of the estimation at three different levels: the household level, the cluster level (consisting of about 9 households) and the level of the dataset. In addition, we calculate the poverty headcount (FGT0), poverty depth (FGT1) and poverty severity (FGT2) as well as the Gini coefficient to capture inequality.

The six proposed estimation techniques presented in the previous section are compared based on 20 simulations with respect to their relative bias and relative standard error. All simulations used the same item assignment to modules using the algorithm as described with parameter $d=3$ (see Table 1 for the resulting consumption shares per module).⁶ The estimation techniques differ considerably in terms of performance. We also compare the techniques to using a reduced consumption module where the same consumption items are collected for all households. The number of items is equal to the size of the core and one optional module implying a comparable face-to-face interview time to the Rapid Consumption methodology.

⁶ We performed robustness checks with different item assignment to modules including setting the parameter $d=1$ and $d=2$. The estimation results are extremely robust to changes in the item assignment to modules.

Table 1: Number of items and consumption share captured per module.

| | Food | | Non Food | |
|----------|-----------------|----------------------|-----------------|----------------------|
| | Number of Items | Share of Consumption | Number of Items | Share of Consumption |
| Core | 33 | 92% | 25 | 88% |
| Module 1 | 17 | 3% | 15 | 3% |
| Module 2 | 17 | 2% | 15 | 3% |
| Module 3 | 15 | 2% | 15 | 4% |
| Module 4 | 17 | 2% | 15 | 3% |

Comparing the reduced consumption approach with the full consumption as reference, the reduced consumption approach suffers from an under-estimation of the consumption (Figure 2 and Table 3 in the appendix). This is not surprising because the approach only collects consumption from a subset of items. Applying the median as a summary statistic also results in an under-estimation of consumption. As consumption distributions have a long right tail, the median consumption belongs to a poorer household than the average household. In the case of Hergeiza, several optional modules have a median of zero consumption. Thus, the median underestimates the consumption similarly to the reduced consumption approach. In contrast, the average consumption of households is larger than the consumption of the median household. Thus, it is not surprising that the technique using the average as summary statistic over-estimates total consumption at the household and cluster level.

The regression techniques have a similar performance with a considerable upward bias at all levels. The tobit regression performs slightly better especially at the global level. In contrast, both multiple imputation techniques perform exceptionally well with a bias of around 1% at the household level, virtually unbiased at the cluster level and a minor downward bias of 0.7% at the global level.

Figure 2: Average Relative Bias at household, cluster, and simulation level for six estimation techniques.

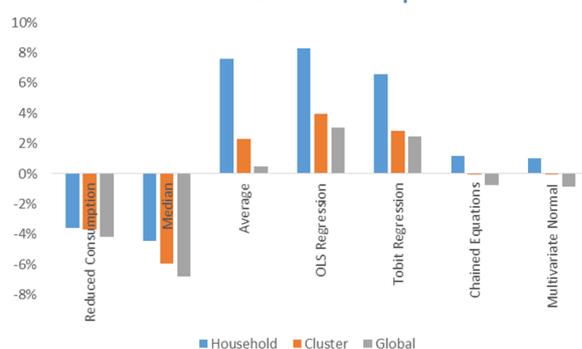
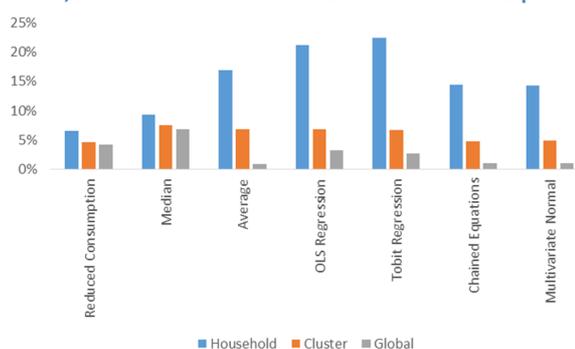


Figure 3: Average Relative Standard Error at household, cluster, and simulation level for six estimation techniques.



While the bias is important to understand systematic deviation of the estimation, the relative standard error helps to understand the variation of the estimation. Except in a simulation setting, the standard error of the estimation cannot be calculated as only one assignment of households to optional modules is available (Figure 3 and Table 3 in the appendix). Thus, it is important that the estimation technique delivers a small relative standard error.

Generally, the relative standard error reduces when moving from the household level over the cluster level to the global level. The relative standard error for the reduced consumption methodology is

smaller than for the summary statistic techniques because the reduced consumption is not subject to the variation from the module assignment to households. The regression techniques have large relative standard errors at the household level of around 20% while the multiple imputation techniques vary around 15%. At the cluster level, the relative standard error drops to 7% for regression techniques and 5% for multiple imputation techniques. At the global level, the relative standard error is around 3% for regression techniques and 1% for multiple imputation techniques.

The distributional shape of the estimated household consumption can be compared to the reference household consumption by employing standard poverty and inequality indicators. The poverty headcount (FGT0) is 24.6% for the reference distribution. Not surprisingly, the reduced consumption and the median summary statistic overestimate poverty by several percentage points due to the under-estimation of consumption (Figure 4 and Table 4 in the appendix). The average summary statistic and the regression techniques underestimate poverty since they overestimate consumption. The multiple imputation techniques over-estimate poverty but only by 0.3 percentage points performing significantly better than the reduced consumption approach with a more than five times larger bias. The reduced consumption and the median summary statistic as well as the multiple imputation techniques deliver good results for the FGT1 and FGT2 emphasizing that not only the headcount can be estimated reasonably well but also the distributional shape is conserved. Except for the median summary statistic, these techniques also perform well estimating the Gini coefficient with a bias of less than 0.5 percentage points. The relative standard errors show similar results as for the estimation of the consumption (Figure 5 and Table 4 in the appendix). While the relative standard error of the reduced consumption for FGT0 is double compared to the multiple imputation techniques, the relative standard errors for FGT1 are comparable but larger for FGT2 and Gini for the multiple imputation techniques.

Figure 4: Average Bias for FGT0, FGT1, FGT2 and Gini coefficient.

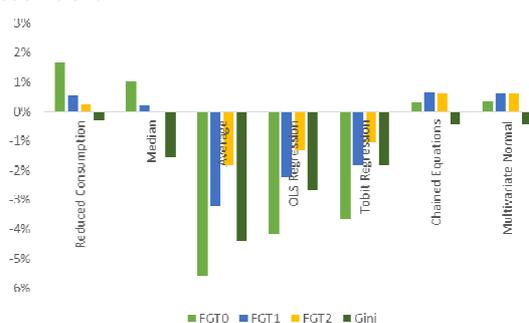
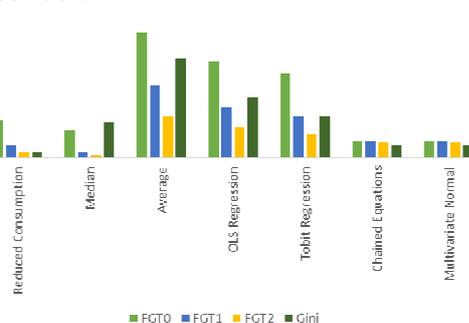


Figure 5: Average Standard Error for FGT0, FGT1, FGT2 and Gini coefficient.



In summary, the average summary statistic and the regression approaches cannot deliver convincing estimations. While the reduced consumption and the median summary statistic perform considerably better, they both over-estimate poverty by construction. Only the multiple imputation techniques can convince in all estimation exercises. Especially in the estimation of the important poverty headcount (FGT0), the multiple imputation techniques are virtually unbiased.

Application to Mogadishu

In late 2014, consumption data using the proposed rapid methodology was collected in Mogadishu using CAPI. The rapid consumption questionnaire did reduce face-to-face time considerably. A household visit

took about 40 minutes on average (median: 35 minutes) including greeting, household roster and characteristics, consumption module as well as a number of perception questions. Nine out of ten interviews took less than 65 minutes.

After data cleaning and quality procedures, 675 households with consumption data were retained.⁷ A welfare model was built to predict missing consumption in optional modules. We test the welfare model on the core consumption (after removing the core consumption as explanatory variable). The model for food consumption retrieves an R2 of 0.24 while non-food consumption is modeled with an R2 of 0.16 (see Table 5). It is important to emphasize that these models give a lower bound of the R2 compared to the models used in the prediction as the prediction models include the core consumption as explanatory variable. Given the assessment of the different estimation techniques in the last section, the multivariate normal approximation using multiple imputations is applied to the Mogadishu dataset.

For the Mogadishu dataset, the assignment of items to modules had to be refined manually.⁸ The refinement has minor impact on the share of consumption per module (Table 2). It is peculiar though that the share of consumption per module is different for Hergeiza and Mogadishu. Using the Hergeiza dataset, 91% of food consumption (76% for non-food consumption) is captured in the core module. In contrast, the core food consumption share is only 64% (for non-food consumption 62%) in Mogadishu before imputing consumption of non-assigned modules. Thus, employing a reduced consumption module based on consumption shares identified in Hergeiza would have crudely under-estimated consumption in Mogadishu without the possibility to evaluate the inaccuracy. In contrast, the rapid consumption methodology allows the estimation of shares for each module while the consumption estimation procedure implicitly takes into account the ‘missing’ consumption shares for each household.

Table 2: the number of items and consumption share captured per module simulated for Hergeiza, estimated for Mogadishu before imputation of non-assignment modules (normalized to 100%) and after imputing full consumption.

| | Food Consumption | | | | Non-Food Consumption | | | |
|----------|------------------|----------------|-----------------|-------------------------|----------------------|----------------|-----------------|-------------------------|
| | Number of Items | Share Hergeiza | Share Mogadishu | Share Mogadishu Imputed | Number of Items | Share Hergeiza | Share Mogadishu | Share Mogadishu Imputed |
| Core | 33 | 91% | 64% | 54% | 26 | 76% | 62% | 52% |
| Module 1 | 19 | 3% | 9% | 16% | 15 | 7% | 9% | 12% |
| Module 2 | 20 | 2% | 14% | 14% | 15 | 5% | 9% | 12% |
| Module 3 | 15 | 2% | 5% | 6% | 15 | 6% | 8% | 9% |
| Module 4 | 15 | 2% | 8% | 9% | 15 | 6% | 11% | 15% |

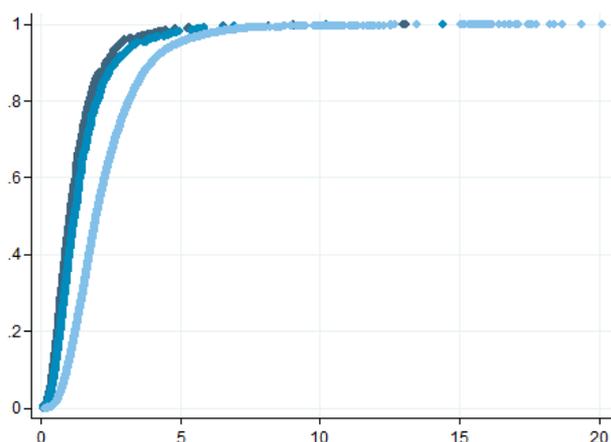
The cumulative consumption distribution can be compared for the consumption captured in the core module, the consumption captured in the core and the assigned optional module and the imputed consumption (Figure 6). By construction, the core consumption shows the lowest consumption per

⁷ While the survey also covered IDP camps, the presented analysis is restricted to households in residential areas excluding IDP camps.

⁸ The manual refinement is necessary to ensure that items like ‘other fruits’ cannot double count type of fruits not assigned to the household. This is implemented by relabeling and manual assignment to modules. In addition, some items grouping several sub-items were split into single items, which is generally preferable for recall and recording as well as calculation of unit values.

household. Adding the consumption from the assigned optional module shifts the cumulative consumption curve slightly. The imputed consumption is shifted even further as the estimated consumption shares from the non-assigned module are added as well.

Figure 6: Cumulative consumption distribution in current USD per day and capita for core module (dark blue), core and assigned optional module (medium blue) and imputed consumption (light blue).⁹



Without a full consumption aggregate available for Mogadishu, we can only show consistency of the retrieved consumption aggregate with other household characteristics to validate the estimates. Consumption per capita usually reduces with increasing household size. Indeed, we find that household size is significantly negatively correlated with estimated per capita consumption (coefficient: -0.04, t-statistic: -2.10, p-value: 0.04).¹⁰ Per capita consumption also decreases with a larger share of children among the household members (coefficient: -0.28, t-statistic: -1.66, p-value: 0.098). The proportion of employed members in the household significantly increases consumption per capita (coefficient: 0.51, t-statistic: 2.77, p-value: <0.01). Thus, the retrieved consumption estimate is consistent and – using the evidence from the *ex post* simulations – highly accurate.

Conclusions

The results from the *ex post* simulation indicate that the rapid consumption methodology can reliably estimate consumption and poverty. At the same time, the experience in Mogadishu showed that the rapid consumption methodology can be implemented in extremely high risk areas while succeeding in limiting face-to-face interview time to less than one hour. While these results are encouraging, the rapid consumption methodology has some limitations.

⁹ Note that the presented consumption aggregate does not include consumption from durables goods.

¹⁰ The reported numbers are corrected against correlation with household characteristics included in the welfare model. As the welfare model for the prediction of consumption includes household size, we have run a robustness check excluding household size from the welfare model used for prediction. The correlation between consumption per capita and household size is still significant (coefficient: -0.03, t-statistic: -2.17, p-value: 0.03).

The rapid consumption questionnaire varies comprehensiveness and order of items in the consumption module between households. The effect of a response bias due to this neither can be estimated from the simulations nor from the data collected in Mogadishu. However, an enhanced design with different optional modules varying in their comprehensiveness of items can shed light on this bias. Comparison between responses for the same item in a comprehensive and an incomprehensive list would indicate a lower bound for response bias. Assuming that the context of a comprehensive list is a better estimate, the response bias could be corrected for.

The rapid consumption survey methodology can increase the gap between capacity at enumerator level and complexity of survey instrument. Capacity at the enumerator level is often low in developing countries – especially in a fragile context. The rapid consumption survey methodology increases complexity of the questionnaire, which can further increase the gap between existing and required capacity at the level of enumerators. However, CAPI technology can seal off complexity from enumerator as software can automatically create the consumption module based on core and optional modules for each household without showing the partition to the enumerator. In Mogadishu, advanced CAPI technology was used generating the questionnaire automatically based on the assignment of the household to an optional module. While enumerators were made aware that different households will be asked for different items, administering the rapid consumption questionnaire did not require any additional training of enumerators beyond standard consumption questionnaires.

Analysis of rapid consumption survey data requires high capacity. Analysis capacity is usually limited in developing – and especially fragile – countries. While the general idea of assignment of optional consumption modules to households will be digestible by local counterparts, poverty analysis based on bootstrapped sample of consumption distribution is likely to overwhelm local capacity. However, even standard poverty analysis is often out of limits for local capacity in fragile countries. Therefore, capacity building usually focuses on data collection skills with a long-term perspective to increase data analysis capacity. In addition, the rapid consumption survey methodology might be the only possibility to create poverty estimates in certain areas, for example Mogadishu.

The results of the ex-post simulation and the application in Mogadishu suggest that the rapid consumption methodology can be a promising approach to estimate consumption and poverty in a cost-efficient and fast manner even in fragile areas.¹¹ A similar ex-post simulation for South Sudan (data not shown) indicates that the rapid consumption methodology can also be applied at the country-level with large intra-country consumption variation.¹² Further research can help further refining the methodology and estimation techniques. A better understanding of the relationship between the number of items in the core module and the number of optional modules with the accuracy of the resulting estimates can help to further optimize the methodology. Also the algorithm for the assignment of items to modules was designed *ad hoc* and can certainly be further improved. The estimation techniques can be optimized utilizing different techniques and more appropriate welfare models, for example including locational

¹¹ Costs for implementing a rapid consumption survey are lower than conducting a full consumption survey due to the reduced face-to-face time allowing enumerators to conduct more interviews per day.

¹² On-going field work employs the rapid consumption methodology currently in South Sudan to update poverty numbers.

random effects. Finally, ultimate validation of the rapid consumption methodology should come from a parallel implementation of a full consumption survey and the rapid consumption methodology to directly compare estimates.

References

- Ahmed, F., C. Dorji, S. Takamatsu and N. Yoshida (2014), “Hybrid Survey to Improve the Reliability of Poverty Statistics in a Cost-Effective Manner”, Policy Research Working Paper 6909, World Bank.
- Beegle, K., J. De Weerd, J. Friedman and J. Gibson (2012), “Methods of household consumption measurement through surveys: Experimental results from Tanzania”, *Journal of Development Economics* 98 (1), 3 – 18.
- Christiaensen, L., P. Lanjouw; J. Luoto and D. Stifel (2010). “The Reliability of Small Area Estimation Prediction Methods to Track Poverty,” Mimeo, Development Research Group, the World Bank, Washington D.C.
- Christiaensen, L., P. Lanjouw, J. Luoto and D. Stifel (2011), “Small Area Estimation-Based Prediction Methods to Track Poverty: Validation and Applications”, *Journal of Economic Inequality* 10 (2), 267 – 297.
- Deaton, Angus (2000), “The Analysis of Household Surveys: A Micro-econometric Approach to Development Policy”, Published for the World Bank, The Johns Hopkins University Press, Baltimore and London (third edition)
- Deaton A. and S. Zaidi (2002). “Guidelines for Constructing Consumption Aggregates for Welfare Analysis”. LSMS Working Paper 135, World Bank, Washington, DC.
- Deaton A. and J. Muellbauer (1986). “On measuring child costs: with applications to poor countries”. *Journal of Political Economy* 94, 720 -44.
- Doudich, M., A. Ezrari, R. van der Weide and P. Verme (2013), “Estimating Quarterly Poverty Rates Using Labor Force Surveys”, Policy Research Working Paper 6466, The World Bank.
- Dorji, C., and N. Yoshida. 2011. “New Approaches to Increase Frequent Poverty Estimates.” Unpublished manuscript.
- Elbers C., J. O. Lanjouw and P. Lanjouw (2002). “Micro-Level Estimates of Poverty and Inequality”. *Econometrica* 71:1, pp. 355 – 364.
- Elbers, C., J. Lanjouw and P. Lanjouw (2002), “Micro-Level Estimation of Welfare”, Policy Research Working Paper 2911, DECRG, The World Bank.
- Elbers, C., J. Lanjouw and P. Lanjouw (2003), “Micro-Level Estimation of Poverty and Inequality”, *Econometrica* 71 (1), 355 – 364.
- Faizuddin, A., C. Dorji, S. Takamatsu and N. Yoshida (2014), “Hybrid Survey to Improve the Reliability of Poverty Statistics in a Cost-Effective Manner”, World Bank Working Paper.
- Foster J., E. Greer and E. Thorbecke (1984). “A class of decomposable poverty measures”. *Econometrica*, Vol. 52, No. 3, pp. 761-766.

Fujii, Tomoki and Roy van der Weide (2013), "Cost-Effective Estimation of the Population Mean Using Prediction Estimators", Policy Research Working Paper 6509, The World Bank.

Haughton, J. and S. Khander (2009). "Handbook on Poverty and Inequality". The World Bank.

Hentschel J. and P. Lanjouw (1996). "Constructing an indicator of consumption for the analysis of poverty: Principles and Illustrations with Principles to Ecuador". LSMS Working Paper 124, World Bank, Washington, DC.

Howes S. and J.O. Lanjouw (1997). "Poverty Comparisons and Household Survey Design". LSMS Working Paper 129, World Bank Washington, DC.

Lanjouw P., B. Milanovic, and S. Paternostro (1998). "Poverty and the economic transition : how do changes in economies of scale affect poverty rates for different households?". Policy Research Working Paper Series 2009, The World Bank.

Newhouse, D., S. Shivakumaran, S. Takamatsu and N. Yoshida (2014). "How Survey-to-Survey Imputation Can Fail", Policy Research Working Paper 6961, World Bank.

Ravallion, Martin (1994), "Poverty Comparisons", Fundamentals of Pure and Applied Economics 56, Hardwood Academic Publishers

Ravallion M. (1996). "Issues in Measuring and Modelling Poverty". Economic Journal, Royal Economic Society, vol. 106(438), pages 1328-43, September.

Ravallion M. (1998). "Poverty Lines in Theory and Practice". Papers 133, World Bank - Living Standards Measurement.

Ravallion M., M. Lokshin (1999). "Subjective Economic Welfare", World Bank Policy Research Working Paper No. 2106.

Appendix

Table 3: Bias and relative error for consumption aggregate at the household, cluster and global level.

| Method | Household | | Cluster | | Global | |
|---------------------|-----------|-------|---------|------|--------|------|
| | Bias | SE | Bias | SE | Bias | SE |
| Reduced Consumption | -3.6% | 6.5% | -3.7% | 4.6% | -4.2% | 4.2% |
| Median | -4.4% | 9.3% | -5.9% | 7.5% | -6.8% | 6.8% |
| Average | 7.6% | 16.9% | 2.3% | 6.8% | 0.5% | 0.9% |
| OLS Regression | 8.3% | 21.1% | 4.0% | 6.8% | 3.1% | 3.2% |
| Tobit Regression | 6.6% | 22.4% | 2.8% | 6.7% | 2.5% | 2.7% |
| Chained Equations | 1.1% | 14.4% | 0.0% | 4.7% | -0.7% | 0.9% |
| Multivariate Normal | 1.0% | 14.2% | -0.1% | 4.8% | -0.9% | 1.1% |

Table 4: Bias and relative error for FGT0, FGT1, FGT2 and Gini for different estimation techniques.

| Method | FGT0 | | FGT1 | | FGT2 | | Gini | |
|---------------------|-------|------|-------|------|-------|------|-------|------|
| | Bias | SE | Bias | SE | Bias | SE | Bias | SE |
| Reduced Consumption | 1.7% | 1.7% | 0.6% | 0.6% | 0.3% | 0.3% | -0.3% | 0.3% |
| Median | 1.0% | 1.2% | 0.2% | 0.3% | 0.0% | 0.1% | -1.6% | 1.6% |
| Average | -5.6% | 5.6% | -3.2% | 3.2% | -1.8% | 1.8% | -4.4% | 4.4% |
| OLS Regression | -4.2% | 4.3% | -2.3% | 2.3% | -1.3% | 1.4% | -2.7% | 2.7% |
| Tobit Regression | -3.7% | 3.7% | -1.8% | 1.9% | -1.1% | 1.1% | -1.8% | 1.9% |
| Chained Equations | 0.3% | 0.8% | 0.6% | 0.7% | 0.6% | 0.7% | -0.5% | 0.6% |
| Multivariate Normal | 0.4% | 0.7% | 0.7% | 0.7% | 0.6% | 0.7% | -0.4% | 0.5% |

Table 5: Test of Welfare Model on core consumption reporting coefficients (t-statistics) for Mogadishu.

| Variable | Core Food Consumption | Core Non-Food Consumption |
|-----------------------------------|-----------------------|---------------------------|
| Core Food Consumption | | |
| ... 2nd Quartile | 0.78 (1.17) | |
| ... 3rd Quartile | 0.09 (1.46) | |
| ... 4th Quartile | 0.52 (7.22) | |
| Core Non-Food Consumption | | |
| ... 2nd Quartile | | 0.07 (1.11) |
| ... 3rd Quartile | | 0.12 (1.77) |
| ... 4th Quartile | | 0.42 (5.81) |
| Household Size | -0.07 (-8.36) | -0.04 (4.34) |
| Household Head Education | 0.16 (3.34) | 0.12 (2.56) |
| Dwelling Characteristics | | |
| ... Shared Apartment | 0.04 (0.59) | -0.13 (-2.12) |
| ... Separated House | -0.14 (-1.13) | -0.19 (-1.55) |
| ... Shared House | -0.07 (-0.81) | -0.14 (-1.52) |
| Water Access | | |
| ... Piped Water | -0.22 (-0.93) | -0.04 (-0.19) |
| ... Public Tap | 0.41 (2.47) | -0.01 (-0.08) |
| Insufficient Food in last 4 weeks | 0.05 (1.49) | -0.05 (-1.50) |
| R2 | 0.24 | 0.16 |
| N | 675 | 675 |

Utilizing Mobile Technology to Innovate CAPI Data Collection in Fragile Contexts

Utz Pape and Johan Mistiaen

16th June 2015

Summary

In recent years, the number of surveys using mobile devices like tablets has increased sharply. However, questionnaire design as well as data collection and monitoring are often replicated from traditional paper-based surveys ignoring the vast potential of mobile technologies. In fragile countries, survey implementation is more challenging and successful implementation requires tapping into this potential by innovating traditional methodologies.

This note argues that tablet technology in conjunction with three innovations can significantly enhance data quality especially in fragile countries. The first innovation is designed to improve data quality for consumption data by implementing a dynamic on-the-fly data validation system. This system flags unusually high or low entries and asks enumerators to confirm the correctness of the entry. The second innovation utilizes available software to monitor and manage tablets remotely. Using GPS tracking software, the location and trajectory of tablets can be determined – even retrospectively when the tablet re-enters 3G/WiFi areas. This helps to determine whether interviews were conducted at the correct location. Using remote management software, errors in the tablet configuration can be detected and solved while tablets can be updated remotely. The third innovation is a near real-time monitoring and analysis system. The monitoring system can identify challenges in the field work as well as weak enumerators early on and mitigate their impact on data quality. The analysis system calculates outcome indicators of the survey, e.g. consumption, education attainment or unemployment, to check incoming data while field work is still ongoing. The availability of the analysis code at the end of the data collection considerably accelerates the process from data collection to publication of results. This is most important in fragile countries where timeliness of data is paramount.

Overview

In fragile and conflict-affected states (FCS), “stresses” originating from social, economic, political and security dynamics can quickly compromise stability, throw countries back into conflict and reverse development progress (World Bank, 2011). Growing dissatisfaction with

economic welfare at the household level as well as other socio-economic indicators can indicate destabilizing pressures in a country. In such contexts, it is vital to monitor and detect potential “stresses” by closely tracking key indicators reflecting economic, social, and security conditions. At the same time, monitoring of data collection is often prohibitively difficult in such countries. This note identifies two major constraints for data collection in fragile states. First, field access is often limited to monitor data collection. This puts data quality at risk if enumerators cannot be sufficiently supervised. Second, the context is highly volatile demanding timely data collection with swift results. This note argues that computer-assisted-personal-interviews (CAPI) can be used to collect high-quality consumption data and proposes advanced survey design, monitoring and analysis techniques to enable data collection in fragile contexts.

Using paper based surveys (PAPI: paper assisted personal interview), enumerators can hardly be reached once they embarked on field work. This restricts monitoring especially in fragile countries where field access can be limited for field coordinators and managers. In addition, collected questionnaires only become available to the analysis team at the end of the field work. This delays analysis and publication of results, which can render collected data outdated even before published in a context with high volatility.

In contrast, CAPI utilizes technology to move the enumerator virtually closer to the field management and the data analysis team. The specific constraints for monitoring and the need for timely data analysis in fragile countries can be addressed by employing CAPI. This note extends lessons learnt from a pilot CAPI survey in South Sudan (World Bank, 2015) by proposing three innovations to deliver high quality household data using CAPI face-to-face interviews.

The first innovation improves data quality through dynamic on-the-fly data validation. The second innovation substitutes physical field monitoring by an online monitoring and management system. The third innovation conducts near real-time analysis improving data quality and shortening the time between end of field work and publication of results. The innovations are generally applicable to CAPI surveys except for the first innovation which is most relevant in the case of CAPI consumption surveys.

Barriers in Fragile Countries

Data collection is generally challenging in most developing countries due to capacity issues and difficult field access. In fragile states, additional challenges are present that are unique to these environments. First, security negatively affects feasibility of sampling, duration of face-to-face interview and capability for field monitoring. Second, volatility requires swift publication of results from data collection efforts.

Barrier 1: Unstable security situations increase risks throughout the data collection process and limits access, making it difficult to adequately monitor data quality.

Security issues can limit the amount of time field teams are able to spend at households to conduct interviews.¹ This is a critical barrier to collect consumption data, which requires face-to-face time of 120 minutes or more (Beegle et al., 2012). However, a newly developed approach can estimate total consumption reliably in such a context (Pape & Mistiaen, 2015). Based on an innovative questionnaire design, the administering time of the consumption module is less than one hour for any households allowing fast and cost-efficient data collection even in areas with high security risks. The questionnaire design selects core consumption items administered to all households while remaining consumption items are algorithmically partitioned into optional modules assigned systematically to households. After data collection, multiple imputation techniques are used to estimate total household consumption.

Monitoring of fieldwork is challenging in fragile environments where field access especially for international staff is limited. Thus, monitoring the conduct and quality of interviews is difficult or impossible. This increases the number of inconsistent or incorrect data entries as well as fake interviews and generally deteriorates data quality. This note presents dynamic on-the-fly data validation to reduce the number of incorrect / inconsistent entries and online monitoring and management to compensate generally for limited field oversight.

Collecting sensitive data in fragile state contexts also raises data security issues. Hence, extra precautions must be taken to safeguard household addresses and GPS locations, contact details and personal information about household members. Tablets offer a convenient way to protect respondent information by encryption. Android allows full encryption of the memory, thus, even extracted memory from a stolen device cannot be deciphered.

Barrier 2: In fragile states, where situations are volatile and circumstances on the ground can change rapidly, governments and development partners need access to relevant and timely data to inform the development of responsive policies and programs.

Questionnaires need to be adapted almost in real time to collect data that is relevant. Using PAPI, the questionnaire is fixed once it is printed and distributed to the enumerators. CAPI technology allows to make changes to the questionnaire, which can be downloaded by enumerators. While this offers the opportunity to adapt the questionnaire in real time, this process must be managed carefully. Adequate testing of the instrument in a pilot is essential and reduces the likelihood considerably that the questionnaire must be changed later. In the case that the questionnaire must be changed while field work started, enumerators must be trained to download questionnaire updates and made aware which questions changed. Otherwise, there is a risk that enumerators, who have learnt many questions by heart, do not reflect the changes when asking the respondent.

¹ The security situation can also limit the feasibility to conduct a listing in enumeration areas. The implications are discussed and solutions offered in Himelein, Eckman & Murray, 2015.

Data also needs to be processed, analyzed and delivered to stakeholders in a timely manner before becoming irrelevant. Processing of paper questionnaires creates a lag of typically a year or more between data collection and its availability to policy makers. By the time the results reach key decision makers, they are often already outdated and it is too late to address identified stress factors. An illustrative example is the last socio-economic welfare survey in South Sudan that took place in 2009. This data is still being used to inform the design of policies and programs in South Sudan, despite the fact that it preceded a period of rapid change in the country, including the country's independence and large-scale economic shocks and conflict events. This note proposes near real-time analysis to fully harvest the opportunities offered by CAPI surveys.

The following exposition assumes that Android devices are used for data collection. Android systems are less costly and offer more flexibility than iOS devices. The proposed innovations can be implemented with Apps and tools already available either at low or no extra cost. Hence, they do not require the development of a complex proprietary system, which can only be modified at high costs. Instead, the proposed modular system based on commercial Apps guarantees that always the most functional Apps can be integrated instead of being locked into a proprietary system.

Dynamic on-the-fly Data Validation

The first innovation improves traditional input constraints (e.g. age of respondent cannot be above 150) by a dynamic on-the-fly data validation procedure. The innovation substitutes time-consuming post-fieldwork data cleaning by real-time data validation. This improves data quality and reduces time required for data cleaning.

In this note, we present dynamic on-the-fly data validation specifically for consumption questions, but in principle the approach can be adopted for other type of questions. Consumption questions aim to measure the consumed quantity of an item as well as the price paid for this item. Since the consumed quantity can be different from the purchased quantity, the structure of the question is usually similar to the following²:

- A. Did you consume any <item> in the last <recall>?
 - [Yes / No]
- B. How much <item> did you consume in the last <recall>?
 - [quantity] [unit]
- C. How much <item> did you purchase in the market in the last <recall>?
 - [quantity] [unit]
- D. How much did you pay for this purchase?
 - [value] [currency]

² A skipping pattern ensures that questions B, C and D are only asked if the answer to question A was 'Yes'. Similarly, question D is only asked if the quantity in question C is greater than 0.

In traditional tablet surveys, questions B, C and D will have constraints to improve data quality. For example, [quantity] in question B must be greater than 0. However, a non-dynamic upper bound for [quantity] is difficult to decide since the choice for [unit] determines the range. For products measured in weight, [unit] can be a choice between ‘kg’, ‘g’ and others. A [quantity] of 0.1g is equally unlikely as a [quantity] of 100kg while 0.1kg as well as 100g is normal. Therefore, static on-the-fly data validation is very limited for these type of questions leading to substantial post survey data cleaning needs.

Our new dynamic on-the-fly data validation approach allows to formulate unit-specific constraints. In addition, it implements flagging of unlikely but not impossible entries. Also, the implied unit value³ of the product is compared with a hidden table to flag entries outside a pre-defined interval. Finally, units like cups with non-standard conversion factors into standard units are estimated on-the-fly by enumerators.

Before the approach can be implemented, additional information about ranges of quantities or values per unit or currency must be defined. The information can easily be captured in three tables (Table 1, Table 2 and Table 3). While this requires additional effort in the design phase of the questionnaire, similar information is necessary to clean data of traditional surveys. Thus, the effort is merely shifted from post-survey forward to the design phase. In practice, this helps to better understand the implications of design decisions like which units should be allowed for which item. In addition, this approach forces to capture the required information in a structured way, which is helpful for documentation and reproducibility.

Table 1: Item table capturing recall, standard unit, unit values and ranges as well as eligible units.

| itemid | itemlabel | recall | unit_std | u_value | u_vmin | u_vmax | unit_others |
|--------|------------|--------|----------|---------|--------|--------|-------------------|
| 10101 | Dura | 7 | kg | 6.92 | 2.73 | 14.12 | basin10000;cup200 |
| 10401 | Fresh milk | 7 | l | 16.70 | 7.95 | 40.00 | cup200 |

Table 2: Unit table capturing conversion, ranges and conversion flag.

| name | short | grams | odk_min | odk_max | max | odk_convert |
|------------------|------------|-------|---------|---------|-----|-------------|
| basin (10 litre) | basin10000 | 10000 | 0.5 | 4 | 50 | 0 |
| cup (200g) | cup200 | 200 | 1 | 10 | 50 | 1 |
| kilogram | kg | 1000 | 0.1 | 10 | 50 | 0 |
| litre | l | 1000 | 0.1 | 10 | 50 | 0 |

Table 3: Currency table including eligible range and conversion factor.

| curr_name | odk_min | odk_max | conversion |
|-----------|---------|---------|------------|
| USD | 1 | 100 | 1 |
| Ssh | 1000 | 2000000 | 0.000045 |

³ The unit value is the price divided by the quantity in standard units. The standard unit often is kg so that the unit value is the price per kg.

Based on this information, the questions B to D are extended based on the following logic:

- B. How much <item> did you consume in the last <recall>?
- [quantity] [unit] with eligible [unit] for item (Table 1) and constraint for quantity to be smaller than {max} for selected [unit] (Table 2)
 - If [quantity] is less than {flag_min} or greater than {flag_max} for selected [unit] (Table 2), request confirmation from enumerator that entries are correct.
 - If {odk_convert} in Table 2 is 1, ask enumerator to estimate the quantity in the standard unit (Table 1)
- C. How much <item> did you purchase in the market in the last <recall>?
- [quantity] [unit] with eligible [unit] for item (Table 1) and constraint for quantity to be smaller than {max} for selected [unit] (Table 2)
 - If [quantity] is less than {flag_min} or greater than {flag_max} for selected [unit] (Table 2), request confirmation from enumerator that entries are correct.
 - If {odk_convert} in Table 2 is 1, ask enumerator to estimate the quantity in the standard unit (Table 1)
- D. How much did you pay for this purchase?
- [value] [currency]
 - If [value] is less than {flag_min} or greater than {flag_max} for selected [currency] (Table 3), request confirmation from enumerator that entries are correct.
 - If calculated unit value (with [quantity] converted by {grams} in Table 2 and [value] converted by {conversion} in Table 3) is less than {u_vmin} or greater than {u_vmax} for item (Table 1), request confirmation from enumerator that entries are correct.

The implemented logic ask enumerators for confirmation whenever an entry is outside an expected range. Above logic is only schematic, the implemented solution specifically tells enumerators why the confirmation is required, e.g. the value is too low. Using this logic, the battery of questions records entries within a reasonable range while outliers carry an extra confirmation from enumerators that they are correct. This ensures that data collection is not biased towards less extreme values by prohibiting larger entries.

In-the-field conversion factors are obtained by aggregating the estimated quantities for non-standard units. The in-the-field conversion factors for non-standard units can be used in the analysis to convert quantities to standard units. In addition, it can be used to analyze the variance across enumerators, across products as well as across regions. The detailed analysis can inform whether quantities are converted by product or/and by region.

The implementation of the questionnaire logic requires a flexible questionnaire design software. In more detail, the implementation requires the ability to define loops over items with table lookups depending on the position in the loop. SurveyCTO, a derivate of the well-

known open source software FormHub, offers this ability.⁴ Equipped with the corresponding tables, the consumption module for multiple hundred items can be coded in less than 50 lines.

Online Monitoring and Management System

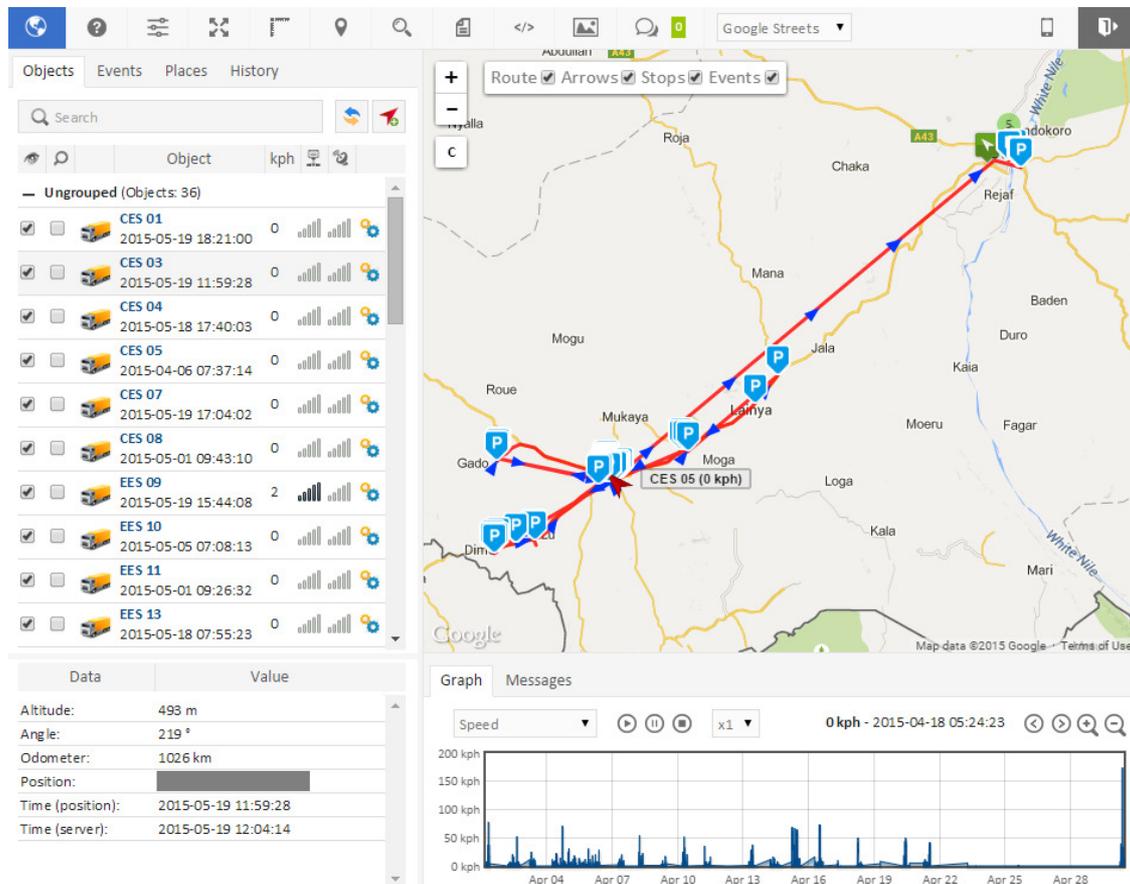
The second innovation proposes an online monitoring and management system to support data collection including near-real time tracking and remote management of tablets. Tracking includes localization of the tablet and its travel trajectory but also recording of audio snippets while interviews are conducted. This mitigates the negative impact of limited field monitoring and control visits. The remote management of tablets helps to solve technical issues with the tablets without a technician physically in the field.

After a screening of different GPS tracking software, we decided to use `gps-server.net` because it offers an inexpensive, easy to maintain and supervise solution to track large teams. An App installed on each tablet tracks the tablet in the background. If the tablet has no Wifi or 3G connection, the position is still tracked but not transmitted. Once the tablet reaches Wifi or 3G, the trajectory of the tablet is transmitted. An online platform can be used to organize teams of enumerators and monitor their current as well as past position (Figure 1 includes an example. Note that the software allows to zoom to higher resolution to locate the tablet with an accuracy of a few meters). The trajectory includes speed of movement, stops and also shows some vital tablet information like mobile signal reception. Especially stops are very useful to monitor whether interviews actually took place within the enumeration area and at the household position (if known). The online map can be overlaid with boundaries of enumeration areas as well as reported household locations to see whether interviews coincide with the actual household location.⁵

⁴ At the time of writing, SurveyCTO charged 200 USD per month per survey.

⁵ While this allows visual inspection, the real-time monitoring includes automatic flags to indicate whether households' locations are located within or outside enumeration areas.

Figure 1: Example of the trajectory of an enumerator in South Sudan tracked in the time of his field work

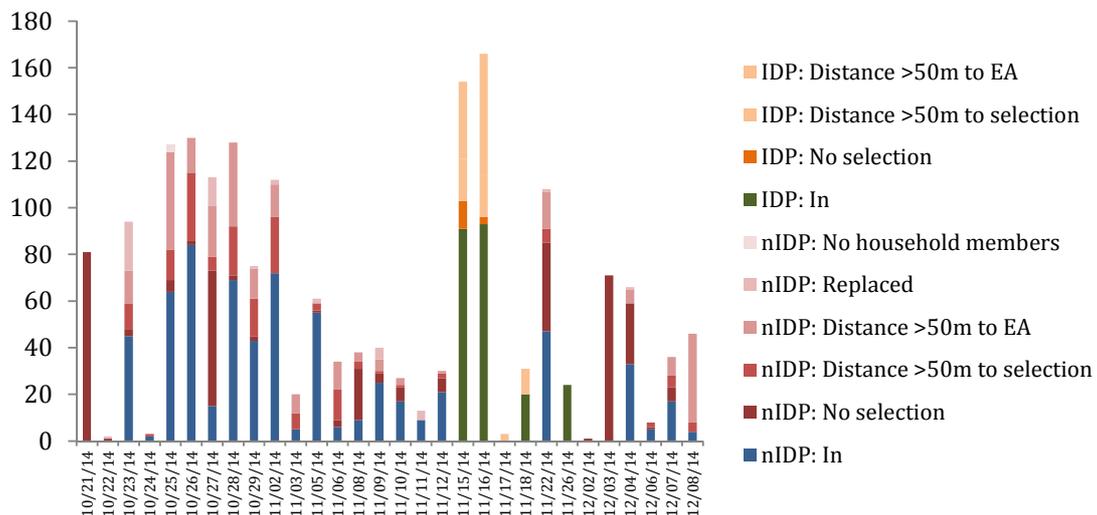


Remote management of tablets is implemented using two additional Apps. First, the Android Device Manager – an Android built-in App – allows remotely locking and wiping tablets. A second app called AirDroid enables managing remotely the tablet. This includes checking the phone call log, changing the configuration of the device as well as changing the content of the file system. While the first App is more important in case a tablet got stolen, the second App allows to inspect the tablet in the field to localize an error, troubleshoot problems as well as updating the software if necessary.

Collected data is instantaneously available for analysis once it is uploaded. Even in remote areas, collected interviews can usually be sent at least once a week. The data can be used to monitor field work based on a few indicators. In a recent pilot, the number of interviews conducted per day were cross-referenced with the location of the interview (Figure 3). For each incoming interview, the distance between the GPS coordinates of the interview and its enumeration area were calculated. Interviews conducted within the enumeration area are assigned a distance of 0m. Interviews conducted farther away than 50m are flagged. Also interviews conducted more than 50m from the household location in the listing are flagged as well as interviews without a corresponding listing entry. Progress of field work can easily be tracked by inspecting the resulting graph. It also allows to take actions in case teams are not following field protocols or operate outside their enumeration areas.

In this particular case, the filter on the GPS location flagged one in four interviews which were conducted more than 50m outside of the enumeration area. In addition, almost 5 percent of interviews were more than 50m away from the recorded household position in the listing. Thus, almost one out of three interviews had to be replaced to achieve the intended sample size. While real-time monitoring ensured that households can be replaced with enumerators still in the field, GPS tracking as discussed above could have prevented replacements altogether by monitoring enumerator’s positions when they arrive in the respective enumeration areas. For this purpose, also Apps for geo-fencing can be used. They require an additional step for enumerators to select the respective enumeration area and, thus, they are ideally complemented by remote GPS tracking.

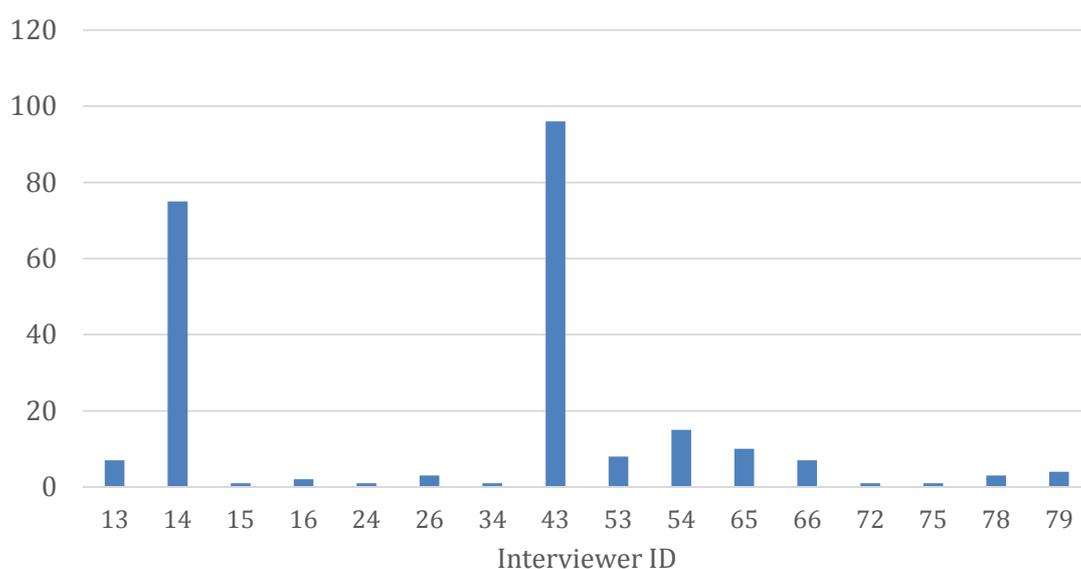
Figure 3: Real-time monitoring of data collection by cross-referencing number of interviews with location. Note that this data comes from a pilot in a challenging environment.



The performance of individual enumerators can also be analyzed by inspecting the frequency of unlikely entries. Enumerators with considerably larger frequency might require additional focused training to improve the quality of their interviews. For example, the number of entries above 50 USD can be investigated (Figure 4). In a poor country, one would expect only

few purchases above 50 USD. More importantly, the frequency of these entries should be approximately uniformly distributed across enumerators. In this example, two enumerators (ID 14 and ID 43) clearly record unusually often entries above 50 USD. A careful discussion with the enumerator – as well as a confirmation of interviews by listening to audio snippets of those two enumerators – can reveal whether enumerators are inventing the numbers or whether, for example, they accidentally touch the ‘0’ after entering a single digit number. Such an early warning system substantially increases data quality.

Figure 4: Number of times enumerators (x-axis) entered a value larger than 50 USD after a few days of data collection.

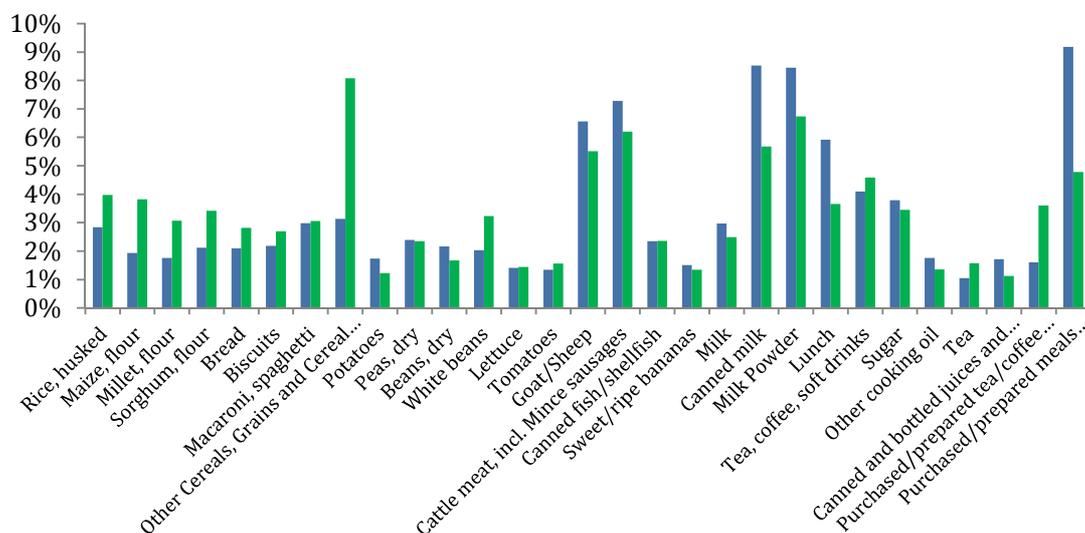


The real-time monitoring also offers monitoring aggregate statistics to detect artifacts.⁶ If this is implemented already at the beginning of the field work, the cause of artifacts can be determined and fixed. While changes in the questionnaire at the time of field work are sensitive, they are sometimes necessary and help to increase the utility of the data. For example, one can monitor the consumption shares and compare between strata, which are assumed to have differences in consumption (Figure 5). First, one can check whether the overall pattern is as expected. Second, one can assess differences between strata relative to expectations. In this example, one strata has a large consumption share of ‘other cereals’. In this case, it can make sense to analyze exactly which type of cereal this refers to (if an option ‘Please specify’ is given in the questionnaire) and consider including this grain explicitly in the list of items.⁷

⁶ Artefacts are patterns in the data, which are not related to the observed phenomenon but to the used methodology.

⁷ Such a decision must be made very carefully since the inclusion of an additional item will bias results. We suggest to consider this option only if it is deemed necessary and ideally observed after 1 – 2 days of field work so that previously collected data can be discarded.

Figure 5: Comparing consumption shares between two different strata.



The real-time monitoring increases data quality. From there, it is a continuum towards real time analysis. Once these statistics are implemented for the field collection, they are immediately available at the end of the field work reducing gap time between field work and the publication of results. Especially in fragile countries, the time between data collection and publication can decide whether data is used at all.

So far, the results of the real-time monitoring and analysis are available in Excel files, which are automatically updated using Stata code. Synchronizing the output files into the cloud makes the analysis accessible to the whole team. In the future, the results will be made available in a Tableau dashboard, which allows the user to customize graphs and interact directly with the data.

Conclusion

This note argues that tablet technology can significantly enhance data quality especially in fragile countries. The note presented three innovations. The first innovation is designed to improve data quality for consumption data by implementing a dynamic on-the-fly data validation system. This system flags unusually high or low entries and asks enumerators to confirm the correctness of the entry. The second innovation utilizes available software to monitor and manage tablets remotely. Using GPS tracking software, the location and trajectory of tablets can be determined – even retrospectively when the tablet re-enters 3G/WiFi areas. This helps to determine whether interviews were conducted at the correct location. Using remote management software, errors in the tablet configuration can be detected and solved while tablets can be updated remotely. The third innovation is a near real-time monitoring and analysis system. The monitoring system can identify challenges in the field work as well as weak enumerators early on and mitigate their impact on the data collection. The analysis system helps to check the design of the questionnaire and whether

the collected data is aligned with expected results. The availability of the analysis code at the end of the data collection considerably accelerates the process from data collection to publication of results. This is most important in fragile countries where timeliness of data is paramount.

CAPI technology in fragile countries can also create new challenges for data collection. The complexity to develop and program questionnaires as well as setting up tablets requires additional capacity by the implementing agency. Training of enumerators must be extended to cover tablets including trouble shooting. Supervisors must be trained to check conducted interviews using tablets. Tablets must be charged and functional in the field. Transmission of data required 3G or WiFi connectivity. Tablets are high value and, hence, increase the risk of theft. While these challenges require careful and intensive management, they are no show stoppers but instead increase capacity in the implementing agency to conduct future CAPI surveys (World Bank, 2015).

References

- Beegle, K., J De Weerd, J. Friedman and J. Gibson (2012), Methods of household consumption measurement through surveys: Experimental results from Tanzania, *Journal of Development Economics* 98 (1), 3 – 18.
- Mobile Data Collection in Humanitarian Context Project (NOMAD). (2011). Analyses of existing mobile data collection against the requirement of organization working in the humanitarian context . NOMAD.
- Caeyers, B., Chalmers, N., & De Weerd, J. (2011). Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment. *Journal of Development Economics* .
- Demombynes, G., Gubbins, P., & Romeo, A. (2013). Challenges and Opportunities of Mobile Phone-Based Data Collection: Evidence from South Sudan. World Bank.
- Dillon, B. (2010). Using Mobile Phones to Collect Panel Data in Developing Countries. Cornell University, Department of Economics. Cornell University.
- Ganesan, M., Prashant, S., & Jhunjhunwala , A. (2012, March). A Review on Challenges in Implementing Mobile Phone Based Data Collection in Developing Countries. *Journal of Health Informatics in Developing Countries* .
- Hewitt, P. C., Erulkar, A. S., & Mensch, B. S. (2003). The Feasibility of Computer-Assisted Survey Interviewing in Africa: Experience From Two Rural Districts in Kenya. Population Council.
- Himelein, K., Eckman, S., & Murray, S. (2015). Second Stage Sampling for Conflict Areas: Methods and Implications. World Bank.
- Pape, U. & Mistiaen, J. (2015). Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu High Frequency Survey. World Bank.
- Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., et al. (2009). The use of mobile phones as a data collection tool: A report from a household survey in South Africa. *BMC Medical Informatics and Decision Making*.
- Trucano, M. (2014, April). Using mobile phones in data collection: Opportunities, issues and challenges. Retrieved March 2015, from www.worldbank.org: www.worldbank.org/edutech
- World Bank (2011). World Bank Development 2011: Conflict, Security and Development. World Bank, Washington.
- World Bank (2015). Challenges and Opportunities of High Frequency Data Collection in Fragile States. World Bank, Washington.