Well-Being and Shocks: Challenges in Poverty Measurement and Analysis

Dissertation to acquire the doctoral degree

from the Faculty of Economic Sciences,

at the Georg-August-Universität Göttingen

Submitted by

Utz Johann Pape

Göttingen, 2023

Supervision and examination committee

First supervisor: Professor Dr. Sebastian Vollmer

Second supervisor: Professor Dr. Thomas Kneib

Third supervisor: Professor Dr. Holger Strulik

Date of disputation: November 8th, 2023

Acknowledgement

This thesis would not have been possible without the support of many people. First of all, I would like to thank my supervisor, Sebastian Vollmer. He put a lot of trust in me and allowed me to progress on my PhD remotely while working at the World Bank. He also provided thoughtful and constructive feedback throughout the process and created opportunities to teach in his lecture and the summer university, and fostered collaborations with other PhD students. I would also like to thank Markus Hennemann for the great team spirit and exciting discussions while co-authoring a paper together.

The World Bank was incredibly helpful in my endeavor, providing funding and space to conduct most of the research projects in this thesis. I have benefitted from the knowledge and experience from a large number of the World Bank – too numerous to list – but would like to especially thank colleagues in the Poverty & Equity Global Practice for their excitement to write academic papers with me – often outside their usual work. I would also like to thank my co-authors outside the World Bank for their interest and engagement to collaborate.

Finally, I would like to thank all my friends and family, who fully supported this path. In particular, my thanks go to Simon Lange, who inspired me to pursue this PhD. My parents were as supportive as one could wish and never had any doubts that I would continue this path. My niece, Marla Bartosch, was instrumental in formatting the final document – even though while traveling through Asia. I am especially grateful to my partner, Zahra Didarali, who patiently supported me through all ups and downs and inspires me daily.

Abstract

Goal 1 of the Sustainable Development Goals calls for a world free of poverty in 2030. Over the last decades, the world has made much progress in reducing poverty. However, COVID-19 has shown that these gains can be reversed when households are affected by shocks – and puts at risk the goal to end poverty by 2030. This is particularly true for households in fragile countries, which are becoming hotspots of poverty, where about two thirds of the poor are projected to live in 2030. To keep the goal of eradicating poverty alive, we need to better understand poverty especially in the context of shocks and fragility. We need good data, appropriate methodologies and solid analysis to measure poverty and the impact of shocks. This is the core challenge at the heart of this thesis.

The first part of the thesis presents seven papers that aim to improve poverty measurement especially in fragile countries. First, we propose a new methodology to reduce administering time of the traditionally comprehensive consumption questionnaire using statistical imputation techniques. Second, we develop a new approach to disaggregate poverty estimates geographically in the absence of a new population census – as is often the case in fragile countries. Third, we tackle the specific issues of poverty measurement for displaced populations, starting with an overview of the diverse challenges to be encountered. Fourth, we shed light on the advantages and disadvantages of different sampling approaches for displaced populations carried out in an Internally Displaced Person (IDP) camp in South Sudan. Fifth, we introduce light-touch nudges into the questionnaire to reduce measurement error, especially for aid-dependent populations. Using a combination of these methodologies, we were able to estimate poverty in Somalia for the first time in three decades and for the first time in South Sudan since independence, also produced the first comprehensive set of micro-data for IDPs and refugees in Africa. The last two papers describe the two applications in Somalia and South Sudan.

The second part presents my work in measuring the impact of conflict and shocks on well-being, based on a second set of seven papers. First, we assess the impact of conflict on livelihoods in South Sudan using a cluster-level difference-in-difference approach. Second, we focus on adolescent girls and amend the analysis to specifically understand their well-being and opportunities. Third, we estimate the short-term impacts of terrorist attacks on livelihoods in Somalia, using a difference-in-difference and instrumental variable approach. Next, we analyze shocks and their impacts. The fourth paper aims to understand the impact of high inflation on livelihoods in South Sudan. The fifth paper assesses the impact of drought on poverty in Somalia. Both papers utilize a difference-in-difference approach. The sixth paper assesses the impact of COVID-19 lockdown measures on mobility and livelihoods in Kenya, based on an instrumental variable approach. Finally, we use the design of our Randomized-Control-Trial for a planned cash transfer program in South Sudan to understand the impact of the cancellation of a program on livelihoods – as the re-emergence of the conflict made it impossible to proceed with cash transfers.

Declaration of Author's Contributions

Part I: Measuring Poverty

Measuring Poverty Rapidly Using Within-Survey Imputations by Utz Pape

Small Area Estimation of Poverty under Structural Change by Simon Lange (SL), Utz Pape (UP) and Peter Pütz (PP)

SL and UP developed the research question, while PP designed and implemented the analysis in discussion with SL and UP. PP provided the first draft of the write-up. All authors jointly interpreted results and finalized the manuscript.

Measuring Poverty in Forced Displacement Contexts by Utz Pape (UP) and Paolo Verme (PV)

UP and PV contributed equally to the manuscript.

Second Stage Sampling and Non-Sampling Errors: IDP camps in South Sudan by Kristen Himelein (KH), Utz Pape (UP) and Michael Wild (MW)

UP and KH contributed equally to the manuscript. MW provided analysis and write-up for the random walk simulations.

A Light-Touch Method to Improve Accurate Reporting of IDP's Food Consumption by Lennart Kaplan (LK), Utz Pape (UP), James Walsh (JW)

LK, UP and JW contributed equally to the manuscript.

Estimating Poverty in a Fragile Context - The High Frequency Survey in South Sudan by Utz Pape (UP) and Luca Parisotto (LP)

UP developed the research question and designed as well as supervised the field work. LP and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

Estimation of Poverty in Somalia Using Innovative Methodologies by Utz Pape (UP) and Philip Wollburg (PW)

UP developed the research question and designed as well as supervised the field work. PW and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

Utz J. Pape

8th February 2023

Signature Date

<u>Part II: Impact of Shocks and Fragility</u>

Impact of Conflict on Livelihoods in South Sudan by Luca Parisotto (LP) and Utz Pape (UP)

UP developed the research question and designed as well as supervised the field work. LP and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

Impact of conflict on adolescent girls in South Sudan by Utz Pape (UP) and Verena Phipps (VP)

UP and VP contributed equally to the manuscript.

Poverty and violence: The immediate impact of terrorist attacks against civilians in Somalia by Gonzalo Nunez-Chaim (GN) and Utz Pape (UP)

GN and UP contributed equally to the manuscript.

Impact of High Inflation on Household Livelihoods in Urban South Sudan by Alvin Etang (AE), Thierry Hounsa (TH) and Utz Pape (UP)

AE, TH and UP contributed equally to the manuscript.

Impact of Drought on Poverty in Somalia by Utz Pape and Philip Wollburg (PW)

UP developed the research question and designed as well as supervised the field work. PW and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

The Labor Market Implications of Restricted Mobility during the COVID-19 Pandemic in Kenya by Markus Heemann (MH), Utz Pape (UP), Sebastian Vollmer (SV)

MH and UP jointly conceptualized the research, obtained and processed the data as well as drafted the manuscript. MH conducted the statistical analysis, while UP and SV provided supervision.

Broken Promises: Evaluating an incomplete Cash Transfer Program by Angelika Müller (AM), Utz Pape (UP) and Laura Ralston (LR)

UP and LR contributed equally, while AM conducted data cleaning, implemented the analysis and provided the first draft of the write-up.

Utz J. Pape

8th February 2023

Signature Date

Declaration for Admission

Declaration for admission to the doctoral examination according to §12 of Prüfungs- und Studienordnung des Promotionsstudiengangs "Wirtschaftswissenschaften" der Georg-August-Universität Göttingen

I confirm

1. that the dissertation "Well-Being and Shocks: Challenges in Poverty Measurement and Analysis" that I submitted was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,
2. that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,
3. that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,
    - No remuneration or equivalent compensation were provided
    - No services were engaged that may contradict the purpose of producing a doctoral thesis
4. that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

Utz J. Pape

8th February 2023

Signature Date

Table of Contents

List of Figures

List of Tables

A. Introduction

Sustainable development aims to incrementally improve well-being in a society. Even though this is a straight-forward statement, it triggers many more questions. What is well-being? Who is part of the society? What do we call improvement? Each of these questions can be answered in different ways. Sometimes the answer makes a normative claim, for example, if we define improvement as an increase in average well-being – or in the well-being of the worst-off. Often, answers strike a judicious trade-off between theoretical concepts like well-being and how it can practically be measured. Understanding the limitations around measurement is key to be able to assess these trade-offs. Furthermore, advances in measurement can reduce the distance between concepts and what we can observe. This is most critical in situations where well-being is most at risk – situations of conflict and shocks – as these are the very situations that require timely and evidence-based decisions to protect livelihoods and create opportunities for sustainable development. This thesis contributes to this undertaking by presenting papers that I have (co-)authored in the area of measuring socio-economic indicators – specifically but not exclusively monetary poverty – in fragile situations and assessing the impact of shocks on livelihoods.

Gross National Income (GNI) is often used to measure development. It reflects the total amount of money earned by a nation's people and businesses. GNI per capita – reflecting average income – is used to classify countries into low-, lower-middle-, upper-middle- and high-income countries. However, it is silent on the distribution of income. A society can grow GNI per capita by enriching the elite while making the rest of the society worse off.

Amartya Sen proposes a very different approach (Sen 1985, Sen 1987). His normative approach focuses on the capability to achieve well-being, addressing some of the shortcomings of a more narrow welfare approach (Ravallion 2020). Following Nussbaum (2011), ten core capabilities must be satisfied for a decent life. They range from life, bodily health and integrity over emotions, practical reason and affiliation to play as well as having control over one's environment. Independent of the specific choices of core capabilities, the approach suffers from the irreconcilable claim of intrinsic value of the chosen capabilities and the obvious contingent negotiated relation with other values like justice, equality and rights (James 2017). Also from a practical perspective, it is inherently difficult to measure capabilities in contrast to outcomes, as capabilities often cannot be directly observed (Brandolini and D'Alessio 2001, Comin, Qizilbash et al. 2008, Vecchi 2017).

In contrast to the capability framework, needs-based approaches define well-being more narrowly as outcomes to gain practical concessions in measuring and applying it. Basic needs usually include health and education, which can readily be operationalized, for example, through indicators like longevity, infant survival, body mass index, literacy, education attainment, etc. As in the case of the Human Development Index, a capability-based motivation can morph into an outcome-based approach to become measurable. However, broad basic need approaches are limited by the contingent choice of indicators and their relative weighting.

A sub-class of needs-based approaches use monetary metrics. While they are excluding non-monetary dimensions, the advantage is a theory-grounded, one-dimensional measure – lending itself to comparisons across time and space – that can practically be applied without prescribing relative importance between different dimensions of needs. Furthermore, the monetary metric often correlates with non-monetary dimensions and, hence, can be a useful starting point for further analysis.

Poverty defined by a monetary metric marries the concept of GNI per capita at the macro-level with its distribution across households and is readily derived from utility theory in welfare economics. Rather than using average income, it considers the distribution of income or consumption in the society. The most popular class of poverty measures are Foster-Greer-Thorbecke (Foster, Greer et al. 1984). They define poverty incidence (FGT0) as the percent of individuals (or households) that live below a poverty line. The poverty line is – usually – derived by a narrow needs-based approach estimating the minimum cost of a food basket to deliver the minimum number of calories needed, taking into account local diet and prices, as well as a non-food component.

Goal 1 of the Sustainable Development Goals (SDGs) calls for a world free of poverty in 2030 with an indicator measuring the percent of individuals living below the international poverty line of US$ 1.90 PPP (2011). The international poverty line is derived from a set of national poverty lines (usually estimated through a needs-based approach as discussed above) from the poorest countries in the world (Ravallion, Chen et al. 2009, Ferreira, Chen et al. 2016). The poverty lines are converted into US$ using purchasing power parity (PPP) estimates from the International Comparison Program (ICP).

Over the last decades, poverty has dropped significantly from above 35 percent in 1990 to below 10 percent in 2017 (World Bank 2020). However, progress has decelerated especially since 2015 and reversed in 2020 because of widespread lockdowns triggered by COVID-19. Instead of dropping further down to 7.5 percent in 2021 as projected prior to COVID-19, the poverty trend reversed from 8.4 percent in 2019 to around 9 percent in 2020 and 2021. More generally, shocks threaten sustainable reduction in poverty. With climate change unaddressed, the number and severity of shocks will increase, potentially pushing more than 100 million people into poverty by 2030 (Hallegatte, Bangalore et al. 2016). In addition to shocks, fragility and conflict are major drivers of poverty. Countries with fragility, conflict and violence (FCV) host only 10 percent of the world's population but contribute more than 40 percent of the global poor (Corral, Irwin et al. 2020). Displaced populations are often among the most affected groups, remaining in poverty for many years (Pape and Sharma 2019).

Getting back on track to eliminate poverty by 2030 requires better preparedness, building resilience and targeted intervention for recovery. Effective design and implementation of any intervention, however, require a high-quality evidence-base to inform decision makers (World Bank 2020). However, such evidence is particularly hard to obtain in fragile and volatile situations, as created by conflict and shocks (Corral, Irwin et al. 2020), as well as for displaced populations (Verme 2016). Though these are exactly the most critical situations where a strong evidence-base can make all the difference.

In our book (Hoogeveen and Pape 2020), we start addressing this data challenge by describing new approaches for data collection in fragile situations, mostly in Africa in the Central African Republic, the Democratic Republic of Congo, Liberia, Madagascar, Mali, Malawi, Nigeria, Senegal, Sierra Leone, Somalia, South Sudan but also in Iraq, Jordan, Lebanon and Yemen. The presented approaches cover a wide range of objectives and methods, including easy-to-implement approaches to gather vital information for peace-inducing recovery planning to more complex imputation techniques to collect comprehensive consumption data to estimate poverty. Through our experiences in identifying innovative ways to collect data, we have learned five lessons. First, it is possible to collect high-quality data in fragile settings. Doing so may require adaptations to the data collection process but situations in which no information can be collected are rare. Second, data collection in fragile contexts does not need to be more expensive than in other settings. In fact, the costs associated with many of the innovations discussed in this book compare favorably to more traditional data collection methods. Third, a careful assessment of the data needs of decision makers is essential. Often relatively easy-to-collect information goes a long way towards meeting their demands, as long as it is provided in a timely fashion. Fourth, technology is not a panacea for all data collection issues and not everything works. Fifth, approaches developed for a fragile situation often are applicable and scalable in non-fragile contexts.

The COVID-19 pandemic emphasized the importance to be able to track socio-economic indicators not only in fragile countries, but also in non-fragile places at the time of crises and shocks. Drawing from approaches and lessons learnt in Hoogeveen and Pape (2020) including the monitoring of Ebola outbreaks (Himelein, Eckman et al. 2016) and assessing livelihoods and food insecurity in Nigeria, Somalia, South Sudan and Yemen (Pape 2020), phone surveys were implemented in dozens of countries providing data to decision makers. While coordinating this effort (Delius, Himelein et al. 2020), we explored the use of real-time indicators to track GDP (Ten, Merfeld et al. 2022) as well as labor market impacts (Newhouse, Swindle et al. 2022) but also tested the use of internet surveys (Soundararajan, Soubeiga et al. forthcoming). In addition to these methodological papers, we assessed the early labor market impacts (Khamis, Prinz et al. 2021) and analyzed differential impacts on welfare in FCV countries, based on a difference-in-difference approach exploiting variation over time and space (Tabakis, Ten et al. 2022).

In Kenya, we implemented several rounds of a nationally representative rapid response survey as early as May 2020. A near real-time dashboard put data and evidence at the fingertips of decision makers.[1] As part of a cross-country study, we described the falling living standards due to COVID-19 in developing countries (Egger, Miguel et al. 2021). In several deep dives, we used the data from the rapid response phone survey. Heemann, Pape et al. (2022) assessed the impact of COVID-19 lockdown measures on mobility and livelihoods (described in more detail as part of this thesis). Vintar, Beltramo et al. (2022) identified specific impacts of COVID-19 for labor market outcomes of refugees vis-à-vis nationals. Biscaye, Egger et al. (2022) used the closing of schools as a natural experiment to understand the labor market implications of child care. Cameron, Delius et al. (2022) described COVID-19 impacts on children and Xu, Delius et al. (2022) explored gender differences in household coping strategies.

The remainder of the introduction consists of a chapter on poverty measurement and a chapter on the impact of shocks and fragility, with the full papers presented in part I and II of this thesis. The first chapter presents approaches to improve poverty measurement specifically in fragile situations, partly drawing from work described in Hoogeveen and Pape (2020). After deriving monetary welfare metrics to introduce the concept and touching upon several measurement challenges, the first section delves deeper into poverty measurement. First, we propose a new methodology to reduce administering time of the traditionally comprehensive consumption questionnaire using statistical imputation techniques (Pape and Mistiaen 2018, Pape 2021). Second, we present a method that allows spatial disaggregation of poverty estimates in the absence of a new Census, which is more applicable for fragile countries (Lange, Pape et al. 2022). In the second section, we tackle the specific issues of poverty measurement for displaced populations, starting with an overview of the diverse challenges to be encountered (Pape and Verme 2023). Next, we shed light on the advantages and disadvantages of different sampling approaches using a benchmarking of different sampling methodologies for displaced populations carried out in an Internally Displaced Person (IDP) camp in South Sudan (Himelein, Pape et al. forthcoming). We also introduce light-touch nudges into the questionnaire to reduce measurement error, especially for aid-dependent populations (Kaplan, Pape et al. 2018). The last section of this chapter shows the application of several of the proposed methodologies to measure

[1] www.kenyacovidtracker.org/rrps

and estimate poverty in particularly fragile countries: South Sudan (Pape and Parisotto 2019) and Somalia (Pape and Wollburg 2019).

The second chapter presents my work in measuring the impact of conflict and shocks on well-being. Using a cluster-level difference-in-difference approach, we assess the impact of conflict on livelihoods (Parisotto and Pape forthcoming) as well as specifically on adolescent girls and their opportunities (Pape and Phipps 2018). Employing a similar approach, we also estimate the short-term impacts of terrorist attacks in Somalia (Nunez-Chaim and Pape 2022). In the second section, we focus on shocks and their impacts on livelihoods. The first example is the impact of high inflation on household livelihoods in South Sudan (Etang, Hounsa et al. 2022), followed by an assessment of the impact of drought on poverty in Somalia (Pape and Wollburg 2019). Both examples utilize a difference-in-difference approach. Another example is an assessment of the impact of COVID-19 lockdown measures on mobility and livelihoods (Heemann, Pape et al. 2022). In the last section, we look into the impact of uncertainty on the design of development interventions (Müller, Pape et al. 2019). In South Sudan, a planned cash transfer program had to be canceled after beneficiaries were selected. We use the Randomized-Control-Trial (RCT) design to assess the impact of the cancellation on livelihoods, including incurred debt and perception of Government institutions.

1. Measuring Poverty

As we have seen in the introduction, a good argument can be made to focus on consumption to measure poverty, even though it reflects only one of many dimensions of potential deprivation (Slesnick 2001). The approach is generally derived from cost-of-living approaches, using the concept of money metric utility (Samuelson 1974). Our exposition follows Deaton and Zaidi (2002) as well as Mancini and Vecchi (2022).

Welfare for an individual is defined through the utility $u = v(q)$ that the individual receives from a bundle of goods $q = (q_1, q_2, \dots, q_n)^T$. Instead of prescribing normatively which goods are desirable, we assume that all individuals are rational maximizing their utility, that all individuals have the same preferences and prices exist for all goods. Based on these assumptions, we do not need to assume a specific shape of the utility function, but can infer utility from the observed bundles that individuals choose to consume. Given a price vector $p = (p_1, p_2, \dots, p_n)$, an individual with a budget of x can consume bundles within their budget constraint: $p \cdot q < x$.

Many different bundles can obtain the same utility u. However, these bundles will differ in the required expenditure $p \cdot q$ to purchase the bundle q. Maximizing utility means that the individual will choose the bundle q^* that provides highest utility u^* and is on the budget constraint x. The equivalent (often called dual problem) is the individual minimizing expenditures $p \cdot q$ to achieve the utility level u^*, which results, again, in choosing the bundle q^*.

The cost function $c(u, p) = x$ captures this idea. It returns the minimum cost that is needed to achieve a utility level u given prices p. To compare across different households, we introduce the household index h and a reference price level p^0 to obtain the money metric utility $u_m^h = c(u^h, p^0)$. The money metric utility defines utility for household h as the minimum cost of reaching utility level u^h at prices p^0. The significance of this equation comes from a simple transformation that allows us to approximate the minimum cost by using observed consumption shares of households:

$$
\begin{aligned}
u_m^h = c(u^h, p^0) &\approx c(u^h, p^h) + (p^0 - p^h)\frac{\partial c(u^h, p^h)}{\partial p^h} \\
&\approx p^h q^h + (p^0 - p^h) q^h \\
&\approx p^0 q^h
\end{aligned}
$$

The first approximation uses a first-order Taylor expansion of $c(u^h, p^0)$ around p^h. The maximized utility u^h for household h at prices p^h is observed through the actual household's choice of consumption bundle q^h, so that the corresponding minimum cost function becomes $p^h q^h$. The partial derivative of the minimum cost can be obtained through Shephard's lemma as the Hicksian demand function (Deaton and Muellbauer 1980). This powerful result states that the money metric utility for household h can be approximated by the cost of the chosen (and observed) bundle q^h evaluated at reference prices p^0.

With the help of the Paasche price index P^h, the money metric utility can be further transformed into its typical form of the consumption aggregate $x^h = p^h q^h$ deflated by P^h. The Paasche index is defined as:

$$
P^h = \frac{p^h q^h}{p^0 q^h}
$$

It compares prices p^h faced by the household with reference prices p^0 by using household consumption shares q^h as weights. With that, we can re-write the money metric utility in its standard form:

$$u_m^h \approx \frac{x^h}{P^h} = x_P^h$$

Thus, we can express the money metric utility for household h by the consumption aggregate deflated with the Paasche index. Practical approaches often use the Laspeyre instead of the Paasche index, which compares price levels at average consumption shares across households rather than the specific household shares. The main practical advantage is that the Laspeyre suffers less from outliers, especially if data quality is limited. However, conceptually, the Laspeyre is not a money metric utility but a welfare ratio (Blackorby and Donaldson 1987), with implications for its interpretation (Deaton and Zaidi 2002).

Using the operationalized money metric utility, we can define monetary poverty. As mentioned in the introduction, Foster, Greer et al. (1984) define the FGT class of poverty measures as the proportion of households living below a defined welfare-level using the money metric utility in a population of n_h

$$FGT_\alpha = \frac{1}{n_h}\sum_{h=1}^{n_h}\left(\frac{z_P - x_P^h}{z_P}\right)^\alpha I\left(x_P^h \leq z_P\right)$$

Where z_P is the poverty line expressed at the same reference price level as the money metric utility and $I\left(x_P^h \leq z_P\right)$ denotes the identify function and is equal to 1 if the parameter is true, in our case if the welfare-level is below the poverty line. The parameter α can be understood as poverty aversion, as higher values weight the poorest more strongly. In the case of $\alpha = 0$, we simply count the number of poor resulting in the poverty headcount rate. Setting $\alpha = 1$ results in the well-known poverty gap, which is the average gap in welfare of poor households relative to the poverty line.

The poverty line z_P can be derived in many different ways. Most commonly used for international comparisons is the US$ 1.90 PPP 2011 poverty line, which is derived from the national poverty lines of a set of poor countries, and uses purchasing power parity to allow comparisons across countries (Ravallion, Chen et al. 2009, Ferreira, Chen et al. 2016). National poverty lines are usually derived from a needs-based approach defining a nutritional minimum caloric intake (often 2,100 kcal or 2,400 kcal per individual) converted into expenditure by using the prevailing local diet and adding a non-food allowance (Deaton 1980, Ravallion 1998, Lanjouw 1999). At the household level, equivalence scales can be used to adjust for gender- and age-specific caloric parameters as well as to incorporate potential economics of scale for larger households.

In practice, an FGT_α poverty indicator is usually estimated based on data obtained from a survey, exposing measured welfare-level x_P^h to sampling and non-sampling errors.[2] Sampling error captures the error in the estimate that originates from using a sample survey rather than a Census. Non-sampling error includes any other error sources, among those are often over- and under-coverage error, non-response error and measurement error. Over- and under-coverage error refer to inclusion and exclusion of sampling units from outside respectively inside the sampling population. Non-response bias can be split into unit non-response, when sampling units refuse to respond to the

[2] Similarly, most other variables in FGT_α are also measured imprecisely. For example, the poverty line involves consumption shares and prices, which are also affected by errors. For the purpose of the thesis, however, we focus on the consumption estimate, since the proposed methodologies improve the measurement of the consumption estimate.

questionnaire, and item non-response, when specific questions are not answered. Finally, measurement error occurs when respondents mis-understand a question or intentionally mis-report.

This remainder of this chapter discusses in more detail how poverty measurement can be improved in specific situations, related to conflict and displacement, by reducing different error sources. First, we propose a new methodology to measure poverty if interview time is severely limited (Pape and Mistiaen 2018, Pape 2021). It uses statistical imputation techniques to compensate for intentionally missing data, reducing unit non-response and measurement error while accepting some model error. Second, we propose a modified small-area-estimation methodology to reduce sampling error on cost of model error to obtain more accurate spatially disaggregated poverty estimates (Lange, Pape et al. 2022). It applies statistical imputations in the situation of an old Census and a new household survey. Third, we discuss poverty measurement for displaced populations. After providing an overview (Pape and Verme 2023), we study sampling and non-sampling errors of different sampling strategies in camps of displaced populations before improving measurement error for aid-dependent populations (Himelein, Pape et al. forthcoming). Finally, we put the methodologies to work in two real case-studies by applying a combinations of the proposed methodologies to obtain accurate poverty estimates in South Sudan (Pape and Parisotto 2019) and Somalia (Pape and Wollburg 2019).

Improving Poverty Estimates

Pape (2021) Measuring Poverty Rapidly Using Within-Survey Imputations

Measurement of consumption is at the core of poverty measurement. However, it has traditionally been very time consuming. A typical household consumption questionnaire contains a series of questions about the price and quantity consumed for each item, and whether it has been purchased, self-produced, or bartered. Usually encompassing more than 200 food and nonfood items, the time required to administer such a questionnaire can often substantially exceed two hours. In addition to high administration costs due to long interview times, measurement errors may become significant towards the end of the questionnaire as enumerators and respondents become fatigued. Respondents might also cancel the interview before it is completed, thus contributing to a higher non-response rate.

Enumerator and respondent fatigue are well documented in the literature (Krosnick 1991, Tourangeau, Rips et al. 2000) and become more pronounced for longer questionnaires (Diehr, Chen et al. 2005, Snyder, Watson et al. 2007, Rolstad, Adler et al. 2011). Enumerator fatigue increases measurement errors often over the course of a day as well as over the time the survey progresses (Baird, Hamory et al. 2008). Especially in consumption surveys, a long list of items can lead to enumerators cutting corners and fabricating data (Finn and Ranchhod 2015, Fiedler and Mwangi 2016) as well as prematurely ending interviews (Deaton and Grosh 2000). Respondents also become fatigued and, for example, learn to say no to consumption of items to evade more detailed follow-up questions (Kreuter, McCullock et al. 2011, Eckman, Kreuter et al. 2014).

To overcome the challenges inherent to measuring consumption poverty, we propose a new methodology that combines an innovative questionnaire design with standard imputation techniques. This new methodology allows us to substantially shorten the consumption questionnaire and reduce the interview time (less than 60 minutes for a standard questionnaire) by imputing deliberately absent consumption values for those items that are not explicitly asked about. Poverty estimates can be derived in this way without compromising the credibility of the resulting estimate. This new methodology is particularly useful in fragile states given the significant risks associated with lengthy interviews. It can also be useful to reduce enumerator and respondent fatigue, or to mitigate the problem of high non-response rates.

The most straightforward way to reduce the expected interview time is to skip rarely-consumed items. Another simple strategy is to ask the respondent about an aggregate amount of spending on an entire category of consumption (e.g., total expenditure on flour) instead of individual items (e.g., expenditure on corn flour, wheat flour, etc.). However, altering the set of items in the questionnaire can result in a nontrivial change in the reported consumption amount (Olson-Lanjouw and Lanjouw 2001). Both approaches are likely to lead to an underestimation of consumption and overestimation of poverty, as was demonstrated in a study in Tanzania that directly compared various methods of measuring consumption (Beegle, De Weerdt et al. 2012).

An alternative approach is to apply methods of cross-survey imputation. In situations where full household expenditure surveys are too costly or impractical to administer, data gaps can be filled using other surveys that have common covariates that are correlated with household expenditure. For example, data from a full consumption survey can be combined with data from shorter and more frequent labor force surveys to generate poverty estimates (Douidich, Ezzrari et al. 2013). While such methods may work well even when there is a rapid economic change (Christiaensen, Lanjouw et al. 2011), the assumption of a stable structural parameter typically cannot be tested and may not be valid, especially in the context of large and systemic shocks, after implementation of projects, or if a substantial amount of time has passed since the baseline survey was implemented. It is also possible to design a survey such that one sample has a full consumption module and another sample has only the covariates of consumption. Consumption can thus be imputed and poverty estimates can be derived at a reduced cost, even though the magnitude of potential cost reduction may be modest (Fujii and van der Weide 2016). In such a setup, however, the sample for the full consumption module must be chosen randomly to avoid biased estimates of the model parameters. Thus, this approach is only of limited usability in the case of fragile countries as it might not be feasible to administer the full consumption module in particular insecure areas, creating a downward bias in poverty estimates for those areas.

The proposed methodology uses statistical imputations to obtain estimates for deliberately absent consumption values. Statistical imputation techniques are widely used to replace missing values in surveys (Ambler, Omar et al. 2007, Van Buuren 2007, Little and Rubin 2019). Straight-forward methods simply replace the missing values with aggregate statistics like a mean. However, this makes the strong and often violated assumption that data are missing at random (Carpenter, Kenward et al. 2007). Model-based approaches can take into account covariates and often use a regression framework to estimate missing values but distort the variance if based on point-estimates. Multiple imputations help to mitigate this by drawing multiple estimates from the posterior distribution using a Bayesian approach (Rubin 2004).

Our new methodology combines an innovative questionnaire design with standard imputation techniques. It substantially shortens the time required to administer a household consumption questionnaire to less than 60 minutes by imputing deliberately absent consumption values for items that are not explicitly asked. The proposed methodology makes it possible to derive poverty estimates without compromising the credibility of the resulting estimate, and it performs considerably better than alternative approaches based on reduced consumption aggregates and cross-survey imputations. This new methodology is particularly useful in fragile states given the significant risks associated with lengthy interviews, as well as to rapidly assess the impact of a shock or of a project. It can also be useful to reduce enumerator and respondent fatigue, or to mitigate the problem of high nonresponse rates.

Lange, Pape et al. (2022) Small area estimation of poverty under structural change

A poverty map is a spatial description of the distribution of poverty in a given country or region. While such a map is useful for policy makers and researchers when small geographic units (e.g., cities, towns, or villages) are discernable, estimates based on household surveys are typically not representative or associated with high uncertainty at such levels of disaggregation. On the other hand, most censuses do not contain information on consumption (or a surrogate such as income or expenditures) required to calculate poverty. To overcome these problems, Elbers, Lanjouw et al. (2003, henceforth called ELL) developed small area estimation poverty maps, a methodology that can be used to combine information from a detailed household survey with that from a comprehensive census. The general methodology usually consists of two steps, calibration of a statistical model based on survey data and application to the comprehensive census data. In the first step, a multiple linear regression analysis is used to estimate a model of household consumption based on survey data (which includes a consumption module). The explanatory variables in the model are restricted to the subset available in both the survey and the census. In the second step, the estimated model parameters are applied to census data. The simulations provide estimates of consumption per capita for every household in the census. Since the regression model predicts the conditional mean of consumption yet one is typically also interested in higher moments of the distribution, for example to obtain accurate estimates of poverty rates, simulation methods are used to introduce a random disturbance term.

Several criticisms have been raised with regard to the ELL estimator and extensions and alternatives have been proposed (Tarozzi and Deaton 2009, Haslett, Isidro et al. 2010, Molina and Rao 2010, Das and Chambers 2017, Marhuenda, Molina et al. 2017). Comprehensive discussions on different small area estimation methods can be found in Guadarrama, Molina et al. (2016) and Haslett (2016). Still, the is arguably the most frequently used poverty mapping approach combining survey and census data. According to Elbers and van der Weide (2014), it has been applied in more than 60 countries.

A key assumption for the applicability of ELL is that the distribution of the explanatory variables is the same in both census and survey. This assumption will often be violated if time has passed between data collection for the census and survey, i.e., only a dated census and a more recent survey are available, a common situation as censuses are usually conducted less frequently than surveys, and something that is particularly often the case in fragile countries. Reasons for a violation of this assumption may include demographic trends, migration, natural disasters, and conflicts. If the population parameters, including the regression coefficients, remain unchanged but the distributions of the explanatory variables change over time, ELL results in an outdated poverty map, namely a poverty map at the time of the census.

The discussed assumptions on the explanatory variables can be relaxed if household characteristics from the census are used to explain consumption values from the survey in the first stage to obtain parameter estimates. These can then be used to predict consumption values using the census data in the second stage. As it is usually impossible to match households between a census and a survey, the estimation needs to be conducted at a higher geographical level, for instance at the level of census enumeration areas. If the assumptions on the explanatory variables hold, this aggregation may worsen the prediction accuracy vis-à-vis ELL, with the magnitude of the loss of precision hinging on the regression model in the first stage.

We present a new method which allows for the estimation of up-to-date small area poverty maps when only a dated census and a more recent survey are available and predictors and structural parameters are subject to drift over time, a situation commonly encountered in practice. Instead of using survey variables to explain consumption in the survey, the new approach uses variables constructed from the census. The proposed estimator has fewer data requirements and weaker assumptions than common small area poverty map estimators.[3]

In the case that at least one of the underlying assumptions of ELL is violated, our new approach still produces up-to-date poverty maps with unbiased poverty estimates. We introduce the key assumption that *aggregate* household characteristics from the old census relate to consumption the same way in clusters covered by the new survey as in clusters not covered by the new survey. This assumption will hold (on average) if clusters are randomly drawn. Note that a similarly weak assumption has to be made for the applicability of the ELL method if the census and survey are conducted at the same time, namely that household characteristics from the survey relate to consumption the same way in clusters covered by the survey as in clusters not covered by the survey.

Using simulation studies, we show the overall good performance of our estimator. If the distribution of explanatory variables changes over time, our new estimator is superior to the most frequently used method for contemporaneous census and survey collection. However, our approach is not immune to issues typically encountered in small area estimation techniques that combine census and survey data. In particular, variable selection and adequate modeling of apparent heteroscedasticity and differences in skewness in the residuals can be challenging. Besides, the key assumption, namely that aggregate household characteristics from the old census relate to consumption the same way in clusters covered by the new survey as in clusters not covered by the new survey, may not hold for the specific welfare estimation exercise at hand. For example, the migration pattern between census and survey collection may vary between clusters and may be correlated with the welfare status which is typically not captured by the model.

Displaced Populations

Pape and Verme (2023) Measuring Poverty in Forced Displacement Contexts

The United Nations High Commissioner for Refugee (UNHCR) estimated that the global number of Forcibly Displaced Persons (FDPs) in the world surpassed 84 million in 2021, up from around 40m in 2010 and accounting for over one percent of the global population.[4] This sharp growth in displaced people during the past decade can be largely attributed to the Syrian conflict started in 2011, the displacement of the Rohingya people since 2017, and the intensification of several conflicts in Sub-Saharan Africa, particularly along the Sahel region. These numbers are unprecedented in the history of displacement when recording started with the establishment of the UNHCR in 1950 and the signature of the Geneva Refugee convention in 1951.

FDPs are not a homogenous group. They include Internally Displaced Persons (IDPs - citizens of a country that have been displaced within the boundaries of their own country due to conflict or security reasons), asylum seekers (displaced people outside their own countries who formally ask for asylum), refugees (people who have obtained asylum in the host country), and other displaced groups that defy simple categorizations. These categories of people fall under the mandate of the UNHCR because they have been displaced "forcibly" because of conflict or violence and because they are in need of international protection. They exclude other categories of displaced people who were not forced to move because of conflict or violence such as economic migrants and victims of natural or environmental disasters. Of course, many people cannot be simply categorized in these groups and

[3] ELL also propose the additional use of census means to explain location effects, i.e. cluster-specific effects. In this regard, our approach can be considered as a variant of ELL without the use of household-level variables included in both census and survey and without reliance on the associated assumptions. When we refer to the ELL method throughout this paper, we have in mind an estimator that combines survey and census variables at the household-level, the central idea of the approach.

[4] https://www.unhcr.org/refugee-statistics/

this makes statistics on FDPs gross estimates, but the growth and relevance of these numbers are undisputed.

The growth in number of FDPs poses a challenge to the measurement of global and national poverty. Those who are forcibly displaced and in need of international protection tend to be persons who have lost their assets, financial resources, and social networks. They are typically very poor with no obvious path out of poverty. For refugees, their number vanish from poverty statistics of their own country because they are no longer counted in the place of origin. Both, IDPs and refugees are also not properly accounted for in the country they reside in. Their numbers – even though high in absolute terms – are often low relative to the non-displaced population (with some exceptions like Lebanon and South Sudan). Hence, they do not explicitly show up in official statistics. Even if – as in some but not all countries – their locations are appropriately included in the sampling frame, they are unlikely to be sampled due to their small proportion relative to the population, in most cases excluding them from official (poverty) statistics.

The number of FDPs is not small in terms of the absolute global poverty count. Before the COVID-19 pandemic, around 10 percent of the global population (780 million people) was estimated to be extremely poor, below 1.9 USD/day in Purchasing Power Parity (PPP) terms. If we make the conservative assumption that about 1 in 2 FDPs are extremely poor (Pape and Sharma 2019), this translates into 41.2 million poor or about 5 percent of the global poor.

However, measuring poverty among displaced population is not straight-forward. Sampling frames, which allow to stratify for FDPs, are usually not available. Furthermore, sampling approaches need to take into consideration local specificities, for example in an IDP camp. Concepts such as income, expenditure, and consumption have a very different content and meaning in the context of FDPs. FDPs may live in camps where shelter and services are provided entirely by the international community. In some very poor countries, FDPs living in camps may be better off than locals, but when it comes to measuring poverty among FDPs in camps one has to reconsider how to measure poverty. For example, it is not obvious whether hand-outs such as food stamps should be considered as income or consumption, or how to quantify housing, education and health services provided in camps free of charge. These issues with its implications for poverty measurement are discussed in Pape and Verme (2023).

Himelein, Pape et al. (forthcoming) Implications of Choice of Second Stage Selection Method on Sampling Error and Non-Sampling Error: Evidence from an IDP camp in South Sudan

The most common sampling approach for cross-sectional household surveys in the developing world is a stratified two-stage design (Grosh and Munoz 1996). Following stratification based on administrative boundaries, clusters are selected in the first stage with probability proportional to size from a national census-based frame. While traditionally clusters are demarcated manually, newer approaches use remote-sensing data and machine learning techniques to obtain or update clusters (Qader, Lefebvre et al. 2019, Qader, Lefebvre et al. 2020, Qader, Lefebvre et al. 2021). In the second stage, a canvasing operation is conducted in the selected clusters to compile an updated list from which households are randomly selected. While this methodology is straight forward to implement in the field and reliably produces unbiased estimates, there are several downsides.

The first downside is cost. The World Bank's Living Standards Measurement Study team, which provides technical assistance on large-scale household surveys around the world, estimates the field listing operation increases the overall budget for data collection by 25 percent. Due to confidentiality concerns, the data collected during a field listing operation, typically the name of the household head and address or location description of dwellings, does not have any analytical applications beyond as a component of the weight calculations. At a time when typical surveys costs are in the USD millions, reducing a significant cost component will increase the financial sustainability of data collection.

The second drawback to the traditional design relates to timeliness. At a minimum, listing operations are usually conducted several days, if not several weeks, before the main fieldwork. As populations shift, the quality of the list degrades as time passes. While this is generally not a major concern for static populations living in villages or cities, it is a major concern for those in IDP and refugee camps. The transient nature of such environments implies building an accurate sampling frame is a complicated process often fraught with inaccuracies. Structures, often tents, for example, can easily be enlarged or split, quickly changing the layout of the camp, potentially invalidating a pre-existing sampling frame.

There are also issues related to the subjectivity in a listing operation. Eckman (2013) found only an 80 percent overlap between the same blocks listed separately by different interviewers in the United States. Undercoverage during the listing operation impacts the representativeness of the final estimates if the undercoverage is non-random. For example, O'Muircheartaigh, English et al. (2007) showed undercoverage in the United States is higher in low-income and rural areas. If this finding extends to the developing world, poverty numbers may be underestimated. In addition, Barrett, Beaghen et al. (2002) find higher undercoverage of households occupied by non-Hispanic black respondents compared with non-Hispanic white or other race respondents. This potential bias introduced by racial differences between the interviewer and respondent is of particular importance in the developing world context when interviewers are often recruited in the capital city and sent to more remote regions for the survey.

In this paper, we build on the work done by Himelein, Eckman et al. (2017) in describing five alternative sampling approaches considered for a household survey in Mogadishu (satellite mapping, segmentation, grid squares, "Qibla method," and random walk). In the previous work, however, the authors used simulations which assumed perfect implementation. Therefore, while it was possible to compare the sampling error of the five methods, it was not possible to consider non-sampling error. This paper goes a step further by using simulations to describe the sampling error and a field experiment in an IDP camp in South Sudan to measure the total survey error of each design compared to a census, allowing for the disaggregation of the total error into sampling and non-sampling components. In addition, we attempt to separate the components of non-sampling error linked to the sample method from those common across all methods, such as interviewers selecting larger households and other issues in properly implementing the household survey protocols.

We find that simulations arrive at the true household size distribution, while all simulations over-estimate household size. This over-estimation is caused by a systematic tendency of enumerators to select larger households because they are more likely to find an adult respondent. Specifically, the North method and the random walk show higher degrees of availability bias than those methods in which the selection can be verified, e.g., satellite mapping where a specific structure is chosen a priori. Also for other indicators, including poverty estimates, we find that simulations obtain unbiased results while the actual experiments are biased, especially for variables correlated with household size. Pooling the analysis across indicators and using satellite mapping as reference, the North Method is unbiased, while the Segmenting and Grid Square methods show minimal bias (0.1 percent and 0.2 percent, respectively). The Random Walk method shows 1.2 percent bias on average across the 14 questions. In conclusion, probability-based methods perform better than non-probability methods like random walk. In addition, implementation of adherence with the survey protocol is extremely important. In practice – in a fragile setting like South Sudan – deviations from the survey protocol, measured as differences between the experiments and the simulations, have large influence on the actual bias of estimates.

Kaplan, Pape et al. (2018) A Light-Touch Method to Improve Accurate Reporting of IDP's Food Consumption

The standard way in which the World Bank and other policy organizations develop statistics is through individuals' responses to questions in economic surveys. Self-reported information, however, is vulnerable to myriad reporting inaccuracies when social scientists ask personal or intrusive questions or when respondents anticipate social or material implications to the answers they provide. This is of particular concern when respondents believe that misreporting may provide relief, both because of the of the sensitivity and the gravity of the policy challenge. In situations where it has been possible to compare survey responses to revealed economic behavior, striking disparities are sometimes found. In one investigation for example, Poterba and Summers (1986) report that misstatements regarding employment status in the Current Population Survey led to an underestimation of the duration of unemployment by up to 80 percent and even greater overestimates of the frequency of labor market entries and exits. In another study, Rosenfeld, Imai et al. (2016) look at voting behaviors in a sensitive anti-abortion referendum held in Mississippi in 2011. They compare actual county level vote shares against survey results from a sample frame of individuals who voted during the election (based on public records). Surveys that used direct questioning led to an underestimation of casting a "no" vote by more than 20 percentage points in the majority of counties.

A number of mechanisms can compromise the validity of self-reported information in surveys. Some inaccuracies result from cognitive biases – for example, acquiescence or "yea-saying" (Bachman and O'Malley 1984, Hurd 1999), extreme responding (Cronbach 1946, Hamilton 1968), and question order bias (Siegelman 1981). One solution to problems such as question order bias is to randomize the order of questions (Warner 2012). Other inaccuracies emerge from conscious but not calculated behavior. Respondents may deliberately misreport information on sensitive subjects not to distort statistics but to maintain their reputation or to abide by political norms (Gilens, Sniderman et al. 1998). A common solution to this is to enable participants to cloak their behaviors or beliefs. List experiments, endorsement experiments, and randomized experiments are commonly used techniques for this purpose (Rosenfeld, Imai et al. 2016).

The explanations above assume that people intend to report accurately but are prevented from doing so due to aspects of the situation. In some contexts, individuals may misreport due to expectations about the implications of the results of the study. For example, individuals may misreport to increase earnings in a study context (Mazar, Amir et al. 2008) or to shape the results of the study if they believe that it will inform policy. In situations where individuals wish to influence a particular research outcome, a guise of anonymity will not shift their behavior. It is important to note that our concern is not with the ethics of individual misreporting – this is a reasonable response to contexts of extreme vulnerability – but rather to ensure that policymakers have access to data that enables them to adequately serve the vulnerable population as a whole.

Behavioral science is increasingly being used as a policy tool to help policymakers create better policy and solve collective action problems more effectively (World Bank 2015). This is based on research illustrating that people make decisions on the basis of both external and internal reward mechanisms (Mazar and Ariely 2006). Even in cases where people have an extrinsic incentive to misreport, this may be overridden by a preference for remaining consistent with their values. One example of this is when individuals' beliefs regarding the consequences of misreporting affects their behavior. In a two-person experiment where one participant can increase her payoff by misreporting but at the expense to her counterpart, Gneezy (2005) finds that individuals' propensity to misreporting is sensitive to the costs it imposes on the other person. Contextual cues affect the salience of internal incentives (or intrinsic motivations) and thus the accuracy of responses. This psychological mechanism has been put to practical use in policy. In multiple contexts, normative messaging has been used to increase tax payments (Hallsworth, List et al. 2017) or reduce littering and environment theft (Cialdini 2003).

In this paper, we apply the tools of behavioral science to investigate the veracity of consumption reports by internally displaced persons, and to propose behavioral nudges to reduce measurement error. In numerous rounds of data collection in Somalia and South Sudan, IDPs report significantly lower levels of consumption than non-IDP households (Pape and Sharma 2019). In previous survey rounds 45 percent of Somali IDP households report food consumption below subsistence levels and approximately 80 percent below recommended levels. While the data may be accurate, there are two reasons to suspect that it is not. First, such high levels of non-consumption would be associated with high rates of mortality due to starvation. Although being high, the mortality rates among IDPs suggest that this is not happening systematically across the country at such a scale (FEWSNET 2018). Second, non-IDP households that are statistically similar on observable characteristics report higher levels of consumption than IDP households. While IDPs and non-IDPs may have different opportunities to generate income, it is unlikely that IDPs choose not to smooth their resources to balance between food and non-food consumption in a way that endangers their life.

If it is the case that survey respondents misreport, the inaccuracies it generates in the data are highly problematic. At best, it makes the data spurious and unusable. At worst, it could lead to misallocations of aid, from more vulnerable areas to less vulnerable areas, or from solutions emphasizing sustainability to immediate relief where immediate relief is unnecessary. Due to the dangerous environment in South Sudan and Somalia, it is not currently possible to do use alternative data collection methods, for example ethnographic research, to investigate this puzzle in the data. The validity of alternative investigative methods such as food diaries is vulnerable to the same incentive to game as surveys.

One way to investigate whether people misreport is to test whether consumption rates change in response to nudges. If these primes are effective, they would be expected to particularly affect potentially underreporting, hence, poor households. Moreover, as vulnerable populations would have higher incentives to underreport, priming should be stronger for IDPs than for comparable non-IDP populations. In this paper, we conduct a randomized-control-trial to assess the impact of primes on IDPs as well as non-IDPs. We find the primes induce higher reporting in lower quintiles of reported consumption. This treatment pattern is driven by aid reliant IDPs and vanishes when considering the comparison group of non-IDPs. The results are especially strong for consumption quantities (items and kilograms), which are most easily subject to intentional misreporting. This suggests that IDPs are indeed misreporting. The paper has two main limitations. First, it can only compare the treated group against an estimate of the "true" consumption rates. Second, the intervention is bundled. For this reason, it is impossible to isolate the causal mechanism affecting the observed changes in reporting. Further work is needed to identify an estimate of the true level of consumption against which to compare the primed individuals and to isolate the causal mechanisms by which people are changing their behavior.

Applications in South Sudan and Somalia

Pape and Parisotto (2019) Estimating Poverty in a Fragile Context -- The High Frequency Survey in South Sudan

Civil war broke out across The Republic of South Sudan in December 2013 only two years after gaining independence on the 9th of July 2011. The South Sudanese conflict has since continued to escalate, resulting in a large-scale humanitarian crisis where more than a third of the population has been forcibly displaced (Pape, Parisotto et al. 2018). Given the extremely difficult context, very little was known about welfare and livelihoods during the early years of the country's independence in 2011. The last nationally representative household survey measuring consumption and poverty was conducted as far back as 2009. To fill this data gap, the High Frequency South Sudan Survey (HFS) conducted several waves of representative surveys across seven of the ten former states between 2015 and 2017. In the period prior to and during the first wave of the HFS in 2015, conflict had primarily been concentrated in the Greater Upper Nile region. This period of relative stability across the remaining Greater Equatoria and Greater Bah El-Ghazal regions allowed the preparation and relatively calm implementation of Waves 1 and 2 of the country in 2015 and early 2016.

In summer 2016, however, clashes broke out in Juba. The escalation of the conflict coincided with the beginning of the implementation of Wave 3 of the HFS, a second urban-rural representative wave measuring consumption and poverty. The third wave of the HFS provides a relatively rare and extremely valuable glimpse of trends in welfare, consumption, and poverty in a country going through a period of upheaval. Indeed, the South Sudanese economy has since displayed all the characteristics of a war economy, including severe output contraction, rapid currency devaluation, and soaring inflation. Unsurprisingly, driven by these powerful shocks the incidence of poverty has risen to extremely high levels. In 2016, the HFS estimated that more than 4 in 5 people across seven of the ten former states were living under the international poverty line of US$ 1.90 PPP 2011 (82 percent). Such high levels of deprivation are not merely a direct result of the crisis but also reflect a history of instability, characterized by a poorly functioning state and a lack of institutional services provision (de Waal 2014, de Vries and Schomerus 2017). In 2017, South Sudan ranked 187 of 189 countries in the Human Development Index, with a life expectancy of merely 57 years.

The HFS was designed with the expectation of potential instability and thus capitalized on recent technological and methodological innovations to obtain reliable national poverty statistics in difficult contexts. Closely monitoring fieldwork is key to implementing such a large project in a risky context. The HFS leveraged the expansion of cellular networks across South Sudan to build a near real-time monitoring system, whereby the data could be uploaded daily to a dedicated server and checked for consistency. Computer Assisted Personal Interviewing (CAPI) also allowed built-in consistency checks, eliminating the need for expensive and potentially dangerous re-visits. Adherence to the sample design can be closely monitored with GPS software, tracking enumerators inside and outside areas with mobile phone coverage. The HFS also leveraged innovations in questionnaire design which permitted reducing the number of consumption items asked to the respondents while still obtaining unbiased poverty estimates through within-survey multiple imputation (Pape and Mistiaen 2018, Pape 2021). The lower amount of time spent collecting consumption data allowed the HFS to devote more time to collecting complementary data. Indeed, the HFS questionnaires contained additional modules covering asset ownership, education, labor market outcomes, perceptions of government performance and provision of public goods and services, psychological well-being, perceptions of violence and safety, allowing a well-rounded depiction of welfare and livelihoods.

The rapid escalation of the conflict in the summer of 2016, including several violent incidents affecting international humanitarian and development staff, led to the closure of the World Bank Office in South Sudan, disrupting the implementation of the third wave of the HFS. Therefore, the survey implementing agency implemented the third wave of the survey more independently relying mainly on remote support. A multitude of challenges had to be met, including large inflation, fuel unavailability, electricity shutdowns, insecurity, delay in payment of staff salaries, high NBS staff volatility, and cash flow limitations. Even though the implementing agency and the World Bank project team managed to mitigate a number of those challenges, the final sample reached only about 50 percent of the intended sample size. Nevertheless, this paper will argue that despite the enormity of challenges faced during fieldwork and the slight methodological departures from established approaches to poverty estimation (Deaton and Zaidi 2002), the data collected by the HFS provide the best-possible insights on welfare and livelihoods during a critical period of the country's history.

Despite initial intentions to expand the HFS across the entire country, continued insecurity made it impossible to reach the former states of Jonglei, Unity, and Upper Nile. To account for this gap in coverage and obtain countrywide poverty rates, a statistical model imputes poverty in inaccessible areas. The resulting poverty predictions are intended as supplemental to the survey estimates and serve as a proof-of-concept for using geo-spatial information alongside on-the-ground data collection. A growing body of research has emerged leveraging the increasing availability of alternative data sources such as satellite imagery and other geo-spatial characteristics. The estimates are derived by exploring the potential correlations between existing spatial data sets as well as custom-derived spatial data with geo-referenced poverty estimates obtained in the HFS. Once a set of spatial correlates were selected several models were trained and evaluated using a cross-validation approach. The final model was used to predict poverty rates at the 100m*100m level into all settled areas of the country including where survey data were not available. To aggregate the estimates at the state and county level, the 100m*100m level are weighted using a newly developed data set of human settlements across South Sudan constructed by combining a variety of publicly available data sources.

The data from the HFS are complemented by video testimonials providing a glimpse of the lives of the people of South Sudan. At the end of the interviews, respondents are offered to provide a short video testimonial where they can share their views and give a sense of their lives. The testimonials capture the dire situation on the ground and provide a much richer qualitative picture that accompanies and complements the quantitative data. While the data may help the government fine tune its policies, the videos may reach a broader audience and depict the sense of powerlessness, the pain of hunger, the stress of hopelessness and the feelings of disappointment that express people's experiences. Overall, this helps to create a more rounded perception of the situation on the ground in South Sudan.[5]

The paper describes in detail the design and analysis of the third wave of the HFS in 2016. [6] We describe the survey design and implementation, including the deviations from the original sample frame presenting consistency-checks used to evaluate potential selection issues that affect representativeness. We detail the process of calculating consumption aggregates and estimating poverty using within-survey multiple imputations, including calculating durables consumption flow and spatial-time deflators. We give a brief overview of the results of the poverty estimation, while a comprehensive assessment of poverty trends is available in Pape, Parisotto et al. (2018). We describe the estimation of poverty rates using satellite data as a proof-of-concept and conclude with a short discussion of main limitations.

Pape and Wollburg (2019) Estimation of Poverty in Somalia Using Innovative Methodologies

Somalia gained independence in 1960. The collapse of Siad Barre's post-independence regime in 1991 led to civil war between local power factions and dismantled the central state completely. Between 1995 and 2000, regional administrations emerged across the country, as security improved and

[5] https://blogs.worldbank.org/voices/giving-voice-poor-adding-human-touch-poverty-data-south-sudan

[6] The data from Wave 3 (2016) of the HFS and the code used to process these data can be downloaded from the World Bank MicroData Library at the following link: http://microdatalib.worldbank.org/index.php/catalog/9584/

economic development accelerated.[7] The formation of the Transitional Federal Government in 2004 and of its successor, the Federal Government of Somalia, in 2012 marked the return of a significant central state institution. After peaceful elections in 2016, a new government was formed in 2017 committed to embark on a development trajectory.

Though Somalia remains one of the world's poorest countries, a vibrant but largely informal private sector sprouted in the absence of government, drove growth in the Somali economy, and took on the provision of services (World Bank 2016). Several economic activities including telecommunications, money transfer businesses, livestock exports, and localized electricity services grew well during this period. Large-scale out-migration of skilled Somalis who sent back part of their earnings made diaspora remittances essential to the Somali economy, equivalent to between 23 and 38 percent of GDP and outweighing both international aid flows and foreign direct investment.

Despite improvements in political stability, Somalia remains fragile. Parts of southern Somalia are inaccessible due to the presence of Al-Shabaab, which also repeatedly carried out terroristic attacks, and violent clashes between various power factions continue to occur throughout the territory. In addition to conflict, the cyclical El Nino phenomenon caused severe droughts in 1991/92, 2011/12, and 2016/17 which exacerbated preexisting vulnerabilities in the Somali population. Both conflict and drought have led to large-scale internal displacement (Pape and Sharma 2019). The recent 2016/17 drought led to the displacement of approximately one million Somalis, adding to an existing population of internally displaced persons of 1.1 million.

As is typical for fragile states, Somalia is highly data-deprived, leaving policy makers to operate in a statistical vacuum (Beegle, Christiaensen et al. 2016). Specifically, years of civil war and ongoing conflict have eroded Somalia's statistical infrastructure and capacity, leading to the lack of key macro- and micro-economic indicators, including the poverty rate (Hoogeveen and Nguyen 2017). The government conducted and published the last full population census in 1975, while Somalia Socioeconomic Survey of 2002 was the last country-wide household survey (UNFPA 2014). Most recent existing data sources are local FSNAU and FAO food and nutrition surveys, while organizations operating within Somalia implemented a range of smaller surveys. In 2014, UNFPA implemented the first nationwide Population Estimation Survey (PESS) in preparation for a national census, finding the total population to be 12.3 million, of which 42 percent are urban, 23 percent rural, 26 percent nomadic, and 9 percent are internally displaced (UNFPA 2014).

Funded by the World Bank, Somaliland carried out a household budget survey (SLHS) in 2013, which generated much-needed indicators, including poverty estimates, but the sample was not representative especially for the rural population and did not cover the nomadic and displaced populations. In spring 2016, the World Bank conducted the first wave of the Somali High Frequency Survey (SHFS), representative of the accessible urban, rural, and IDP population in 9 of 18 prewar regions as well as Mogadishu, providing a baseline dataset for monitoring poverty and contributing to other key statistical indicators. However, in addition to large inaccessible areas, the sample excluded nomadic population and households in insecure areas. Furthermore, the rural sampling frame had to be derived ad-hoc with only limited representativeness. Wave 2 of the SHFS, implemented in December of 2017, significantly expanded coverage to urban and rural areas in central and southern Somalia and included the nomadic population for the first time, while a newly derived sampling frame enhanced overall representativeness.

[7] Somaliland self-declared independence in 1991.

This paper describes how the specific context of insecurity and lack of statistical infrastructure in Somalia posed a number of challenges for implementing a household survey and measuring poverty, which provided the evidence base for a comprehensive Poverty Assessment (Pape and Karamba 2019). First, in the absence of a recent census, no exhaustive lists of census enumeration areas along with population estimates existed, creating challenges to derive a probability-based representative sample. Second, while some areas remained completely inaccessible due to insecurity, even most accessible areas held potential risks to the safety of field staff and survey respondents, so that time spent in these areas had to be minimized. Third, poverty in completely inaccessible areas had to be estimated by other means. Finally, the non-stationary nature of the nomadic population required special sampling strategies. Next, the paper outlines how these challenges were overcome in wave 2 of the SHFS through methodological and technological adaptations.

First, geospatial techniques and high-resolution imagery were used in the SHFS to model the spatial population distribution, build a probability-based population sampling frame, and generate enumeration areas in an effort to overcome the lack of a recent population census (Qader, Lefebvre et al. 2021). The SHFS sampling strategy bears resemblance to the strategy proposed by Muñoz and Langeraar (2013), which relies satellite imagery and grid cells to build a sampling frame in Myanmar.

Second, risks to the safety of field staff required spending as little time in enumeration areas as possible. One strategy to address this issue is to call or message respondents on their mobile phones and not visit dangerous areas at all. A growing body of literature explores the use of mobile technology in this context (Dillon 2012, Demobynes and Sofia 2016, Firchow and Mac Ginty 2016). However, administration of necessary consumption modules to estimate poverty is not feasible via phone surveys. To address security concerns, the SHFS adapted logistical arrangements, sampling strategy, and questionnaire design to limit time on the ground. In logistical arrangements, a detailed and timely security assessment ensured that the enumeration areas to-be-visited were safe on the day of fieldwork. Concerning sampling strategy, it was not feasible to conduct a full listing of all households in an enumeration area, as this was too time-intensive and may have raised suspicion. Instead, a micro-listing approach was used, which required enumeration areas to be segmented into smaller enumeration blocks using satellite imagery (Himelein, Eckman et al. 2016). Complete food and nonfood consumption modules result in an overall questionnaire length that is prohibitive in areas with high insecurity. Hence, the Rapid Consumption Methodology (Pape and Mistiaen 2018, Pape 2021) was used to significantly reduce the length of the survey's consumption modules.

Third, the SHFS relies on correlates derived from satellite imagery and other geo-spatial data to estimate poverty in areas that remained completely inaccessible as a result mainly of insecurity. A growing field of research is dedicated to predicting a range of outcomes based on a diverse set of such data sources. Early applications use night-time lights data to predict economic activity. These data are particularly successful at predicting GDP at the country-level (Henderson, Storeygard et al. 2012, Pinkovskiy and Sala-i-Martin 2016), but appear less well-suited for measuring income and when variation in welfare is desired at a highly disaggregated level (Mellander, Lobo et al. 2015, Engstrom, Hersh et al. 2017). More recently, deep learning techniques applied to daytime imagery in order to classify such objects as roof types, roads, tree coverage, and crops has led to advances in measuring welfare at more disaggregated levels (Krizhevsky, Sutskever et al. 2012, Neal, Burke et al. 2016). In the SHFS, estimating poverty in inaccessible areas relied on a linear model with the objective of creating reliable and transparent poverty measures.

2. Impacts of shocks and fragility

It is important to measure impacts of shocks on livelihoods and socio-economic variables to design the right kind of responses. It is also important to document such shocks to argue for resilience measures, especially in the case of natural hazards. Blattman and Miguel (2010) provide an extensive overview covering the impact of poverty and other factors on conflict as well as the impact of conflict on livelihoods and human capital, which is our focus here. The problem of identifying the impact of a shock or conflict on potential outcomes like livelihoods is a more general problem of identifying a causal impact. Inheriting the vocabulary from health research, we talk about 'treatment' for subjects exposed to the shock (or individuals as being treated), and 'control' if they were not exposed to the shock. In theory, the causal effect is defined using a counter-factual. We are interested in the difference of the outcome of interest in the presence of a shock compared to the outcome in the absence of a shock. We can denote the outcome of interest for an individual i as Y_{1i} if the individual was treated (exposed to the shock), and as Y_{0i} if the individual was not treated. The causal effect is $Y_{1i} - Y_{0i}$.

The difficulty of measuring the causal effect is the fact that we cannot observe the same individual treated and non-treated at the same time. Therefore, we rely on comparing the expected outcome across different individuals. Following the exposition in Angrist and Pischke (2009), we can capture this idea by using the expected value and write the average treatment effect on the treated as $E[Y_{1i} - Y_{0i}|D_i = 1]$ where we indicate that individual i was in fact treated as $D_i = 1$. As before, we still cannot observe Y_{1i} and Y_{0i} for the same individual. However, we can decompose the average treatment effect on the treated as

$$E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_i|D_i = 1] - E[Y_i|D_i = 0] - (E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])$$

Where we define $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$, so that $Y_i = Y_{0i}$ if $D_i = 0$ and $Y_i = Y_{1i}$ if $D_i = 1$. We can interpret $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$ as the observed difference in the outcome between treat and non-treat individuals, which becomes particular apparent when plugging in the definition for Y_i. In our case, this would be the observed average outcome of individuals exposed to the shock minus the observed average outcome of individuals, who are not exposed to the shock. Even though this is readily observed, this is not the same as the average treatment effect on the treated, given the second part of the equation $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$, which is called the selection bias. In other words, the observed difference of outcomes is a biased estimator for the average treatment effect on the treated, because of the selection bias. The selection bias is the difference of the unobservable expected untreated outcome for an individual, who has been treated, and the observable expected outcome for an untreated individual. Again, the problem is that we cannot observe the counter-factual. The selection bias can be positive or negative, and can become very large (in absolute terms) so that the observed difference of outcome in most cases is not a reasonable estimator for the average treatment effect on the treated.

We can overcome the challenge of selection bias by randomly assigning the treatment. This makes Y_{0i} independent of D_i, as it allows us to swap $D_i = 0$ in the last expected value with $D_i = 1$:

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\ &= E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}] \end{aligned}$$

Thus, an experiment, where treatment can be truly randomly assigned, can use the observed difference in the outcome between the treated and the non-treated (control) as an unbiased estimator for the average treatment effect on the treated. This is the reason for the popularity of randomized control trials (RCTs) as they are factually the gold standard to identify causal effects. An example of an RCT from the first part of the thesis is the impact assessment of the behavioral nudges to reduce misreporting, by randomly assigning respondents to be treated by nudges or to be in the control group (Kaplan, Pape et al. 2018).

To assess the impact of a shock, however, such a design is in most cases unethical. Nobody would want to expose individuals randomly to major negative shocks, including conflict and violence. The rare exception are natural experiments where an approximately random assignment determines exposure to shocks (for an example, see Abadie and Gardeazabal 2003). We use such a natural experiment, which is discussed further below, in our study to assess the impact of a canceled program on its beneficiaries. The study exploits an RCT setup for an impact evaluation of a cash-transfer program, which had to be canceled because of the re-emergence of conflict (Müller, Pape et al. 2019). Hence, we were able to use the random assignment of beneficiaries to the program to assess the impact of the shock: the cancellation of the program.

To better understand how we can deal with the selection bias in the absence of random assignment, it is helpful to rewrite the causal effect model with a regression framework. Assuming that all individuals have the same treatment effect, we can write $Y_{1i} - Y_{0i} = \rho$ without using an index i for ρ as we assume constant treatment effects. Denoting $E[Y_{0i}] = \alpha$ and $Y_{0i} - E[Y_{0i}] = \eta_i$, we obtain from the definition of Y_i and expanding with $E[Y_{0i}]$

$$\begin{aligned} Y_i &= E[Y_{0i}] + (Y_{1i} - Y_{0i})D_i + Y_{0i} - E[Y_{0i}] \\ &= \alpha + \rho D_i + \eta_i \end{aligned}$$

Where η_i captures the random part of Y_{0i}, often called the error term. Based on this regression equation, we can evaluate the difference of the outcome Y_i conditional on being treated and non-treated

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= \alpha + \rho + E[\eta_i|D_i = 1] - (\alpha + E[\eta_i|D_i = 0]) \\ &= \rho + E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0] \end{aligned}$$

Where ρ reflects the constant treatment effect and the last two terms represent the selection bias as before. In other words, the selection bias is the correlation between the error term η_i and the regressor D_i. In the context of an RCT, the random assignment of the treatment D_i ensures that the error term (the random part of Y_{0i}) is not correlated with the treatment D_i itself, confirming the result above that the selection bias vanishes.

Even in the absence of a random treatment, we can theoretically assess the treatment effect without selection bias. Assume that the true causal model is

$$Y_i = \alpha + \rho D_i + X_i^T \gamma + e_i$$

Where X_i^T is a (transposed) vector of individuals' characteristics that also affect the outcome Y_i, and γ the corresponding regression coefficient. If this vector X_i^T of characteristics is comprehensive in a sense that it includes all characteristics that affect the outcome Y_i, then the error term e_i is independent of the treatment D_i so that we can directly identify the causal effect by estimating ρ. However, we usually face two challenges. First, we do not always know which characteristics are affecting the outcome. Second, even if we know which characteristics, they might be unobservable.

Failing to include all relevant characteristics into the regressions leads to the famous omitted variable bias. For simplicity, let us assume that we fail to include all characteristics X_i^T. Hence, the omitted variables will be captured in the error term $\eta_i = X_i^T \gamma + e_i$. The omitted variable bias depends on the influence of the omitted variables on the outcome and the treatment. In this case, the estimated regression coefficient of the regression where X_i^T is omitted becomes

$$\frac{Cov(Y_i, D_i)}{V(D_i)} = \rho + \gamma^T \delta_{X,D}$$

Where $\delta_{X,D}$ is the vector of coefficients from regressions of the omitted variables on the treatment. Thus, the estimator is biased by the influence of the omitted variables on the outcome γ and by the correlation of the omitted variables on the treatment $\delta_{X,D}$. Thus, only variables that both influence the outcome and are correlated with the treatment matter for the omitted variable bias.[8] In most cases, it is not possible to predict the direction of the bias. Thus, the regression with omitted variables cannot be interpreted in a causal way.

Several techniques can be used to improve the identification of a causal effect. Here, we will discuss instrumental variables as well as fixed effect and difference-in-difference identification, as those are used in the studies presented in this thesis. However, other techniques like regression discontinuity design (Cook 2008) and interrupted time series analysis can also be helpful (see for an overview: Pape, Millett et al. 2013, and for an example: Pape, Huckvale et al. 2015).

An instrumental variable is a variable that is correlated with the causal variable of interest D_i and does not affect the outcome Y_i except through the treatment. The second requirement is called the exclusion restriction. Conceptually, the instrumental variable focuses the regression on the variance of the treatment explained by the instrumental variable. Given the exclusion restriction, this guarantees that the coefficient ρ only captures the causal impact of D_i. However, it is important that the variance explained by the instrumental variable is sufficiently large.[9]

Denoting the instrumental variable as Z_i, the exclusion restrictions means that $Cov(\eta_i, Z_i) = 0$, or equivalently that Z_i is uncorrelated with X_i^T and e_i. Hence, the coefficient of interest ρ can be expressed as:

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} = \frac{Cov(Y_i, Z_i)/V(Z_i)}{Cov(D_i, Z_i)/V(Z_i)}$$

Thus, the causal coefficient ρ is the ratio of the regression coefficients from regressing Y_i on Z_i (called the reduced form) and regressing D_i on Z_i (called the first stage). Instrumental variables are frequently applied in determining the impact of conflict (e.g., Montalvo and Reynal-Querol 2007, Miguel and Roland 2011). We apply this concept in Nunez-Chaim and Pape (2022) and Heemann, Pape et al. (2022).

Without the feasibility of an RTC and no appropriate instrumental variables, identifying causal impacts becomes more challenging. However, with a few assumptions on the nature of the causal impact and panel data (multiple data points for each unit of observation), we can incorporate individual fixed effects. These fixed effects capture any time-invariant differences between individuals, so that the remaining variance is only due to time-variant differences and the treatment. Assuming that individual

[8] Note that it can also be advantageous to include other co-variates that are correlated with the outcome but not with the treatment. As they will contribute to the explanation of the outcome, they will reduce the error term and, hence, improve precision. However, bad controls are variables that can be outcomes of the experiment themselves. They change the interpretation of the causal coefficient rendering it non-comparable between treatment and control. To avoid this, good controls are ideally determined before the treatment, for example, in a baseline survey.

[9] This is often captured in the guidance that the F-statistic of the regression of the treatment on the instrumental variable (also called first stage) is at least 10.

time-variant differences are negligible[10], this leads us directly to the identification of the causal effect. This assumption can be denoted as

$$E[Y_{0it}|X_i, t, D_{it}] = E[Y_{0it}|X_i, t]$$

Where we denote the unobservable but time-invariant individual characteristics as X_i, with t denoting the time period. This definition also assumes that the treatment is assigned as good as random, given the unobservable individual characteristics. Hence, we obtain the following regression equation:

$$E[Y_{0it}|X_i, t] = \alpha + \lambda_t + X_i^T \gamma$$

Assuming that the causal effect ρ is additive $E[Y_{1it}|X_i, t] = E[Y_{0it}|X_i, t] + \rho$, we obtain:

$$E[Y_{it}|X_i, t, D_{it}] = \alpha + \lambda_t + \rho D_{it} + X_i^T \gamma.$$

Note that we obtain this equation only because of more restrictive assumptions, including the absence of time-variant individual characteristics (though observable time-variant individual characteristics can readily be integrated into this framework) as well as the linear (additive) and constant nature of the causal impact. Equipped with this equation though, we can conveniently write

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + \varepsilon_{it}$$

Where $\varepsilon_{it} = Y_{0it} - E[Y_{0it}|X_i, t]$ and $\alpha_i = \alpha + X_i^T \gamma$. Hence, the non-observable time-invariant characteristics $X_i^T \gamma$ get absorbed in the individual fixed effect α_i. Instead of attempting to measure the non-observables, we use individual dummy variables (binary 0/1 variables) to capture the within-individual variance, such that λ_t captures the constant time effect and ρ the causal impact. While the number of dummy effects can quickly become large as it scales with the sample size, they do not necessarily need to be estimated since the equation can be expressed in differences from means or in differences across time points where the individual dummies get cancelled from the final equation.[11] Hence, fixed effects can be seen as differences across time within individuals.

The final extension to identify causal effects relates to an even more specific setup, where the treatment is constant within groups of individuals. For shocks, this is often the case, for example, when they have clear geographic boundaries. This brings us to the difference-in-difference approach, where the first difference is across time as before, and the second difference between treated and control groups (for an overview, see Bertrand, Duflo et al. 2004).

In the difference-in-difference framework, we assume an additive structure of the outcome, such that membership to group g and time t completely determine the outcome in the absence of treatment:

$$E\left[Y_{0igt}|g, t\right] = \gamma_g + \lambda_t.$$

However, in specific time periods t only specific groups g will be treated. In general, we use this structure to identify the causal impact by taking the difference between groups and time periods. We can express the individual outcome as:

$$Y_{igt} = \gamma_g + \lambda_t + \rho D_{gt} + \varepsilon_{igt}$$

Where D_{gt} is a dummy for belonging to the treated group in the time periods of the treatment, and $\rho = E\left[Y_{1igt} - Y_{0igt}|g, t\right]$ being a constant treatment effect, and $E\left[\varepsilon_{igt}|g, t\right] = 0$. For the remainder of the section, we assume – as in most applications – that we only have two groups (called treatment

[10] Note that observable time-variant individual characteristics can be included in the following exposition by adding the corresponding term to the equations. For simplicity, we assume that no time-variant individual characteristics exist.
[11] However, it remains important to obtain accurate estimates for standard errors taking into account the panel nature of the dataset.

and control) and two periods of time, such that the treatment group gets treatment only in the second period. We define the treated group as $g = 1$ and the time period of the treatment similarly as $t = 1$. Thus, we obtain:

$$E\left[Y_{igt} | g = 0, t = 1\right] - E\left[Y_{igt} | g = 0, t = 0\right] = \lambda_1 - \lambda_0$$

$$E\left[Y_{igt} | g = 1, t = 1\right] - E\left[Y_{igt} | g = 1, t = 0\right] = \lambda_1 - \lambda_0 + \rho$$

The difference of both equations identifies the causal impact ρ. It is important to note though that the key identifying assumption in this framework is the assumed parallel trend for groups g where only the treatment changes the trend for the treated group, while all other differences are captured in the fixed group effect. The parallel trend can be observed if more than two time periods are available.

The first known application of difference-in-differences can be found in Snow (1855) to provide evidence that cholera spreads through contaminated water (rather than air). In the context of conflict, the difference-in-difference approach has been widely applied (e.g., Angrist and Kugler 2008, Bundervoet, Verwimp et al. 2009, Shemyakina 2011). For example, Angrist and Kugler (2008) estimate the economic impact of a shift of coca-production from Peru and Bolivia to Colombia. We apply this framework in South Sudan to estimate the impact of conflict on livelihoods (Parisotto and Pape forthcoming) and on adolescent girls (Pape and Phipps 2018) as well as the impact of terrorist attacks in Somalia (Nunez-Chaim and Pape 2022). For shocks, we utilize the difference-in-difference approach to determine the impact of high inflation in South Sudan (Etang, Hounsa et al. 2022) and of droughts in Somalia (Pape and Wollburg 2019).[12]

The remainder of this chapter is structured into three sections about impact of conflict, shocks and implications for program design. The first section starts with the assessment of the impact of conflict on livelihoods (Parisotto and Pape forthcoming) and on adolescent girls (Pape and Phipps 2018). The section ends with the analysis of the short-term impacts of terrorist attacks in Mogadishu, Somalia (Nunez-Chaim and Pape 2022). The second section describes the impact of high inflation on livelihoods in South Sudan (Etang, Hounsa et al. 2022) and the impact of a drought on livelihoods in Somalia (Pape and Wollburg 2019). The final section uses the canceled cash transfer program with its accompanying RCT to learn lessons how to setup programs in fragile places (Müller, Pape et al. 2019).

Impact of Conflict

Parisotto and Pape (forthcoming) Impact of Conflict on Livelihoods in South Sudan

Civil wars and violent conflict are inextricably linked with poverty. The 19 countries classified by FAO as being in a protracted food crisis in 2017 were all experiencing violent conflict. Similarly, 60 percent of the 815 million people who are undernourished and 79 percent of the 155 million stunted children worldwide live in countries affected by violent conflict (FAO, IFAD et al. 2017, Brück and d'Errico 2019). A large macro-level literature documents the empirical association between conflict and poverty (Collier and Hoeffler 2007, Blattman and Miguel 2010). Poverty is both a strong predictor for the onset of conflict and the incidence of conflict is associated with heightened deprivation and poverty, at least in the short run. More recently, the increasing availability of comprehensive micro-data from post-conflict regions has led to the emergence of a new literature examining the consequences of conflict exposure and its mechanisms at the household level (Justino 2009, Justino 2012, Martin-Shields and Stojetz 2019).

[12] In the absence of panel data at the individual level, we often have to retreat to determine the impact at a higher geographical level like survey clusters.

Some of the more salient findings that have emerged from this micro-level literature concerns the persistence of the impact of conflict exposure on human capital within countries. It manifests itself primarily through lower anthropometric and health outcomes, which can be observed even long after the end of the fighting (e.g., Bundervoet, Verwimp et al. 2009, Shemyakina 2011, Minoiu and Shemyakina 2014). This body of evidence rests on the well-established empirical observation that exposure to adverse events during critical developmental years can have long-lasting consequences on individuals' future outcomes (Almond and Currie 2011). For example, children born in conflict-exposed regions had significantly lower height-for-age z-scores and higher rates of stunting than those who were not (Bundervoet, Verwimp et al. 2009, Akresh, Verwimp et al. 2011). Akresh, Bhalotra et al. (2012) show that, even 40 years later, Nigerian women exposed to the Biafran civil war were shorter than their counterparts on average, especially for those exposed between 13-16 years old. Similarly, Camacho (2008) first showed that Colombian women exposed to violent conflict during the first three months of their pregnancy gave birth to children with lower birth weights. A smaller but related literature looks more specifically at the impact of conflict on poverty and food security, arguably the primary driver of these differences in human capital, and finds a strong association between conflict exposure and greater food insecurity and lower consumption levels (Dabalen and Paul 2012, D'Souza and Jolliffe 2013, Mercier, Ngenzebuke et al. 2016).

Due to the lack of micro-data collected during or shortly after conflict exposure, much of this literature is forced to rely on ex-post data typically collected several years after the end of the conflict, and as such remains relatively silent on the short-term impacts of conflict exposure. While there is an implicit link between conflict and food insecurity, strong evidence documenting this relationship is still lacking (Martin-Shields and Stojetz 2019). This paper contributes to this literature by leveraging representative consumption expenditure data collected during the most recent conflict in South Sudan.

South Sudan gained its independence in July 2011, it was only two years later, in December 2013, that clashes broke out in Juba between factions of soldiers loyal to President Salva Kiir and factions loyal to former vice-president Riek Machar. This was followed by a wave of violence sweeping throughout the country. Although a peace agreement was signed in August 2015, a constant state of violence largely prevailed throughout the country. The conflict intensified in July 2016 after renewed clashes in Juba, and by the end of 2017 the conflict had escalated into a large-scale humanitarian crisis, with almost 4.5 million people forcibly displaced and 6 million facing heightened food insecurity – out of a population of about 12 million (UNHCR 2018). It was during this period of intense violence, between late 2016 to early 2017, that the High Frequency Survey in South Sudan (HFS) conducted a representative consumption and expenditure survey. The HFS was explicitly designed to be conducted in a context of insecurity and many households were thus be interviewed very recently after having been exposed to the conflict – up to less than a month following exposure (Pape and Parisotto 2019).

This study combines the HFS 2016-17 data with data collected before the conflict began during the National Baseline Survey in 2009 to estimate a repeated cross-section difference-in-differences model of the impact of conflict exposure on welfare, as proxied by average consumption levels, the poverty headcount, and poverty gap. By differencing across time and across groups, DID estimation nets out both group specific heterogeneities and overall time trends, which is key in dealing with the non-random incidence of exposure to conflict and the macro-economic crisis driven by the rapid devaluation of the South Sudanese Pound over the later phase of the conflict. Given the repeated cross-section empirical setup this study relies on an external measure of conflict exposure, derived from geo-coded event data collected by the Armed Conflict Event & Location Data Project (ACLED). Given this estimation strategy, results are capturing the broader impact of residing in an area exposed to conflict and insecurity. This includes households which are directly subject to conflict or violent events like looting, as well as households which are not. Our study shows that the conflict led to a large decline in consumption levels and a corresponding increase in poverty across the entire country. However, households residing in areas that were exposed to more intense violence experienced greater declines in average consumption, higher poverty incidence as well as deeper poverty. The results are driven by households residing in areas exposed to high-intensity conflict related violence, proxied by total conflict fatalities.

Pape and Phipps (2018) Impact of conflict on adolescent girls in South Sudan

Conflict and displacement escalated dramatically after the civil war in South Sudan in December 2013. The December 2013 conflict between President Salva Kiir and former Vice President Riek Machar quickly became an ethnically-charged conflict particularly between the Dinka and Nuer ethnic groups. Skirmishes as well as brutal violence against civilians were reported in dozens of locations. In the days following the start of the conflict, incidences were more isolated with violence against Nuer civilians in Juba, attacks by Nuer on Dinka and other civilians in these areas as well as incidences of armed groups of different ethnic backgrounds launching revenge attacks on community members. The civil war with high rates of violence resulted in high mortality and displacement, as well as worsening livelihoods, poverty and food insecurity (Shankleman 2011, World Bank 2014, World Bank 2015).

More than 50,000 civilians have been killed since the resurgence of conflict in December 2013, in addition to various severe crimes including extrajudicial killings, abductions, rape, and torture. More than 2.2 million people have also fled the country or have been displaced internally, and it is believed that 4.8 million are at risk of famine (FAO 2017). The conflict has severely impacted welfare indicators and cost the country an estimated 6.3 percent of its GDP (World Bank 2016).

Violent conflict and instability affect men and women in heterogeneous ways, including differentiated impacts on economic, social, physical and mental well-being. Research highlights that boys and men often confront direct, first-round effects of conflict, including death and morbidity, while conflict contributes to indirect impacts on women and girls, including health-related impacts like malnutrition, exposure to disease and lack of access to health services (Buvinic, Das Gupta et al. 2012). Children's health and access to education are often severely affected by exposure to conflict.

In many countries, women and children frequently account for the majority of populations displaced by conflict. In South Sudan, for example, 53 percent of the 2.43 million externally displaced due to the 2013 conflict are female while 63 percent of those displaced are children under the age of 18 (UNHCR 2018). While displacement generally contributes to a critical loss in assets, including housing, land and property and other productive assets, women confront particular constraints extending from social norms that restrict women's ownership rights over land and other assets, and contributes to their exclusion from decision-making processes (Cagoco-Guiam 2013). Displacement also often gives rise to or exacerbates serious protection challenges including increased exposure to gender-based violence.

Violent conflict often changes the demographic composition of households, contributing to a rise in female-headed households due to the extended absence of males either due to conflict or abnormal migration. These shifts impact traditional gendered division of tasks through its impacts on household composition, often increases women's participation in labor markets and augmenting responsibilities of women within households (Annan, Blattman et al. 2009, Brück and Vothknecht 2011, Justino, Ivan et al. 2012, Menon and Rodgers 2013). At the same time, data on whether women's greater market participation and shifts in household responsibilities contributes to wider welfare gains and long-term social empowerment, however, is more ambiguous (Bozzoli and Brück 2009, Justino, Ivan et al. 2012). There are data to suggest that the economic and social gains women may have achieved due to the absence of men during conflict periods can erode during post-conflict periods due to a reversion in pre-conflict norms and do not always result in a comparable increase in social empowerment or improved bargaining power (Justino 2009). Non-material well-being, such as marriage outcomes and happiness, has also been negatively impacted by conflict and displacement in some cases (Wang and Weina 2016). Robust evidence also exists on the positive correlation between rates and incidence of varying forms of gender-based violence (GBV), including sexual and physical assault, intimate partner violence, trafficking and early and forced marriage, as well as exposure to conflict (Annan, Blattman et al. 2009, Dijkman, Catrien et al. 2014, Ostby 2016). Lastly, studies have found that women are more vulnerable to developing anxiety disorders and struggling with psychosocial distress in conflict-affected settings (Murthy and Lakshminarayani 2006, Roberts, Ocaka et al. 2008, Farhood and Dimassi 2012, Luitel, Jordans et al. 2013, Ayazi, Lien et al. 2014).

The devastating nature of the recent conflict in South Sudan and the grim reality of its gendered effects provides the motivation for this study. The conflict has affected millions of South Sudanese people, among those also of a particularly vulnerable group: adolescent girls. Economic, social, and mental impacts at an early age tend to be long-lasting and should be addressed before they worsen and persist. Therefore, this study aims to measure the impact of this conflict on adolescent girls across a set of welfare indicators to inform and guide appropriate intervention strategies.

There is growing consensus that studying conflict cannot be dissociated from how it is experienced and perceived by individuals affected by armed violence. Econometric research on the various channels through which conflict affects women, however, and the impact of conflict on gender dynamics is relatively nascent (Ibáñez, Calderón et al. 2011, Justino, Ivan et al. 2012). Within the literature on the intersection between conflict and gender dynamics, there is scant research on non-combatant adolescent girls.[13] This study contributes to this literature by offering one of the first efforts to empirically quantify the impact of violence and conflict on educational attainment, labor market behavior, and social empowerment for non-combatant adolescent girls.

This study utilizes survey data emerging from a World Bank-administered pilot project in South Sudan to contribute to the existing conflict and gender literature on several fronts. First and foremost, it uses a cluster-level difference-in-difference analysis to identify the impact of the conflict in South Sudan on girls aged 15-24. Given the high levels of mobility in South Sudan, these surveys are repeated cross-sections. Second, the study contributes new knowledge on the impact of conflict on welfare, poverty, and aspirations by offering one of the first analyses of data on adolescent girls, a generally under-researched demographic. Finally, this research contributes to a growing body of evidence examining the impacts of earlier-life environment on later life outcomes, and is closely related to a large body of literature on subjective perceptions of well-being linked to significant and potentially traumatic life events.

The analysis builds on two rounds of survey data that were collected for an impact evaluation of an adolescent girls program. The first round of data was collected between August and October 2010 and the second round of data was collected between January and February 2015. The two surveys measure the same indicators except that the endline survey has an additional module on conflict exposure. We use the data from the conflict exposure module to obtain self-reported measures of cluster-level exposure to the conflict and examine the impact of conflict-related victimization on adolescent girls. For robustness, we also use external data on conflict events to examine the impact of the conflict

[13] In this conflict, adolescent girls and young women did not constitute a significant number of participating combatants.

exogenously. This analysis tests the hypothesis that girls exposed to the conflict had statistically different welfare outcomes than girls who were not exposed to the conflict.

The results suggest that girls from clusters more affected by the conflict had statistically different outcomes compared with girls from less affected clusters. Specifically, there is strong evidence that the conflict negatively affected outcomes related to income opportunities, aspirations, marriage, and household characteristics, but increased self-reported empowerment and entrepreneurial potential scores. The results indicate that impacts on labor supply, personal motivation, household conditions, and other forms of victimization are important channels through which conflict negatively impacts adolescent girls.

Nunez-Chaim and Pape (2022) Poverty and violence: The immediate impact of terrorist attacks against civilians in Somalia

Somalia was plagued by violence and conflict for decades. In 2004, an interim central state was established with the aim of bringing political stability across Somali regions. The political transition culminated with the establishment of the Federal Government of Somalia in 2012 and a first electoral process in 2017. The elected government has aimed to improve national security conditions, yet the opportunity to ensure a development trajectory still faces many challenges. Somalia remains one of the poorest countries in Sub-Saharan Africa with 69% of the population living under the standard international poverty line of US$ 1.90 in 2017-18 (Pape and Karamba 2019). Progress in poverty reduction faces many challenges in Somalia, one of them is continued violence through terrorist attacks.

At a first glance, the consequences of a terrorist attack might seem small and contained given that they usually affect a small fraction of the population and the economy. Yet, several studies suggest sizeable effects on economic outcomes (Abadie and Gardeazabal 2008). Further, nearly two-thirds of the poor around the world are projected to live in conflict-affected countries by 2030, including Somalia (World Bank 2020). Therefore, it is important to shed light and improve our understanding on the links between conflict and poverty.

This paper estimates the immediate (within a week) impact of terrorist attacks from Al-Shabaab against civilians in Somalia using micro-data from two waves of the Somali High Frequency Survey (SHFS), combined with geo-tagged information on attacks. We exploit the spatial and time variation of interviews through a difference-in-difference identification strategy that compares outcomes of control and exposed households, before and after terrorist incidents. We also derive a shift-share instrument using changes in the number of US air/drone attacks against Al-Shabaab and employ an instrumental variables approach. We provide evidence to support the validity of our identification strategies and that our estimates are robust to different specifications, samples considered and several sensitivity checks.

The literature models terrorists as rational actors, with terrorism having large consequences on economic outcomes, besides the loss of life, damage to persons and negative psychological effects. Conflict can also lead to sharp increases in poverty and vulnerability and other adverse outcomes (Pape, Parisotto et al. 2018, Parisotto and Pape forthcoming). We contribute to the literature on the intersection between poverty and adverse shocks in developing countries, as well as to the policy debate by quantifying the impact of terrorist attacks on consumption and poverty, describing which households are affected by such incidents and the mechanisms through which this is likely to occur. Most of the empirical literature on the effects of terrorism on economic outcomes has relied on data aggregated at some geographical level (district, region or country), while the growing body of research exploiting micro-data to understand the effect of various shocks on poverty has not analyzed the effect of terrorism. To our knowledge, this is the first study to measure the causal impact of terrorism on consumption and poverty using household-level data in a fragile and conflict-affected country.

Our results suggest that consumption of households exposed to terrorist incidents decreases by 33 percent, mainly driven by a decline on food consumption. The reduction in consumption increases poverty and the depth of poverty among the poor. The impact on consumption seems to be associated to a smaller share of household members (aged 15 to 50) working and earning income after an attack. In addition, we document that the negative impact on consumption is clustered within a 4-kilometer radius from the incident and has an heterogenous impact, not affecting households in the top 20 percent of the consumption distribution. The perception of police competence also worsen as a result of a terrorist incident.

Impact of Shocks

Etang, Hounsa et al. (2022) Impact of High Inflation on Household Livelihoods in Urban South Sudan

In the years from 2015 to 2017, the South Sudanese economy displayed all characteristics of a war economy, including severe output contraction, rapid currency devaluation, and soaring inflation. Oil dependency has tied the fate of the nation to the volatility of global commodity prices. Widespread fighting and large-scale displacement over several consecutive planting seasons have disrupted many households' normal agricultural activities, resulting in increasingly large production deficits each year and widespread food insecurity. Compounding on this, falling international oil prices triggered the rapid devaluation of the local currency driven by pressures from a low domestic supply of foreign currency, exacerbated by concurrent high domestic demand for foreign currency due to the need to supplement domestic production shortages with imported food. Falling oil prices also meant a collapse of Government revenues, which resorted to financing its deficit by printing money and incurring a growing stock of debt. Combined, these shocks have led to rapidly rising food prices, with the year-on-year CPI inflation reaching its peak at 549 percent in September 2016. While the level of inflation almost reached hyper-inflation, it remained – on an annual basis – still below the threshold of hyper-inflation.

An important and inevitable question is how inflation is affecting household livelihoods in South Sudan, particularly the poor. High inflation can have negative impacts on household livelihoods due to increased prices for consumed goods and services with lagging wage and social assistance increases. However, households that produce goods like food are usually less affected by high inflation as they are shielded from market prices. In fact, they can benefit from inflation if they sell products in the market. Also, other characteristics like product types and market access can influence how much a household loses or benefits. For non-agricultural households, type of employment, level of education and other factors can render households more resilient against shocks.

The theoretical causes and impacts of hyperinflation are well known, and provided in the seminal work of Cagan (1956) and Nordhaus (1973). A more recent review and update was conducted by Fischer, Sahay et al. (2002). A historic overview can be found in He (2020). Recent studies focus on the causes and policy options (e.g., Acemoglu, Johnson et al. 2003, Reinhart and Savastano 2003), and historic dimensions often in the context of Zimbabwe (for example, Coomer and Gstraunthaler 2011). Given the dearth of micro-data in countries with high- or hyperinflation, only very few studies look at the direct welfare impacts of high- or hyperinflation. Fajardo and Dantas (2018) study the impact of hyperinflation on investment behavior in Brazil. However, they do not touch on welfare or livelihood impacts. Larochelle, Alwang et al. (2014) uses a small-area-based approach with an asset index and finds that also rural poverty increased in Zimbabwe's hyperinflation period. In contrast, Kurasha (2021) uses micro-data from several years before and after hyperinflation, but finds that rural poverty fell while urban poverty increased, while asset inequality dropped during the hyperinflationary period. Health indicators worsened for both urban and rural as well as access to electricity, safe drinking water, improved toilets and healthcare. In contrast,

In our study, we assess the shorter-term impacts of high inflation on household livelihoods in urban South Sudan. Longitudinal micro-data for a representative sample of households is used to understand the changes in livelihoods between 2015 and 2017, accompanied by continuous price data collected across South Sudan. The novel datasets based on a set of innovative high frequency surveys allow the use of a difference-in-difference approach providing a stronger identification than can currently be found in the literature.

We find that inflation has a strong negative impact on urban poverty between 2015 and 2017, mainly driven by the increase of non-food prices. Food price inflation had a negative and statistically significant impact on girls' primary and secondary school attendance, while proximity to school is very important for girls' school attendance. Increases in food prices led to a decline in labor force participation, increasing unemployment among urban residents. Inflation is exacerbating food insecurity and hunger, particularly for the poorest households who are more vulnerable to hunger. Inflation has also negatively affected households' perceptions of welfare. These changes in welfare are mostly explained by the period of near hyper-inflation in 2017.

Pape and Wollburg (2019) Impact of Drought on Poverty in Somalia

Understanding the magnitude and importance of income shocks in causing and perpetuating poverty is critical to designing measures aimed at building resilience, contributing towards the goal of ending poverty. A growing body of literature provides empirical evidence of the micro-level impacts of adverse shocks in developing countries (Dercon and Hoddinott 2004). Dercon (2004), and Porter (2012) find that For example, weather shocks have a negative and long-lasting effect on consumption outcomes in rural Ethiopia (Dercon and Krishnan 2000, Dercon 2004, Porter 2012). Drought and price shocks reduce consumption and especially farm income, while increasing vulnerability to poverty in rural Ethiopia (Hill and Porter 2016) and Malawi (Makoka 2008). Similarly, Alem and Soderbom (2012) conclude that high food prices adversely affect households in urban Ethiopia, especially those relying on casual work and with low asset levels. Hill and Mejia-Mantilla (2017) find negative effects of drought, conflict, and prices on poverty levels in Uganda, and Parisotto and Pape (forthcoming) find a large and significant impact of conflict on poverty in South Sudan. Hoddinott and Kinsey (2001) and Alderman, Konde-Lule et al. (2006) show the causal relation between rainfall shocks and reduced human capital formation.

This paper contributes to the existing literature, by focusing on the impact of drought on poverty in Somalia. Four consecutive seasons of poor rains between April 2016 and December 2017 resulted in a severe drought across Somalia (FEWSNET 2018). The drought exacerbated preexisting food insecurity, as half of the population faced acute food insecurity in mid-2017. The drought threatened the livelihoods of many Somalis. Lack of water and pasture led to high livestock deaths and low birth rates, and induced distress selling caused the 26 percent of Somalis relying on livestock for their livelihoods to lose between 25 and 75 percent of their herds in the first half of 2017. Households depleted productive assets and food stocks to cope with the rising food and water prices, while weak demand for labor in the agricultural sector led to lower wage levels. As a result, the drought displaced close to one million people between 2016 and 2017. Large-scale humanitarian interventions provided critical relief to up to 3 million people to reduce the risk of famine.

Using data from two waves of the Somali High Frequency Survey (SHFS), this analysis employs a regression framework to measure the micro-level impact of the 2016/17 drought on poverty. It exploits spatial variation in the intensity of drought that different households experienced and compares consumption before and after the drought. Households' level of drought exposure is measured by using the Normalized Difference Vegetation Index (NDVI). The temporal difference is provided by the timing of the first two waves of the SHFS. The first wave took place before the onset of the drought in early 2016, while the second wave surveyed households in late 2017, when the drought had surpassed its peak.

The drought is found to have a sizable effect on poverty, consumption, and hunger in rural areas, where agricultural households and those lacking access to infrastructure and basic services are most severely affected. A renewed drought shock could lead to an increase in poverty of 9 percentage points. The findings underscore the importance of investing in rural resilience, especially among agricultural households.

Heemann, Pape et al. (2022) The Labor Market Implications of Restricted Mobility during the COVID-19 Pandemic in Kenya

SARS-CoV-2, the virus that causes COVID-19, has reached nearly every corner of the world, resulting in millions of deaths (Dong, Du et al. 2020). To decrease the spread of the virus, governments around the world have implemented lockdowns and mobility restrictions. Governments introduced peak stringency immediately following the onset of the COVID-19 pandemic in the first quarter of 2020, after which there was a mild decline in stringency. The pandemic dramatically slowed economic activity as governments implemented lockdown measures, individuals reacted by reducing both their mobility and economic activity, and firms' production processes were disrupted. These broader shifts in the economy affected both firms' demand for labor and workers' ability and willingness to work. In developed countries where data are readily available, early labor market impacts varied considerably across countries, depending on initial economic and labor market conditions and variations in policy responses (Khamis, Prinz et al. 2021). Unfortunately, however, most of the countries with post-crisis data are high-income countries, and there is little systematic knowledge about the labor market impacts of the crisis in developing countries, with the exception of the description of livelihood impacts in several developing countries (Egger, Miguel et al. 2021).

Kenya's first case of COVID-19 was recorded in March 2020. Since then, reported infections have considerably increased, peaking on October 31, 2020, with 1,395 new infections per day. Following Kenya's first case of confirmed COVID-19 in March 2020, the Government of Kenya quickly put in place multiple policies and measures to contain the spread of the virus. In March 2020, for instance, the Government of Kenya introduced a series of restrictions ranging from the closure of educational institutions to directing public and private sector workers to home-based work, except for essential workers. Entry into Kenya was limited to citizens and residents but required quarantine for 14 days while local air travel was suspended and resumed on July 15. These measures were followed by fast reductions in average mobility outside of residential areas but with an increase in residential movement.

Kenya's Rapid Response Phone Survey was deployed immediately after COVID-19 became a global pandemic. From May-June 2020 until April-June 2021, we interviewed 6,343 households consisting of both refugees and Kenyans over five survey waves. Our data is unique in at least four dimensions: i) it leverages the high cell phone penetration and coverage throughout the country, including the refugee communities, to reach households during lockdowns when face-to-face interviews are impossible to conduct; ii) its longitudinal nature allows not only to assess the first order impact of the COVID-19 shock but also its longer-term implications for recovery; iii) interviews cover refugees and nationals over the same period and are conducted in the same modality, allowing for a comparison between both communities; iv) the survey is nationally representative of both the national and the refugee population. It tracked socio-economic indicators over time, published a real-time dashboard[14] and results briefs (e.g., World Bank 2021). A detailed analysis of changes in socio-economic indicators for Kenya are published in Pape, Delius et al. (2021), succeeded by a follow-up report by Pape, Delius et al. (2021). More detailed deep dives identify specific impacts of COVID-19 for labor market outcomes of refugees vis-à-vis nationals (Vintar, Beltramo et al. 2022), the closing of schools and their implications for labor markets (Biscaye, Egger et al. 2022), the impacts on children (Cameron, Delius et al. 2022) and specifically on women (Xu, Delius et al. 2022).

As a response to the pandemic many governments have imposed two types of measures. Firstly, measures aimed at restricting mobility and social interaction to reduce the speed of further infection as well as, secondly, measures to mitigate the economic consequences on businesses and households. The consequences from the pandemic and restrictions on personal mobility have severely disrupted economic activities, as between one and four in five workers reside in countries with required workplace closures (ILO 2021).

Particularly for households in developing countries, the labor market implications of the pandemic can be dire. The lack of economic safety nets especially in the informal sector but also increased risk of infection and related expenses, especially for poor people living in high density areas with daily hands-on income, can exacerbate the consequences of losing parts of the income or the job entirely (Bargain and Ulugbek 2021, Gupta, Bavinck et al. 2021). Given the additional challenges households in developing countries face in coping with the crisis, it is elementary for policy makers to understand which socio-economic consequences any countermeasures aimed at curbing the spread of the virus may have. As governments react and impose restrictions to save lives, people subsequently change their behavior (e.g., reduce mobility) and this in turn affects labor markets. Therefore, a better understanding of the causal relationships between human behavior and labor market outcomes is vital to crafting better, more effective and targeted policies in future situations in which there is the joint goal of slowing down everyday life to save lives while minimizing the negative economic and societal effects.

Mobility is an outcome of labor market activity as well as something that drives labor market activity, for example by providing jobs in the transportations sector. Likewise, the ability to move determines whether people have access to markets to sell their goods, as well as whether customers can attain the goods that they would like to have. Finally, supply chains as well as trade rely on frictionless mobility, which in turn may impact production and thus labor markets further downstream (Espitia, Mattoo et al. 2022). Given that mobility was severely impacted by policy to curb the spread of the virus in Kenya, it is an interesting shock-like mechanism driving labor market outcomes to look at. We quantify the changes in labor market outcomes that were driven by changing mobility levels over the course of the pandemic in Kenya by applying instrumental variable analyses.

To better understand mobility levels as mechanism that drives labor market outcomes, it is important to better understand what drives policy adherence of citizens in the respective setting. Many studies have looked at determinants of mobility restriction and COVID-19 guidelines. However, most of them were either placed in developed countries (Al-Hasan, Yim et al. 2020, Coroiu, Moran et al. 2020) or lacked a representative sample size (Ahmed, Siewe Fodjo et al. 2020, Usman, Ssempijja et al. 2020). Given the importance of policy adherence to understand mobility levels, we complement our analysis by determining which factors were associated with respondents self-reported mobility reduction in Kenya over the course of the pandemic.

[14] www.kenyacovidtracker.org/rrps

We add to the literature by examining labor market effects driven by changing mobility levels that can be attributed both to the measures imposed by the Kenyan government as well as people's adherence to these policies, combining data on policy restrictions with insights from Google Mobility Reports and large-scale household surveys. As far as we are aware, this is the first paper to investigate the causal effects of changing mobility levels on labor market outcomes over the course of the pandemic in a developing country, especially in a nationally representative setting in a developing country with panel data reaching into early 2021.

Our findings show that a 10 percent recovery of mobility led to a 12 percentage points increase in labor force participation and a 9 percent points increase in household members being employed. At the same time, a 10 percent recovery of mobility caused an increase of 11 wage hours per week (formal and informal). Among the factors influencing self-reported mobility-reducing behavior, trust in the government's ability to deal with the pandemic correlates with less self-reported mobility reduction, while people who knew someone with an infection tend to reduce mobility less. Finally, countrywide policy stringency levels clearly reduce self-reported mobility. Given the demonstrated adverse impacts of reducing mobility on economic indicators, it would be advantageous for the government to explore options to limit the economic fall-out while protecting citizens from infections, for example, by using partial or geographically constrained lockdowns.

Implications for Program Design

Müller, Pape et al. (2019) Broken Promises: Evaluating an incomplete Cash Transfer Program

An increasing share of the world's poor live in fragile states, which poses new challenges to programs that seek to raise their incomes. One major risk associated with an insecure and fragile context is the unintended and unplanned interruption or cancellation of the program. Despite the prevalence of these cases, little is known about the effect of a program cancellation on intended beneficiaries. However, knowing about these risks would help policy makers make informed decisions about the costs and benefits of an intervention a priori. In addition, information on the consequences of failed implementation can help reduce detrimental impacts at the program design stage.

To our best knowledge, this study is the first to analyze what happens if an intended intervention is canceled. The Youth Startup Business Grant Program in South Sudan that was canceled due to erupting violence in 2016 provides us with the opportunity to study the impacts on socio-economic, behavioral and psychological outcomes on intended beneficiaries. In particular, we are interested in understanding effects on participants who were promised to receive a cash grant but did not ultimately receive it. Economic theory lends multiple reasons why outcomes for these participants could differ from outcomes in the absence of the program. Overall, our results suggest that the impact of failed interventions is mixed and depends on the gender of participants and their ex post treatment status. In this instance, on average across all participants, the invention was largely ineffective, but some sub-groups were negatively affected. Given that applicants for the intervention were on average more educated than the average youth in South Sudan, the average population might have displayed reduced skills to cope with the program cancelation. In that sense, findings present a lower bound.

The Youth Startup Business Grant Program consisted of an unconditional cash grant combined with a business and life skills training exercise and was particularly targeted at young women. South Sudan has suffered from political instability and latent conflict since its inception in 2011. In this context, the youth struggled with declining livelihoods and a lack of economic opportunities. This put them at risk of participating or becoming victims of criminal or violent activities. Young women were at particular risk. In response, the program was designed by the World Bank in collaboration with the Ministry of Commerce to offer a cash grant worth US$ 1,000. Existing evidence suggests that injections of capital are the most effective means of raising income in poor and fragile states (Blattman and Ralston 2015). Beneficiaries could access the grants denominated in local currency through a commercial bank account. Although the cash grant was aimed towards promoting (self-) employment and business development, beneficiaries were free to decide on its use. The program also entailed a one-week business and life-skill training, which participants needed to attend in order to access the grant.

In late 2014, the program randomly selected 1,200 beneficiaries out of a pool of more than 6,000 applications to receive the grant. More than 60 percent of the grants were awarded to young women. A similarly sized control group was selected to enable the assessment of the program in a rigorous impact evaluation. Baseline data from both treatment groups were collected before grant beneficiaries received their business and life skill training in April and May 2015. Almost all selected beneficiaries attended the 1-week training. After the training, participants were asked to open a commercial bank account in which the grant would be deposited.

Escalating violence at the end of 2015 forced the program to terminate the disbursement of the grants before all participants had accessed them. Completion of the program was first postponed and finally canceled to mitigate the perceived risk for beneficiaries to become the target of crime. In addition, there was the risk that the conflict might be exacerbated if grant money got into the wrong hands and was used to purchase arms. Delays in communication and in the processing of the grants meant that the timing at which disbursement was stopped varied across regions and bank branches.

Interventions in highly fragile and insecure states are often at risk of failing to be rolled out as originally planned. Obvious ethical objections make it impossible to study this effect in the form of a randomized-controlled trial. This study takes advantage of the circumstances under which the Youth Startup Business Grant Program was canceled to identify the socio-economic and behavioral consequences of projects that fail to be implemented as intended. Those originally assigned to the treatment group but who did not end up receiving grants show few systematic differences, except their location, from those who accessed the grants. We exploit this natural variation in location in interaction with the original assignment to the treatment group as an instrument for those who obtained the grants versus those who did not.

Hence, this study distinguishes between two *de facto* treatments. “Training but no grant” consists of having participated in the business skills training and been informed of receiving a US$ 1,000 grant, but later having to experience that the grant disbursement was stopped. To assess the treatment effect, this group will be compared to the control group of the original intervention who was informed of not having been selected to receive the grant. In addition, this study analyzes the effect of the originally planned intervention. “Training and grant” consists of having participated in the life-skills training and successfully having accessed the cash grant.

On average, across all participants most socio-economic, and behavioral and psychological indicators were neither negatively nor positively affected by the intervention. However, when considering ex post treatment groups and gender, some groups were detrimentally affected by the intervention. For example, participants who only received the training, but expected the grant also, seem to have experienced small consumption declines relative to the control group. Female participants among this group also showed a strong reduction in their trust level. We also found some evidence that these women were less likely to migrate. Given that large shares of the population in South Sudan migrated in the period of our analysis to escape conflict affected areas, it is possible that women who expected the grant stayed back who would have migrated in the absence of the intervention. While we do not have direct information on this unintended consequence, one could be concerned of the potential detrimental outcomes to these participants.

Positive impacts were only detected on some outcomes and only to those who received the grants. For example, consumption, savings and reductions in debt, as well as reported levels of psychological well-being increased among the participants receiving both the training and the grant. These positive effects seemed to be independent of gender. Given these results, we argue that greater concern should be taken when planning programs in these volatile environments, as there is at least some evidence from our results on unintended negative consequences on program participants who did not receive the full set of benefits anticipated at the program outset.

B. Part I: Measuring poverty

1. Measuring Poverty Rapidly Using Within-Survey Imputations[15]

Utz Pape

Introduction

Poverty is an indicator of paramount importance for gauging the socioeconomic well-being of a population. Especially during or after a shock, poverty estimates are invaluable for understanding the situation, as well as for assessing the severity of the impact and for identifying which parts of the population were most affected. Especially in the developing world, consumption-based monetary poverty measures are used, defining the poor as those households with consumption levels that fall below a set poverty line (Deaton and Zaidi 2002). The poverty line is usually set at a consumption level adequate for sustaining the minimum level of welfare required for healthy living (Ravallion 1998). Consumption-based poverty measures are widely used in development contexts and play a critical role in policy decisions (e.g., Beegle, Christiaensen et al. 2016).

The measurement of consumption, however, has traditionally been very time consuming. A typical household consumption questionnaire contains a series of questions about the price and quantity consumed for each item, and whether it has been purchased, self-produced, or bartered. Usually encompassing more than 200 food and nonfood items, the time required to administer such a questionnaire can often substantially exceed two hours. In addition to high administration costs due to long interview times, measurement errors may become significant towards the end of the questionnaire as enumerators and respondents become fatigued. Respondents might also cancel the interview before it is completed, thus contributing to a higher non-response rate.

Enumerator and respondent fatigue are well documented in the literature (Krosnick 1991, Tourangeau, Rips et al. 2000) and become more pronounced for longer questionnaires (Diehr, Chen et al. 2005, Snyder, Watson et al. 2007, Rolstad, Adler et al. 2011). Enumerator fatigue increases measurement errors often over the course of a day as well as over the time the survey progresses (Baird, Hamory et al. 2008). Especially in consumption surveys, a long list of items can lead to enumerators cutting corners and fabricating data (Finn and Ranchhod 2015, Fiedler and Mwangi 2016) as well as prematurely ending interviews (A. Deaton and Grosh 2000). Respondents also become fatigued and, for example, learn to say no to consumption of items to evade more detailed follow-up questions (Kreuter, McCullock et al. 2011, Eckman, Kreuter et al. 2014).

To overcome the challenges inherent to measuring consumption poverty, we propose a new methodology that combines an innovative questionnaire design with standard imputation techniques.[16] This new methodology allows us to substantially shorten the consumption questionnaire and reduce the interview time (less than 60 minutes for a standard questionnaire) by imputing deliberately absent consumption values for those items that are not explicitly asked about. Poverty estimates can be derived in this way without compromising the credibility of the resulting estimate. This new methodology is particularly useful in fragile states given the significant risks associated with lengthy interviews. It can also be useful to reduce enumerator and respondent fatigue, or to mitigate the problem of high non-response rates.

---

[15] Corresponding author: Utz Pape (upape@worldbank.org). The findings, interpretations and conclusions expressed in this paper are entirely those of the author, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. The author would like to thank Johan Mistiaen for discussions and support in pursuing the idea and Chris Elbers for help with the statistical properties of the new methodology, as well as Kathleen Beegle, Tomoki Fujii, Kristen Himelein, Dean Jolliffe, Peter Lanjouw, Emmanuel Skoufias, Shinya Takamatsu, Roy Van der Weide and Nobuo Yoshida for discussions. The author is also grateful to the Kenyan National Bureau of Statistics for implementing the methodology in a pilot survey.

[16] A precursor methodology based on the same principle was previously published in Pape, U., and J Mistiaen (2015). Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu. Washington DC: World Bank. Proceedings of ABCA. Washington DC: World Bank.

The most straightforward way to reduce the expected interview time is to skip rarely-consumed items. Another simple strategy is to ask the respondent about an aggregate amount of spending on an entire category of consumption (e.g., total expenditure on flour) instead of individual items (e.g., expenditure on corn flour, wheat flour, etc.). However, altering the set of items in the questionnaire can result in a nontrivial change in the reported consumption amount (Olson-Lanjouw and Lanjouw 2001). Both approaches are likely to lead to an underestimation of consumption and overestimation of poverty, as was demonstrated in a study in Tanzania that directly compared various methods of measuring consumption (Beegle, De Weerdt et al. 2012).

An alternative approach is to apply methods of cross-survey imputation. In situations where full household expenditure surveys are too costly or impractical to administer, data gaps can be filled using other surveys that have common covariates that are correlated with household expenditure. For example, data from a full consumption survey can be combined with data from shorter and more frequent labor force surveys to generate poverty estimates (Douidich, Ezzrari et al. 2013). While such methods may work well even when there is a rapid economic change (Christiaensen, Lanjouw et al. 2011), the assumption of a stable structural parameter typically cannot be tested and may not be valid, especially in the context of large and systemic shocks, after implementation of projects, or if a substantial amount of time has passed since the baseline survey was implemented. It is also possible to design a survey such that one sample has a full consumption module and another sample has only the covariates of consumption. Consumption can thus be imputed and poverty estimates can be derived at a reduced cost, even though the magnitude of potential cost reduction may be modest (Fujii and van der Weide 2016). In such a setup, however, the sample for the full consumption module must be chosen randomly to avoid biased estimates of the model parameters. Thus, this approach is only of limited usability in the case of fragile countries as it might not be feasible to administer the full consumption module in particular insecure areas, creating a downward bias in poverty estimates for those areas.

The proposed methodology uses statistical imputations to obtain estimates for deliberately absent consumption values. Statistical imputation techniques are widely used to replace missing values in surveys (Ambler, Omar et al. 2007, Van Buuren 2007, Little and Rubin 2019). Straight-forward methods simply replace the missing values with aggregate statistics like a mean. However, this makes the strong and often violated assumption that data are missing at random (Carpenter, Kenward et al. 2007). Model-based approaches can take into account covariates and often use a regression framework to estimate missing values but distort the variance if based on point-estimates. Multiple imputations help to mitigate this by drawing multiple estimates from the posterior distribution using a Bayesian approach (Rubin 2004).

This paper is organized as follows. We first present the proposed methodology with its statistical properties as well as the data in Section 2. In Section 3, we apply the methodology to different scenarios showing the trade-offs between the performance and parameters of the approach, and then compare it to a reduced consumption approach as well as a more sophisticated reduced consumption approach adjusting the poverty line. The section ends with a real-world example based on a pilot survey in Kenya, assessing the performance of the new methodology and comparing it to a cross-survey imputation approach. The paper finishes with Section 4 concluding the findings and discussing some of the limitations of the new approach.

Methodology

Overview

The rapid approach being proposed here applies a split-questionnaire design to the consumption module of a household survey, thereby generating systematically absent data that can be conveniently imputed. While the split-questionnaire design is more popular in other disciplines such as psychology (Graham, Hofer et al. 1996), the approach has not yet been applied to large-scale household-based surveys, nor with the goal of reducing the time required to estimate consumption or poverty. Instead of having all households report on all consumption items, important items are assigned to a core module and the remaining items are split into two or more optional modules.[17] Each household then answers the questions in the core module and in only one of the optional modules. This approach reduces average interview time considerably, down to 45 to 60 minutes per household for a standard household consumption survey. The cost of this efficiency gain is that data are deliberately absent for those optional modules that were not administered to certain households. We can however offset this cost by estimating the deliberately absent data for each household based on the data collected from other households for that module. While this approach utilizes a structural model for the imputation of the deliberately absent data, the model is estimated within the survey rather than between two surveys, thereby circumventing the problem of biased structural parameters due to having different sample populations or considering the same population at different points in time.

The rapid approach starts by defining the number of core items and the number of optional modules for the non-core consumption items. The smaller the number of core items and the greater the number of optional modules used, the faster the questionnaire can be administered, as fewer items need to be asked for each household. However, having fewer core items and more optional modules also increases the uncertainty in the estimation as less consumption information is available. Thus, the choice of these two parameters can be informed by simulations on a previous or similar survey to gauge the performance of the estimation vis-à-vis the time savings in administering the questionnaire. Another consideration is that it is beneficial to balance the number of households for each optional module, ideally at the cluster level of the survey.

The next step is to select core consumption items. Although consumption in any given country will exhibit some variability, data on a few dozen key items will usually be sufficient to capture the majority of consumption. Important consumption items can be identified using average consumption share per household or across households, as estimated by previous consumption surveys in the same context or recorded consumption shares in neighboring and/or similar countries. While a good choice of core items will improve the performance of the estimation, the methodology still works if no core items are used, e.g. in a context without any prior information. The identified key items are then assigned to the core module that will be administered to all households.

Finally, non-core items are randomly partitioned into optional modules. It is important to note that the conceptual distinction between core and optional items should not be reflected in the layout of the questionnaire. Instead, all items per household need to be grouped into categories of consumption items(e.g. meat, fruits, vegetables, cereals) and different recall periods. It is therefore recommended to use CAPI (Computer-Assisted Personal Interviewing) technology, which makes it possible to hide the modular structure of the consumption questions within the layout of the questionnaire.

Once the core and optional modules have been defined and the design has been finalized, the survey can be implemented. The assignment of optional modules to households is performed randomly and

[17] As is shown below, the core module is not strictly necessary, further reducing the interview time.

is stratified by enumeration area, thus ensuring an appropriate representation of all optional modules in each enumeration area. Once the data have been collected and cleaned, household consumption is estimated by imputation. The average consumption of each optional module can be estimated based on the sub-sample of households assigned to that optional module.

Theoretical Properties

Consumption for a household *i* is the sum of the consumption for each item in each module

$$y_i = \sum_k y_{ik} = \sum_k \sum_j y_{ikj}$$

where y_{ikj} denotes the consumption of item *j* in module *k*.[18] Applying the rapid approach, we only observe a subset of modules y_{ik}, specifically for each household k=0 and one other module where k>0. We can formalize this by using a binary (0,1) variable b_k, which is independent of y_{ik}, where $P(b_k=1) = \pi_k$. In practice, the assignment of optional modules can be done more systematically to ensure a balanced design at the cluster-level, which does not invalidate the assumed independence of b_k from consumption. The expected consumption of a household is:

$$Ey_i = E\sum_k y_{ik}\frac{b_k}{\pi_k} = \sum_k Ey_{ik}$$

We obtain a consistent and unbiased estimator for expected consumption if we can find consistent and unbiased estimators for expected module consumption. This also holds for regressions assuming b_k and household characteristics x_i are independent:

$$E(y_i|x_i) = E(\sum_k y_{ik}\frac{b_k}{\pi_k}|x_i) = \sum_k E(y_k|x_i)$$

Furthermore, the second moment can be estimated as follows:

$$E{y_i}^2 = E(\sum_k y_{ik})^2 = E\sum_k y_{ik}^2 + 2E\sum_{k\neq l} y_{ik}y_{il} = \sum_k \frac{b_k}{\pi_k}Ey_{ik}^2 + 2\sum_{k\neq l}\frac{b_k}{\pi_k}\frac{b_l}{\pi_l}Ey_{ik}y_{il}$$

Similarly, higher moments can be constructed. Thus, the complete distributional information of *y* can theoretically be recovered from sufficiently large samples if the design of the split questionnaire allows for the estimation of correlations between modules.

Consumption Estimator

Distinguishing between administered core module $k = 0$, the administered optional module k_i^* and the non-administered remaining optional modules $0 < k \neq k_i^*$, we obtain as estimator for consumption

$\hat{y}_i = y_{i0} + y_{ik_i^*} + \sum_{k\neq k_i^*}\hat{y}_{ik}$.

As shown above, the estimator is unbiased for Ey_i as

$$Ey_i = \sum_k Ey_{ik} = Ey_{i0} + Ey_{ik_i^*} + \sum_{k\neq k_i^*} E\hat{y}_{ik} = E\hat{y}_i$$

The variance of consumption can be decomposed as

[18] Note that we assume consumption to be per-capita throughout the paper.

$$Var(y_i) = Var(y_{i0}) + \sum_k Var(y_{ik}) + 2\sum_k Cov(y_{i0}, y_{ik}) + \sum_{k \neq l} Cov(y_{ik}, y_{il})$$

$$\geq Var(y_{i0}) + \sum_k Var(\hat{y}_{ik}) + 2\sum_k Cov(y_{i0}, \hat{y}_{ik}) = Var(\hat{y}_i)$$

with the inequality given by the assumption of positive correlation between optional modules.[19] The variance is thus underestimated, as we cannot measure correlation between modules and so assume them to be independent $Cov(y_{ik}, y_{il}) = 0$ for all optional modules *k* and *l*. The more optional modules are used, the higher the under-estimation of the variance. Contrarily, the larger the fraction of the variance captured in the core module, the lower the underestimation of the variance. This suggests using a low number of optional modules with a large number of items in the core module. This represents the fundamental trade-off between the accuracy of the estimator and time savings, which are higher with more optional modules and fewer items in the core module.

We apply the Foster–Greer–Thorbecke measures of poverty (Foster, Greer et al. 1984) to the consumption aggregate defined as

$$FGT_{\alpha,z} = N^{-1} \sum_{i:y_i<z} \left(\frac{z - y_i}{z}\right)^{\alpha}$$

where N denotes the number of households, z is the poverty line, and y_i is consumption for a given household. By selecting the coefficient α we can produce different poverty measures: $\alpha = 0$ for the poverty headcount, $\alpha = 1$ for poverty depth and $\alpha = 2$ for poverty severity. Given that we are underestimating the variance of y_i, this implies that the estimator $\widehat{FGT}_{\alpha,z}$ will be underestimated for a poverty line *z* smaller than the mode of y and overestimated for larger poverty lines.

Estimation

The optional module consumption can be estimated in the log-space conditional on strictly positive consumption:

$$\log y_{ik} = \beta X_i + \varepsilon_{ik} \mid y_{ik} > 0$$

where X_i denotes a vector of household characteristics and ε_{ik} the error term. This is implemented as a two-step estimation procedure with the first step utilizing a logit regression to estimate whether $y_{ik} = 0$ and the second step using an OLS regression. We use the framework of multiple imputations to obtain several point estimates by drawing from the error distribution to ensure accurate tails of the consumption distribution.

The household characteristics X_i are selected based on a step-forward algorithm minimizing the AIC by regressing household characteristics on the observed core and assigned non-core consumption including a fixed effect for the assigned module:

$$\log\left(y_{i0} + y_{ik_i^*}\right) = \beta X_i + D_{ik} + \varepsilon_{ik} \mid y_{i0} + y_{ik_i^*} > 0$$

where D_{ik} represents *k* dummy variables with the *k*th variable equal to 1 if household *i* is assigned to module *k* and equal to 0 otherwise.

[19] Even though the consumption aggregate consists of complements and substitutes, a random allocation of items into optional modules will tend to make the correlation between modules positive except in the unlikely case of two modules sharing a large number of complements with one module capturing, for example, all the items typically consumed by the poor. Thus, the optional modules can be assumed to be positively correlated.

Performance Assessment

We assess the performance of the estimation based on the bias and the coefficient of variation (CV). The bias is defined as the expected value of the absolute percentage difference of $\widehat{FGT}_{\alpha,z}$ estimated using the rapid approach and $FGT_{\alpha,z}$ estimated based on full consumption data. Using the additional index $1 \leq s \leq 20$ for the simulation, we obtain

$$E_z E_s \left| \widehat{FGT}_{\alpha,z,s} - FGT_{\alpha,z,s} \right|$$

Each simulation uses random allocations of items to optional modules and random assignments of households to optional modules.[20] We average the bias over all possible poverty lines *z* so that 1 percent, 2 percent, et cetera, and 99 percent of the population are defined as poor based on the full consumption distribution. The integration over all possible poverty lines makes the resulting performance measures independent of the poverty line, while the absolute difference in the definition of the bias avoids canceling out errors.

Accordingly, the coefficient of variation is defined as the average ratio of the standard deviation and the mean of the FGT measure over all possible poverty lines:

$$E_z \frac{\sqrt{E_s\left(\widehat{FGT}_{\alpha,z,s} - FGT_{\alpha,z,s}\right)^2}}{E_s FGT_{\alpha,z,s}}.$$

Data

We applied this method to recent household consumption data from Kenya. Kenya's source for official poverty estimates is the Kenyan Integrated Household Budget Survey (KIHBS). The two last rounds were implemented in 2005/6 and 2015/16. The 2005/6 round used a representative sample of households in Kenya stratified by 7 provinces and 69 districts split into urban and rural. The sample size of the cleaned data set includes 12,695 households in 1,338 clusters. The 2015/16 round used a representative sample of households in Kenya stratified by 47 counties split into urban and rural. The sample size of the cleaned data set includes 21,585 households in 2,387 clusters. For both surveys, data collection was carried out over a period of 12 months.

The 2015/16 round was accompanied by a CAPI pilot implementing the rapid approach. The pilot used the same sampling frame as the 2015/16 round and interviewed up to an additional 6 households in the same clusters, resulting in a sample size of 12,662 households (as not all clusters and all intended households were interviewed due to non-response).[21] The questionnaire was derived from the KIHBS 2015/16 questionnaire but was considerably simplified across all modules. Specifically, the consumption module was administered according to the rapid approach. We are using the data set that was constructed with 5 food and 5 non-food items in the core (selected based on the highest consumption share in KIHBS 2005/6), with the remaining 128 food and 76 non-food items partitioned into 3 optional modules.[22] Thus, the expected time saving was about 30 percent.

In the remainder of the paper, we use KIHBS 2015/16 with the full consumption module as a benchmark. The previous round of 2005/6 is used to define the core items, and for two of the alternative approaches it is used to adjust the poverty line and to build the consumption model for the cross-survey imputation. The 2015/16 pilot is used as a real-world example for the implementation

---

[20] The constraints in these allocations are to ensure that items are uniformly distributed among optional modules and that each optional module is assigned equally often to households within each cluster.

[21] Balance mean tests are indicating similar households in both surveys with similar although not always statistically indistinguishable characteristics.

[22] The assignment of optional modules to households was balanced with 4,222 households assigned to module 1, 4,192 to optional module 2, and 4,248 to optional module 3.

of the rapid approach. We harmonize the data sets to ensure comparability across the different surveys. The harmonized data sets include 133 food and 81 non-food items. The consumption aggregate is exclusively based on these 214 items. In addition, we use harmonized location and household characteristics for the various models, including a binary and a categorical location variable as well as 6 additional binary, 9 additional categorical and 9 continuous variables for household characteristics (Table B1-6 in the Appendix).

Consumption shares differ markedly between the 2005/6 and 2015/16 surveys, which is unsurprising given the 10-year gap (Kenya National Bureau of Statistics 2018). For example, the top 10 food items in 2005/6 capture 58 percent of the food consumption share, but only 39 percent of the consumption share in 2015/16. Non-food consumption changed to a lesser degree. The top 10 items in 2005/6 represent 64 percent of non-food consumption shrinking to 59 percent in 2015/16. These differences will impact the performance of those approaches that strongly rely on data from previous surveys, e.g. adjusting poverty lines and cross-survey imputations.

Results

We assess the performance of the rapid approach vis-à-vis alternative approaches. The long questionnaire of the full-consumption approach can increase unit and item non-response but is nevertheless used as benchmark due to its *de facto* standard for consumption surveys. The rapid approach compromises on the long list of items by introducing a subset of core items and distributing remaining items in optional modules, reducing the questionnaire length with beneficial impacts on unit and item non-response but at the cost of additional estimation error. The reduced approach further decreases questionnaire length by simply dropping items altogether, creating substantial bias in the resulting estimates. The bias can be minimized by adjusting the poverty line based on data from a previous consumption survey, called the adjusted reduced approach below. Finally, the largest time savings are generated by completely abandoning consumption data and using a structural model to impute consumption from a baseline survey using common co-variates. The time gap with the baseline survey as well as shocks and other structural changes, for example from the implementation of a project to reduce poverty, invalidate the structural model leading to substantial bias in the estimates.

Table B1-1: Comparison of consumption methodologies and sources of error

	Unit non-response	Item non-response	Implementation issues	Error inherent to method
Full consumption	Elevated levels of non-response, particularly among urban and wealthy households (Korinek, Mistiaen et al. 2006, Osier 2016)	Long list of items increase item non-response and measurement error (Finn and Ranchhod 2015, Fiedler and Mwangi 2016) as well as unfinished interviews (A. Deaton and Grosh 2000)	Relatively limited issue as questionnaire is straightforward though length may be issue	Theoretically unbiased if implemented correctly with full response
-Rapid approach	Non-response is an issue but not as much as in full consumption due to the shorter questionnaire	Less of an issue than full consumption due to shorter questionnaire	Could be substantial issue with paper questionnaires but almost completely mitigated with CAPI	Trade-off between length of questionnaire and accuracy of poverty and inequality estimates due to underestimation of variance attenuated by less core items and more optional modules.
Reduced consumption	Non-response is an issue but not as much as in full consumption due to the shorter questionnaire	Less of an issue than full consumption due to shorter questionnaire	Limited issue as questionnaire is straightforward	Substantial bias in total consumption attenuating inequality.
Cross-Survey (X-Survey) imputations	Non-response is an issue but not as much as in full consumption due to the shorter questionnaire	Very small issue as no consumption section	Very small issue as questions are simple	Reliance on old data introducing bias in structural model leading to biased poverty estimates.

Using the full consumption data from KIHBS 2015/16 as benchmark, in this section, first, we investigate the empirical trade-off between the number of core items and the number of optional modules for the rapid approach with respect to the performance of poverty indicators. Second, we compare the rapid approach with a traditional reduced consumption aggregate, without any adjustment of the poverty line. Third, we again use a reduced consumption aggregate, but adjust the poverty line based on previously observed consumption from 2005/6. Fourth, we compare the application of the rapid approach in the pilot in Kenya with a cross-survey imputation approach also using the 2005/6 data as baseline.

Trade-off with Number of Core Items and Optional Modules

As discussed in the methodology section, the rapid approach creates a trade-off between the number of core items and the number of optional modules, as a larger number of core items (smaller number of optional modules) will improve the performance of the estimation. A larger number of core items will capture a larger fraction of the variance in the core module, minimizing the estimation error for the variance. Similarly, a smaller number of optional modules reduces the estimation error of the covariance between modules. However, the time savings by the rapid approach are larger for a smaller number of items in the core module and a larger number of optional modules as fewer items are asked about for each household.

Table B1-2: FGT measures by number of modules and core items measured by bias and cv.

	Opt. Mod.	Bias: Core Items 0	1	3	5	10	20	CV: Core Items 0	1	3	5	10	20
fgt0	2	0.021	0.019	0.015	0.012	0.009	0.006	0.081	0.075	0.051	0.043	0.029	0.017
	4	0.033	0.032	0.024	0.019	0.014	0.007	0.125	0.122	0.085	0.07	0.051	0.026
	6	0.039	0.036	0.027	0.021	0.015	0.01	0.139	0.14	0.099	0.081	0.06	0.037
	8	0.042	0.042	0.031	0.026	0.021	0.009	0.151	0.155	0.105	0.096	0.08	0.035
fgt1	2	0.015	0.014	0.01	0.008	0.005	0.003	0.128	0.12	0.081	0.071	0.051	0.028
	4	0.024	0.023	0.016	0.013	0.009	0.005	0.19	0.189	0.129	0.114	0.087	0.045
	6	0.027	0.026	0.019	0.015	0.011	0.007	0.207	0.214	0.15	0.131	0.103	0.062
	8	0.029	0.029	0.02	0.018	0.015	0.006	0.224	0.232	0.156	0.151	0.132	0.063
fgt2	2	0.011	0.01	0.007	0.006	0.004	0.002	0.166	0.157	0.108	0.098	0.075	0.04
	4	0.017	0.017	0.012	0.01	0.007	0.004	0.241	0.241	0.166	0.154	0.123	0.064
	6	0.02	0.019	0.014	0.011	0.008	0.005	0.259	0.272	0.192	0.174	0.144	0.085
	8	0.021	0.022	0.015	0.013	0.011	0.005	0.279	0.29	0.198	0.198	0.179	0.09
gini	2	0.005	0.001	0.001	0	0.003	0.002	0.014	0.004	0.004	0.004	0.009	0.005
	4	0.004	0.001	0.003	0.001	0	0.003	0.013	0.009	0.01	0.004	0.004	0.01
	6	0.009	0.005	0.005	0.008	0.006	0.008	0.025	0.014	0.014	0.021	0.015	0.021
	8	0.001	0.008	0.011	0.003	0.006	0.007	0.011	0.022	0.03	0.01	0.015	0.018

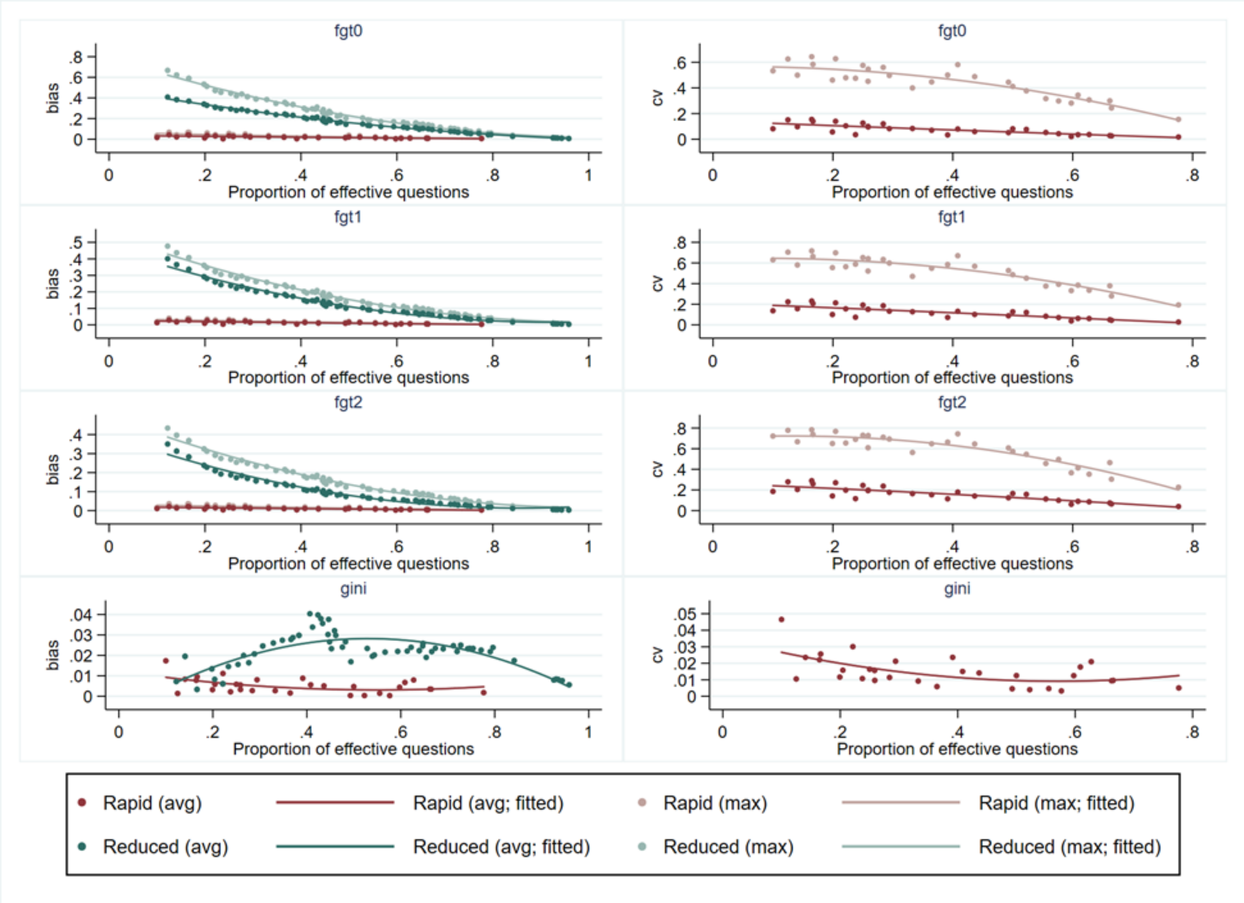
Based on simulations using KIHBS 2015/16, we estimate the bias and coefficient of variation to estimate FGT0, FGT1 and FGT2 as well as the Gini using as reference the full consumption aggregate from the survey.[23] As expected, all performance measures deteriorate for a smaller number of core items, as well as for a larger number of optional modules (Table B1-2). Using the minimum number of optional modules (2), we obtain an average bias of 0.021 for FGT0 using no core items, but a considerably smaller bias of only 0.006 (a more than 70 percent reduction) if using 20 core items. An increase of the number of optional modules from 2 to 8 almost doubles the bias for FGT0 from 0.021 to 0.042 using 0 core items. We observe similar trade-offs for FGT1 and FGT2. The trade-offs for the Gini coefficient are less clear as independent of the number of core items and modules the bias is consistently extremely low, almost always below 0.01.

Rapid vs. Reduced Estimation

Traditionally, time savings in administering consumption modules are achieved by reducing the number of consumption items included in the questionnaire. Here we compare the rapid approach with a reduced approach based on the effective time savings achieved. Assuming that only consumed items require substantial interview time, we estimate the average number of items that were administered and consumed by households relative to the total number of items in the full consumption module. The smaller the number of items consumed and administered in the questionnaire, the larger the time savings. This measure takes into account that effective time savings are smaller if fewer but often-consumed items are administered to a household, compared to a larger number of items, which are rarely consumed. The results show simulations with varying numbers of core items and optional modules for the rapid approach, and varying numbers of items included in the reduced module.

[23] If not noted otherwise, each core and optional module configuration is run 20 times, each using 50 multiple imputations. Note that the same survey data are used to define the core module items (based on highest consumption shares).

Figure B1-1: Absolute bias and coefficient of variation (cv) of rapid vs. reduced poverty estimation.



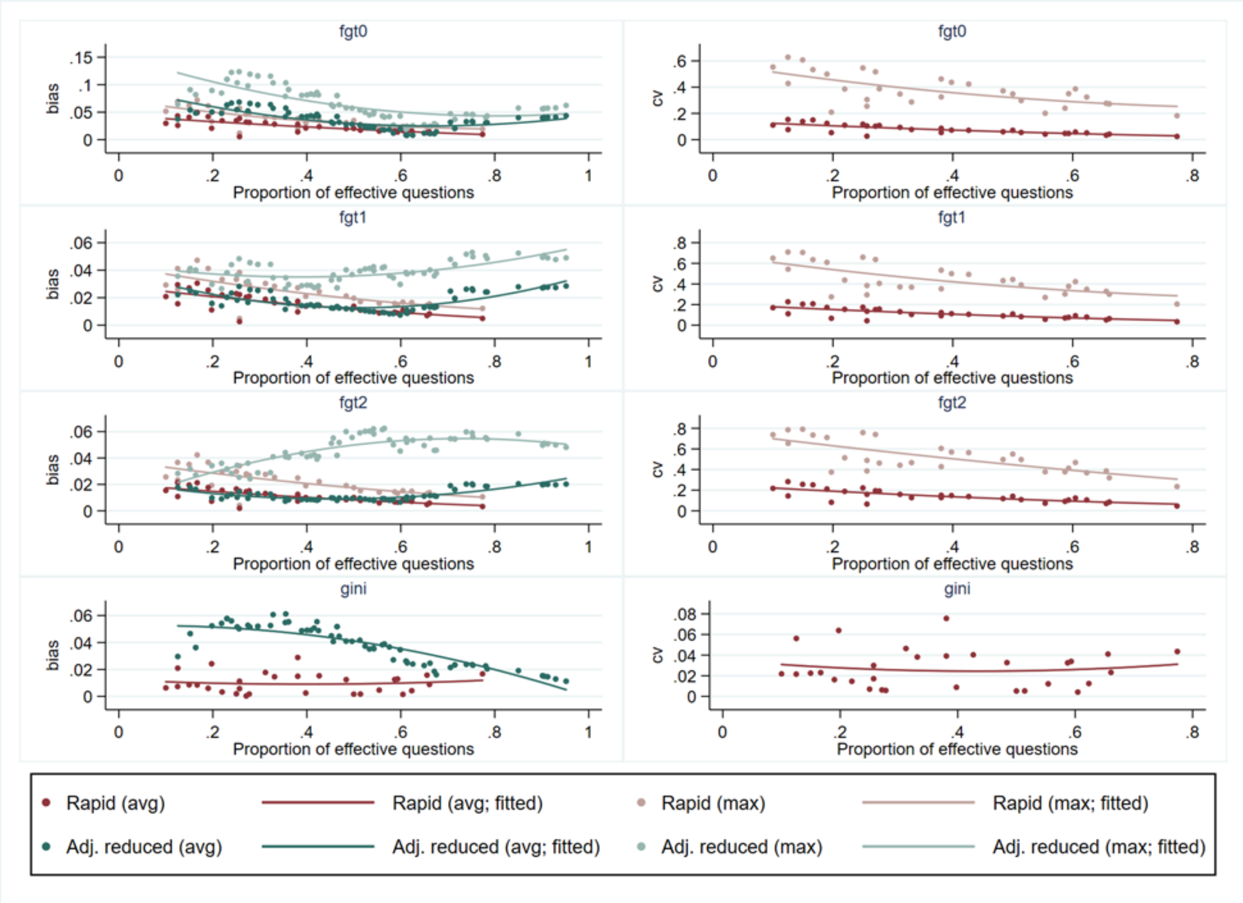
Based on KIHBS 2015/16 using the full consumption aggregate as the reference, we compare the average and maximum bias across all poverty lines for FGT0, FGT1 and FGT2 as well as the Gini (Figure B1-1). The results clearly show the superiority of the rapid approach for any time saving larger than 10 percent. Generally, the bias and coefficient of variation are increasing for larger time savings, except for the Gini which is generally low with a bias of usually less than 0.01 for the rapid approach while the reduced approach over-estimates the Gini by up to 0.04. The maximum bias for FGT0 remains below 7 percent in 95 percent of the simulations. The average bias generally remains below 5.3 percent. The average coefficient of variation slightly increases for larger time savings, but generally remains below 20 percent. Note that the coefficient of variation is only meaningful for the rapid approach, as the reduced approach is deterministic across simulations.

Time savings of 50 percent can be achieved with the rapid approach by accepting an average bias of 1.8 percent and a maximum bias of 3 percent for FGT0 for 0 items in the core and 2 optional modules. For similar time savings, the reduced approach would need to consist of the 20 percent of items with largest consumption, which will be consumed by most households, but resulting in an average and maximum bias of 15 percent and 21.9 percent, respectively. A 75 percent time saving is possible with 0 core items and 4 optional modules. It comes at the cost of an average and maximum bias of 3.3 percent and 4.6 percent respectively.

Rapid vs. Adjusted Reduced Estimation

The reduced approach from the previous section can be improved by adjusting the poverty line based on a previous survey (Olson-Lanjouw and Lanjouw 2001). To simulate this case, we use the KIHBS 2005/6 survey to re-estimate the poverty line for the reduced approach and use the survey for the definition of core items for the rapid approach (optional module items are randomly assigned). The re-estimated poverty line for the reduced approach and the core module assignment is then applied to the KIHBS 2015/16 survey (Figure B1-2). As before, we only show the coefficient of variation for the rapid approach, as the adjusted poverty line approach is deterministic for each simulation.

Figure B1-2: Absolute bias and coefficient of variation (cv) of rapid vs. OLL poverty estimation.



We observe a very similar performance for the rapid approach as the performance does not change for any number of optional modules with zero core items. However, the best performance is now achieved with some core items, as it helps to isolate more of the variance from the estimation. For example, a time saving of 75 percent can now be achieved with a core module of 5 items and 10 optional modules, resulting in an average and maximum bias of only 0.5 percent and 1.2 percent respectively for FGT0. This is a considerable reduction in bias as compared to using 0 core items and 4 optional modules (reported in the previous subsection), although in both cases the time savings are the same. Thus, it is useful to include a few core items with highest consumption shares, even if they are selected from a rather outdated survey as in Kenya with a 10-year gap.

The approach of using an adjusted poverty line performs significantly better than the simple reduced approach presented in the previous section, but at the cost of the distributional shape captured in the Gini coefficient. For the FGT measures, it is, thus, generally advisable to adjust the poverty line if a reduced approach must be used, even if the adjustment is based on outdated consumption shares from an old survey. For FGT0, the average bias of the adjusted reduced approach is usually around or above the maximum bias of the rapid approach. Furthermore, the maximum bias is considerably larger than the average bias for the adjusted, reduced approach compared to the rapid approach. For FGT1, the average bias of the adjusted, reduced approach becomes more comparable with the rapid approach, but the maximum bias is still significantly larger than the maximum bias of the rapid approach. For FGT2, the performance of the adjusted, reduced approach becomes more difficult to interpret. The Gini coefficient is not well conserved approaching a bias of 0.05 for highest time savings.

As the results show, the adjusted, reduced approach has larger variation across poverty lines (as shown by the larger difference between average and maximum bias) as well as across the different number of items (implying different time savings). The definition of the estimator explains this. The estimator depends strongly on the items selected for the reduced approach, as the accuracy relies on the adjustment factor of the poverty line, which is the share captured by those items. The error can be decomposed into two components. The first component is the change of the distributional shape between the survey used for the adjustment and the application of the adjusted poverty line. The second component depends on the difference in the share captured by those items between the two surveys. Both errors become zero if the approach is implemented for the same population at the same time. In the usual case though, neither the population nor the time point is the same. In these cases, especially the second component of the error leads to large variation of the performance. Even though the consumption shares of the items can change, the changes might cancel out leading to a good performance of the approach. However, adding one more item to the consumption module can void the cancellation, leading to a worse performance. Without knowing the share of the items from total consumption (which is not measured), it is impossible to predict how many items should be selected for a good performance of the approach.

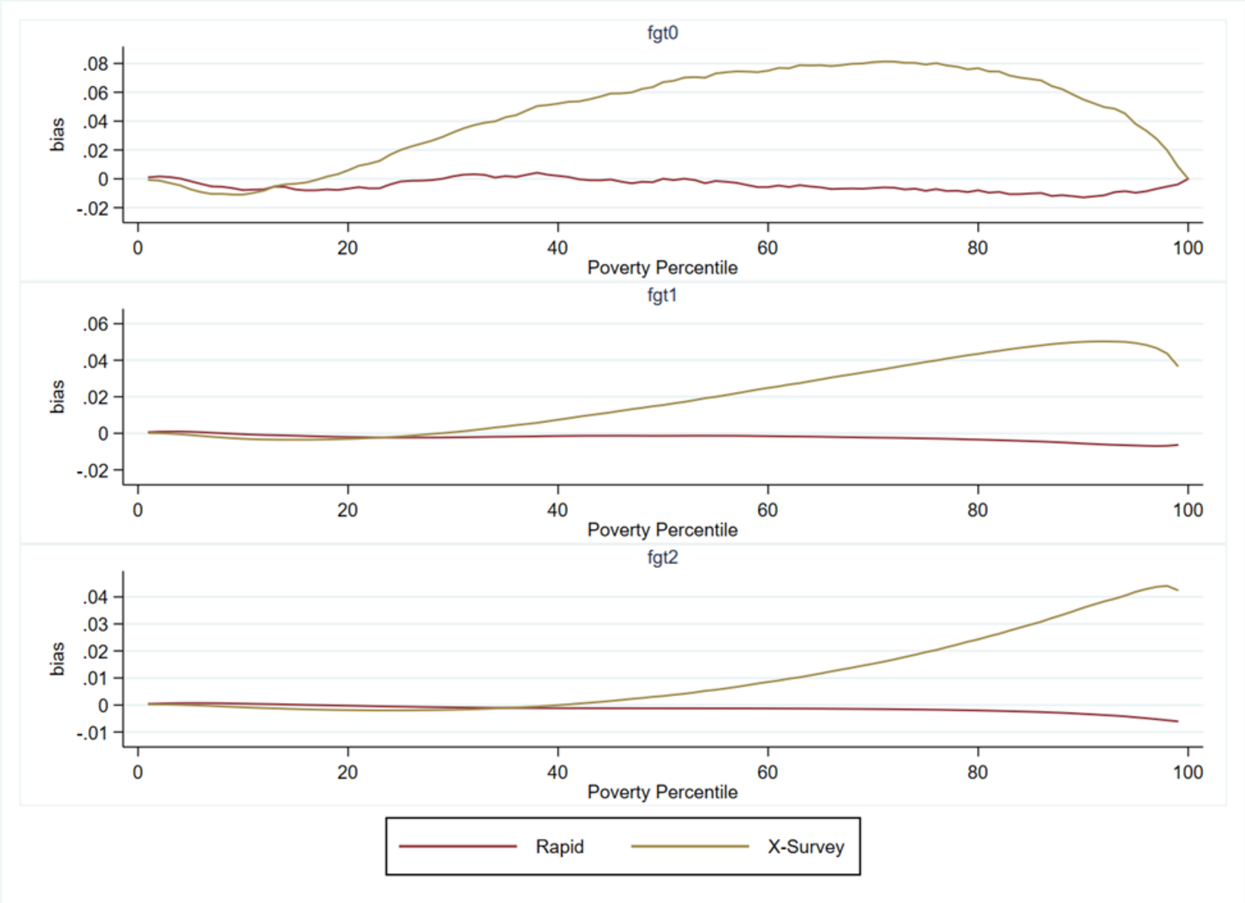
Application to Kenya: Rapid vs. Cross-Survey Estimation

In 2015/16, a CAPI pilot was implemented alongside KIHBS 2015/16 using the rapid approach. While the configuration is conservative with only 30 percent time savings, the results show impressive performance for all FGT measures (Figure B1-3) in comparison with the full consumption as estimated for KIHBS 2015/16. Across all potential poverty lines, the rapid approach has a bias of below 1.3 percentage points for FGT0, 0.7 percentage point for FGT1 and 0.6 percentage point for FGT2. The Gini has a bias of only 0.012.

We compare the performance with a cross-survey imputation of consumption using a structural model built based on the KIHBS 2005/6 data set and applied to KIHBS 2015/16, ignoring the collected consumption data in 2015/16. The performance of the structural model is then assessed against the KIHBS 2015/16 full consumption aggregate (Figure B1-3).[24] The cross-survey imputation cannot provide convincing results. FGT0 is under-estimated by up to 8.1 percentage points, FGT1 by up to 5.0 percentage points, and FGT2 by up to 4.4 percentage points. The Gini is off by 0.036. This is not surprising given the 10-year gap between the parameters of the structural model from 2005/6 and inference of poverty for 2015/16. In such long timeframes, not only do consumption patterns change but also structural drivers or correlates with poverty. To the further detriment of cross-survey imputations, it is in practice not possible to estimate the error, making it difficult to recommend its usage.

[24] The cross-survey imputation is based on the best model minimizing the AIC using a step-forward algorithm on the variables from KIHBS 2005/6. The results of the model selection are provided in the Appendix. The imputation is performed in log-space with 50 multiple imputations.

Figure B1-3: Bias of rapid vs. cross-survey (X-Survey) poverty estimation, by poverty percentile.[25]



Conclusions

The rapid approach proposed in this paper can be used to achieve significant time savings, while only introducing a small bias into poverty estimates. The choice of the number of items in the core module and the number of optional modules allows for a precise calibration of time savings. The results show that it is helpful to utilize a previous survey from the same or a similar population to assign a few key items to the core module. In the best case, the selected items are still highly consumed and will improve the estimates of poverty. In the worst case, the selected items are no longer important in which case they will hardly affect the time savings compared to a design with zero core items and also equal its performance. In countries with large variation in diet across regions, subnational core modules can potentially improve estimates even further.

This paper also demonstrates the difficulty of achieving convincing results using alternative methods. Simply removing items from the consumption aggregate to create time savings, but without adjusting the poverty line, can lead to considerable bias in the poverty estimates. Better results can be achieved by adjusting the poverty line based on a previous survey in order to accommodate the reduced number of items, but without conserving the shape of the consumption distribution measured by the Gini. Furthermore, the potential for large changes in consumption patterns, which cannot be determined under this approach, creates considerable uncertainty in the resulting estimates. Similarly, a survey based only on covariates and their structural relationship with poverty estimates from a previous survey introduces a large bias in the estimates, at least in the studied case with a 10-

[25] Note that this figure shows the bias while previous figures showed the absolute bias to avoid canceling negative and positive bias across percentiles.

year gap between surveys. The proposed rapid approach outperforms all these approaches, only at the cost of increased complexity.

The rapid approach introduces additional complexity into both the questionnaire design and the estimation of poverty. The capacity of enumerators is often low in developing countries. While the rapid approach increases the complexity of the questionnaire, CAPI technology easily solves this problem. Survey software can automatically compile a single consumption module based on the core and optional modules for each household, without making the partition explicit to the enumerator or demanding the execution of complex skip patterns. Furthermore, advanced CAPI technology can be used to generate the questionnaire automatically based on the assignment of the household to an optional module. While enumerators should be made aware that different households will be asked for different items, administering a rapid questionnaire does not require any additional training of enumerators beyond the standard skills for consumption questionnaires.

Conversely, the analysis of data using the rapid approach requires high analytical capacity, something that is usually lacking in developing countries. While the general concept of the assignment of optional consumption modules to households can usually be explained to local partners, poverty analysis based on a bootstrapped sample of the consumption distribution can potentially be too demanding for local capacity. However, even standard poverty analysis often surpasses the limits of local capacity, especially in conflict or post-conflict settings. Therefore, capacity building tends to focus on data collection skills with the long-term perspective of creating data analysis capacity. In addition, the rapid approach might be the only possibility to create poverty estimates in certain areas. For example in the case of Somalia, the rapid approach limited overall questionnaire administering time to less than 60 minutes for more than 90 percent of households as required by security considerations for enumerators (Pape and Mistiaen 2018, Pape and Wollburg 2019).

The rapid approach administers different consumption modules to different households. In theory, this can create a response bias if households report differently on a consumed item depending on the type and number of items previously asked. Unfortunately, we cannot estimate such a response bias in the available data. However, implementation of the rapid approach with an enhanced design with different optional modules varying in their comprehensiveness of items can in general shed light on this bias. Comparison between responses for the same item in a comprehensive and a non-comprehensive list can also 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. However, it is expected that this type of response bias is very small in comparison to general measurement and estimation errors.

The main source of bias for the rapid approach is created by the assumption of zero co-variance between optional modules. Further research can help to estimate co-variance between modules within the survey and adjust the consumption estimates accordingly. Using a random assignment of items to optional modules, the co-variance between groups of items within an optional module can potentially be used to estimate the co-variance between optional modules. Or administering optional modules that share items might also be helpful to estimate the co-variance between modules.

The rapid approach reduces administering time considerably. While this creates opportunities to include additional questionnaire modules on different topics (e.g. remittances or health), it also has the potential to reduce the non-response rate (A. Deaton and Grosh 2000). The KIHBS 2015/16 survey suffered from a high non-response rate specifically in wealthier areas. For example, the capital city Nairobi had a response rate of 77 percent compared to 92 percent in the rest of the country. The highest non-response rates were specifically observed in clusters in wealthier areas within Nairobi. A high correlation of non-response with welfare status can lead to biased poverty estimates. The considerably shorter pilot survey – which not only used the rapid approach for consumption but generally shortened the questionnaire across modules and was carried out using tablets rather than paper – did not show the same pattern of lower response rates in wealthier areas. The pilot response rate of 99 percent in Nairobi was considerably higher than the standard KIHBS and was the same as in the rest of the country.[26] Thus, the rapid approach can help to contribute to shorter questionnaires mitigating concerns around low response rates, especially if correlated with welfare status. In addition, the rapid approach is likely to reduce enumerator and respondent fatigue based on the documented impact of fatigue on (consumption) estimates in the literature (Diehr, Chen et al. 2005, Snyder, Watson et al. 2007, Rolstad, Adler et al. 2011, Finn and Ranchhod 2015, Fiedler and Mwangi 2016).[27]

The rapid approach might also be particularly useful in the context of evaluating shock or project impacts on poverty. In these cases, reliance on structural models estimated between surveys is dangerous. Shocks are likely to distort structural relationships between household characteristics and poverty. For example, a light shock is often mitigated by the household reducing its consumption, rather than selling assets or moving into another dwelling. A structural model estimated before the shock will not be able to capture the reduced consumption, thereby underestimating the impact of the shock. Similarly, project impacts cannot be adequately estimated using a structural model from before the project. For example, the distribution of metal sheets as rooftop materials is unlikely to change consumption patterns, but a structural model might use the absence of metal roofs to help predict poverty. While administering a full consumption module is often not feasible, especially in the case of shocks or in fragile settings, the rapid approach can readily be applied without relying on the assumptions of a structural model that would likely be violated.

Appendix A: Performance of Estimation Techniques

Consumption of non-assigned optional modules can be estimated by different techniques. In addition to the two-step approach presented in the main text, simple summary statistics and simple regression models can be used.

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 non-administered modules. For each module *k*, a summary statistic $F\left(\{y_{jk} \mid j{:}\, k_j^* = k\}\right)$ can be computed based on households *j* to which the module *k* was administered so that consumption for household *i* can be estimated as

$$\hat{y}_{ik} = F\left(\{y_{jk} \mid j{:}\, k_j^* = k\}\right).$$

Using this approach, each household is assigned the same consumption per non-administered module. The summary statistics *F* can be, for example, a simple average or the median. 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.

Using this approach, each household is assigned the same consumption per non-administered module. The summary statistics *F* can be, for example, a simple average or the median. The median has the

[26] The change in the response rate is unlikely to be explained by the transition from PAPI to CAPI Banks, R. and H. Laurie (2000). "From Papi to Capi: The Case of the British Household Panel Survey." Social Science Computer Review **18**(4): 397 - 406, Schräpler, J.-P., Schupp, Jürgen , and Gert G. Wagner (2010). "Changing from PAPI to CAPI: Introducing CAPI in a Longitudinal Study." Journal of Official Statistics **26**(2): 239 - 269.

[27] Estimating the reduction of fatigue on measurement error would require a specifically designed survey.

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.

Regression (OLS and Tobit regression)

Module-wise estimation applies a regression model for each module and exploits the differences in observed household characteristics

$$y_{jk} = \beta_k X_j + \varepsilon_{jk} \mid j\colon k_j^* = k$$

so that the deliberately absent consumption can be estimated as

$$\hat{y}_{ik} = \hat{\beta}_k X_i$$

with $\hat{\beta}$ representing the estimated OLS coefficient. Given the impossibility of negative consumption, a Tobit regression with a lower bound of 0 is used instead of a standard OLS regression approach. For the OLS regression, negative imputed values are set to zero.

Multiple Imputation

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. Thus, the regression can also be embedded in a multiple imputation framework taking into account the variation absorbed in the residual term estimated via bootstrapping so that the resulting estimate becomes

$$\hat{y}_{ik} = \hat{\beta}_k X_i + \hat{\varepsilon}_k$$

where $\hat{\varepsilon}_k$ are repeated draws from the modeled residual distribution.

Performance Comparison

The comparison of the different estimation techniques reveals that the two-step estimation works well with highest consistency across different numbers of core items and different numbers of optional modules outperforming also the simple regression approach.

Table B1-3: Performance by number of core items and estimation technique, using 2 optional modules.

	core items	0		1		3		5		10		20	
		bias	cv	bias	cv	bias	cv	bias	cv	bias	cv	bias	cv
FGT0	avg	0.16	0.55	0.16	0.55	0.14	0.51	0.12	0.45	0.1	0.4	0.06	0.29
	med	0.1	0.38	0.1	0.37	0.08	0.32	0.06	0.28	0.05	0.21	0.02	0.11
	mi_2cel	0.02	0.08	0.02	0.08	0.01	0.05	0.01	0.04	0.01	0.03	0.01	0.02
	mi_reg	0.05	0.4	0.05	0.4	0.05	0.39	0.04	0.33	0.04	0.31	0.03	0.24
	reg	0.03	0.09	0.03	0.09	0.02	0.06	0.01	0.04	0.01	0.03	0.01	0.03
	tobit	0.02	0.1	0.02	0.08	0.02	0.07	0.02	0.06	0.01	0.03	0.01	0.02
FGT1	avg	0.1	0.72	0.1	0.72	0.09	0.68	0.08	0.63	0.07	0.57	0.05	0.45
	med	0.05	0.5	0.05	0.48	0.04	0.44	0.03	0.4	0.03	0.32	0.01	0.18
	mi_2cel	0.02	0.13	0.01	0.12	0.01	0.08	0.01	0.07	0.01	0.05	<0.01	0.03
	mi_reg	0.06	2.57	0.06	2.53	0.06	2.38	0.05	1.79	0.04	1.58	0.03	1.09
	reg	0.02	0.12	0.02	0.1	0.01	0.06	0.01	0.04	0.01	0.05	<0.01	0.04
	tobit	0.01	0.15	0.01	0.1	0.01	0.09	0.01	0.08	<0.01	0.03	<0.01	0.01
FGT2	avg	0.08	0.81	0.08	0.81	0.07	0.78	0.06	0.74	0.05	0.68	0.04	0.56
	med	0.04	0.61	0.04	0.6	0.03	0.55	0.03	0.52	0.02	0.43	0.01	0.26
	mi_2cel	0.01	0.17	0.01	0.16	0.01	0.11	0.01	0.1	<0.01	0.08	<0.01	0.04
	mi_reg	0.12	13.98	0.12	13.52	0.11	12.16	0.08	7.92	0.07	6.6	0.05	3.94
	reg	0.01	0.14	0.01	0.1	0.01	0.07	0.01	0.05	<0.01	0.07	<0.01	0.06
	tobit	0.01	0.24	0.01	0.14	0.01	0.14	<0.01	0.12	<0.01	0.04	<0.01	0.01
Gini	avg	0.19	0.5	0.19	0.5	0.17	0.43	0.14	0.36	0.12	0.3	0.08	0.2
	med	0.15	0.39	0.15	0.39	0.13	0.32	0.1	0.27	0.08	0.21	0.05	0.12
	mi_2cel	0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	0.01
	mi_reg	0.02	0.04	0.02	0.04	0.01	0.02	<0.01	0.01	<0.01	<0.01	0.01	0.02
	reg	0.03	0.08	0.03	0.08	0.02	0.06	0.02	0.04	0.01	0.03	0.01	0.02
	tobit	0.01	0.03	0.01	0.03	0.01	0.02	<0.01	0.01	<0.01	0.01	<0.01	<0.01

Table B1-4: Performance by number of optional modules and estimation technique, using 0 core items.

	opt. modules	2		4		6		8	
		bias	cv	bias	cv	bias	cv	bias	cv
FGT0	avg	0.16	0.55	0.23	0.67	0.25	0.7	0.26	0.71
	med	0.1	0.38	0.15	0.49	0.18	0.54	0.2	0.58
	mi_2cel	0.02	0.08	0.03	0.12	0.04	0.14	0.04	0.15
	mi_reg	0.05	0.4	0.07	0.48	0.07	0.48	0.07	0.5
	reg	0.03	0.09	0.04	0.11	0.05	0.16	0.05	0.17
	tobit	0.02	0.1	0.04	0.13	0.06	0.16	0.09	0.26
FGT1	avg	0.1	0.72	0.13	0.8	0.14	0.82	0.14	0.82
	med	0.05	0.5	0.06	0.52	0.06	0.52	0.07	0.53
	mi_2cel	0.02	0.13	0.02	0.19	0.03	0.21	0.03	0.22
	mi_reg	0.06	2.57	0.08	3.53	0.09	3.6	0.09	3.78
	reg	0.02	0.12	0.02	0.14	0.03	0.19	0.03	0.21
	tobit	0.01	0.15	0.02	0.2	0.03	0.18	0.05	0.32
FGT2	avg	0.08	0.81	0.09	0.88	0.09	0.89	0.1	0.89
	med	0.04	0.61	0.04	0.62	0.04	0.59	0.04	0.57
	mi_2cel	0.01	0.17	0.02	0.24	0.02	0.26	0.02	0.28
	mi_reg	0.12	13.98	0.18	22.71	0.19	23.59	0.2	25.36
	reg	0.01	0.14	0.01	0.16	0.02	0.21	0.02	0.23
	tobit	0.01	0.24	0.01	0.32	0.02	0.2	0.04	0.35
Gini	avg	0.19	0.5	0.28	0.73	0.31	0.81	0.33	0.85
	med	0.15	0.39	0.24	0.61	0.27	0.69	0.29	0.74
	mi_2cel	0.01	0.01	<0.01	0.01	0.01	0.02	<0.01	0.01
	mi_reg	0.02	0.04	0.02	0.06	0.03	0.07	0.03	0.07
	reg	0.03	0.08	0.04	0.1	0.05	0.13	0.06	0.15
	tobit	0.01	0.03	0.01	0.03	0.03	0.07	0.04	0.11

Appendix B: Additional Tables

Table B1-5: Consumption shares of the top 20 items for KIHBS 2005/6 and 2015/16.

	Food				Non-Food			
Rank	KIHBS 2005/5		KIHBS 2015/16		KIHBS 2005/5		KIHBS 2015/16	
1	Milk - fresh unpacketed	8.7%	Maize Flour - loose	8.3%	city bus / matatu fares	17.8%	household soap / bar soap	12.3%
2	Sugar + Sugar cane	7.6%	Milk - fresh unpacketed	7.8%	household soap / bar soap	14.4%	city bus / matatu fares	11.2%
3	Maize Grain - Loose	7.0%	Sugar + Sugar cane	6.0%	water	6.7%	boda boda fare	8.0%
4	Maize Flour - loose	6.6%	Beef - with bones	4.5%	batteries (dry cells)	5.9%	water	7.4%
5	Beans	5.3%	Hotel and restaurants (food + beverages)	4.0%	petroleum jelly	4.4%	hair dressing (women)	6.7%
6	Beef - with bones	4.8%	Kale + Traditional Vegetables	3.9%	detergents	3.6%	country bus fare	4.6%
7	Maize Flour - sifted	3.6%	Beans	3.7%	hair dressing (women)	3.3%	detergents	4.5%
8	Cooking Fat	3.4%	Bread (White + Brown)	3.3%	hair cut (men)	3.2%	hair cut (men)	4.3%
9	Rice Grade 2	3.1%	Maize Flour - sifted	3.0%	match box	2.9%	petroleum jelly	4.0%
10	Cakes	2.7%	Rice-Grade1-Pishori / Basmati	2.7%	country bus fare	2.1%	sanitary pads	2.4%
11	Potatoes	2.7%	Cooking oil	2.7%	fever / pain killers eg panadol	2.0%	toilet paper	2.3%
12	Kale + Traditional Vegetables	2.5%	Tomatoes	2.4%	shoe polish / cream	2.0%	Cell phone	2.1%
13	Cooking banana	2.4%	Potatoes	2.2%	toothpaste	1.8%	toothpaste	2.1%
14	Tomatoes	2.3%	Maize Grain - Loose	2.2%	sanitary pads	1.8%	batteries (dry cells)	1.9%
15	Chicken	1.8%	Wheat Flour (White + Brown)	2.1%	toilet soap	1.8%	toilet soap	1.7%
16	Tea Leaves	1.8%	milk - fresh packeted unflavoured	2.0%	medicine anti-malaria	1.8%	match box	1.7%
17	Bread (White + Brown)	1.6%	Rice Grade 2	2.0%	books	1.4%	medicine anti-malaria	1.4%
18	Mutton / Goatmeat	1.6%	Chicken	1.8%	Cell phone	1.4%	body lotion	1.4%
19	Onion / Leeks	1.3%	Mutton / Goatmeat	1.7%	mattress	1.2%	cups and saucers	1.3%
20	milk - fresh packeted unflavoured	1.3%	Banana	1.6%	cold tablets / cough syrup	1.2%	cooking sufuria	1.1%
	Total top 20 items	**72.1%**	**Total top 20 items**	**68.1%**	**Total top 20 items**	**80.5%**	**Total top 20 items**	**82.6%**

Table B1-6: Harmonized household variables

Category	**Variable**	**Type**
Location	strata	categorical
	urban	binary
Household Characteristics	owns house	binary
	wall type	categorical
	roof type	categorical
	floor type	categorical
	improved drinking water source	binary
	improved sanitation facility	binary
	access to electricity	binary
	asset index from PCA	continuous
	quartiles of asset index from PCA	categorical
	number of rooms in household	continuous
	quartiles of number of rooms	categorical
	number of persons in household	continuous
	number of children in household	continuous
	proportion of children in household	continuous
	number of adults in household	continuous
	proportion of adults in household	continuous
	number of seniors in household	continuous
	proportion of seniors in household	continuous
	dependency ratio by intervals	categorical
	at least one member is literate 15+	binary
	male household head	binary
	household head age group	categorical
	household head education level	categorical
	household head employment type	categorical

Table B1-7: Balance tests for KIHBS 2015/16 and CAPI pilot.

Characteristics	KIHBS 2015/16	CAPI Pilot	Difference	Characteristics	KIHBS 2015/16	CAPI Pilot	Difference
number of rooms in household	2.501	2.767	0.266***	Household floor type: other	0.229	0.247	0.019**
	(0.026)	(0.035)	(<0.001)		(0.007)	(0.009)	(0.031)
Owns house	0.595	0.606	0.011	Household dependency ratio: >0.2 & <0.5	0.228	0.263	0.035***
	(0.009)	(0.010)	(0.173)		(0.005)	(0.006)	(<0.001)
Improved drinking water source	0.765	0.629	-0.136***	Household dependency ratio: >=0.5 & <0.67	0.348	0.357	0.009
	(0.007)	(0.009)	(<0.001)		(0.005)	(0.007)	(0.217)
Improved sanitation facility	0.657	0.662	0.004	Household dependency ratio: >=0.67 & <=1	0.109	0.111	0.002
	(0.008)	(0.008)	(0.594)		(0.003)	(0.004)	(0.567)
HH has access to electricity	0.435	0.414	-0.022***	Male household head	0.677	0.680	0.003
	(0.009)	(0.010)	(0.002)		(0.005)	(0.006)	(0.739)
Number of children in household	1.632	1.716	0.084***	Household head age group: 30 - 44	0.392	0.402	0.010
	(0.020)	(0.025)	(<0.001)		(0.005)	(0.007)	(0.276)
Proportion of children in household	0.312	0.323	0.012***	Household head age group: 45 - 59	0.228	0.229	0.001
	(0.003)	(0.004)	(0.003)		(0.004)	(0.005)	(0.820)
Number of adults in household	2.195	2.253	0.059***	Household head age group: 60+	0.173	0.173	0.001
	(0.015)	(0.019)	(0.009)		(0.004)	(0.005)	(0.880)
Proportion of adults in household	0.625	0.606	-0.020***	Household head education level: primary	0.458	0.457	-0.001
	(0.004)	(0.004)	(<0.001)		(0.006)	(0.007)	(0.935)
Number of seniors in household	0.158	0.162	0.005	Household head education level: secondary	0.355	0.283	-0.071***
	(0.004)	(0.006)	(0.410)		(0.006)	(0.008)	(<0.001)
Proportion of seniors in household	0.063	0.060	-0.003	Household head education level: tertiary	0.052	0.126	0.075***
	(0.002)	(0.003)	(0.225)		(0.005)	(0.006)	(<0.001)
At least one member is literate 15+	0.900	0.734	-0.167***	Household head employment category: self-employed	0.478	0.496	0.019**
	(0.003)	(0.008)	(<0.001)		(0.006)	(0.007)	(0.024)
Asset index from PCA	0.007	0.008	0.000	Household head employment category: unemployed	0.082	0.176	0.094***
	(0.024)	(0.028)	(0.988)		(0.003)	(0.005)	(<0.001)
Household wall type: stone	0.175	0.221	0.046***	Household head employment category: other	0.036	0.017	-0.019***
	(0.009)	(0.011)	(<0.001)		(0.002)	(0.002)	(<0.001)
Household wall type: wood	0.100	0.195	0.096***	Quartiles of number of rooms: 2nd	0.222	0.207	-0.015**
	(0.004)	(0.007)	(<0.001)		(0.005)	(0.006)	(0.027)
Household wall type: brick	0.081	0.070	-0.011*	Quartiles of number of rooms: 3rd	0.205	0.332	0.127***
	(0.006)	(0.003)	(0.066)		(0.005)	(0.008)	(<0.001)
Household wall type: other	0.323	0.299	-0.023**	Quartiles of number of rooms: 4th	0.218	0.152	-0.065***
	(0.011)	(0.011)	(0.015)		(0.005)	(0.006)	(<0.001)
Household roof type: grass	0.082	0.088	0.006*	Quartiles of asset index: 2nd	0.316	0.274	-0.042***
	(0.004)	(0.004)	(0.074)		(0.005)	(0.007)	(<0.001)
Household roof type: other	0.095	0.101	0.006	Quartiles of asset index: 3rd	0.205	0.256	0.052***
	(0.007)	(0.008)	(0.281)		(0.005)	(0.007)	(<0.001)
Household floor type: earth	0.294	0.269	-0.025***	Quartiles of asset index: 4th	0.200	0.156	-0.045***
	(0.007)	(0.008)	(<0.001)		(0.006)	(0.007)	(<0.001)
Observations	21,585	12,662	34,247	Observations	21,585	12,662	34,247

Note: Standard errors for means and p-value for the difference annotated in brackets, based on an adjusted Wald test taking the survey design into account.

Table B1-8: Model selection for rapid approach and cross-survey estimation.

	rapid		cross-survey	
dataset	KIHBS 2015/16 pilot		KIHBS 2005/6	
urban	0.0382	(1.71)	0.148***	(7.17)
owns house			-0.0453*	(-2.15)
wall type (category 2)			0.0454*	(2.24)
wall type (category 3)	0.0419*	(2.16)	0.112***	(5.77)
wall type (category 4)	-0.117***	(-5.18)		
wall type (category 5)	-0.0506*	(-2.46)	0.0730**	(2.74)
roof type (category 2)	-0.143***	(-4.27)	-0.106***	(-5.18)
roof type (category 3)	0.0985**	(3.08)	-0.0610*	(-2.54)
floor (category 2)	-0.157***	(-7.70)	-0.166***	(-9.33)
floor (category 3)	0.0673**	(2.98)	-0.111**	(-2.75)
improved drinking water source	0.0398**	(2.64)	0.0532***	(3.88)
improved sanitation facility	-0.0541**	(-3.02)	0.0671***	(5.17)
access to electricity	0.0383	(1.63)	0.0991***	(4.25)
asset index from PCA	0.0442**	(3.14)	0.110***	(19.98)
quartiles of asset index from PCA (2nd quartile)	0.0665**	(2.89)	0.0293	(1.46)
quartiles of asset index from PCA (3rd quartile)	0.118***	(3.69)	0.0349*	(2.37)
quartiles of asset index from PCA (4th quartile)	0.105	(1.95)		
number of rooms in household	0.0380***	(4.05)	0.0382***	(8.62)
quartiles of number of rooms (2nd quartile)	0.0524*	(2.22)	0.0263*	(1.97)
quartiles of number of rooms (3rd quartile)	0.0534	(1.76)		
quartiles of number of rooms (4th quartile)	0.0126	(0.25)		
number of persons in household	-0.00942	(-0.20)	0.194	(1.34)
number of children in household	-0.0562	(-1.17)	-0.226	(-1.56)
proportion of children in household	0.475*	(2.13)	-1.024***	(-16.40)
number of adults in household	-0.148**	(-3.06)	-0.352*	(-2.43)
proportion of adults in household	1.148***	(5.18)		
number of seniors in household	-0.179***	(-3.30)	-0.379**	(-2.60)
proportion of seniors in household	1.162***	(5.04)		
dependency ratio by intervals (2nd interval)			0.0539**	(3.07)
dependency ratio by intervals (3rd interval)			0.0237	(1.37)
at least one member is literate 15+			0.0528*	(2.44)
male household head	-0.0384*	(-2.30)		
household head age group (category 2)			-0.0413*	(-2.08)
household head age group (category 3)			-0.0322	(-1.40)
household head age group (category 4)			-0.0770**	(-2.68)
household head education level (category 2)	0.117***	(4.52)	0.0628**	(3.15)
household head education level (category 3)	0.134***	(4.54)	0.132***	(5.52)
household head education level (category 4)	0.186***	(5.24)	0.351***	(7.27)
household head employment type (category 2)	0.0632***	(3.30)		
household head employment type (category 3)	-0.0418	(-1.58)	-0.0778***	(-3.89)
household head employment type (category 4)	-0.135	(-1.34)	0.0227	(1.28)
assigned 2nd module	0.0224	(1.26)		
assigned 3rd module	-0.140***	(-7.88)		
constant	-0.587**	(-2.62)	1.352***	(34.65)
N	12658		12695	
R-sq	0.373		0.511	
adj. R-sq	0.371		0.509	
AIC	21502.0		20232.4	

2. Small Area Estimation of Poverty under Structural Change[28,29]

Simon Lange,[30] Utz Johann Pape[31] and Peter Pütz[32]

Introduction

A poverty map is a spatial description of the distribution of poverty in a given country or region. While such a map is useful for policy makers and researchers when small geographic units (e.g., cities, towns, or villages) are discernable, estimates based on household surveys are typically not representative or associated with high uncertainty at such levels of disaggregation. On the other hand, most censuses do not contain information on consumption (or a surrogate such as income or expenditures) required to calculate poverty. To overcome these problems, (Elbers, Lanjouw et al. 2003) developed small area estimation poverty maps, a methodology that can be used to combine information from a detailed household survey with that from a comprehensive census. The general methodology usually consists of two steps, calibration of a statistical model based on survey data and application to the comprehensive census data. In the first step, a multiple linear regression analysis is used to estimate a model of household consumption based on survey data (which includes a consumption module). The explanatory variables in the model are restricted to the subset available in both the survey and the census.[33] In the second step, the estimated model parameters are applied to census data. The simulations provide estimates of consumption per capita for every household in the census. Since the regression model predicts the conditional mean of consumption yet one is typically also interested in higher moments of the distribution, simulation methods are used to introduce a random disturbance term.

Several criticisms have been raised with regard to the ELL estimator and extensions and alternatives have been proposed. See (Tarozzi and Deaton 2009), (Haslett, Isidro et al. 2010), (Molina and Rao 2010), (Das and Chambers 2017) and (Marhuenda, Molina et al. 2017). Comprehensive discussions on different small area estimation methods can be found in (Guadarrama, Molina et al. 2016) and (Haslett 2016). Still, ELL's is arguably the most frequently used poverty mapping approach combining survey and census data. According to (Elbers and van der Weide 2014), it has been applied in more than 60 countries. Some examples for the application of ELL, including in areas other than poverty mapping, are (Healy, Jitsuchon et al. 2003), (Demombynes and Özler 2005), (Elbers, Fujii et al. 2007), (Araujo, Ferreira et al. 2008), (Agostini, Brown et al. 2010), (Bui and Nguyen 2017) and (Gibson 2018).

A key assumption for the applicability of ELL is that the distribution of the explanatory variables is the same in both census and survey. This assumption will often be violated if time has passed between data collection for the census and survey, i.e., only a dated census and a more recent survey are available, a common situation as censuses are usually conducted less frequently than surveys. Reasons for a violation of this assumption may include demographic trends, migration, natural disasters, and conflicts. If the population parameters, including the regression coefficients, remain unchanged but the distributions of the explanatory variables change over time, ELL results in an outdated poverty map, namely a poverty map at the time of the census.

[28] SL and UP developed the research question, while PP designed and implemented the analysis in discussion with SL and UP. PP provided the first draft of the write-up. All authors jointly interpreted results and finalized the manuscript.

[29] Findings, interpretations and conclusions expressed in this paper are entirely those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. The authors would like to thank Pierella Paci and Nobuo Yoshida for valuable comments on earlier drafts as well as Chris Elbers, Peter Lanjouw and Roy Van der Weide for discussions.

[30] World Bank, Poverty and Equity Global Practice, Africa.

[31] World Bank, Poverty and Equity Global Practice, Africa. Corresponding author. E-mail: upape@worldbank.org.

[32] Economics Department, University of Göttingen.

[33] The ELL estimator requires relevant explanatory variables for the model predicting consumption to be measured in a comparable way both in the census and in the survey. Differences in coding schemes or even the way the interview was conducted can prevent reasonable harmonization between census and survey variables. See also Tarozzi and Deaton (2009) for a brief discussion.

The discussed assumptions on the explanatory variables can be relaxed if household characteristics from the census are used to explain consumption values from the survey in the first stage to obtain parameter estimates. These can then be used to predict consumption values using the census data in the second stage. As it is usually impossible to match households between a census and a survey, the estimation needs to be conducted at a higher geographical level, for instance at the level of census enumeration areas. Throughout this paper, we will refer to the generic term of clusters as the lowest level at which census and survey information can be matched. If the assumptions on the explanatory variables hold, this aggregation may worsen the prediction accuracy vis-à-vis ELL, with the magnitude of the loss of precision hinging on the regression model in the first stage. Note that ELL also propose the additional use of census means to explain location effects, i.e. cluster-specific effects. In this regard, our approach can be considered as a variant of ELL without the use of household-level variables included in both census and survey and without reliance on the associated assumptions. When we refer to the ELL method throughout this paper, we have in mind an estimator that combines survey and census variables at the household-level, the central idea of the approach.

In the case that at least one of the underlying assumptions of ELL is violated, our new approach will still produce up-to-date poverty maps with unbiased poverty estimates. The key assumption we introduce is that *aggregate* household characteristics from the old census relate to consumption the same way in clusters covered by the new survey as in clusters not covered by the new survey. This assumption will hold (on average) if clusters are randomly drawn. Note that a similarly weak assumption has to be made for the applicability of the ELL method if the census and survey are conducted at the same time, namely that household characteristics from the survey relate to consumption the same way in clusters covered by the survey as in clusters not covered by the survey.

In a different scenario, a recent census and only dated survey data may be available. Reliable predictions of poverty measures at the time of the recent census can only be obtained under the additional strong assumption of non-changing structural parameters (including the regression parameters linking explanatory variables to consumption) over time (e.g. Kijima and Lanjouw 2003). This holds for both ELL and our estimator. If both structural parameters and the distribution of the explanatory variables change over time, ELL results in biased estimates. In contrast, linking census covariate means to predict survey consumption would remain a valid method to generate a poverty map at the time of the survey. In the remainder of this paper, we will focus on the practically more relevant case of a dated census and a recent survey.

Although monitoring poverty over time is of eminent interest to economists (see, for instance, Deaton and Kozel 2005), little attention has been paid to updating small area estimation approaches which combine dated census and recent survey data. (Emwanu, Hoogeveen et al. 2006) require panel data with one wave collected at the time of the census. While structural changes in the explanatory variables over time are not an issue in such a setting, the remaining assumptions of the ELL method as described above are still required. Furthermore, availability of panel data over a longer time span without substantial attrition is rare, especially in developing countries. The National Statistical Coordination Board of the Philippines uses only explanatory variables deemed time-invariant to estimate inter-censal poverty measures. Whether variables change over time is not assessed formally but rather based on *impromptu* assumption. This approach still relies on similar assumptions as the ELL method, even though changes in the distribution of the explanatory variables are ruled out by choosing time-invariant variables. One may also test whether the distribution of potential predictors changed over time and then restrict the set of predictors in the first stage to only those that exhibit no drift.[34] However, severe shocks and extended time periods between survey and census will tend to quickly exhaust the set of viable predictors to do so. And it is exactly in those settings in which the demand for an updated poverty map is likely to be high. (Isidro 2010) and (Isidro, Haslett et al. 2016) propose to fit a model on simultaneously collected survey and census data first, for instance by ELL, and update the resulting estimates using a more recent survey. Their Extended Structure Preserving Estimation (ESPREE) approach does not require panel data but contemporaneous surveys and census collection with common variables. The ESPREE method relies on updating multi-way contingency tables which is computationally tractable only for a limited number of categorical explanatory variables and an outcome indicator which is a proportion, for instance the number of people who live below the poverty line.

In the remainder of this paper, we show that our proposed method has comparably low data requirements and weak assumptions. Although our outcome variables will be measures of welfare, our method is applicable to a wide range of outcome measures and research questions beyond poverty mapping. Section 2 presents the idea of the approach in detail. Section 3 describes the properties of the resulting poverty estimator. Simulation studies on artificial and real data are presented in Sections 4 and 0, respectively. Section 6 concludes.

Estimating poverty measures under structural change

Assume that the target population is a village v. While the proposed method is applicable to essentially all measures which can be derived from consumption (or any other dependent variable measuring welfare), for instance inequality measures such as the Gini coefficient, assume for now that the measures of interest are poverty measures of the FGT family (Foster, Greer et al. 1984):

$$W_{\alpha v} = \frac{1}{N_v}\sum_{j=1}^{N_v} W_{\alpha v j} \tag{1}$$

with

$$W_{\alpha v j} = \left(\frac{z-y_{vj}}{z}\right)^{\alpha} I(y_{vj} < z), \quad \alpha = 0, 1, 2.$$

Here, N_v is the size of the village population, y_{vj} is the consumption for individual j in village v, z is the poverty line and $I(y_{vj} < z)$ is an indicator function which equals one if the consumption of an individual is below the poverty line and zero otherwise. Poverty headcount ratio, poverty gap and poverty severity are obtained for $\alpha = 0, 1$ and 2, respectively.

The consumption model

Usually, consumption values are observed at the level of the household, not the level of the individual. As most household consumption values are unobserved in a village, one needs a model which predicts those values for all households H_v. Let y_{cht} be the consumption of household h in cluster c at time t. Then, the model of consideration is

[34] This has been suggested for an update of the Bangladeshi poverty maps by researchers from Bangladesh Bureau of Statistics, World Bank and United Nations World Food Programme (2010). Updating Poverty Maps: Bangladesh Poverty Maps for 2005.

$$y_{cht} = \boldsymbol{x}'_{c.,t-1}\boldsymbol{\beta} + u_{ch} = \boldsymbol{x}'_{c.,t-1}\boldsymbol{\beta} + \eta_{ct} + e_{cht}, \quad h = 1, \dots, H_c, \quad c = 1, \dots, C,$$
$$\eta_{ct} \sim iid\, \mathcal{F}_1\left(0, \sigma_\eta^2\right), \quad e_{cht} \sim iid\, \mathcal{F}_2\left(0, \sigma_e^2\right), \tag{2}$$

which relates the (potentially transformed) consumption variable linearly to a vector $\boldsymbol{x}_{c.,t-1}$ containing dated census means of covariates over the cluster c from time point $t-1$.[35] The two error components are the cluster effects η_{ct} and the household errors e_{cht} which follow the distributions $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, and are assumed to be independent of each other. It is possible to allow for heteroscedasticity in the household error by modeling its variance to covariates. Such covariates may include the census means used in the main regression, but also higher moments such as the variance. Furthermore, geographic information and the fitted values of the first-stage regression may be used. The ELL method describes one option to model heteroscedasticity within the framework discussed here, while (Pinheiro and Bates 2000) provide a more comprehensive discussion.

Model estimation based on survey consumption values

In the first stage, model (2) is estimated using all household consumption values which are available for the village of interest in the survey. The estimation can be done within the maximum likelihood framework or by weighted or (feasible) generalized least squares.[36] As the estimates are used to predict consumption values for the census, the aim is to find a model with high predictive power. Thus, one should find a parsimonious model containing only covariates which explain a substantial share of the variation in the dependent variable. Due to averaging over the cluster, means over candidate variables should exhibit variation across clusters.

Bootstrapping census consumption data

In the second stage, model (2) is used to predict consumption values for each household in the village of interest based on the census. Note that, to be consistent with the first-stage model using the consumption values from the survey, the explanatory variables in the second stage are also averaged within clusters, i.e., all households in the same cluster have the same value for each explanatory variable. Using the estimated regression coefficients $\widehat{\boldsymbol{\beta}}$ from model (2) yields predictions $\hat{y}_{cht} = \boldsymbol{x}'_{c.,t-1}\widehat{\boldsymbol{\beta}}$, i.e. predicted conditional means. To account for the deviations of the observed consumption values from these means, random disturbance terms have to be added by simulation. Assume that the aim is to estimate a poverty measure W, where the indices from (1) are dropped for notational convenience.

A bootstrap procedure is applied to generate R pseudo censuses and resultant poverty measures:

1. Draw all model coefficients from their respective sampling distribution estimated by the model in the first stage, including regression coefficients, random term variances and possible heteroscedasticity parameters. Multivariate normal distributions with first-stage estimates for the means and variance-covariance matrix are used to draw the regression coefficients and the heteroscedasticity parameters.[37]

[35] In practice, one could use additional secondary information to explain consumption, e.g. geographic information which is typically available in poverty mapping exercises. In this paper, we restrict ourselves to information that is available in the census. Besides, time-invariant explanatory variables on the household level $\boldsymbol{x}_{cht}$ could be easily added to the consumption model. As discussed in Section 0, we do not assume many time-invariant variables to be available in practice.

[36] The chosen estimation method depends on whether and how the survey design, potential heteroscedasticity and the clustering nature of the data are taken into account.

[37] One may also assume a distribution for the error components' variances such as the gamma distribution, e.g., but in many cases it is reasonable to treat their estimates from the first stage as fixed, especially if the numbers of enumeration areas and households in the survey

2. Conditional on the parameters describing the error components' distributions from the first step, cluster effects and household errors are drawn from their respective distributions. One option is to use a parametric bootstrap, i.e., to assume certain parametric distributions for which the estimates from the first stage regression might give some indication. However, a nonparametric bootstrap procedure is a valid alternative or supplement. In this case, a cluster effect can be estimated as the mean of the deviations between observed and predicted values in one cluster, i.e. $\widehat{\eta}_{ct} = 1/H_c \sum_h^{H_c}(\hat{y}_{cht} - \boldsymbol{x}'_{c.,t-1}\widehat{\boldsymbol{\beta}})$, while the household residuals are computed as those deviations minus the cluster effects, i.e., $\widehat{e}_{cht} = (\hat{y}_{cht} - \boldsymbol{x}'_{c.,t-1}\widehat{\boldsymbol{\beta}}) - \widehat{\eta}_{ct}$. There are different strategies to draw from these sampling distributions. One may draw with replacement from all estimated cluster effects and all household residuals. Alternatively, the household residuals may be drawn only from the location to which the cluster effect belongs. This strategy generally allows the estimated two error components to be related in a nonlinear way, even though they are by construction uncorrelated.
3. Calculate the predicted consumption values for all households and all individuals as well as the poverty measure $\widehat{W}^{(r)}$ derived from those values.
4. Repeat steps 1 to 3 R times.

For the poverty measure W, the (simulated) expected value is then given by

$$\tilde{\mu} = \frac{1}{R}\sum_{\mathrm{r=1}}^{\mathrm{R}} \widehat{W}^{(r)}$$

and its variance by

$$\tilde{V} = \frac{1}{R}\sum_{r=1}^{R}(\widehat{W}^{(r)} - \tilde{\mu})^2. \tag{3}$$

Due to the bootstrap procedure, the variance contains uncertainty from the first-stage model (step 1, referred to as model error in the next section) and the unobservable part of consumption (step 2, referred to as idiosyncratic error in the next section).

Properties of the estimator

In the following, we will investigate the properties of our welfare estimator presented in the previous section.

As described in ELL, the prediction error, the difference between the true poverty measure W for a target population, say a village, and our estimator $\tilde{\mu}$ of its expectation $E(W) = \mu$, is given by

$$W - \tilde{\mu} = (W - \mu) + (\mu - \hat{\mu}) + (\hat{\mu} - \tilde{\mu}). \tag{4}$$

Here, the third component is the computation error which is the difference between our estimator $\tilde{\mu}$ and its expectation $\hat{\mu}$. In the following, we assume the computation error to be negligible by applying a sufficiently high number of bootstrap simulations.

are large since then there is not much uncertainty in the variance estimators. The household error variance estimator is usually very precise as it is based on the (large) number of households in the survey. The amount of enumeration areas in the survey is smaller but the uncertainty in the variance estimator of the enumeration area effects is often still negligible. In practice, one may check whether the estimated variances of the error components' variances are small enough in order to treat them as fixed in all bootstrap replications.

The first term on the right-hand side of equation (4), $(W - \mu)$, is the idiosyncratic error arising from the unexplained part of consumption of which the poverty measure is a function. Due to the stochastic nature of consumption, the true poverty measure differs from its expected one. Note that the population in the small area of interest is finite and can be seen as a realization from an infinite population. Hence, all asymptotic results for the idiosyncratic error of the poverty measure from ELL carry over to the new approach presented here: the idiosyncratic error vanishes asymptotically for growing population size, including additional clusters and individuals.

The second part of equation (4), $(\mu - \hat{\mu})$, is the model error, which originates from the estimation of (unknown) population parameters. The expectation of the model error equals zero if the poverty estimator is an unbiased estimator for the expected value of the true poverty measure. Whether this is the case hinges on the regression model selected for the survey data.[38] What is crucial is that the assumptions of zero mean, independence, and homoscedasticity for the error components, namely the cluster effects and the household errors, hold. Likewise, if the error components are assumed to follow certain distributions and these parametric assumptions are used for the generation of simulated census data sets (see Section 2.3), they also have to hold. Note that these assumptions may be valid even if dated census data are used for predicting survey consumption values. Thus, one crucial part is the diagnosis of the estimated error components from the first-stage regression. If plots or statistical tests on the estimated cluster effects and residuals suggest violations of distributional assumptions, one should adjust the model accordingly. More specifically, heteroscedasticity, serial correlation, and non-normality can be detected and accounted for, for instance by choosing different predictor specifications, transforming the dependent variable, or explicit modeling of heteroscedasticity as discussed in Section 2.1. The variance of the model error also depends fully on the properties of the first-stage estimators. It decreases in survey sample size.

If the assumptions of the ELL method hold and the models are correctly specified, the ELL estimator will usually exhibit a smaller variance of the prediction error than our estimator. The reason is that the latter is a between estimator that ignores variation within clusters. Intuitively, both estimators would only be similarly efficient if the explanatory variables differed distinctly more between clusters than within clusters. In practice, another exception might occur if there are many missing values in the explanatory variables in the survey. Without imputation methods that are subject to estimation uncertainty, the ELL first-stage estimator would be based on a smaller sample than our estimator.

In practice, the variance components of the idiosyncratic and model error are not estimated separately. Rather, the entire variance of the prediction error is obtained from the variation of the simulated poverty estimates in equation (3). Hence, under correct distributional assumptions on the random components, the bootstrap procedure allows to draw valid inferences, i.e., to build confidence intervals which include the true poverty measure with a predetermined probability. For instance, bootstrap percentile intervals, which can be constructed directly from the bootstrap estimates (see Section2.3), can be used for inference.

Another potential issue in practice is multicollinearity. Note that the fundamental unit of the predictors in the first stage is a cluster, not a household, and that the number of parameters that can be included in (2) is hence restricted to the number of clusters. However, household budget surveys that are used to estimate poverty incidence typically cover 500 clusters or more, with some covering substantially more. Hence, we believe that our estimator could be based on a moderate number of

[38] Note that it is neither intended nor necessary to establish causal or direct effects of explanatory variables on consumption. Thus, the regression coefficients in model (3) need not be estimated unbiasedly or consistently with regard to the direct effects of the explanatory variables. In contrast, asymptotical unbiasedness of $\hat{\mu}$ can be obtained for several models, even if a single parameter in such a model might capture the effect of several correlated variables.

regressors that would be sufficient to accurately predict household consumption which is assumed to differ between clusters.[39]

Simulation experiments

A simulation study is conducted to compare the performance of our approach, ELL, and a purely survey-based estimator in predicting FGT poverty measures. We focus on the poverty headcount ratio and the poverty gap with three generic poverty lines that render 25%, 50%, and 75% of the population poor. The simulation setting is based on (Tarozzi and Deaton 2009). In particular, the target population in the census is a village with $N = 15{,}000$ households, divided into 150 clusters $k_c \in \{1, \dots, 150\}$, each of size 100. In each simulation run, an artificial household survey is drawn from the census by selecting randomly ten households from 100 randomly selected clusters. First, both data sets are generated by the following process with homoscedastic errors:

$$y_{ch} = \beta_0 + \beta_1 x_{ch} + \eta_c + e_{ch} = 20 + x_{ch} + \eta_c + e_{ch}$$

$$x_{ch} = 5 + 0.01k_c + w_{ch} - t_{ch}, \qquad w_{ch} \sim N(0,1), \quad t_{ch} \sim U(0,1),$$

$$\eta_c \sim N(0,0.01), \quad e_{ch} \sim N(0,1).$$

Note that the explanatory variable is generated so that it differs in expectation between clusters. Such a situation with large and systematic differences in the averages of covariates across clusters (e.g., average levels of education or dwelling characteristics) is frequently observed in practice. This setting is ideal for the ELL method, which exactly models the data generating process. A linear regression based on the target population yields an R^2 of 0.55 while the new method with an R^2 of 0.08 has considerably lower explanatory power.

A second setting mimics a real-word situation where the census is dated and a more recent household survey (with an underlying true census which is not observed) is available. Here the model which explains consumption in the same way as the first setting for both the census and the survey, but the explanatory variable for the more recent survey is generated by

$$x_{ch} = 5 + 0.01k_c + w_{ch}, \qquad w_{ch} \sim N(0,1),$$

where the sampled 100 clusters in the survey have the same values for k_c as they have in the old census. For both estimators, the R^2 obtained from the first-stage regression for all generated surveys is on average similar to the R^2 based on the census in the first setting.

Note that in both settings, estimators purely based on the survey have desirable properties as the surveys are representative of the respective village population at the time of data collection. In real-word situations, however, a survey is not necessarily representative at the village-level.

All results are based on 300 Monte Carlo replications with 500 bootstrap census data sets generated in each replication for the two methods which use census data. The bootstrap procedure to sample the error components uses a simple nonparametric version, i.e., both cluster effects and household errors are independently sampled with replacement from their sample analogs from the first-stage regression. See Section Bootstrapping census consumption data for details.

In the first setting, the root mean squared error is, as expected, smallest for the ELL method, followed by our estimator and an estimator solely based on the survey (Table B2-1). Although the R^2 from the first-stage regression for the ELL method is seven times as large as for our new method, the root mean

[39] One commonly used rule-of-thumb is to restrict the number of predictors to the square root of observations. While our results in Sections 0 and 0 are based on 100 clusters and less than ten variables, 500 clusters would allow the analyst to base the first-stage estimation on more than 20 census averages (or other summary statistics computed at the cluster-level).

squared errors only differ by a factor of about 1.5 or two-thirds, respectively. The coverage rates of the two methods are close to the nominal one of 95% and the bias is negligible.

In the second and more interesting setting, the ELL method naturally is the worst in terms of prediction and generates invalid confidence intervals (Table B2-2). The upward bias originates from the data generating process above: as the expected values of x_{ch} and thus y_{ch} are greater in the recent survey and its underlying population than in the dated census, using the dated census data to predict current poverty statistics necessarily underestimates the current values of y_{ch} and hence overestimates the magnitude of poverty. In contrast, the new method results in valid confidence intervals. It also results in a lower mean squared error in comparison to the purely survey-based estimate since additional census information is exploited. The last result typically holds on average if the model assumptions are fulfilled (as it is the case in this simulation setting) and census and survey size differ distinctly. The latter is often true in practice.[40]

Table B2-1: Setting 1 – simultaneous census and survey collection.

		New estimator			ELL estimator			Survey Est.
	True value	Bias	RMSE	Coverage	Bias	RMSE	Coverage	RMSE
W_0 (.25)	0.2500	-0.0043	0.0114	0.9767	-0.0030	0.0085	0.9767	0.0025
W_0 (.50)	0.5000	-0.0047	0.0135	0.9867	-0.0047	0.0100	0.9667	0.0043
W_0 (.75)	0.7500	0.0023	0.0117	0.9767	0.0013	0.0084	0.9867	0.0092
W_1 (.25)	0.0092	-0.0002	0.0007	0.9400	-0.0001	0.0005	0.9533	0.0002
W_1 (.50)	0.0239	-0.0004	0.0011	0.9767	-0.0003	0.0008	0.9700	0.0003
W_1 (.75)	0.0482	-0.0005	0.0014	0.9767	-0.0005	0.0011	0.9733	0.0004

The RMSEs is the root of the mean squared deviations of the estimates from the true value over 300 replications. Coverage rates are calculated for 95% bootstrap percentile intervals.

Table B2-2: Setting 2 – dated census and recent survey, explanatory variable changes over time

		New estimator			ELL estimator			Survey Est.
	True value	Bias	RMSE	Coverage	Bias	RMSE	Coverage	RMSE
W_0 (.25)	0.2500	-0.0021	0.0114	0.9700	0.1152	0.1157	0.0000	0.0026
W_0 (.50)	0.5000	0.0029	0.0132	0.9800	0.1286	0.1289	0.0000	0.0089
W_0 (.75)	0.7500	0.0034	0.0122	0.9667	0.0890	0.0893	0.0000	0.0084
W_1 (.25)	0.0088	0.0000	0.0006	0.9433	0.0065	0.0066	0.0000	0.0002
W_1 (.50)	0.0232	-0.0001	0.0010	0.9767	0.0112	0.0112	0.0000	0.0004
W_1 (.75)	0.0460	0.0000	0.0013	0.9800	0.0149	0.0150	0.0000	0.0007

The RMSEs is the root of the mean squared deviations of the estimates from the true value over 300 replications. Coverage rates are calculated for 95% bootstrap percentile intervals.

[40] Note that under the stated conditions, our estimator performs better only in predicting the true value on average. In a single sample, the pure survey mean is superior to our approach if the sample mean is by chance equal or very close to the census mean. An extreme example includes the limiting case in which the recent survey is equal to the underlying census. Then, the survey mean is trivially the census mean, that is, there is no error at all. But our new method is still prone to idiosyncratic and (small) simulation error, even under correct model specification.

Application to census data from Brazil

In order to test the proposed method in a real-world example, we use data extracts from the 2000 and 2010 Brazilian censuses provided by the Integrated Public Use Micro Sample (Minnesota Population Center 2017). While the data sets provide no information on consumption, the preferred basis of welfare measurement in developing countries, they include information about monthly income at the level of the individual. In addition, the data sets provide information that is potentially useful in explaining incomes, including the location in which the household resides (urban / rural), the number of household members, ownership of specific assets, and employment status. This allows us to generate artificial surveys from the more recent census and predict income by dated census data. The poverty measures derived from the predicted income values can then be compared to the true ones based on the entire recent census.

The data sets are extracts from the respective censuses. Roughly ten million individuals are included in each data set, corresponding to 6 and 5 percent of the population in 2000 and 2010, respectively. The country is divided into 25 states and 1,980 municipalities. These municipalities constitute the smallest geographical unit which can be matched between 2000 and 2010. Accordingly, we consider them as clusters in the terminology used in the previous sections. Thus, we use averages over municipalities for the 2000 census to predict household incomes in 2010. Household incomes are calculated as the sum of individual incomes of all household members, adjusted for the household size according to the OECD-modified scale.[41] The poverty line is set to \$5.5 in 2011 PPP per person and day.[42] For the sake of illustration, we focus on one single Brazilian state, Minas Gerais. In comparison to other states, it features a large number of municipalities (282) which we can match over the two censuses. The data sets comprise 259,096 and 350,696 observed households in 2000 and 2010, respectively. Roughly maintaining the ratio of number of households, we sample randomly about 20,000 households (year 2000) and 26,000 (year 2010) from the respective censuses and treat the resulting data sets as new censuses. The reason for that is not only computational convenience but also the fact that the state of Minas Gerais is the small area of interest and should therefore exhibit a population size similar to common empirical applications in small area estimation. The true headcount ratios in these artificial censuses change substantially over time, from 0.27 percent in 2000 to 0.11 percent in 2010.

As variables with sufficient variation between municipalities and power to explain variation in income we use location (urban or rural), number of household members, availability of a phone as well as employment status and level of schooling completed of the person with the highest educational attainment in the household. When all households from the 2010 census are used, a linear regression with these explanatory variables yields an R^2 of 0.095. The estimates of the regression coefficients can be found in Table B2-3: New estimator using all households from 2010 census. We also added squares of the variables, interactions and many other variables to this simple model without obtaining a substantially higher predictive ability measured by the Akaike Information Criterion. The estimated cluster effects variance in a linear mixed effects model based on the 2010 census is 0.02 and small compared to the estimated household residual variance of 0.88.

[41] http://www.oecd.org/eco/growth/OECD-Note-EquivalenceScales.pdf.

[42] The World Bank calculates poverty rates at three poverty lines for Brazil, see http://databank.worldbank.org/data/download/poverty/B2A3A7F5-706A-4522-AF99-5B1800FA3357/9FE8B43A-5EAE-4F36-8838-E9F58200CF49/60C691C8-EAD0-47BE-9C8A-B56D672A29F7/Global_POV_SP_CPB_BRA.pdf. We chose the highest one since otherwise there are very few households below the other two poverty lines in both years. Our main aim is to illustrate the method's applicability even in settings in which the time span between the data sets is large and relevant changes in the welfare status have occurred over time.

Table B2-3: New estimator using all households from 2010 census

Dependent variable: Income	Coefficient estimate	95% confidence interval
Phone	0.448	[0.318; 0.579]
Employment status	-0.518	[-0.668; -0.367]
Urban	0.233	[0.126; 0.340]
Education	0.335	[0.248; 0.422]
Household members	-0.159	[-0.188; -0.130]
Constant	2.655	[2.449; 2.861]
Number of census households	21,543	
Number of municipalities	282	
R^2	0.0950	

We draw artificial surveys from the 2010 census by first sampling randomly without replacement 100 municipalities and then sampling without replacement 10 households randomly from each of those municipalities, resulting in an overall survey sample size of 1,000 households. As the number of households differs between municipalities, the estimation at the first stage has to account for these differences by using appropriate weights. Note that this requires knowledge of the number of households in the municipalities at the time of the survey. In practice, when no recent census is available, the number of households at the cluster level can be obtained from a listing exercise which is usually also needed for the sampling scheme for the household survey.

We use a weighted linear regression in the first stage. Means of the explanatory variables over municipalities for the year 2000 are used to explain household per capita income in 2010. To remove apparent right-skewness in the dependent variable, a log-transformation is applied after adding one to the household income values. The latter is done due to the non-negligible amount of zero income values.[43]

In the second-stage bootstrap procedure, the regression coefficients are sampled from a multivariate normal distribution where the expected values are the first stage estimates and the robust variance-covariance matrix accounts for correlation within the clusters. The error components are generated by a nonparametric bootstrap. In particular, cluster effects are drawn with replacement from the 100 first-stage estimates. The household errors are drawn with replacement from the first-stage residuals belonging to this specific cluster. See also Section Bootstrapping census consumption data.

For computing an overall state-level poverty measure, it is crucial to know at least approximately the distributions of households over municipalities in the population at the time of the recent survey: The proposed approach imputes poverty measures for the municipalities by using the dated census households. Clearly, a composite measure of those single poverty measures has to account for the number of households in the municipalities at the time of the recent survey.

We compare the performance of our estimator for the headcount ratio[44] in the state of Minas Gerais with the ELL estimator and a simple (weighted) mean based solely on the recent survey. Note that the sample is, in contrast to many real-world applications, representative and rich at the small-area level

[43] The proportion of all households in the 2010 census data with an income of zero amounts to 3.16 percent.
[44] We also estimated the poverty gap in the same simulation setting and obtained qualitatively similar results.

such that the weighted survey mean is an unbiased poverty estimator by construction. For the ELL first-stage regression, the same explanatory variables are used, yet on the household level and using the 2010 survey data. In a regression based on all households from the 2010 census, this simple model specification already yields an R^2 of 0.33. We conduct 300 Monte Carlo simulations with 200 bootstrap census data sets generated in each replication.

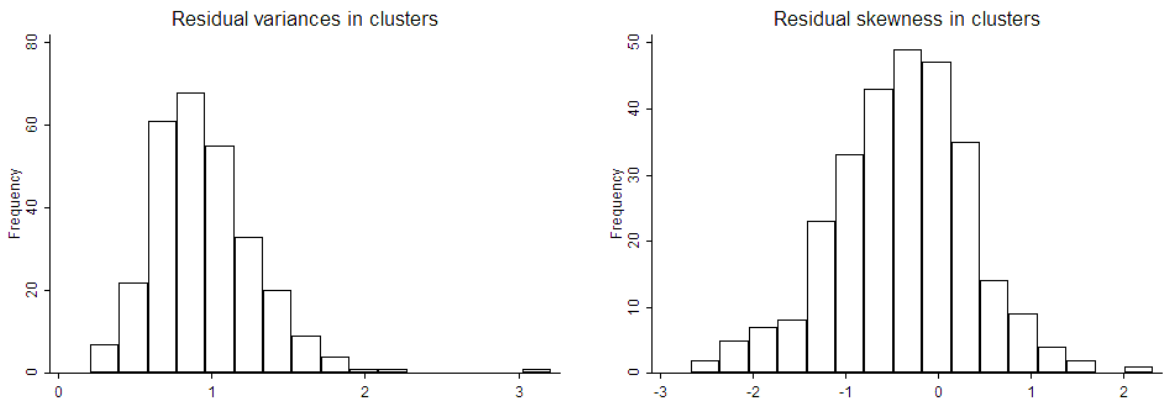
For our estimator, the coverage of the confidence intervals is below the nominal one of 95% (Table B2-4). The estimator is slightly biased which may be because of unmodeled heterogeneity in the error distribution, e.g. between clusters. In a regression based on all households from the 2010 census, variances and skewness of the residuals differ considerably between clusters (Figure B2-1). However, we found no clear pattern with respect to the fitted values from a first-stage regression or other explanatory variables. As the number of clusters is relatively small, already one cluster with an extreme behavior of its errors can potentially have a large effect on estimates of welfare measures. In practice, it can be challenging to detect and model such peculiarities in the error distribution. Potential remedies are discussed in Section 0.

Table B2-4: State level headcount ratio at household-level

		New estimator			ELL estimator			Survey estimator	
	True v.	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE
W_0 (5.5)	0.1076	0.0098	0.0138	0.8900	0.1020	0.1038	0.0000	-0.0015	0.0137

Due to the bias in the headcount ratio estimator, a comparison with a (weighted) mean purely based on the survey yields a comparable, even slightly superior performance of the latter in terms of the root mean squared error. Since the distribution of the explanatory variables has changed from 2000 to 2010 (e.g. the share of households owning a phone increased from 67% to 70%), the ELL estimator is severely biased.

Figure B2-1: Household residual variances and skewness in clusters



So far, the poverty measures have been calculated at the household-level, while one is typically also interested in poverty measures at the individual-level, e.g. the percentage of poor people and not households in a small area. In principle, one could conduct the first-stage regression at the individual level which is equivalent to replicating the household entries in the data sets by the respective household sizes.[45] However, when calculating an overall poverty measure from the simulated income

[45] This is due to the fact that both the household equivalent income and all explanatory variables are the same for all household members.

values in the second stage, one then needs to know the number of individuals in each cluster at the time of the recent survey. The required information may be available from a previous listing exercise.

A second option starts with the first-stage regression on the household-level as described above. The smallest unit to match between the census and the survey are the municipalities. In fact, the same value of consumption is predicted on average for all households in the same municipality. For a single bootstrap simulation, they only differ by the simulated household error. Since a relationship between household size and income is assumed on the household level, typically that bigger households are poorer, one cannot randomly assign household sizes to the households. Hence, one possible remedy is to save the household sizes from the survey households and residuals from the first-stage regressions and draw them together in the bootstrap procedure in the second stage.

Another approach would impute the individual poverty measure based on its relationship with the household poverty estimators. This relationship may be hypothesized on the basis of prior knowledge or estimated from the data set at hand. Though, if the relationship between household sizes and income differs between municipalities, the latter two methods do not yield unbiased state-level poverty estimators in general.

In our application, we follow the second approach, i.e. we run the regression on the household level and sample residuals together with household sizes. The results indicate similar conclusions as the analyses at the household-level.

Table B2-5: State level headcount ratio on individual level

		Our estimator			ELL estimator			Survey estimator	
	True v.	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE
W_0 (5.5)	0.1259	0.0054	0.0126	0.9600	0.1249	0.1270	0.0000	-0.0026	0.0179

Conclusions

In this paper method to generate poverty maps has been presented.[46] While ours is a valid approach to combine simultaneously collected census and survey data, it also allows analysts to obtain up-to-date poverty maps when only a dated census and a more recent survey are available. In contrast to existing approaches, it has low data requirements and weak assumptions. Simulation studies showed an overall good performance. If the distribution of explanatory variables changes over time, our new estimator is superior to the most frequently used method for contemporaneous census and survey collection.

However, our approach is not immune to issues typically encountered in small area estimation techniques that combine census and survey data. In particular, variable selection and adequate modeling of apparent heteroscedasticity and differences in skewness in the residuals can be challenging. Besides, the key assumption, namely that aggregate household characteristics from the old census relate to consumption the same way in clusters covered by the new survey as in clusters not covered by the new survey, may not hold for the specific welfare estimation exercise at hand. For example, the migration pattern between census and survey collection may vary between clusters and may be correlated with the welfare status which is typically not captured by the model.

Violations of the assumptions on the error term may be partly solved by allowing for more distributional flexibility in the response variable or the error term. (Rojas-Perilla, Pannier et al. 2017) and the references therein provide various transformations of the response variable to achieve the

[46] Software code in Stata and R for the implementation of our proposed method are available on request from the authors.

validity of the assumption of identically and normally distributed error terms. A more comprehensive approach would be the application of Generalized Additive Models for Location, Scale and Shape (Rigby and Stasinopoulos 2005). This framework not only includes a huge variety of potential response distributions, but also allows to link all parameters of those distributions to explanatory variables. This allows for a straightforward way to model heteroscedasticity and skewness simultaneously in one coherent model. Moreover, nonlinear and spatial effects can be integrated into the GAMLSS framework. Although model choice is also a challenging task, it might be a very interesting direction for future research to combine GAMLSS and existing small area approaches, irrespective of the time span between census and survey collection.

3. Measuring Poverty in Forced Displacement Contexts[47,48]

Utz Pape and Paolo Verme[49]

Introduction

The United Nations High Commissioner for Refugee (UNHCR) estimated that the global number of Forcibly Displaced Persons (FDPs) in the world surpassed 84 million in 2021, up from around 40m in 2010 and accounting for over one percent of the global population.[50] This sharp growth in displaced people during the past decade can be largely attributed to the Syrian conflict started in 2011, the displacement of the Rohingya people since 2017, and the intensification of several conflicts in Sub-Saharan Africa, particularly along the Sahel region. These numbers are unprecedented in the history of displacement when recording started with the establishment of the UNHCR in 1950 and the signature of the Geneva Refugee convention in 1951.

FDPs are not a homogenous group. They include Internally Displaced Persons (IDPs - citizens of a country that have been displaced within the boundaries of their own country due to conflict or security reasons), asylum seekers (displaced people outside their own countries who formally ask for asylum), refugees (people who have obtained asylum in the host country), and other displaced groups that defy simple categorizations. These categories of people fall under the mandate of the UNHCR because they have been displaced "forcibly" because of conflict or violence and because they are in need of international protection. They exclude other categories of displaced people who were not forced to move because of conflict or violence such as economic migrants and victims of natural or environmental disasters. Of course, many people cannot be simply categorized in these groups and this makes statistics on FDPs gross estimates, but the growth and relevance of these numbers are undisputed.

The growth in number of FDPs poses a challenge to the measurement of global and national poverty. Those who are forcibly displaced and in need of international protection tend to be persons who have lost their assets, financial resources, and social networks. They are typically very poor with no obvious path out of poverty. For refugees, their number vanish from poverty statistics of their own country because they are no longer counted in the place of origin. Both, IDPs and refugees are also not properly accounted for in the country they reside in. Their numbers – even though high in absolute terms – are often low relative to the non-displaced population (with some exceptions like Lebanon and South Sudan). Hence, they do not explicitly show up in official statistics. Even if – as in some but not all countries – their locations are appropriately included in the sampling frame, they are unlikely to be sampled due to their small proportion relative to the population, in most cases excluding them from official (poverty) statistics.

The number of FDPs is not small in terms of the absolute global poverty count. Before the COVID-19 pandemic, around 10 percent of the global population (780 million people) was estimated to be extremely poor, below 1.9 USD/day in Purchasing Power Parity (PPP) terms. If we make the conservative assumption that about 1 in 2 FDPs are extremely poor, this translates into 41.2 million

[47] UP and PV contributed equally to the manuscript.
[48] This is an expanded version of the previously published short note 'Poverty Measurement for Forcibly Displaced Populations: Challenges and Prospects of a New Field'.
[49] Authors listed alphabetically. UP: The World Bank and Georg-August-University Göttingen, upape@worldbank.org. PV: World Bank, pverme@worldbank.org. The findings, interpretations and conclusions expressed in this paper are entirely those of the author, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. This work is part of the program ``Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership''. The program is funded by UK aid and was established in partnership with the United Nations High Commissioner for Refugees (UNHCR). The scope of the program is to expand the global knowledge on forced displacement by funding quality research and disseminating results for the use of practitioners and policy makers. This work does not necessarily reflect the views of the UK, WBG or UNHCR.
[50] https://www.unhcr.org/refugee-statistics/

poor or about 5 percent of the global poor. The UN population division uses UNHCR's FDPs estimates to adjust national population figures by adding FDPs to population statistics. However, when the national and global poverty figures are estimated, FDPs are often missing from the poverty count. This practice inflates the denominator (population) but does not correct the numerator resulting in an underestimation of the poverty rate.

Low interest, lack of proper microdata on FDPs, measurement issues and political reasons are some of the obstacles that prevent a proper measurement of poverty among FDPs. Before the Syrian conflict, poverty measurement among FDPs was largely confined to occasional exercises carried out by local or international NGOs on behalf of humanitarian agencies such as the UNHCR and the World Food Program (WFP). These exercises often lacked the data sophistication and academic rigor that characterizes poverty measurement for regular populations in high, middle, or low-income countries, typically undertaken by development agencies such as the World Bank, UNDP or regional development banks. The UNHCR, for example, did not collect survey data on income, consumption, or expenditure for refugees systematically as the focus and mandate of the agency is on humanitarian protection rather than poverty alleviation. It only started to collect data on income more systematically when budget restrictions forced the agency to start targeting cash programs. WFP did collect data on consumption regularly but this effort was largely focused on food consumption for food security and nutritional assessments rather than poverty measurement. With little data and visibility, the economics profession and poverty specialists across the social sciences also largely neglected FDPs (Verme 2016). It is only in the past decade that development agencies took an interest in FDPs and, together with humanitarian agencies, started to consider how to gather data and measure poverty among FDPs continuously and rigorously.

Measuring poverty among displaced population has also its own specificities. FDPs are not a representative sample of the population of origin or destination country. They tend to be poorer than most host populations, tend to have higher shares of females and children, have very few assets and only occasional incomes, they may not be allowed to work, have more limited access to public services, have a higher incidence of people with physical and psychological disorders and tend to rely almost entirely on aid provided by the international community (Verme 2016). Concepts such as income, expenditure, and consumption - the three monetary metrics that are typically used to measure poverty – have a very different content and meaning in these contexts. FDPs may also live in camps where shelter and services are provided entirely by the international community. In some very poor countries, FDPs living in camps may be better off than locals, but when it comes to measuring poverty among FDPs living in camps one has to reconsider how to measure poverty. For example, it is not obvious whether hand-outs such as food stamps should be considered as income or consumption, or how to quantify housing, education and health services provided in camps free of charge.

This paper discusses the emerging new literature on FDPs, issues related to the measurement of poverty among FDPs, and future prospects for this new area of study.

A New Field of Study

Economics has traditionally paid very little attention to Forced Displacement (FD), generally considering this issue a sub-set of migration studies. Partly because of lack of microdata, partly because of the small size of the population studied, and partly because refugees were relegated to the sphere of humanitarian issues, economists rarely invested in researching this topic. This changed in the aftermath of the Syrian displacement crisis.

Consider the literature on the measurement of the impact of forced displacement on host communities. This question has been the main preoccupation of European and North American countries for decades, and the first studies on forced displacement were largely focused on this question. A seminal study in this respect is David Card's 1990 study on the impact of the 1981 Mariel boatlift of Cubans from Cuba to Miami (Card 1990), a study that focused on the impact of the newly arrived Cubans on the wages and employment of the local Miami population. Despite this interest, between 1990 and 2010, only an average of one article per year appeared in economics journals on this topic and none of these articles measured the impact on poverty of refugees or host communities. A few studies on poverty among FDPs appeared before 2010, including a few studies on the well-being of Palestinian refugees (Hejoj 2007), and occasional studies on refugees in upper income countries (Bollinger and Hagstrom 2004, Kriechbaum-Vitellozzi and Kreuzbauer 2006), but these studies were extremely rare and relegated to specialized journals, or specialized agencies' reports. This changed after 2010 with tens of studies now published on refugee poverty and household well-being of host populations in top economics journals (Verme 2021).

The surge in general interest on the topic, the media exposure of the Syrian and European crises and the growing preoccupation of donor countries for the political and economic implications of refugees, led development organizations and poverty specialists to start considering FDPs more closely and collaborate with humanitarian organizations. The UNHCR and the World Bank, for example, started to collaborate very closely around 2014 with the first joint study on the welfare of Syrian refugees in Jordan and Lebanon (Verme 2016). This study used UNHCR's administrative and survey data to measure poverty among refugees contributing to improve the targeting system of the UNHRC's cash assistance program. It found that - using the international extreme poverty line of 1.9 USD/day in Purchasing Power Parity (PPP) - 7 in 10 refugees were poor, a figure that rose to 9 in 10 refugees if the national poverty lines of the host countries were used.

This seminal study led the World Bank to reconsider its role in working with refugee populations and encouraged this organization to strengthen cooperation with the UNHCR leading to the establishment of a joint research program ("Building the Evidence on Forced Displacement"), country level cooperation on data collection in several countries, a first round of analytical studies on refugees' well-being (Pape and Mistiaen 2018, Cuevas, Kaan et al. 2019, Pape 2019a), the establishment of a Joint Data Centre between the two organizations, and the implementation of joint rapid poverty assessments for FDPs in the course of the COVID-19 pandemic (see World Bank 2021 for example). This collaboration was instrumental in equipping the World Bank with improved knowledge on refugee populations and the UNHCR with improved knowledge on data collection and poverty measurement among FDPs.

Better and more data and knowledge on FDPs, in turn, attracted significant investment in FD research on the part of donors generating a real boom in research on the topic which quickly became mainstream, even in disciplines that traditionally disregarded FD such as economics. By 2021, most of the top economics journals had published articles on FD including the Quarterly Journal of Economics, the Journal of Political Economy, and the Review of Economics and Statistics with special issues being published by development journals such as the Journal of Development Economics. This literature, in turn, is providing hard evidence on highly debated topics such as the impact of FD on host communities. A review of this specific literature has shown, for example, that most studies find a positive or non-significant effect of FD on hosts' employment, wages and household well-being, a finding that disputes much of the popular perceptions on this question (Verme 2021).

This new research area is also generating significant innovations with the potential to expand research methods in the poverty measurement field. In the area of targeting based on means-tests, for example, a study has shown that Receiving Operations Characteristics (ROC) curves can be an effective decision making tool for humanitarian assistance programs (Verme 2019) while another study found that poverty differences in prediction methods for targeting purposes among refugees are attributable to few data fields suggesting that refugee homogeneity can make poverty predictions and targeting easier as compared to regular populations (Altindag, O'Connell et al. 2021).The existence of the UNHCR refugee registration system, which can be regarded as a live census of refugees, has encouraged others to use cross-survey imputation techniques to estimate poverty among refugees even in the absence of income or consumption data (Beltramo, Dang et al. 2021, Dang and Nguyen 2021). The mobile nature of refugees and IDPs also lends itself to experimenting with new methodologies to measure poverty with alternative methods such as mobile phones (Blumenstock, Cadamuro et al. 2015, Pape, Baraibar Molina et al. 2020, Wieser 2021), or satellite imagery and remote sensing data (Abelson, Varshney et al. 2014, Neal, Burke et al. 2016). For example, night lights or type of infrastructures captured by satellite imagery can provide gross estimates of poverty. In essence, poverty measurement among FD populations benefits from decades of developments in the poverty measurement field, but also provides new opportunities to expand the field because of the atypical characteristics of these populations.

Measurement Issues

Lack of Microdata

One of the important factors that has limited research on FDPs in the past was the chronic shortage of quality microdata, something that is quickly changing. Today, microdata on these populations can be found in three main publicly available repositories: The World Bank microdata library, the UNHCR microdata library recently established in collaboration with the World Bank, and the Humanitarian Data Exchange (HDX), a data repository managed by the United Nations Office for Humanitarian Affairs (OCHA) on behalf of a consortium of humanitarian organizations. An analysis of the WB and UNHCR microdata libraries (as of 23 April 2021) finds 273 data sets related to refugees covering 73 countries with 249 of these data sets collected after 2011, whereas the HDX data repository shows 57 survey data sets on refugees, all of them dated after 2015. On IDPs, the situation is less encouraging as on-going efforts are still largely focused on counting IDPs rather than surveying them. The WB and UNHCR microdata libraries combined have only 37 data sets on IDPs, all administered after 2010, whereas the HDX shows 39 data sets, all of which have been administered after 2015. This shortage of microdata can be addressed if survey instruments specifically designed for FDPs are developed and the development of these instruments requires, in turn, addressing questions such as sampling, choice of unit of observation, unit of measurement and poverty line tailored to FDPs' characteristics. These are some of the major challenges that scholars and practitioners are currently facing.

Sampling strategies

There are some technical reasons that make surveying refugees and IDPs complex. Sampling is one of these reasons. Refugees and IDPs are mobile populations in that a significant share of these populations tend to move in the host location, even if they are initially residing in camps. When they settle outside camps, they often lack a formal address or reside at addresses nominally occupied by local residents. These factors make the inclusion of these populations in national master samples for surveys difficult. For refugees, this problem is partly overcome by the UNHCR registration system (proGres) which requires all registered refugees to provide a set of basic information including location and socio-economic characteristics of the household's members. However, using the proGres registration system to sample refugees has its own challenges. This system is available in some countries but not in others where a system may be missing or managed by the local authorities and not available to others. Where the system exists, many refugees are not registered, some are registered but their information is outdated or missing, while identifying the unit of observation (household, family, case) may be challenging. In other words, extracting a representative sample of refugees in a country, even when the proGres registration system is available, is not simple.

Sampling IDPs can be even more complex because they no longer reside in the place where they are registered in the census and master sample of their own country. International organizations that collect information on IDPs, such as the International Organization for Migration (IOM), tend to collect information on IDP communities rather than individuals or households. National registration systems that some countries put in place for the monitoring of IDPs do not collect comprehensive information on individuals or households, with only some exceptions such as Colombia. Security reasons also hamper proper sampling and data collection as many destination areas for IDPs are themselves unsecure locations. In essence, the sampling infrastructure that is typically available for national populations, and even the basic infrastructure that is available for refugees is lacking for IDPs.

In general, two classes of sampling approaches are used for FPDs: area-based sampling and list-based sampling. The area-based sampling partitions the area of interest into smaller enumeration areas. In the first stage of the sampling, enumeration areas are randomly selected. All households in selected enumeration areas are listed so that a given number of households can be randomly chosen from the list to be interviewed. The advantage of this approach is that it only requires knowledge of any FDP residing in an enumeration area, so that this person can be included in the sampling frame. The more accurate the number of FDPs in an enumeration area before listing households, the more efficient is the resulting sample. However, this approach requires appropriate enumeration area maps that can be easily defined through satellite images for camps,but might not be available across the country for FDPs living in host communities. Implementation can be expensive especially with limited knowledge about the number of FDPs in enumeration areas. However, area-based sampling has the big advantage of not requiring FDPs to register. In contrast, list-based sampling uses an existing list of all FDPs and randomly chooses a sample among them. If additional characteristics are available, like location or country of origin, the sample can be stratified. However, the sample will only be representative of registered FDPs. Hence, it is rarely used in the context of IDPs.

To partly address sampling issues, the UNHCR and the World Bank have been cooperating to use the UNHCR or national refugee registration systems as initial master samples to conduct consumption surveys, similarly to what is done with censuses of regular populations. Initial experiments in this respect have been conducted in Jordan, Lebanon, Iraq, Uganda, Kenya, Ethiopia, and South-Sudan resulting in consumption surveys and poverty analyses of refugees and IDPs. Both the UNHCR and the World Bank are now also working with statistical agencies in multiple countries across the Middle East and North Africa and Sub-Saharan Africa regions to include FDPs in national sampling frames with initial efforts conducted in countries such as Jordan, Kenya and Uganda (e.g. World Bank 2021).

However, the political economy and data privacy can make these efforts challenging. For host countries, FDPs may be only one of the many marginalized groups they may be concerned with. With limited budgets, statistical agencies need to justify prioritizing one group above another. Furthermore, data privacy is particularly relevant for FDPs. National sampling frames are constructed based on census information, including personal information such as addresses, phone numbers, names as well as GPS locations in some cases. Given the protection mandate of UNHCR, it is not obvious whether and how this information can be shared with a national statistical agency. At the same time, national statistical agencies protect the national sampling frame and cannot share, for example, the underlying cartography that would be required for UNCHR to amend the sampling frame with counts of FDPs. Even if both agencies would be able to create a trusting relationship allowing close collaboration, questions remain on whether FDPs would need to agree with sharing their information with a Government agency. This becomes even more relevant in the context where a Government might be a real or perceived contributor to displacement or is harboring an overt or covert policy to reduce the number of refugees in the country.

With limited results in including FDPs in national sampling frames, alternative approaches remain necessary. The UNHCR, the World Bank and numerous scholars and practitioners worldwide are experimenting with satellite images and phone surveys to try detecting refugee and IDPs populations that may escape the UNHCR and national registers with some initial encouraging results. In Lebanon, Jordan and the Kurdistan region of Iraq, Aguilera, Krishnan et al. (2020) designed sampling strategies for Syrian refugees with known ex-ante selection probabilities. They used a variety of data sources, including data collected by humanitarian agencies, and also employed geospatial segmenting to create enumeration areas where they did not exist. Systematic field experiments are also underway to test different sampling approaches for IDPs living in camps. For example, Himelein, Pape et al. (forthcoming) compare the performance of five alternative sampling approaches (satellite mapping, segmentation, grid squares, "Qibla method," and random walk). Different indicators are assessed including household size, consumption, poverty and ownership of assets. Using empirical evidence from a field experiment in an IDP camp in South Sudan, the total survey error of each sampling approach is compared to a census, allowing for the disaggregation of the total error into sampling and non-sampling components. One of the main findings is that the implementation of all approaches suffers from over-estimation of household sizes caused by a systematic tendency of enumerators to select larger households because they are more likely to find an adult respondent. Such studies with a focus on ground truthing remain important to validate sampling approaches for displaced communities.

Sampling FDPs outside camps faces a different challenge, especially when a list-based approach is not possible. FDPs can be spread in low density across a larger population requiring an area-based adaptive sampling approach. The approach continues to list additional enumeration areas systematically selected, e.g., by selecting neighboring enumeration areas, until a sufficient number of FDPs are covered. Adaptive sampling creates analytical complexities for the calculation of sampling weights. It also adds to the cost as more enumeration areas must be listed – often with only very few or possibly no FDPs present – and it is more difficult to implement given the uncertainty about the number of required enumeration areas that need to be visited to satisfy a given sample size.

Finally, fatigue from over-surveying is often cited as an anecdotal challenge in the context of FDPs, leading to survey non-response. Even though FDPs are often subject to intensive surveying by multiple agencies, survey non-response is usually less of a problem than for regular populations given lower opportunity costs to participate as well as the expectations of indirect benefits. In addition, the likelihood for a household to be interviewed multiple times is limited. Only censuses interview everyone in the population. However, they are often prohibitively expensive and suffer more from low data quality outstripping their advantage of large sample size. Thus, only few censuses – with very short questionnaires – are necessary for verification exercises related to direct aid delivery. In contrast, surveys usually have limited sample size around a few thousand participants, which is only a small fraction of the local FDP population, which often is in the 10s or 100s thousands. Hence, the chance of multiple interviews in a short period of time is low.

Unit of observation

The definition of the unit of observation in a survey (household, family, case, individual) is a crucial choice that has direct impacts on the measurement of poverty because it determines how household income, consumption or expenditure is measured in per-capita terms, and has implications for other indicators of well-being such as housing, rents, or crowding. The Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS) and Living Standards Measurement Surveys (LSMS) are the most common household surveys used to produce comparative statistics on well-being across time and countries. The DHS and MICSs define household members as (i) usual residents or people who slept in the dwelling the previous night and who (ii) share living arrangements and (iii) share food (IFC 2012, UNICEF 2013). The LSMSs define household members as (i) people who slept in the dwelling three or more months of the last 12 months and (ii) share food (Grosh and Munoz 1996).[51] While the definitions differ, they are defined on similar concepts and are unlikely to lead to major differences in key household characteristics.

The UNHCR has definitions for household that resemble the definitions used by MICSs and LSMSs but uses the concept of "case" as unit of observation. The UNHCR defines a case as: "A processing unit similar to a family headed by a Principal Applicant. It comprises (biological and non-biological) sons and daughters up to the age 18 (or 21) years, but also includes first degree family members emotionally and/or economically dependent and for whom a living on their own and whose ability to function independently in society/in the community and/or to pursue an occupation is not granted, and/or who require assistance from a caregiver." (Verme 2016) This definition is different from the DHS, MICS and LSMS, may pose challenges when one compares cases with households in surveys, and lends itself to exploitation on the part of users, for example by spreading different household members across different cases to maximize benefits.

Some socio-economic surveys for FDPs do not rely on the UNHCR family definition as the unit of socio-economic analysis – in addition to the individual. Instead, they employ a household definition either from DHS, MICS or LSMS, or from the established national household survey implemented by the national statistical agency, which often is similar to the traditional definitions. When using a list-based sampling approach with UNHCR's registration data, this creates the challenge of translating the family (case) definition used as a sampling unit to a traditional household definition for the interviews and the analysis. In most cases, not all members of a proGres family are members of the household, while the household usually also has members from other proGres families. Different approaches can be used to overcome this challenge.[52]

First, the sampling can be based on individuals rather than families. By definition, each household with at least one FDP will have a positive probability to be part of the sample creating a representative sampling frame. Since larger households have a higher chance of being sampled, sampling weights need to correct for that. The sample will be representative but not very efficient as estimates for smaller households have lower accuracy than larger households.

Second, the sampling can be done based on proGres families. In this case, the interview must determine the household for each proGres family member. If family members in one proGres family are from multiple households, the household to be interviewed must be selected randomly among them. The household interview must include all household members, also if they are not part of the proGres family. Since the sample is drawn randomly among all proGres families, all households with FDPs have a positive likelihood to be selected. However, a household with members from several proGres families has a higher chance to be selected. Hence, sampling weights must be adjusted accordingly by the number of members from different proGres families.

A correct implementation of the household definition is crucial even though difficult. Enumerators and respondents can reduce interview time by reporting fewer household members. However, household-level information like assets or consumption will usually still be reported across the

[51] While the DHS and MICS define households mutually exclusive, the LSMS definition suffers from individuals potentially belonging to multiple households leading to double-counting.

[52] If not explicitly mentioned, we assume that the universe of FDPs is defined as FPDs registered in proGres.

household. Hence, households will appear richer as consumption and rooms are shared by fewer reported household members. It is important to note that this is a more important challenge than the discussed bias of area-based sampling methods in camps to select larger households. Selection of larger households creates a bias by moving the average estimate towards larger households, e.g. increasing the average poverty rate if larger households tend to be poorer. However, a misreporting of the number of household members biases the characteristics of households, often making them appear richer than they are as resources are shared by fewer reported members.

Finally, mixed households consist of FDPs and non-FDPs. All household members independent of their displacement status must be equally considered in the interview, to allow accurate estimates for household-level indicators that depend on household size. Mixed households are usually more prevalent outside camps and often exhibit distinctly different characteristics, e.g., in terms of deprivation and labor market access. Comparing camp and non-camp households must consider the presence of mixed households outside camps. Results indicating that non-camp households are less deprived and have better access to labor markets might simply be driven by the presence of non-FDPs in non-camp households.

Survey design and administration

Designing a socio-economic survey including consumption measures for non-displaced households is challenging as the interview time is limited and one has to be very selective to restrict the number of questions in surveys. For FDPs, the challenge is exacerbated given the need to understand not only their present socio-economic well-being, but also their displacement trajectory as well as their aid dependency. In this context, spending 90 or more minutes on an elaborate consumption module might not be a priority. Instead, detailed information on displacement and aid is crucial and something that is not properly assessed in non FDPs surveys.

FDPs are not only particularly vulnerable and need specific protection, they also often originate from a different region in the case of IDPs or country in the case of refugees. This creates several challenges in interviewing FDPs. The respondent might not speak the official language of the country. This can be overcome by assigning purposefully enumerators to specific households to ensure that they speak the same language. However, enumerators should also have a similar cultural background as respondents. Di Maio (2020) use a large-scale experiment in Uganda in which enumerators and respondents are randomly paired to explore for which types of questions a significant enumerator effect may exist. While the enumerator effect is minimal for many questions, it is large for specific perception questions, for which it can account for over 30 percent of the variation in responses. Such a bias can also occur for sensitive questions to FDPs including questions on their displacement trajectory, current needs as well as perception questions.

Specifically, in the context of a statistical agency conducting a survey including FDPs, enumerators need to be carefully selected. Refugees might see the statistical agency simply as Government that might pass information to other Government agencies potentially threatening their livelihoods. IDPs in the context of an internal conflict involving the Government might similarly be reluctant – or even feel personally threatened – if interviewed by a Government official. These risks can be mitigated if enumerators are hired among FDPs from the same background. However, this can create legal challenges if FDPs do not have work permits. In some cases, it might also be necessary to have UNHCR or IOM officials accompanying enumerators or even conducting the interviews themselves..

The implementation of FDP-specific socio-economic surveys can be done in parallel with UNHCR's verification exercises. The verification exercise is a census that visits all registered refugees to update their data. In Kenya, an additional socio-economic questionnaire was administered to a random subset of refugees participating to the verification exercise (Pape, Beltramo et al. 2019b). The parallel implementation reduced costs substantially and improved data quality as interviews were conducted on the spot rather than having enumerators search for addresses, which can be complicated in large camps. However, the implementation requires close coordination between the verification exercise and the socio-economic survey. Ideally, no redundant data would be collected but in reality most questions from the verification exercise are asked slightly differently than in the socio-economic survey, which usually aims to be consistent with the latest national survey, necessitating different instruments for the same concept to ensure comparability.

Poverty measures

Poverty measures the level of deprivation in a population. The most widely used concept defines as poverty headcount the number of individuals living below a threshold, usually called the poverty line (Foster, Greer et al. 1984). Since the poverty headcount only reflects the proportion of poor people, it is usually complemented by a measure of the depth of deprivation, the poverty gap, which estimates the average gap between poor individuals and the poverty line. All poverty measures constructed in such a way are based on an underlying welfare metric. In most cases, the metric is defined at the household level and needs to be transformed into an individual measure. The transformation either divides household welfare by household size providing a per capita measure or, in a more sophisticated way, takes into account differences in household composition (usually age and gender) leading to a per-adult equivalent measure.

To measure household welfare, different metrics have been proposed and can generally be classified either into monetary or non-monetary metrics. Monetary metrics equal levels of well-being to a monetary indicator of utility (Samuelson 1974), usually income, consumption or expenditure.. Non-monetary metrics define normatively dimensions of deprivations, e.g., access to education and clean water, and aggregate them either by using dimension-specific thresholds or an aggregated threshold. The most commonly used approach in this class is the Multidimensional Poverty Index by UNDP and Oxford University (OPHI 2018). While it is not uncommon to find experts with a strong preference for one measure over the other, both classes of poverty measures are in most cases complementing each other. While the monetary metric has the advantage of its theory-grounded definition without normative choices, it is complex to measure and does not capture acute deprivations in health or education, as the OPHI MPI does. For example, a third of those experiencing multi-dimensional poverty are not captured by the monetary headcount ratio (World Bank 2020). Hence, it is not surprising that efforts are underway to combine the advantage of both measures like the World Bank's Multi-Dimensional Poverty Measure (Nguyen, Wu et al. 2021).

In the case of FDPs, monetary and multi-dimensional measures of poverty are both indicated and, as it is often the case in other settings, they complement each other. In this respect, there is not much difference with other types of populations. However, the money metrics used for monetary indicators may be more challenging to estimate for FDPs than regular populations, and the composition of multi-dimensional indexes should be adapted to FDPs characteristics.

Moreover, especially in a context of very high deprivation, a poverty headcount might not be able to reveal the actual gravity of the situation and qualitative information should be sought to complement standard questionnaires and possibly inform the development of future questionnaires. Collection of socio-economic data is a passive process where respondents are asked pre-formulated questions. This constrains the respondents in sharing their own narratives and emphasizing what they feel is important. The implementation of household surveys can be seen as an opportunity to collect qualitative information and transform a one-sided narrative into one that provides voice to the poor. Pape (2020), for example, describes how short, voluntary video testimonials with informed consent from people living in South Sudan and Somalia[53] can be used to empower the poor in voicing their own needs. While this is a practice that can be developed in any context, it is particularly promising in contexts where measuring poverty is still relatively new and the specific challenges associated with this measurement are still little known, as in FDPs contexts.

Poverty metrics

Income, consumption or expenditure can be used as monetary metrics to estimate poverty. The correct measurement and classification of income can be a real problem in low-income countries with large informal economies (Deaton and Zaidi 2002) and, especially, for FDPs. FDPs have rarely regular labor income and tend to rely on occasional, informal income, or have no labor income at all. They have various forms of in-kind and cash assistance that vary from household items such as blankets and kitchen items to food vouchers and cash assistance. The combination of these income sources is not always well captured in surveys because many of these items are provided occasionally and by multitudes of donors. Food vouchers are often traded and can function more as cash than food. FDPs may also produce products that they exchange as they try to use their crafts to supplement incomes, and some of the items produced may be consumed. Many FDPs store wealth in jewelry or gold to carry these values with them, and it is unknown how much of this wealth is used for regular or occasional expenditures. These facts make the estimation of income for FDPs rather hard so that most often consumption-based measures are used.

However, measuring consumption among FDPs has its own challenges including misreporting and questionnaire fatigue. It relies on a list of all potential products consumed by a household. For each consumed item, the household needs to provide specific information on the consumed quantity. Usually, additional questions are asked about the amount that was purchased including the outlet and price, to be able to estimate the monetary value of consumption. However, FDPs have a strong incentive for misreporting if they are aid-dependent and believe that their responses affect future aid. Kaplan, Pape et al. (2018) find in numerous rounds of data collection in Somalia and South Sudan that IDPs report significantly lower levels of consumption than non-IDP households. For example, 45 percent of Somalia IDP households report food consumption below subsistence levels and approximately 80 percent below recommended levels. While the data may be accurate, there are two reasons to suspect that it is not. Such high levels of below-subsistence consumption would be associated with high rates of mortality due to starvation, which is not borne out in mortality data. Kaplan, Pape et al. (2018) use a randomized experiment to test the effectiveness of a bundled nudge including a truth primer and more investigative reporting controls. They find that the bundled nudge induces higher reporting in lower quintiles of reported consumption. This treatment pattern is driven by aid reliant IDPs and vanishes when considering the comparison group of non-IDPs. This suggests that IDPs – and possibly more generally aid-dependent populations – are indeed misreporting. However, the study does not yield a 'true' estimate of consumption, which is needed to assess the overall level of misreporting and the different approaches that could be used to mitigate misreporting.

Questionnaire fatigue is another important issue. Administering consumption modules often takes more than 90 minutes and can require multiple household visits. Particularly in a deprived setting like an IDP or refugee camp, it can be challenging for enumerators to justify spending so much time on asking about items that the respondent cannot afford. However, reducing the number of items either by aggregating or by removing items will bias estimated consumption (Beegle, De Weerdt et al. 2012). This explains the frequent use of imputed consumption methodologies.

[53] www.thepulseofsouthsudan.com and www.thesomalipulse.com.

Scholars have experimented with alternative methods to do that while striving to maintaining accuracy. For example, Pape (2021) refines a methodology (Pape and Mistiaen 2018) that combines an innovative questionnaire design with standard survey imputation techniques without relying on a previous baseline survey. Consumption items are partitioned into modules. Each household is administered only one randomly assigned module, creating significant time savings and making it possible to administer a full questionnaire including household and individual questions in less than 60 minutes. The full consumption estimate is obtained by imputing the deliberately absent consumption values for items that are not explicitly asked for a specific household, but was administered to other households. The methodology makes it possible to derive poverty estimates without compromising the credibility of the resulting estimate, and it performs considerably better than alternative approaches based on reduced consumption aggregates and cross-survey imputations. The methodology has been widely applied especially in the context of FDPs in Ethiopia, Kenya, Nigeria, Somalia, South Sudan and Sudan (Pape 2019a, Pape, Beltramo et al. 2019b) to allow for shorter interview time, creating space for in-depth displacement-specific questions while reducing enumerator and respondent fatigue.

An alternative way to obtain accurate consumption estimates without limiting the time of the interview can be borrowed from small-area estimation methods (Elbers 2002, Molina and Rao 2010). This approach administers the full consumption module only to a small subset of households, while the majority of households do not receive any consumption questions. Instead, within-survey imputations are used to impute consumption for this part of the full sample. This approach avoids the pitfalls of cross-survey imputations while improving accuracy as compared to assigning subsets of consumption items across all households.[54] However, the subset of households with full consumption modules might suffer from questionnaire fatigue increasing measurement error, which can ultimately exceed the model error in approaches like Pape (2021).

After consumption data is obtained, questions arise on how to value consumption. FDPs often receive aid in the form of in-kind transfers as well as cash vouchers. The standard consumption questions will capture consumption of in-kind transfers like food as well as goods and services that were purchased using cash vouchers. Usually, market prices are used to value consumption. However, market prices in a camp can be distorted or absent for specific products if they are mainly bartered. Traders accepting food vouchers might add a premium on prices given their monopoly power in accepting food vouchers. The premium will be reflected in a higher value of consumption leading to higher welfare, which is – of course – grossly misleading. More generally, households in camps can face significantly different prices than households outside camps. In non-FDP contexts, a similar challenge occurs between rural and urban households or other regionally disparate populations facing different prices (e.g. Ravallion and Lokshin 2006, Boom, Halsema et al. 2015). Deflators can be used to adjust for such price differences. However, they are sensitive to methodological choices (Ravallion and Benu 1994) and, usually, require large amounts of high-quality data on quantities and prices.

If income or consumption data for FDPs are missing from the data at hand but available in other data representative of the same population, one can also use cross-survey imputation techniques to estimate consumption using proxies of well-being. This method uses a baseline survey inclusive of consumption data to model consumption using a regression model inclusive of easily measurable household characteristics as independent variables. The estimated coefficients from the model are then used to predict consumption for FDPs using census data (usually FDPs registration data) that lack

[54] Accuracy improves compared to Pape (2021) because distributing consumption modules across households ignores correlation between consumption modules in the imputation. The small-area estimation approach, in contrast, estimates full consumption for a subset of households automatically considering any within correlations.

consumption but include the same household characteristics used with the baseline survey model. This method has been tested with refugees in Jordan and Chad (Beltramo, Dang et al. 2021, Dang 2021) providing encouraging preliminary results. These works showed that accurate poverty estimations for refugees can be obtained with a relatively small number of proxies of well-being which are usually already available in the UNHCR proGres registration system. The quality of the imputed poverty estimates depends on the similarity of the population surveyed at baseline and the population used for imputations. Both populations might differ because of the time passed between baseline and imputation survey, and because of different characteristics if both surveys are not representative for the same populations. It is therefore essential for these types of exercises to use survey and census data from the same time period and with very similar predictors.

The treatment of aid is particularly challenging. By definition, aid is included in consumption aggregates and, hence, will reduce estimated poverty among FDPs. This provides a fair assessment of the current situation and is particularly helpful in the context of comparison with host communities to ensure that they are not receiving less assistance while being more deprived. However, it can also easily be misinterpreted that FDPs do not require assistance as they are not poor. To avoid this misinterpretation, a second consumption estimate can be constructed that excludes aid from consumption to reflect livelihoods of FDPs in the absence of aid. A precise estimate of consumption excluding aid is not easily feasible. Even if consumed quantities for each item are split into aid vs. non-aid, in-kind assistance that is sold would make attribution difficult, while an income-focused angle would have similar problems as discussed before. A light-handed approach that asks for each item whether it was mainly provided for free can produce a consumption aggregate excluding aid that can be useful in re-thinking targeting of aid (Pape and Sharma 2019).

Poverty lines

How to establish a relevant poverty line for FDPs is another important question. The choice is between international poverty lines such as 1.90 USD/day at PPP value (Ravallion, Chen et al. 2009), national poverty lines of host countries or contextual poverty lines specific to FDPs. International and national poverty lines can be considered after the use of appropriate deflators. These choices can be controversial for regular populations and carry additional complexities when FDPs are considered.

One question is how to adapt poverty lines for FDPs living inside and outside camps. FDPs in camps often receive shelter, education and health assistance that is not – or not in the same way – provided to the population residing outside camps. This can create a paradox that refugees that pay rents in urban areas would appear much richer than refugees living in camps as their shelter is free of charge and not necessarily accounted for in the consumption aggregate. In traditional poverty estimations, a similar challenge occurs when one considers poverty lines for urban and rural populations since the rental market is often absent in rural areas making it impossible to estimate or impute rent. To allow for the inclusion of rent for urban populations, separate urban and rural poverty lines are estimated, called contextual or regional poverty lines. Only the urban poverty line and its corresponding consumption aggregate would include rent.

Similarly, other services like education and health are extremely hard to value given that cost data for camps' services are incomplete and grossly available only in aggregate form. Hence, they would also need to be excluded in the contextual poverty line for camps. Given the difficulty to estimate contextual poverty lines, however, this approach is usually not feasible in FDP contexts, while it also adversely affects the comparability of estimated poverty rates. Furthermore, a comparison between camp and non-camp population would ignore differences in housing, health and education between these populations, substantially limiting the value of a comparison. These questions are still largely unresolved and the search for poverty lines adapted to FDPs contexts is still in its infancy.

Comparisons across time and populations

Understanding how poverty among FDPs evolves over time and how this compares to other populations is another essential pillar of poverty measurement.

Tracking FDPs over time can be particularly helpful to monitor changes in livelihoods and changes in displacement status, location as well as the end of displacement.[55] Panel surveys in this respect are useful instruments for FDPs as they would allow to track the same individuals over time and understand their livelihood and residency trajectories. However, classic panel surveys typically rely on home addresses and they are particularly difficult to administer when populations are highly mobile. This has encouraged scholars working on FDPs to develop new instruments to track people over time. Etang and Himelein (2020) developed a survey known as the "Listening to Displaced People Survey (LDPS)", a survey that tracked living conditions of displaced people over time in Mali with a face-to-face baseline survey complemented by monthly follow-up mobile phone interviews for a period of 12 months. These data have been used by Hoogeveen, Hoogeveen, Rossi et al. (2019) to study patterns of return of the displaced and understand the factors that contribute to return.

Phone interviews have also increased in popularity with the COVID-19 pandemic, which made it necessary to conduct interviews without face-to-face contact. During this period, the UNHCR and World Bank have launched bi-monthly monitoring surveys of the impact of COVID-19 on the well-being of refugees in several countries across the MENA, SSA and Latin American regions using phone interviews. These resulted in panel surveys that now offer the possibility to assess the impact of COVID-19 on refugees over time and across countries in a comparable manner. Vintar, Beltramo et al. (2022), for example, provide an example of how to use these data to understand the differential labor impacts of COVID-19 on refugees and non-refugees. UNCHR's proGres database includes phone numbers for refugee family heads that can be utilized as a sampling frame. However, data privacy concerns need to be addressed if the phone survey is conducted by a firm. An easily implementable solution is sending text messages to selected respondents through UNHCR to ask for permission to share phone numbers with a contractor.

Comparisons between FDPs and host populations is also an essential exercise to conduct in the context of FDPs poverty measurement. These comparisons are important for FDPs, host country and international organizations given that resentment against FDPs is often fueled by a perception that FDPs receive special assistance that is not available to locals. Several surveys have now been conducted in Jordan, Lebanon, Iraq and a few Sub-Saharan African countries to compare the well-being of FDPs and their hosts. These comparisons, while important, are complex because host populations have full access to the labor and consumer markets and government services that are often not available to FDPs, whereas FDPs rely on aid from the international community that is not available to local residents. It is difficult to compare health services in camps, for example, to those provided to the host population by the government, or other social protection services such as unemployment insurance or paid leave that do not exist for FDPs. Again, these are new and largely under-researched issues among poverty specialists.

Future prospects

There are now a wide variety of survey instruments that are being designed for or adapted to measure income or consumption among FDP populations. Home visits initially designed by the UNHCR to question FDPs for protection purposes are being revised to ask questions on income and consumption becoming in this way viable instruments for poverty measurement. Large home visits exercises are conducted every year in countries such as Jordan and Lebanon and these data have been used to

[55] While IOM uses the Displacement Tracking Matrix (DTM), it only provides an aggregated view of the location of FDPs.

conduct poverty assessments of Syrian refugees (Verme 2016). The WFP conducts vulnerability assessments that have been used by the WFP, UNHCR and World Bank to make gross poverty estimates using the consumption modules of these surveys, even if these modules are typically very short, with few items. The UNHCR conducts Multi-sector Needs Assessment such as the one conducted in Cox's Bazar, Bangladesh in 2018, Socio-economic assessments such as the one conducted in Zimbabwe in 2017, or nutrition surveys such as the one conducted in Tanzania in 2017.[56] All these surveys contain some information on income, consumption, or expenditure that is being used to assess well-being of FDPs. The World Bank has conducted welfare assessments of Venezuelans in various Latin American countries including Colombia, Peru and Ecuador with a forthcoming study expected for Chile. The welfare of Afghans refugees has now been studied in their main host country Pakistan and in Afghanistan upon return providing elements to compare the living conditions of this population in the two locations. Again, all these surveys and studies are either very recent or were not used before for poverty/welfare analyses.

Whereas individual and household data on refugees are now systematically collected, data collection for IDPs remains extremely scarce when compared to refugees, mostly limited to head counting. Refugees and IDPs are very different in that they have a different legal status (asylum seekers vs. citizens), which leads to different access to public services, labor markets, government and international assistance. Surveys for IDPs are often more difficult to conduct because the host country might be linked to the cause of displacement, and a registration system like UNHCR's proGres is absent making it more challenging to obtain a representative sample. Given the large number of IDPs (58 percent of all FDPs), the real challenge will be to collect microdata on IDPs systematically in all those countries that are home of large numbers of IDPs. So far, the only country that collects data on IDPs systematically is Colombia and this country has shown that, when quality microdata are available, research on IDPs flourishes.

While many of the discussed issue need more attention and research, some processes are in place that should lead to the establishment of guidelines that address some of these questions. A recent process initiated by the United Nations led to the preparation of guidelines for refugee and IDPs statistics including poverty statistics. The United Nations Statistical Commission (UNSC) established in 2016 an international Expert Group on Refugee and Internally Displaced Persons Statistics (EGRIS) made of national and international organizations and several expert statisticians. This process has produced two main documents: The International Recommendations on Refugee Statistics (EGRIS 2018), and the International Recommendations on IDP Statistics (EGRIS 2020). The first document recommends measuring the proportion of population below the international poverty line among the social inclusion indicators (indicator 1.1.1), and the second document recommends measuring the proportion of population living below the national poverty line among the livelihoods and economic self-reliance indicators (indicator 1.2.1). While not specifically focused on poverty, these documents recognize the importance of measuring poverty among FDPs and provide general indications on the indexes to measure and the data necessary to measure such indexes. As already mentioned, the UNHCR has started to introduce standards for the measurement of well-being of FDPs while the new World Bank-UNHCR Joint Data Center established in Copenhagen should produce standards for data collection and questionnaires. These processes have also encouraged other organizations such as WFP and IOM to develop their own standards for surveying FDPs and measuring various well-being indicators.

[56] All these surveys can be found in the World Bank microdata library at: https://microdata.worldbank.org/index.php/catalog?sort_by=rank&sort_order=desc&sk=refugees

Conclusion

The paper provided a first insight into the state of the literature on the measurement of poverty among FDPs. We argued that the economics profession and poverty specialists across the social sciences largely neglected these populations for a combination of factors including lack of interest and microdata. This changed with the beginning of the Syrian conflict in 2011 and the peak of the European migration crisis in 2015. These events have generated partnerships between development and humanitarian organizations that contributed to boost microdata collection among FDPs, poverty and welfare studies worldwide, and is also sparking a virtuous cycle of poverty measurement innovations. As FDPs require special solutions to questions such as sampling, consumption measurement, and targeting, this search is also producing innovative approaches that can serve the poverty measurement community at large.

Poverty measurement for FDPs remains at a very early stage. One should remember that the global poverty measurement spearheaded by the World Bank in the 1980s has required several decades to develop into a structured system of microdata collection at the country level and a theoretical and empirical body of knowledge that could make use of these data. Microdata collection for FDPs can build on this infrastructure but also requires its own specific surveys, measurement methods, and theoretical and empirical adjustments. Multilateral organizations such as the World Bank and UNHCR have started this process and this has encouraged the research community to follow but this is only the beginning of a long process.

4. Second Stage Sampling and Non-Sampling Errors: IDP camps in South Sudan[57]

Kristen Himelein, Utz Pape and Michael Wild[58]

Introduction

The most common sampling approach for cross-sectional household surveys in the developing world is a stratified two-stage design (Grosh and Munoz 1996). Following stratification based on administrative boundaries, clusters are selected in the first stage with probability proportional to size from a national census-based frame. In the second stage, a canvasing operation is conducted in the selected clusters to compile an updated list from which households are randomly selected. While this methodology is straight forward to implement in the field and reliably produces unbiased estimates, there are several downsides.

The first downside is cost. The World Bank's Living Standards Measurement Study team, which provides technical assistance on large-scale household surveys around the world, estimates the field listing operation increases the overall budget for data collection by 25 percent. Due to confidentiality concerns, the data collected during a field listing operation, typically the name of the household head and address or location description of dwellings, does not have any analytical applications beyond as a component of the weight calculations. At a time when typical surveys costs are in the USD millions, reducing a significant cost component will increase the financial sustainability of data collection.

The second drawback to the traditional design relates to timeliness. At a minimum, listing operations are usually conducted several days, if not several weeks, before the main fieldwork. As populations shift, the quality of the list degrades as time passes. While this is generally not a major concern for static populations living in villages or cities, it is a major concern for those in IDP (Internally Displaced People) and refugee camps. The transient nature of such environments implies building an accurate sampling frame is a complicated process often fraught with inaccuracies. Structures, often tents, for example, can easily be enlarged or split, quickly changing the layout of the camp, potentially invalidating a pre-existing sampling frame.

There are also issues related to the subjectivity in a listing operation. Eckman (2013) found only an 80 percent overlap between the same blocks listed separately by different interviewers in the United States. Undercoverage during the listing operation impacts the representativeness of the final estimates if the undercoverage is non-random. For example, O'Muircheartaigh, English et al. (2007) showed undercoverage in the United States is higher in low-income and rural areas. If this finding extends to the developing world, poverty numbers may be underestimated. In addition, Barrett, Beaghen et al. (2002) find higher undercoverage of households occupied by non-Hispanic black respondents compared with non-Hispanic white or other race respondents. This potential bias introduced by racial differences between the interviewer and respondent is of particular importance in the developing world context when interviewers are often recruited in the capital city and sent to more remote regions for the survey.

This paper builds on the work done by Himelein, Eckman et al. (2017) in describing five alternative sampling approaches considered for a household survey in Mogadishu (satellite mapping, segmentation, grid squares, "Qibla method," and random walk). In that paper, however, the authors used simulations which assumed perfect implementation. Therefore, while it was possible to compare

[57] UP and KH contributed equally to the manuscript. MW provided analysis and write-up for the random walk simulations.

[58] Authors in alphabetical order. All views are those of the authors and do not reflect the views of their employers including the World Bank or its member countries. The authors would like to thank Mario Gronert, Luca Parisotto and Marina Tolchinsky for providing research assistance and Toan Do, … for comments on the project proposal and earlier drafts of this paper.

the sampling error of the five methods, it was not possible to consider non-sampling error. This paper goes a step further by using simulations to describe the sampling error and a field experiment in an IDP camp in South Sudan to measure the total survey error of each design compared to a census, allowing for the disaggregation of the total error into sampling and non-sampling components. In addition, we attempt to separate the components of non-sampling error linked to the sample method from those common across all methods, such as interviewers selecting larger households and other issues in properly implementing the household survey protocols.

The next section briefly describes each method and highlights the literature as it relates to the relevant selection methods. Section 3 describes the data set and protocols for each method included in the experiment, followed by Section 4, which discusses implementation issues. Section 5 reports the results of the analysis, and section 6 concludes with further discussion of the overall performance and areas for future research.

Description of methods

This paper compares five alternatives of second stage selection (satellite mapping, segmentation, grid squares, “Qibla” (or “walk north”) method, and random walk) to a human canvasing operation. We consider the human canvasing to be the gold standard of listing methods, though acknowledge Eckman (2013) has identified the limitations mentioned in the previous section.

Satellite listing

While using satellite data to construct a sampling frame is common in land and agricultural surveys, household surveys are a more limited, though growing, application of the technology. In satellite listing, structures are identified from satellite imagery using either manual demarcation or an automated “computer vision” algorithm. Structures are then selected using simple random sampling, teams provided with GPS coordinates and maps to locate selected households. The main benefit of satellite listing is, if properly implemented, the results would match the precision of the gold standard of manual canvasing. The drawbacks include potential difficulties in identifying selected households due to the margin of error in the GPS machines, if structure identification is done using outdated maps, or if an automated model is trained on a context that is does not readily translate into the current application. In addition, imaging cannot always consistently distinguish between dwelling structures and accessory or commercial structures, leading to overcoverage issues. If a non-household dwelling structure is identified, the selected point is declared out-of-scope and a replacement is used. If there are a large number of these points, however, assumptions would be required to adjust the denominator of the probability of selection calculations or the resulting weights could be biased. Large numbers of out-of-scope structures were not anticipated to be an issue in South Sudan though since the camps were predominantly residential. Another potential issue with automated algorithms is undercoverage resulting from a failure by the algorithm to identify a structure, such as if the roof is constructed of an organic material not sufficiently distinct from the ground cover. While this problem can be mitigated by improved imagery, inaccuracies may still remain depending on the context in which the model was trained.

Examples from the literature include a study measuring disparities in health in Bobo-Dioulasso, Burkina Faso, in which Kassié, Roudot et al. (2017) used satellite images and the cadastral map of the town for random sampling through a supervised classification method. Escamilla (2014) used Google earth imagery and GIS software to manually digitize structures for a sampling frame for household survey in Lilongwe, Malawi. A random sample was then drawn from the list of households and interviewers used hand-held GPS devices to locate and interview households. A similar approach was used by Wampler (2013) for an ethnographic and water quality survey in Haiti. Specifically related to conflict, Lin and Kuwayama (2016) used high-resolution satellite imagery and manual identification to develop a sampling frame of man-made structures for their health survey in the Kerenik Camp in Darfur. Structures were then manually selected, and interviewers used hand-held GPS devices to navigate to the selected locations to conduct interviews.

The probability of selection for this method is simply the sampling fraction $\frac{j}{N_j}$ where j is the number of selected structures and N_j is the total number of structures. In the cases where it was necessary to select from multiple households within the dwelling, there would be an additional probability of selection for the household $\frac{i}{N_{ji}}$ where i is the number of households selected and N_{ji} is the total number of households in structure j. Therefore, the weight for this method can be represented as $w_i' = \frac{N_j N_{ji}}{ji}$. In the case of the experiment, the form simplifies to $w_i' = \frac{N_j N_{ji}}{j}$ as only one household was selected per structure.

Segmentation

Segmentation is a well-established practice of addressing primary sampling units (PSU) that are too large to list and can be done either prior to or after selection. Dividing large PSUs prior to selection is more efficient statistically because it keeps the selection to two-stages, but costlier, particularly if there are substantial numbers of large PSUs in the frame (Kish 1965, p. 156). In addition, this approach only works if a reasonably updated frame exists. Field segmentation is the more common approach, which allows the larger PSUs to be selected and then does the segmentation as part of the fieldwork. This approach is less costly as it only requires segmenting the selected clusters and can be used if unexpectedly large clusters are found in the field but does reduce statistical precision due to the additional level of selection. Regardless if segmentation is done pre- or post-selection, the segments should be approximately equal sized and boundaries should follow identifiable landmarks on the ground to facilitate accurate implementation by field teams.

Assuming field segmentation, the weight for this method is based on the probability of selection at each of the stages. Probability p_1 is the probability of selection of a PSU from the total number of PSUs N_k, or $\frac{k}{N_k}$, where k is the number of PSUs selected. Probability p_2 is the probability of selection of a segment from the total number of segments within the selected PSU N_{kl}, or $\frac{l}{N_{kl}}$, where l is the number of segments selected. Probability p_3 is the probability of selection of a structure from the total number of structures within the selected segment N_{klj}, or $\frac{j}{N_{klj}}$, where j is the number of structures selected. As above, there would be an additional layer of selection for households if there are multiple households within a structure, the probability for which can be represented as $\frac{i}{N_{klji}}$. The weight would therefore be $w_i' = \frac{(N_k)(N_{kl})(N_{klj})(N_{klji})}{klji}$.

Grid squares

The grid squares approach breaks selected areas down into smaller units for listing, but instead of manually drawing boundaries or using an algorithm, a uniform grid is imposed on the area. This approach can be used either for PSU, in which case it would be similar to segmentation, or to the area as a whole, in which case each grid square would act like a PSU. The benefit is a decrease in pre-survey preparation time, but at the cost of greater difficulties in implementation if grid lines do not follow landmarks, as well as greater difficulties in calculating weights for households overlapping grid squares

(Himelein, Eckman et al. 2017). Elangovan, Elavarsu et al. (2016)used a grid sampling methodology in the study of the health impacts of hard stone crushers in a residential neighborhood of Chennai. The authors found that 65 of the 300 selected grid squares were empty land, despite having excluded forests, bodies of water, etc. ex ante. In a mortality study in Iraq, Galway, Bell et al. (2012)used GIS and Google earth imagery for household sampling. The method used gridded population data for selection of clusters. The first cluster sampling stage of their study used the 'Create Spatially Balanced Points' (CSBP) function in the ArcGIS (v10) software.

Assuming the grid square method is applied to the area itself rather than a selected PSU, the weights for the grid method are similar to those for segmentation, where the cells are the PSUs, but without the additional step of selecting segments. The weights can therefore be represented as $w_i' = \frac{(N_k)(N_{kj})(N_{kji})}{kji}$.

North method

The "Qibla method" described in (Himelein, Eckman et al. 2017), or what is called in this paper the "North method" method, is an attempt to assign probability weights to random point selection methods. Several random point selection methods can be found in the literature, particularly in relation to epidemiological studies. Grais et al. (2007) used a methodology in which the closest household to a randomly selected point is selected for a study of vaccination rates in urban Niger, though did not attempt to calculate probabilistic sampling weights. Similar approaches were used by Kondo, Bream et al. (2014) in a study of the city of Sanitiago Atitlán, Kumar (2007) in urban India, and Kolbe and Hutson (2006) in Port-au-Prince, Haiti. Shannon, Hutson et al. (2012) also used such a method to select points in a study of violence in Southern Lebanon in 2008 but used the radius of a circle to define an area to be field listed, and from which buildings and then households were selected for enumeration. The circle area and building density were used to calculate probability weights. The main difference between most random point selection methods and the North Method described here is that the North Method attempts to accurately estimate the probabilities of selection.

To accurately calculate weights for the North Method, the area of possible random selection points (RSPs) leading to the selection of a structure must be measured or calculated. A structure is chosen if the RSP falls within it or if walking north from the RSP the structure is encountered. The area of all points from which one and the same structure is selected, its selection area, is made up of the structure and the *shadow* it casts to its south without interference of any other structure. Shadow here refers to the union of all points south of the structure that by protocol should lead to its selection (Figure B4-6).

Since a structure with a larger selection area is more likely to have a random RSP landing within it than one with a smaller area, weights are required for unbiased estimates. As all starting points fall within the camp area and the camp area itself is made up of the selection area of the structures, the weight given to an observation is proportional to the ratio of the selection area to the total area of the camp. The weights for the North Method therefore require the calculation of the area of valid RSPs that lead the enumerator to select the structure to determine its selection probability.

Let the selection area be labelled as A_i, then the weight is $w_i' = \frac{N_{ji}}{\left(1-\left(1-A_j/K\right)^n\right)}$, where K represents the sum over structures j of all areas A_j in the camp, and n is the number of RSPs. The inverse selection probability is multiplied by the number of households N_{ji} in structure j if there are multiple households within the same structure. See Särndal and Wretman (2003) and Himelein, Eckman et al. (2017) for further discussion.

Random walk

Random walk (or random route) surveys are extremely common in the developing world as a method to control costs when representative sampling frames are not readily available. These designs, however, are non-probabilistic and have been shown in the literature to generate biased estimates even under perfect implementation (Bauer 2014, Bauer 2016, Himelein, Eckman et al. 2017). The assumption of perfect implementation, however, is quite strong as interviewers have shown a preference for selecting respondents willing to participate in the survey (Alt, Bien et al. 1991), and a number of other studies found that data collected with random walk designs exhibit differences from known population statistics on gender, age, education, household size, and marital status (Bien, Bender et al. 1997, Hoffmeyer-Zlotnik 2003, Blohm 2006, Eckman and Koch 2016).

Probabilities of selection inherently cannot be calculated in a random walk sample design as no information is collected on how many structures are in the camp, or how likely it was that a given structure was the x^{th} structure along any path. Random walk must then assume all structures have the same selection probability, implying constant sampling weights. Therefore, the only component of the weights for the random walk is the sub-sampling of households within a selected structure: $w_i' = \frac{N_{ji}}{i}$.

Comparison of methods

As mentioned above, stratified cluster samples with the canvasing of selected clusters is the most common sample design used to collect official socioeconomic statistics in the developing world, but in other disciplines it is relatively rare. A review of published public health literature by Chen, Hu et al. (2018) found most surveys use probabilistic designs in the first stage, but random walk or similar methods in the second stage. Lupu and Michelitch (2018) suggest that the combination of random walk and quota sampling is the common approach for political science-themed surveys conducted in the developing world, with 77 percent of respondents to their expert survey using a variation on this design. Diaz de Rada and Martín (2014) compare a combination of random walk and quota sampling (based on age and gender) to probability designs and find a more accurate estimation of age and educational attainment in the combined method than in the probability methods, but that the probability methods perform better for measuring unemployment. The authors cite the replacement protocols for the probability methods as a reason for the bias, and attribute the use of quota sampling for the success in estimating age and education, compared to the gold standard of a high-quality probability sample design.

There are also a limited number of papers which directly compare two or three of the methods, but none that consider this wide range of alternatives. Chew, Amer et al. (2018) use a baseline convolutional neural network model on a gridded population sampling frame to select a sample of households in Nigeria and Guatemala. The authors found this technique to be on par with human canvasing in terms of accuracy, and to outperform other machine learning models based on crowdsource or remote sensing data. Grais, Rose et al. (2007) compared an unweighted random point selection methodology to a random walk in their study of vaccination rates in urban Niger. The authors do not find statistically significant differences between the methods, though the sample size was limited and both methods were non-probabilistic.

Design & Field Protocols

Experiment Design

This paper makes use of a dataset from the purposefully designed methodology experiment conducted in one section of the Protection of Civilians site 1 (PoC1, Figure B4-1), one of the largest IDP camps in Juba, South Sudan. To generate a gold standard as the basis of comparison, a household census was conducted between August and September 2017. During this exercise, 2,655 households were interviewed using a questionnaire designed to collect demographic information, dwelling characteristics, household consumption, and perception data. At the end of each census interview, households received a unique barcode that could be used to identify them later in the experiment.

As to avoid changes in camp composition, immediately following the completion of the census fieldwork, the interviewers returned to the field to implement the experiment. Teams used each of the sample selection methods to identify which households would have been selected had that method been used for a survey. To avoid respondent fatigue, instead of re-asking the questionnaire, the interviewers simply scanned the unique bar code of the selected household. Once scanned, the barcodes created an observation in the method-specific dataset with the information captured in the census. Each sampling technique targeted about 322 interviews so that comparisons could be made between the methods using an identical sample size. There was, however, some non-response for each method if interviewers were not able to contact a household member who could provide access to the barcode, if the barcode had not been retained by the household, or if the barcode was not scanned correctly. Protocols for each individual method are listed below.

Satellite Mapping

The Satellite Mapping method used a geo-referenced listing of structures in the PoC camp based on imagery from March 13th, 2017, approximately five months before the start of fieldwork. The interviewer team was given 322 randomly selected structures as well as a list of replacement structures from a list of all geo-referenced structures in the PoC camp. Interviewers navigated to selected points using the GPS coordinates of the structure. Non-residential structures were substituted with replacement points. If there was more than one household residing within the selected structure, one household was randomly selected.

Segmentation

The objective of segmentation is to decrease the listing burden, generally for speed, financial, or security concerns. For this experiment, the PoC camp was divided into 19 clusters each containing 12 blocks of approximately 9 to 12 structures (Figure B4-3). The size of the blocks varied because, to the extent possible, segment boundaries followed easily discernible landmarks. Since the segmentation was done using satellite maps, it was not possible to distinguish between administrative or residential structures. To select the households for the survey, 16 of the 19 clusters were selected, then 10 of the 12 blocks within the cluster. To select individual households, the enumerators conducted a listing of all structures within the selected blocks and randomly selected two structures be interviewed. This selection method yields a sample size of 320 households. Similar to satellite mapping, if there were multiple households within the structure, one was randomly chosen for the interview.

Grid Method

The grid method is similar to segmentation, but instead of purposefully-drawn, approximately equal population segments, the PoC camp was overlaid with a grid of cells measuring 50 meters by 50 metres. For the fieldwork, 27 grid cells were selected. After a listing of structures in each selected cell, 12 structures were randomly selected in each cell. Within each structure, a random household was selected in the case of multiple households per structure. Structures that fell into more than one cell were assigned to a single cell hosting the majority of the area of the structure. This determination was made in the field.

The loss of control over the number of households within the primary sampling unit, which in this case is the grid cell, complicates the selection process. If all grid squares contained at least 12 households, then it would be necessary to select (with equal probability) about 27 squares to reach the target sample size of 322 households. The number of households in a given grid square varied from 1 to 136 households, with 13 grid squares containing fewer than 12 structures (Figure B4-4). To reach the target sample size, grid squares were randomly ordered and the first 27 are selected. The expected sample size is then calculated by assuming that 12 structures are selected from each grid square in the case of grid squares containing more than 12 structures, and all households are selected if there are less than 12 structures in the grid square. If the expected sample size is 310 or less, an additional grid square is selected up to a maximum expected sample size of 328. The result is that while on average the total sample size was 322, the simulations gave a range between 317 and 328.

North Method

The North Method uses RSPs to determine the selected households. RSPs are chosen from the universe of all possible points with the boundaries of the PoC camp. To implement the North Method, 322 RSPs along with replacement RSPs were chosen. These points were random geo-coordinates within camp borders (Figure B4-5). If the RSP lay within a structure, the corresponding structure was selected. If not, starting at the selected RSP, enumerators walked directly north, using the compass application on their tablet, until a structure was encountered. If the structure was residential, the structure was chosen to be interviewed. In the case of multiple households present in the structure, one household was randomly chosen. If the structure was not residential or if the enumerator reached the boundary of the camp, a replacement RSP was used.

As it would be extremely difficult to determine the area of the shadow in the field, satellite imagery is used for these calculations. In the case of this experiment, the selection areas are calculated using Google Earth imagery taken on December 22nd, 2017, approximately one month after the census of households in the PoC camp. Given the dependence of the North Method on having current satellite imagery for accurate calculations, the availability of this imagery is a major consideration for this method. The weights would be over-estimated if new structures had been built in the shadow since the imagery was taken.

Random Walk

Random Walk obtains a sample by randomly selecting starting points for enumerators with generic but unambiguous instructions to select households at regular intervals on their path. For this experiment, enumerators conducted random walks using 21 RSPs (Figure B4-8). Starting as near as possible to the RSP, the supervisor chose any random point (like a street corner or a school). From this point, four enumerators walked each in one of the four cardinal directions. Walking in their designated direction away from the RSP, they counted structures on both the right and the left and each selected the fifth structure for interview. Enumerators were instructed to start with the buildings on the right if two buildings were opposite to each other. To select the next structure, enumerators continued along the cardinal path, and selected the next fifth structure. If the enumerator could not proceed on its cardinal path because she had reached the boundary of the PoC camp, enumerators were instructed to turn right at a 90-degree angle and continue counting until finding the fifth dwelling. Enumerators had to conduct six interviews along their paths.

Implementation Issues

Failure to follow survey protocols

As noted above, even if field protocols are perfectly implemented, the estimates generated from Random Walk designs are likely to be biased. Enumerators furthermore often were unable or unwilling to follow the protocols. Streets and paths were not necessarily aligned with cardinal directions and obstacles further impeded the ability to follow a straight path. Additionally, since the selection method requires enumerator judgment, it is not replicable and therefore allows enumerators greater discretion to choose which households are "selected." Figure B4-9 shows the paths taken by two teams of enumerators from random starting points. The team starting from point 16 more or less followed the field protocols, traveling a straight line until reaching the edge of the camp, and then making a right turn. The team starting from point 11 had more difficulty following the protocols. The enumerator traveling west actually travelled in a south-westerly direction and the enumerator traveling east followed a jagged path. These deviations can further increase error in a method which is already known to deliver biased results.

Structure identification issues

A key challenge in using GPS-based sampling strategies is to efficiently match the information from the satellite maps to the information collected on the ground. In the Satellite Mapping method, the interviewers must be able to match the GPS coordinates generated on the satellite map to actual structures on the ground. In the case of the experiment fieldwork, interviewers were not able to match the GPS coordinates to a structure in 15 of 322 cases. In addition, in one case the structure was out of scope, identified as a shipping container being used as a school.

When using the North Method, there is the opposite issue of matching the GPS coordinate captured at the time of the interview to a structure on the satellite map. To calculate the weights for the North Method, the analyst must be able to identify the interviewed structure and calculate its 'shadow.' However, it was not possible to match the selected structure captured by a GPS reading at the time of the interview to a household in the satellite map in 10 of 322 cases in the experiment and 132 out of 2,655 households in the complete census of the camp. Due to GPS error, outdated maps, or interviewer error, the GPS positions of those interviews were not located within a structure in the satellite imagery. In these cases, the sampling weight of the closest (or the average for multiple closest) household(s) was used for the respective household.

Non-response

The protocol for conducting interviews stipulated that household had to be visited three times if no knowledgeable adult was present. If, after the third visit, still no person was available to be interviewed, it was marked as a case of non-response. The cases of non-response were replaced for all methods except the census. For the methods which rely on simple random sampling (satellite mapping) or random point selection (North method and random walk), replacements consisted of additional random selections. For segmentation and the grid method, additional households were selected from the segment or grid square listing.

In the census, non-response was low, with interviewers unable to conduct interviews in only 36 out of 2,655 households, or 1.4 percent, due mainly to refusals or no adult being present in the household at the time of the repeated interview requests. For the sampling methods, the replacement rates were 1.1 percent for segmentation, 3.9 percent for the North method, 5.7 percent for satellite mapping, and 10.1 percent for grid square selection. The replacement rate for segmentation may, however, be artificially low as enumerators would have the incentive to not list households in the listing that they knew would not be home to respond. The North method may also be artificially low as random point selection similarly gives the enumerator the possibility to evade the strict protocol. Enumerators could unofficially replace a non-responding household by going to an adjacent structure rather than obtaining a new random point.

Weights must be adjusted for all sampling methods to compensate for non-response. Hence the final weight is

$$w_{i,m} = w'_{i,m} * nr_i$$

Here, $w'_{i,m}$ is the selection weight for household i using method m, and nr_{im} the non-response weight.

Multiple households per dwelling

In all methods except segmentation, structures are selected instead of households. In the case where a structure is occupied by only one household, there are no further stages of selection and the interviewer proceeds with the questionnaire. If there are multiple households, however, the interviewer must randomly select one for interview. This additional selection increases the potential for non-sampling error as the interviewer must implement the randomization procedure correctly in a setting where it is difficult, if not impossible, to verify. If randomization is done correctly, there will be no additional bias, but the extra stage will decrease the efficiency of the estimate and increase its standard error.

The frequency of selecting structures with multiple households varies by method. The impact is the lowest for grid squares (1.06), segmentation (1.08), and random walk (1.08). The percentage was much higher for the North Method (1.17) because larger structures have larger footprints and often have larger shadows, and, thus, are more likely to both be selected and to contain multiple households. The higher probabilities of selection, however, are accounted for in the weight calculations, and therefore the resulting statistics are unbiased, assuming the first stage of selection was implemented without bias. The highest percentage of structures containing multiple households, however, was found with the satellite mapping method (1.25). Since structures were randomly selected from a list of all structures, there is no theoretical reason why there should be more multiple household structures with satellite mapping, so this observation may be related to availability bias for larger households with available respondents.

Results

The objective of our analysis is to compare multiple sources of error and uncertainty in each of the five methods. In terms of the sources of error, we examine the bias inherent in the method design; non-sampling error common to all five methods; and non-sampling error specific to the method. To examine bias inherent to the method, we use simulations assuming perfect implementation based on the census data. Non-sampling error common to all methods is mainly availability bias (Cuddeback, Wilson et al. 2004), which we explore by comparing estimated household size and other measures correlated with household size. We look also at variables uncorrelated with household size to explore method-specific error. In terms of uncertainty, we look at both the overall design effects as well as decompose those effects into the unequal weight effect (UWE) and the cluster effect (our survey has no stratification) to further understand how much of the observed uncertainty is related to the method and how much is specific to the somewhat unique circumstances of the South Sudan refugee camp context (Liu, Iannacchione et al. 2002). Finally, we also look at the mean square error (MSE) as this measure takes into account both bias and uncertainty.

Bias

Household size & other correlated demographic variables

All simulation results generate estimated average household sizes which contain the true mean within the confidence interval. Compared with a census mean of 4.28, the simulation results for grid squares, North method, satellite mapping, and segmenting were all within 0.08 percent of the true mean, while the random walk results were almost thirty times higher at 2.2 percent - clearly biased compared to the probability methods. The experimental results all statistically significantly overestimated household size compared to the census mean. The survey methods yield means with biases of 12.4 percent for satellite mapping, 8.8 percent for segmenting, 13.6 percent for grid squares, 16.2 percent for the north method and 15.5 percent for random walk. This over-estimation is caused by a systematic tendency of enumerators to select larger households because they are more likely to find an adult respondent (Cuddeback, Wilson et al. 2004). As larger structures often have more rooms, the results are further confirmed by a similar upward bias for the number of rooms in the experimental results. All of the experimental methods overestimate the number of rooms compared to the census mean by between 8.2 and 12.2 percent, with the largest overestimation generated by the methods for which the probability of inclusion is higher for physically larger structures: 10.3 percent for grid squares, 10.4 percent for the north method, and 12.2 percent for random walk.

Table B4-2 shows the distribution of household size by method and uses a likelihood test to check for differences from the census distribution. Satellite mapping is not statistically significantly different from the census distribution, grid squares and segmenting are weakly significantly different, and random walk and the north method are significantly different. Figure B4-10 shows the distribution of household by household size in the census and from the North method, with the latter showing the highest degree of bias compared to the census mean. The North method captures less than half the percentage of single member households as were found in the census (8.0 percent compared to 16.8 percent). Though similar patterns are found for all methods, the methods which do not select a specific structure, the North method and the random walk, show higher degrees of availability bias than those methods in which the selection can be verified.

Other demographic variables, including adult equivalent household size and the adult-to-member ratio of household members, are highly correlated with household size and therefore show similar patterns as household size. In the simulations, the results for the adult equivalent household size were within 0.05 percent of the census mean for all methods except random walk, which generated a bias of 1.4 percent. In the case of the adult ratio, the differences were less stark, with bias estimates ranging between 0.01 percent and 0.31 percent for the probability methods, compared to 0.41 percent for random walk. The experimental results also overestimate the mean for the adult equivalent household size measure, with the largest overestimation found for segmenting and random walk. For the ratio of adults to total household members, all methods underestimate the ratio compared to the census. This underestimation is again related to the tendency of the experimental methods to select larger households, which have larger numbers of children and consequentially lower adult ratios.

Variables Correlated with Household Size

Other variables considered in the analysis that are positively but not highly correlated with household size are if the household head had ever attended school (correlation = 0.191) and if the household head can read in any language (correlation = 0.183). While the simulation generates unbiased estimates for the probability methods (ranging from 0.03 percent to 0.44 percent), the random walk shows substantially higher bias with 1.63 percent for school attendance and 1.77 percent for being able to read. All experimental methods show lower percentages of school attendance (from 4.92 percent to 15.21 percent) and literacy (from 0.12 percent to 13.92 percent) than the census. This finding could reflect more difficulties in finding work for those heads with lower levels of education and, therefore, a higher likelihood of being found at home by the interviewer.

Variables Uncorrelated with Household Size

We consider three variables, which are uncorrelated with household size. The variables are whether the respondent / household owns a mobile phone (correlation = 0.043), owns a mattress (correlation = 0.031), and wants to leave this location (correlation = 0.015). The simulation confirms that the probability methods yield largely unbiased results for wanting to move, ranging between 0.13 percent and 0.24 percent, but ownership of a mattress and mobile phone are slightly more biased ranging from 0.28 percent to 0.70 percent, again excluding random walk. For the random walk simulations, the bias was 0.71 percent for mobile phone ownership, 1.37 percent for desire to move location, and 3.02 percent for owning a mattress. The estimates from the experiments vary between methods and indicators. The satellite mapping generates the largest bias (ranging from 4.88 percent to 12.61 percent) followed by North method (from 1.93 percent to 7.53 percent) and segmenting (from 0.61 percent to 9.18 percent) with best results yield by grid squares (from 1.40 percent to 6.74 percent) and random walk (from 1.01 percent to 7.84 percent).

Poverty Variables

Simulations using the probability methods confirm largely unbiased results for consumption (total, per capita and per adult equivalent), with bias ranging from 0.01 percent to 0.2 percent for the satellite mapping, segmenting, and north methods, and being slightly higher for the grid squares methods, ranging from 0.38 percent to 0.82 percent. Bias was slightly higher for poverty measures (per capita and per adult equivalent), with the probability methods ranging from 0.07 percent to 0.54 percent. For the random walk method, the biases from the simulations were generally above 1 percent, and as high as 2.35 percent for per capita poverty. The results from the experiments show an upward bias for total consumption around 7 percent, except for satellite mapping with a bias of almost 25 percent, and a largely unbiased estimate from the grid squares methodology. Per capita and per adult equivalent consumption are largely biased downwards due to the upward bias in household size, except for satellite mapping in which the upward bias in household size is more than offset with a large upward bias in total consumption. Accordingly, per capita and per adult equivalent poverty measures are biased upwards (from 7.57 percent to 19.96 percent, respectively) with the exception for satellite mapping (downwards by 4.44 and 5.33 percent, respectively). The experimental random walk results are also upwardly biased at around 10 percent, consistent with what was found in other methods.

Meta-analysis of bias

To better understand the different factors that impact the accuracy of the five sampling methods being studied here, we undertook analysis of the simulated and observed results pooled across the five methods, controlling for difference characteristics of the particular questions and clustering the standard errors at the question level. The dependent variable for this analysis is the absolute value of the normalized mean of the bias, or the (observed value – census value) / census value. In addition to the type of sampling method, three of question-level measures are included in the analysis: coefficient of variation, correlation with household size, and four versions of the Moran's I spatial dispersion. The coefficient of variation is a measure of the inherent variability of responses across the census values for a particular question and was included to control for higher variation variables being more prone to sampling error, which would show up as bias in the results. The correlation with household size was included as household size is a variable known to be impacted by availability bias, which is common across all five methods. Finally, four versions of the Moran's I spatial dispersion statistics (at 4m, 8m, 16m, and 32m) were included to understand the impact of clustering within the PSU. If the spatial dispersion index were zero, there is no relationship between measured value of a certain household and those of their neighbors. In a case where an attribute is completely randomly distributed throughout the population, then the impact of the selection of sampling method is limited as you could speak with any 12 households and have a random set of responses.

Columns 1-5 in Table B4-3 in the appendix show the results for the pooled regressions for the simulated results. The results on method are consistent across all four specifications. Compared to the reference method of satellite mapping, the North Method is unbiased, while the Segmenting and Grid Square methods show minimal bias (0.1 percent and 0.2 percent, respectively). The Random Walk method shows 1.2 percent bias on average across the 14 questions. The additional controls for the coefficient of variation, correlation with household size, and spatial dispersion are also not significant.

Columns 6-10 in Table B4-3 show the results for the same specification using the observed results. The R2 for these models is substantially lower than for those with the simulated results. The base model including only the sampling methodologies has an R2 of 0.593 for the simulated results compared to only 0.097 for the observed results and none of the variables for the sampling method are statistically significant in the observed models – indicating there is much more noise in the observed models compared to the simulated. The variables that contribute additional explanatory power also vary between the simulated and observed results. In both cases the addition of the coefficient of variation yields only a negligible increase. In the simulated models, there are similar small changes when the correlation with household size and the spatial dispersion measures are introduced. In the observed models, the R2 more than doubles to 0.215 when the correlation with household size variable is added, and the variable itself is strongly significant. This finding demonstrated the strong influence of availability bias across the five methods. The R2 also nearly doubles with the addition of the four spatial dispersion variables though the coefficient is weakly significant only on the 8m measure and not significant at all on the 4m, 16m, or 32m measures. Also, the coefficient on 8m is the only one of the four that is positive – indicating higher levels of bias for variables that have more spatial correlation with households within 8m of their location. The interpretation here is less clear, though one hypothesis is that for households that are very similar to their close neighbors (within 4m) that there is little impact of an interviewer breaking field protocols to switch households to one more likely to yield a respondent. At a slightly higher increased distance though, the impact on methods such as random walk, where the interview can "get trapped" in a corner is larger.

Normalized Root Mean Squared Error

The root mean squared error (RMSE), or the square root of the average of the deviation of the estimated mean from the true mean, accounts for both bias and uncertainty. Lower bias and lower variance yield lower values of RMSE and therefore lower RMSEs are preferable to higher RMSEs in evaluating methodologies. See section 7.3 in the appendix for more detail. In this application, because we have both continuous and dichotomous variables, we use the normalized RMSE (nRMSE) to facilitate comparability. Figure B4-11 and Figure B4-12 below compare the nRMSEs across the methods and questions. On average segmenting has the lowest nRMSE (1.67), followed by random walk (1.63), satellite mapping (1.70), grid squares (1.79) and the north method (2.14). As shown in Figure B4-11, however, the results for satellite mapping are skewed by one outlier value on total weekly household consumption. Excluding that value, the nRMSE for satellite mapping is 1.49. Overall, the method that perform the best is segmenting, which has the lowest or second lowest nRMSE for 12 of the 14 questions, and the highest or second highest for only two questions, followed by satellite mapping, which gives the best or second best results in 9 questions and the worst or second worst results for 4 questions. The methods that performs the worst is the North Method, which does not give the best or second best results for any of the questions and gives the worst of second worst results for 9 of 14 questions, followed by the random walk, which give the best or second best results for 4 questions and the worst or second worst results for 7 questions. The final method, grid squares, gives the best or second best results for 3 questions and the worst or second worst results for 6 questions.

Discussion

We find that simulations arrive at the true household size distribution, while all simulations over-estimate household size. This over-estimation is caused by a systematic tendency of enumerators to select larger households because they are more likely to find an adult respondent. Specifically, the North method and the random walk show higher degrees of availability bias than those methods in which the selection can be verified, e.g., satellite mapping where a specific structure is chosen a priori. Also for other indicators, including poverty estimates, we find that simulations obtain unbiased results while the actual experiments are biased, especially for variables correlated with household size. Pooling the analysis across indicators and using satellite mapping as reference, the North Method is unbiased, while the Segmenting and Grid Square methods show minimal bias (0.1 percent and 0.2 percent, respectively). The Random Walk method shows 1.2 percent bias on average across the 14 questions. In conclusion, probability-based methods perform better than non-probability methods like random walk. In addition, implementation of adherence with the survey protocol is extremely important. In practice – in a fragile setting like South Sudan – deviations from the survey protocol, measured as differences between the experiments and the simulations, have large influence on the actual bias of estimates.

Appendix

Simulation and Frame

To compare the efficiency of the different sampling frames and designs, we will apply an empirical sampling simulation. In this type of (Monte-Carlo style) simulation, either a true or synthetic population is used as the target population. By applying a specific sampling design, and repeated sampling (usually 1000 repetitions) under this design, we can compare the resulting population estimates with the known true population values for each run of the simulation.

The resulting distribution of these estimates is called the sampling distribution, and the average squared deviation from the underlying population value is the Mean Squared Error (MSE) or when taking its square root, the Root MSE (RMSE). To facilitate the comparison, we use the relative version expressed in percentage deviation.

Empirical sampling simulations can be considered as the "[…] ultimate tool for investigators who want to know if one sampling strategy will work better than another for their population." (Thompson, 2013). However, this requires the underlying simulation population to replicate as realistically as possible the target population.

Quality Metrics

A standard Measure in the assessment of particular sampling designs is the Root Mean Squared Error (RMSE) and calculated as:

$$RMSE = \frac{\sum_{sim=1}^{1000} RMSE_{sim}}{1000} = \left[\frac{1}{1000} \times \frac{\sqrt{(\hat{Y} - Y)^2}}{Y}\right] \times 100$$

Expressed here as percentage deviation from the population mean Y and calculated for each parameter of interest. Table [..] compares this for the different approaches discussed above.

Equation .. is only the empirical representation though and a result of rearranging the definition of the Mean squared Error,

$$MSE(\hat{Y}) = E(\hat{Y} - Y)^2 = E[(\hat{Y} - \tilde{Y}) + (\tilde{Y} - Y)]^2 = E(\hat{Y} - \tilde{Y})^2 + 2E(\hat{Y} - \tilde{Y})(\tilde{Y} - Y) + (\tilde{Y} - Y)^2$$

And decomposing it into

$$MSE(\hat{Y}) = Var(\hat{Y}) + Bias(\hat{Y})$$

with $\hat{Y}$, $\tilde{Y}$ and Y being the estimate from the sample, the mean of this estimate and the true value in the population respectively. *Var* is the corresponding variance, and *Bias* the resulting bias component, which is defined as:

$$Bias\,(\tilde{Y}) = \tilde{Y} - Y$$

If the mean of the estimator and the population mean is the same, the bias is 0. And the MSE would be equal to the variance of the estimate, which is only a result of the sample size. However, in a real survey situation, the population mean is commonly unknown, the resulting MSE therefore captures both, the variance and the bias. Since a sampling frame which is not covering the target population well, is likely to produce a different mean for the variable of interest than its true population mean, we may expect the bias to be different from 0.

Figures

Figure B4-1: IDP Camp PoC1 in Juba, South Sudan.

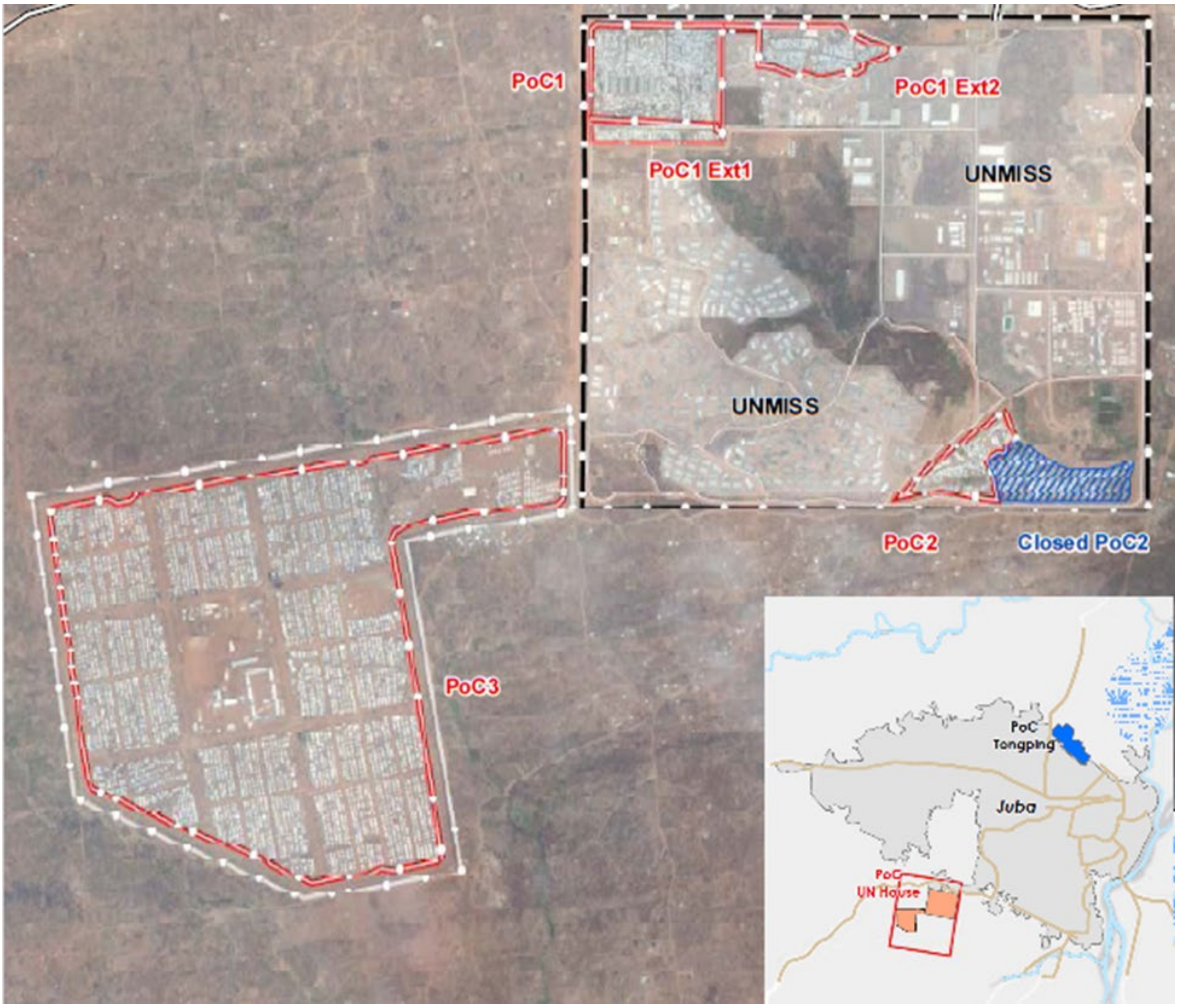


Figure B4-2: Satellite mapping of residential structures.

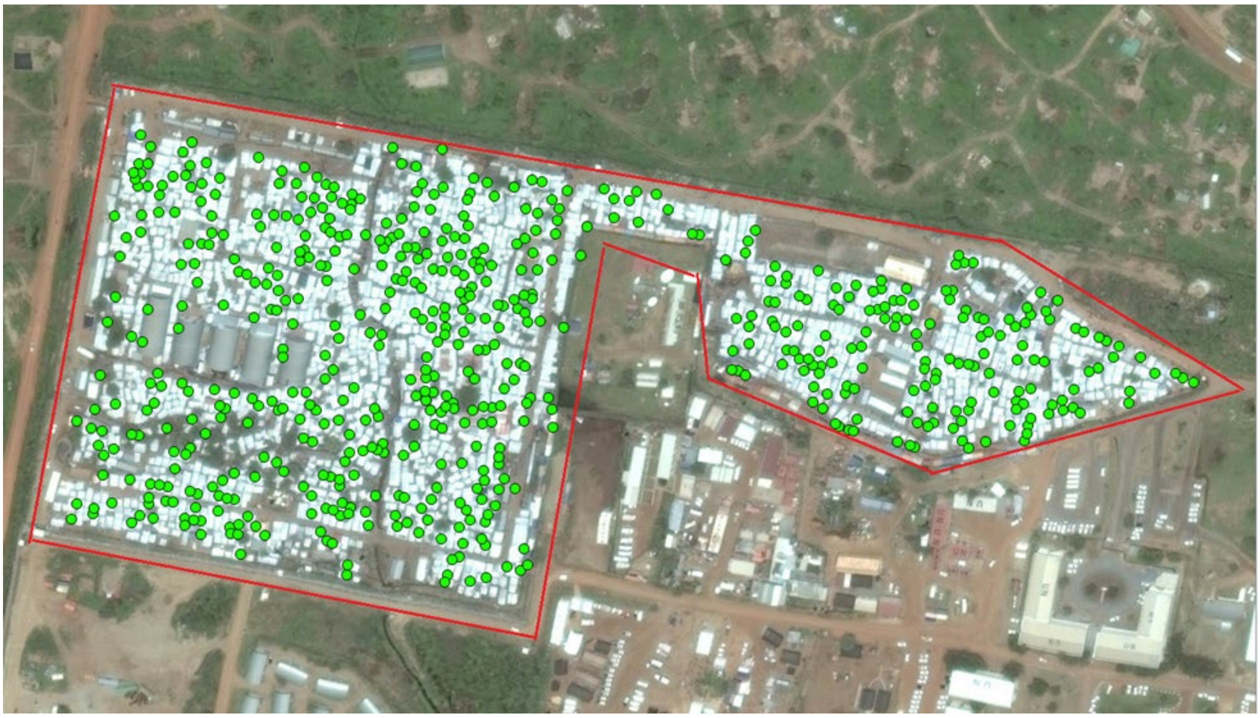


Figure B4-3: Enumeration Areas (blue) and Blocks (orange) for segmenting method.

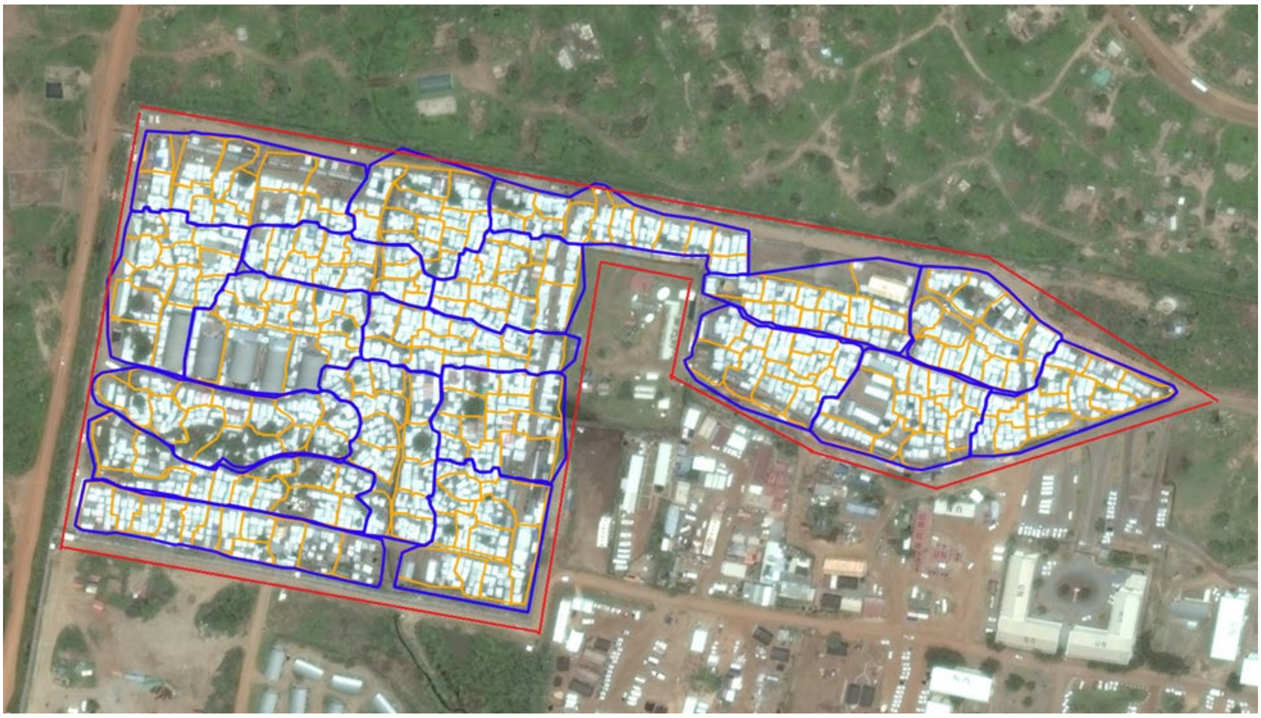


Figure B4-4: Grid overlay over the camp.

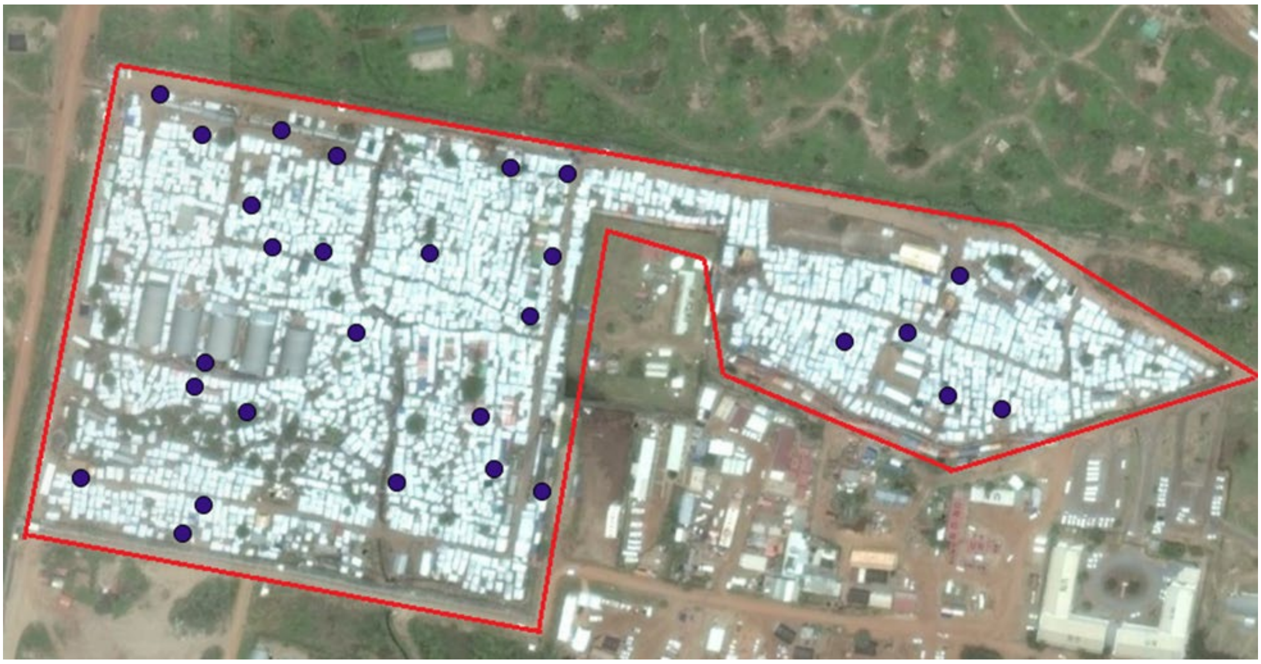


Figure B4-5: Random coordinates for the North Method.

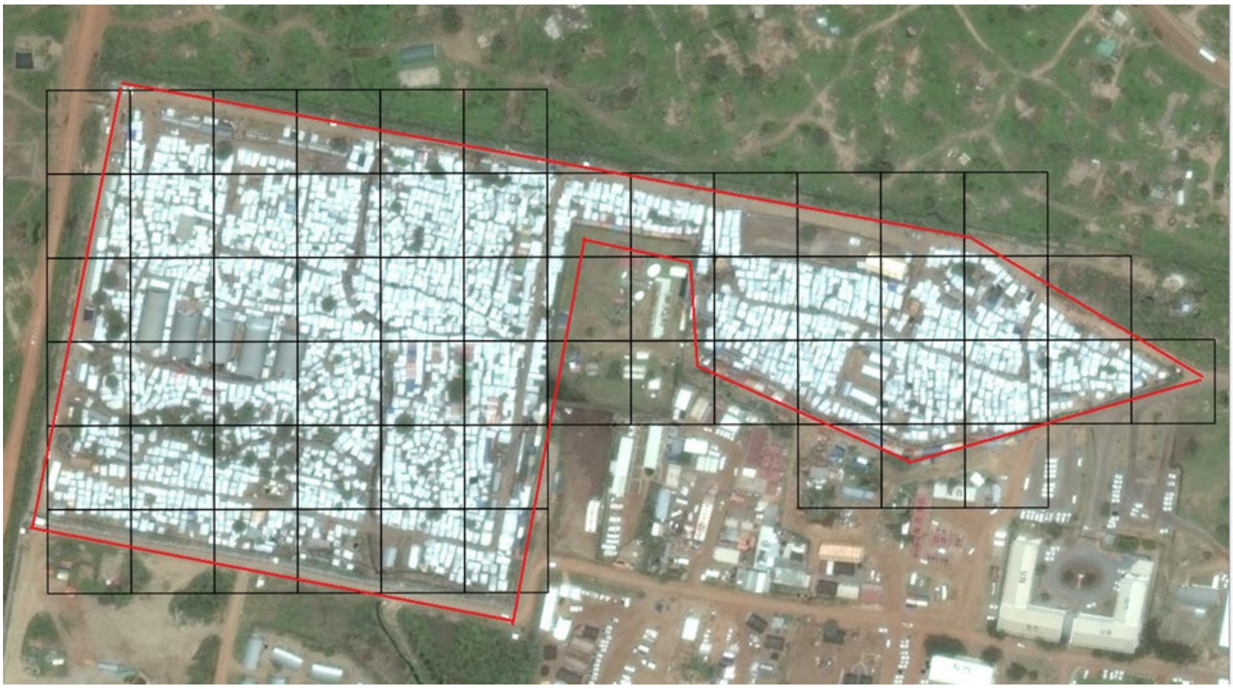


Figure B4-6: Example of the selection area of a structure.

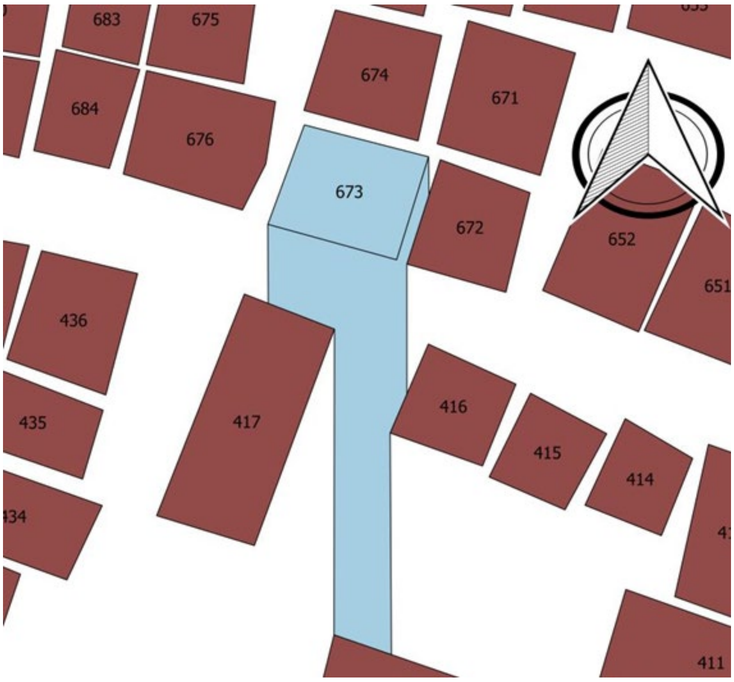


Figure B4-7: Areas leading to selection of given household in North method.

Figure B4-8: Selected RSP for Random Walk.

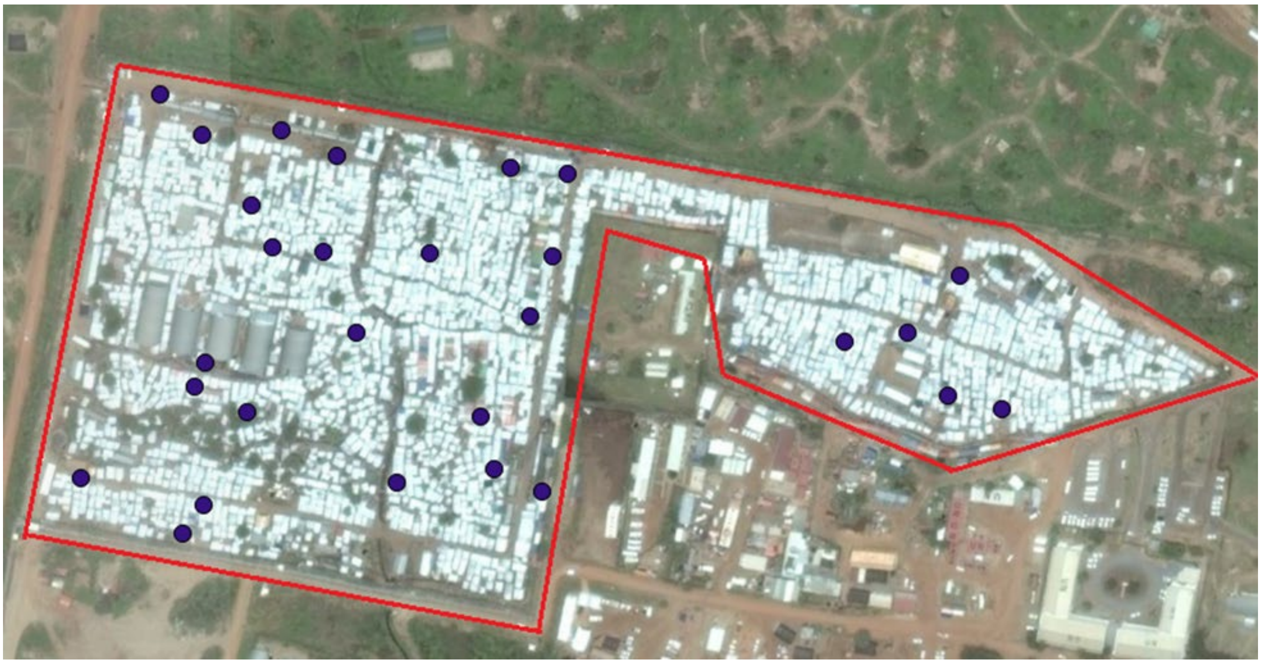


Figure B4-9: Examples of a correct and an incorrect Random Walk path.

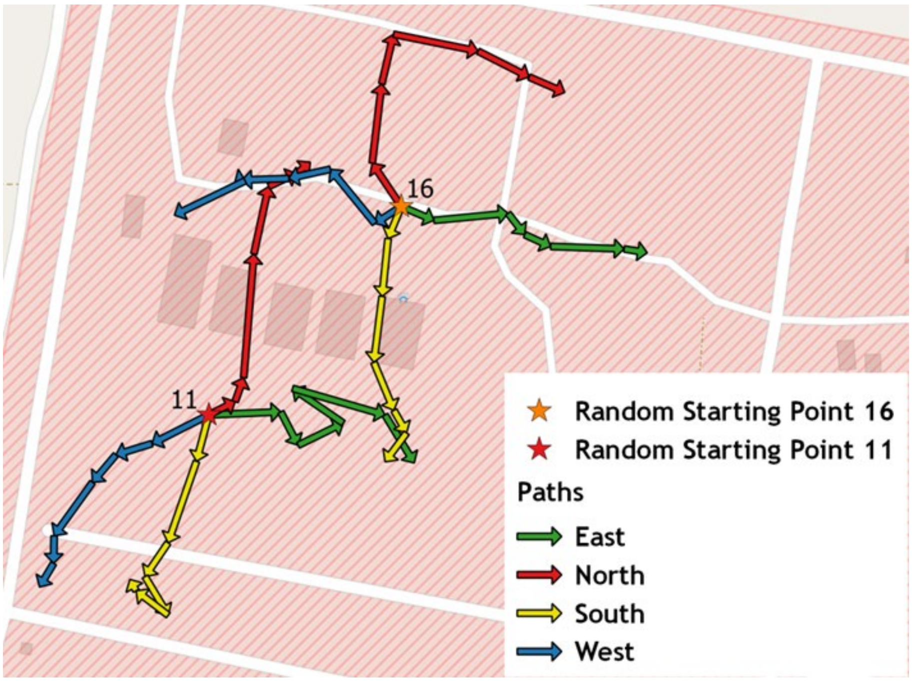


Figure B4-10: Household size.

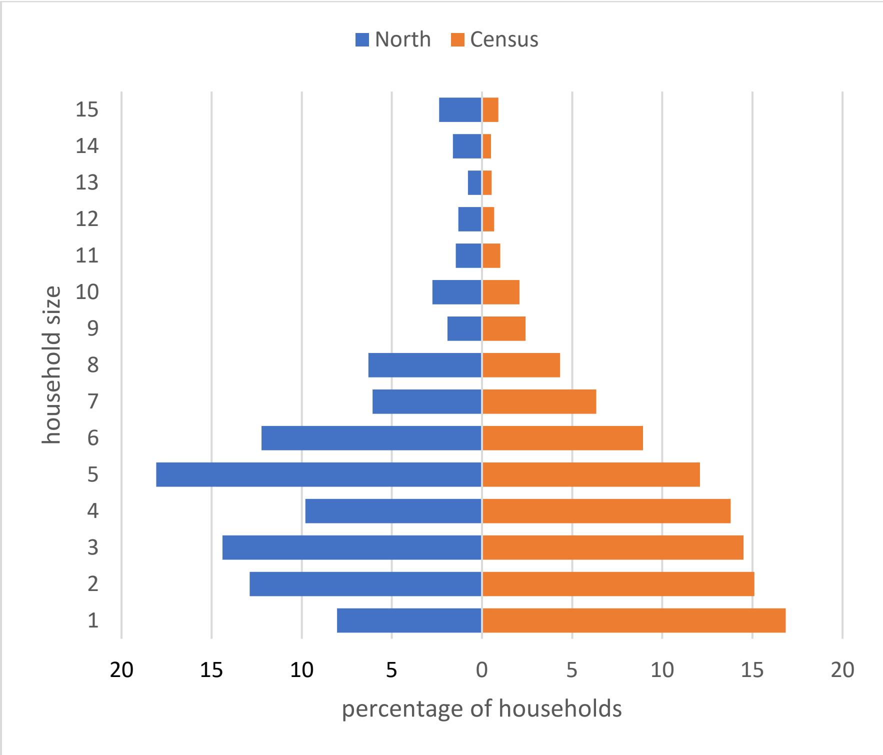


Figure B4-11: Normalized Mean Square Error by Method.

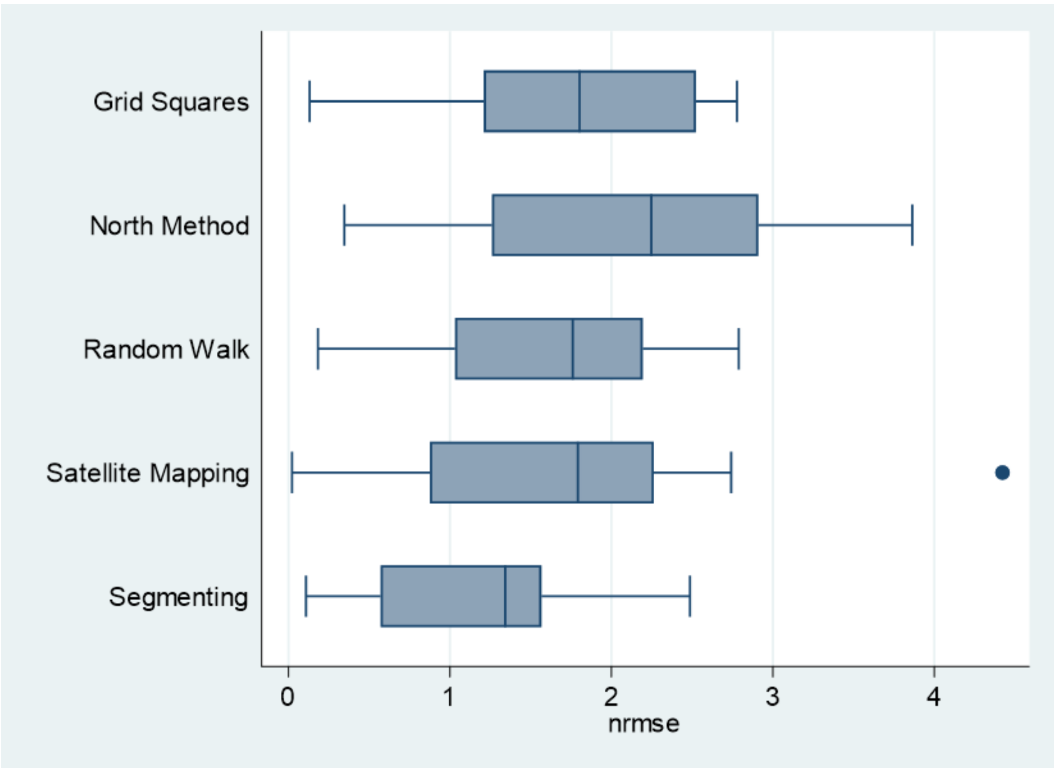


Figure B4-12: Normalized Mean Square Error by Method and Question.

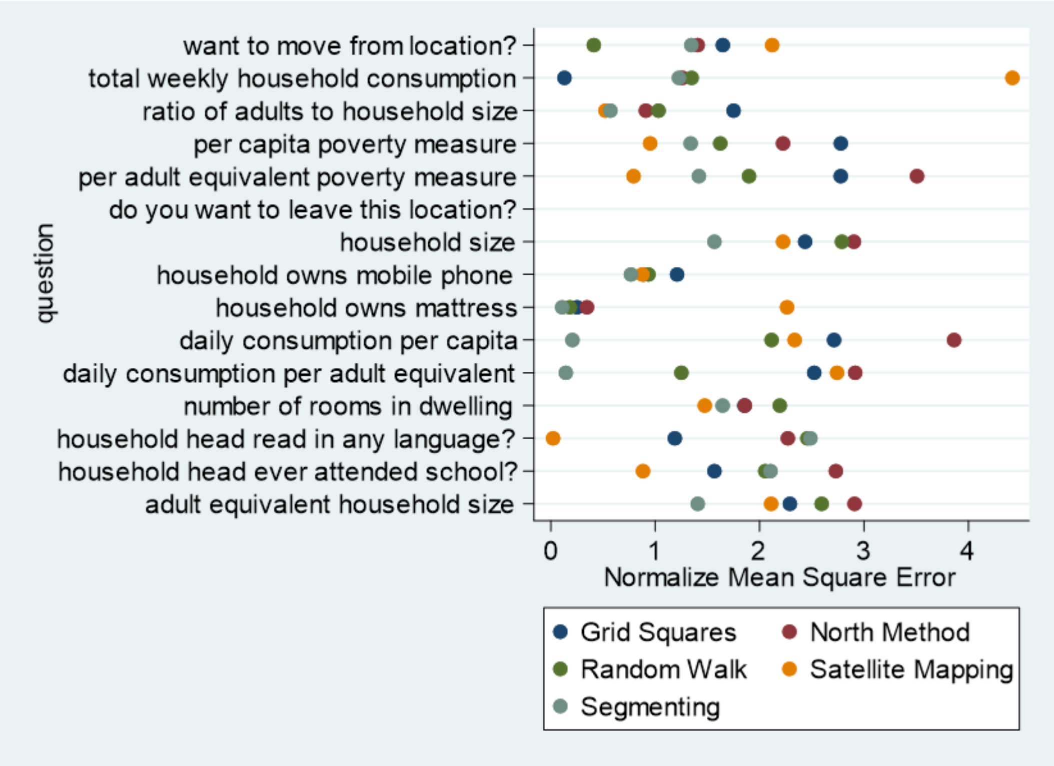


Table B4-1: Replacement Rate and Mean Number of Households per surveyed structure.

method	Replacement Rate (%)	Average households per structure in sample
Census	1.4	--
Grid Squares	10.1	1.06
North Method	3.9	1.17
Random Walk	--	1.08
Satellite Mapping	5.7	1.25
Segmenting	1.1	1.08

Table B4-2: Household size by method, compared to Census.

	Census	Grid Squares	North Method	Random Walk	Satellite Mapping	Segmenting
1	16.8	17.1	10.3	11.2	13.2	12.0
2	15.1	8.4	13.4	10.6	12.5	10.2
3	14.5	10.2	15.4	16.4	15.2	11.9
4	13.8	12.3	8.0	13.5	11.5	15.5
5	12.1	19.2	16.5	9.8	12.2	13.9
6	8.9	11.4	11.5	8.6	9.7	10.8
7	6.3	7.5	4.9	10.1	9.5	8.7
8	4.3	4.2	6.2	6.9	5.2	5.7
9	2.4	2.2	1.7	4.0	2.0	3.5
10+	5.7	7.4	12.1	8.9	9.0	7.9
likelihood-ratio chi2(9)		16.7	34.5	33.7	9.4	16.4
pr		0.053	0.000	0.000	0.397	0.060

Table B4-3: Pooled regression analysis of indicators across methods.

	Census		Grid Squares		North Method		Satellite Mapping		Segmenting		Random Walk	
method			simulation	experiment	simulation	experiment	simulation	experiment	simulation	experiment	simulation	experiment
ratio of adults to household size	0.694	bias (%)	0.23%	-9.76%	-0.12%	-5.07%	0.01%	-2.91%	-0.31%	-3.19%	-0.41%	-5.75%
		deff	1.53	1.74	1.35	1.29	1.00	1.30	1.24	1.28	1.16	1.07
total weekly household consumption	2,870	bias (%)	0.82%	-0.74%	0.01%	7.17%	0.20%	24.80%	0.13%	6.93%	1.02%	7.57%
		deff	1.49	0.98	1.26	1.19	1.00	1.29	1.07	1.25	0.97	0.99
daily consumption per adult equivalent	164.5	bias (%)	0.42%	-14.35%	-0.24%	-16.59%	0.09%	15.37%	0.02%	0.81%	0.12%	-7.02%
		deff	1.51	1.11	1.30	1.192	1.00	1.349	1.07	1.200	1.12	1.078
daily consumption per capita	127.8	bias (%)	0.38%	-15.43%	-0.37%	-21.98%	0.03%	13.10%	-0.05%	-1.17%	-1.09%	109.45%
		deff	1.51	1.08	1.30	1.225	1.00	1.335	1.10	1.231	1.10	1.094
adult equivalent household size	2.86	bias (%)	0.00%	11.76%	-0.02%	12.76%	-0.04%	7.88%	-0.01%	14.46%	1.41%	16.21%
		deff	1.51	1.25	1.32	1.492	1.00	1.129	1.23	1.019	1.18	1.318
household size	4.28	bias (%)	-0.08%	13.57%	-0.05%	16.18%	-0.06%	12.39%	0.04%	8.77%	2.21%	15.54%
		deff	1.53	1.53	1.34	1.29	1.00	1.26	1.23	1.17	1.18	1.02
household head ever attended school? (%)	0.571	bias (%)	-0.44%	-8.73%	-0.15%	-15.21%	-0.03%	-4.92%	-0.08%	-11.78%	-1.63%	-11.44%
		deff	1.52	1.49	1.36	1.295	1.00	1.310	1.25	1.238	1.19	1.066
household head read in any language? (%)	0.587	bias (%)	-0.29%	-6.62%	-0.02%	-12.65%	-0.15%	-0.12%	-0.17%	-13.92%	-1.77%	-13.70%
		deff	1.54	1.51	1.36	1.29	1.00	1.29	1.25	1.28	1.20	1.067
household owns mobile phone (%)	0.528	bias (%)	0.50%	6.74%	0.57%	-4.96%	0.66%	4.88%	0.69%	4.29%	0.71%	5.21%
		deff	1.54	1.38	1.34	1.28	1.00	1.30	1.22	1.23	1.31	1.076
per adult equivalent poverty measure	0.430	bias (%)	0.41%	15.80%	0.14%	19.96%	0.28%	-4.44%	0.54%	8.02%	1.69%	10.66%
		deff	1.51	1.38	1.34	1.30	1.00	1.28	1.18	1.25	1.07	1.071
per capita poverty measure	0.604	bias (%)	0.07%	15.80%	0.10%	12.65%	0.14%	-5.33%	0.47%	7.57%	2.35%	9.11%
		deff	1.54	1.26	1.33	1.32	1.00	1.33	1.18	1.27	1.22	1.077
number of rooms in dwelling	1.38	bias (%)	0.12%	10.33%	-0.07%	10.36%	-0.03%	8.21%	-0.11%	9.21%	0.94%	12.22%
		deff	1.60	1.73	1.31	1.14	1.00	1.46	1.19	1.42	1.43	1.02
household owns mattress (%)	0.44	bias (%)	0.28%	-1.40%	0.69%	-1.93%	0.65%	12.61%	0.70%	-0.61%	3.02%	1.01%
		deff	1.56	1.20	1.34	1.28	1.00	1.31	1.15	1.25	1.02	1.071
want to move from location? (%)	0.56	bias (%)	0.24%	2.29%	0.14%	-7.53%	0.13%	11.81%	0.14%	9.18%	-1.37%	7.84%
		deff	1.53	1.08	1.35	1.22	1.00	1.22	1.06	1.31	1.15	1.28

Table B4-4: Multivariate Regressions on pooled simulated and observed results.

	simulations					observed results				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se
Reference: Satellite Mapping										
Grid Squares	0.002***	0.002***	0.002***	0.002***	0.002**	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.032)	(0.032)	(0.032)	(0.033)	(0.034)
North Method	0.000	0.000	0.000	0.000	0.000	0.015	0.015	0.015	0.015	0.015
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)
Random Walk	0.012***	0.012***	0.012***	0.012***	0.012***	-0.014	-0.014	-0.014	-0.014	-0.014
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.028)	(0.028)	(0.028)	(0.029)	(0.030)
Segmenting	0.001*	0.001*	0.001*	0.001*	0.001*	-0.043	-0.043	-0.043	-0.043	-0.043
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.028)	(0.029)	(0.029)	(0.029)	(0.030)
Coefficient of Variation		0.004			0.012***		-0.006			0.006
		(0.003)			(0.004)		(0.041)			(0.037)
Correlation with HH Size			-0.003		0.001			0.071***		0.119***
			(0.002)		(0.003)			(0.020)		(0.023)
Moran's I @ 4m				-0.006	0.077**				-0.126	-0.750**
				(0.037)	(0.032)				(0.380)	(0.323)
Moran's I @ 8m				-0.037	-0.304***				1.741*	2.394***
				(0.085)	(0.098)				(0.971)	(0.899)
Moran's I @ 16m				0.105	0.417***				-2.051	-1.437
				(0.116)	(0.151)				(1.378)	(1.218)
Moran's I @ 32m				0.022	-0.216				-3.094	0.628
				(0.121)	(0.149)				(2.644)	(1.307)
constant	0.002***	-0.001	0.003***	0.001	-0.007*	0.109***	0.114**	0.085***	0.118**	0.022
	(0.001)	(0.002)	(0.001)	(0.002)	(0.004)	(0.023)	(0.048)	(0.023)	(0.048)	(0.053)
n	70	70	70	70	70	70	70	70	70	70
R2	0.593	0.615	0.613	0.607	0.684	0.097	0.098	0.215	0.190	0.338

note: *** p<0.01, ** p<0.05, * p<0.1

5. A Light-Touch Method to Improve Accurate Reporting of IDP's Food Consumption[59]

Lennart Kaplan, Utz Pape, James Walsh

Introduction

Accurate data on the key economic variables affecting people who have been forcibly displaced, such as consumption and assets, are essential to understanding their situation and to developing evidence-based policies to support them. Poor information can lead to flawed diagnostics or incorrect assessments of impact. Data inaccuracies may lead policy makers to allocate funds to the wrong people or to the wrong programs. The standard way in which the World Bank and other policy organizations develop statistics is through individuals' responses to questions in economic surveys. Self-reported information is vulnerable to myriad reporting inaccuracies when social scientists ask personal or intrusive questions or when respondents anticipate social or material implications to the answers they provide. [60] This is of particular concern when respondents believe that misreporting may provide relief, both because of the of the sensitivity and the gravity of the policy challenge. In situations where it has been possible to compare survey responses to revealed economic behavior, striking disparities are sometimes found. In one investigation for example, Poterba and Summers (1986) report that misstatements regarding employment status in the Current Population Survey led to an underestimation of the duration of unemployment by up to 80 percent and even greater overestimates of the frequency of labor market entries and exits. In another study, Rosenfeld, Imai et al. (2016) look at voting behaviors in a sensitive anti-abortion referendum held in Mississippi in 2011. They compare actual county level vote shares against survey results from a sample frame of individuals who voted during the election (based on public records). Surveys that used direct questioning led to an underestimation of casting a "no" vote by more than 20 percentage points in the majority of counties.

There are a number of mechanisms through which the validity of self-reported information in surveys can be compromised. Some inaccuracies result from cognitive biases – for example, acquiescence or "yea-saying" (Bachman and O'Malley 1984, Hurd 1999), extreme responding (Cronbach 1946, Hamilton 1968), and question order bias (Siegelman 1981). One solution to problems such as question order bias is to randomize the order of questions (Warner 1965). Other inaccuracies emerge from conscious but not calculated behavior. Respondents may deliberately misreport information on sensitive subjects not to distort statistics but to maintain their reputation or to abide by political norms (Gilens, Sniderman et al. 1998). A common solution to this is to enable participants to cloak their behaviors or beliefs. List experiments, endorsement experiments, and randomized experiments are commonly used techniques for this purpose (Rosenfeld, Imai et al. 2016).

The explanations above assume that people intend to report accurately but are prevented from doing so due to aspects of the situation. In some contexts, individuals may misreport due to expectations about the implications of the results of the study. For example, individuals may misreport to increase earnings in a study context (Mazar, Amir et al. 2008) or to shape the results of the study if they believe that it will inform policy. In situations where individuals wish to influence a particular research outcome, a guise of anonymity will not shift their behavior. It is important to note that our concern is not with the ethics of individual misreporting – this is a reasonable response to contexts of extreme

[59] LK, UP and JW contributed equally to the manuscript.

[60] This is of particular concern, for example when asking about race as in Kuklinski, J. H. (1997). "Racial Prejudice and Attitudes Toward Affirmative Action." American Journal of Political Science **41**(2): 402 - 419.

vulnerability – but rather to ensure that policymakers have access to data that enables them to adequately serve the vulnerable population as a whole.

Behavioral science is increasingly being used as a policy tool to help policymakers create better policy and solve collective action problems more effectively (World Bank 2015). This is based on research illustrating that people make decisions on the basis of both external and internal reward mechanisms (Mazar and Ariely 2006). Even in cases where people have an extrinsic incentive to misreport, this may be overridden by a preference for remaining consistent with their values. One example of this is when individuals' beliefs regarding the consequences of misreporting affects their behavior. In an two-person experiment where one participant can increase her payoff by misreporting but at the expense to her counterpart, Gneezy (2005) finds that individuals' propensity to misreporting is sensitive to the costs it imposes on the other person. Contextual cues affect the salience of internal incentives (or intrinsic motivations) and thus the accuracy of responses. This psychological mechanism has been put to practical use in policy. In multiple contexts, normative messaging has been used to increase tax payments (Hallsworth, List et al. 2017) or reduce littering and environment theft (Cialdini 2003).

In this paper, we apply the tools of behavioral science to investigate the veracity of consumption reports by internally displaced persons (IDPs). In numerous rounds of data collection in Somalia and South Sudan, IDPs report significantly lower levels of consumption than non-IDP households. In previous survey rounds 45 percent of Somali IDP households report food consumption below subsistence levels and approximately 80 percent below recommended levels. While the data may be accurate, there are two reasons to suspect that it is not. First, such high levels of non consumption would be associated with high rates of mortality due to starvation. Although being high, the mortality rates among IDPs suggest that this is not happening systematically across the country at such a scale. Second, non-IDP households that are statistically similar on observable characteristics report higher levels of consumption than IDP households. While IDPs and non-IDPs may have different opportunities to generate income, it is unlikely that IDPs choose not to smooth their resources to balance between food and non-food consumption in a way that endangers their life.[61]

If it is the case that survey respondents misreport, the inaccuracies it generates in the data are highly problematic. At best, it makes the data spurious and unusable. At worst, it could lead to misallocations of aid, from more vulnerable areas to less vulnerable areas, or from solutions emphasizing sustainability to immediate relief where immediate relief is unnecessary. Due to the dangerous environment in South Sudan and Somalia, it is not currently possible to do use alternative data collection methods, for example ethnographic research, to investigate this puzzle in the data. The validity of alternative investigative methods such as food diaries is vulnerable to the same incentive to game as surveys. One way to investigate whether people misreport is to test whether consumption rates change in response to nudges. If these primes are effective, they would be expected to particularly affect potentially underreporting, hence, poor households. Moreover, as vulnerable populations would have higher incentives to underreport, priming should be stronger for IDPs than for comparable non-IDP populations. We find the primes induce higher reporting in lower quintiles of reported consumption. This treatment pattern is driven by aid reliant IDPs and vanishes when considering the comparison group of non-IDPs. The results are especially strong for consumption quantities (items and kilograms), which are most easily subject to intentional misreporting. This suggests that IDPs are indeed misreporting. The paper has two main limitations. First, it can only compare the treated group against an estimate of the "true" consumption rates. Second, the intervention is bundled. For this reason, it is impossible to isolate the causal mechanism affecting the

[61] The underlying survey data of this study discussed at a later stage actually indicates that IDPs have a more calorie intensive food consumption profile.

observed changes in reporting. Further work is needed to identify an estimate of the true level of consumption against which to compare the primed individuals and to isolate the causal mechanisms by which people are changing their behavior.

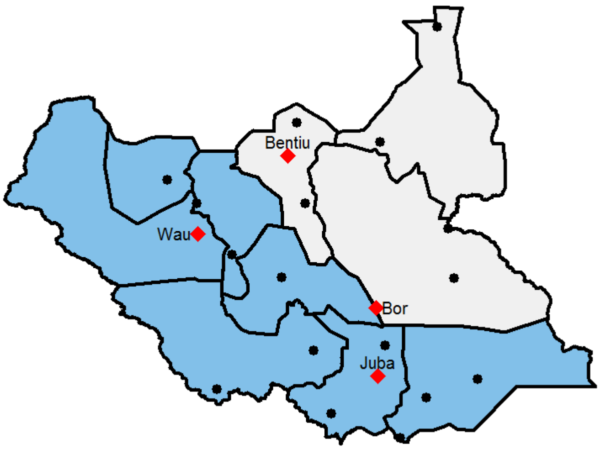
The paper proceeds as follows. Section 2 provides an overview about the underlying context and the compiled data. Section 3 provides an overview about the underlying methods, while Section 4 introduces the empirical approach, which builds the foundation for the results in Section 5. This is complemented by an assessment of robustness and potential channels in Section 6. Finally, findings are discussed and summarized in Section 7.

Context and data

On July 9 2011 South Sudan became the 55th African independent state after seceding peacefully from the Republic of Sudan. Facing a history of a 50 year lasting conflict South Sudan slid back to instability after its peaceful independence process. This led to an internal displacement of circa two million, more than 15 percent of South Sudan's population(2017). Moreover, the conflict contributed to a deterioration of South Sudanese economic outcomes, with poverty rates reaching 82 percent in 2016, widespread severe food shortages and famine being declared in some counties in 2017 (Devi 2017). This makes well-targeted crisis response and aid allocation highly important.

The experiment sample includes 4145 IDP and 781 non-IDP households interviewed in 2017 in South Sudan across the High Frequency South Sudan Survey (HFSSS), the Crisis Recovery Survey (CRS), and the IDP Census and Sampling Study (IDPCSS). The CRS interviewed a representative sample of IDPs in IDP camps across South Sudan. In the same period the HFS conducted interviews across urban centers in seven of the ten former states (Figure B5-1). The IDPCSS conducted a census of all households in Juba POC1. The consumption modules in questionnaires administered to respondents in the three surveys were built in exactly the same manner so as to ensure comparability, and the fieldwork was implemented by the same organization. The only difference across the three surveys is the population that was sampled.

Figure B5-1: HFS and CRS coverage.



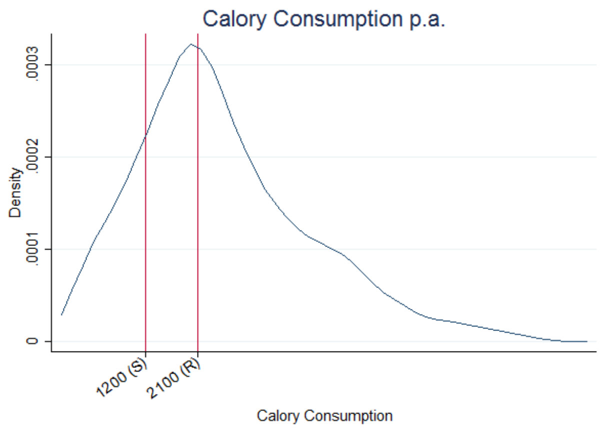
Note: The HFS interviewed a representative sample of households in urban centers in the states colored in blue in the map above. The CRS interviewed households in 4 of the largest IDP camps in South Sudan, denoted by red diamonds in the map. Major urban areas are indicated via black dots. The IDPCSS was conducted in the Juba POC1.

The conditions in camps do not allow for standard household surveys, hence, an alternative survey approach based on the Rapid Consumption Methodology was applied (Pape 2015). Here, only 30 / 25 food and non-food items are administered to all households. Additional 20 food and non-food items vary between households. More specifically, households are pre-assigned to one out of four sub-modules for food and non-food consumption (each containing 20 items). Neither the enumerators nor the respondents see the structure of the sub-modules, but the assigned items are asked in a categorically meaningful way (like cereals, fruits, etc).[62]

The data is used to construct four outcome measures. The surveys collect information on quantities in terms of (i) number of consumption items and (ii) kilograms. The quantities can be used to construct measures of (iii) monetary and (iv) caloric food consumption scaling the quantities with data on average prices and energy levels.[63] Though we are mainly interested in evaluating the impact of the nudges on the total consumption value - both in terms of money and food intake - these variables are difficult for respondents to falsify because these are *second-order* values that are calculated as a function of other variables, including consumption quantities and calories or prices that are in turn deflated. All of this adds noise to the answer provided by the respondent, and they depend in part on variables over which the respondent has no control. The consumption quantity in kilograms is a more direct measure of the quantity consumed as expressed by the respondent, and may lead to more accurate estimation of the impact of the nudges. Finally, counting the number of items may lead to an even more accurate measure, since the variable does not undergo any cleaning at all and is taken at face value. Furthermore, omitting an item is likely to be the easiest and quickest way for respondents to reduce the true value of the household's consumption.[64]

Poverty amongst IDP households is high, and 9 in 10 IDP households across South Sudan live under \$1.90 USD PPP (2011) per capita per day in 2017. IDP households in the sample interviewed for the experiment consume on average 333 SSP (2017) per capita per day. IDP households reported on average 6.63 core consumption items. These figures represent about 20 percent of core items asked to the households. Figure B5-2 visualizes that 39 percent of households report consumption below the recommended daily intake of 2,100 kcal (R) and 16 percent below the subsistence level of 1,200 kcal (S) (Ravallion and Benu 1994).

Figure B5-2: Density plot of value of core food consumption.



[62] Due to the survey method applied CRS surveys contain the core consumption module and one additional consumption module. The share in imputed consumption is on average 99.9 percent. IDP surveys contain due to the previously outlined time constraints only core consumption items. However, by design these items capture the lion's share of consumption (on average approx. 94 percent of total consumption in more comprehensive CRS surveys).
[63] For a description of the caloric intake measure, please consult the appendix.
[64] Note that the number of consumption items is not reported per capita as it does not increase proportionally with household size.

Note: Estimates presented in the figure above are not weighted and are representative only of IDP and non-IDP households surveyed in the study sample[65]

Approach and randomization

Our light-touch method introduce exogenous variation into the consumption module to try and tease out whether consumption might be underreported in IDP households. A prime is an environmental cue that unconsciously induces a subsequent cognition or behavior. For example, in studies with prisoners and bankers, participants who engage in activities that prime their identity report less accurately in behavioral experiments than participants who have not participated in priming activities (Cohn, Maréchal et al. 2010, Cohn, Fehr et al. 2014).[66] Nudges have been found to elicit more accurate responses during questionnaires (Rasinski, Visser et al. 2005, Vinski and Watter 2012).

To investigate whether consumption might be underreported in IDP households, we introduce exogenous variation into the consumption module. Households are randomly exposed to a bundle of light-touch measures. These include an emphasis on the importance of accurate answers at the beginning of the survey, a short fictional scenario which will require passing judgment on the behavior of one of the characters, and additional questions to tell when was the last time their household had a meal, forcing the respondents to explicitly report that they have not eaten in the last week.

1

- **Appeal to Honesty**
  The importance of accurate answers and the purpose of the survey will be reiterated.
- "Thank you for taking the time to speak to us. We really appreciate the time you are giving to participate in the survey. We encourage you to provide honest information. By participating in the survey and by providing accurate information, you are playing an important role in helping us understand the situation in South Sudan."

2

- **Prime to encourage more accurate reporting**
  At the beginning of the survey module concerning food consumption, the respondent will be given a short fictional scenario which will require passing judgement.
- "John asks his good friend Deng if he has some money that he can lend him to help him pay for medicine for his sick son. Deng has money but was planning to buy cigarettes with it. He lies and tells John that he has none. Is it okay for Deng to lie to John?"

3

- **Investigative Probing:**
  At the start of the survey module concerning food consumption, the respondent will be asked to tell when was the last time their household had a meal. This question will then also be asked for each of four major food categories: 'Bread and Cereals', 'Meat', 'Fruits', 'Pulses and vegetables'.
- E.g., "When was the last time that any of the household members had Bread and Cereals?"

Households are randomly exposed to behavioral treatments, in the form of a prime for more accurate reporting and investigative probing, to try and elicit more truthful answers from respondents. This

[65] We do not use weights throughout the study as the research hypothesis relates not to the average treatment effect, but particularly the primes' effectiveness at the tails of the distribution.
[66] Questionnaires confirmed that participants associate their identity with dishonesty.

way they do not constrain the choice frame, but rather alter the anchoring towards more truthful reporting (World Bank 2015).

The bundle of primes addresses different behavioral processes. (1) Appeals to honesty are a standard tool in surveys to increase data accuracy by relying on social approval (Talwar, Arruda et al. 2015). (2) Primes to encourage more accurate reporting induce unconscious cognitions, which are intended to affect subsequent behavior. When facing incentives to misreport, respondents would answer more accurately to sustain self-consistency. (3) Investigative probing puts a higher salience on the question. By asking for broader categories first, subsequent sub categories are put under more scrutiny. Self-consistency is reinforced by relating to a longer recall period of seven days.[67] While the appeal to honesty and the prime target intentional misreporting, investigative probing is addressing classical measurement error.

The sample was randomly selected into each treatment arms in two groups of approximately 50 percent, with 2,467 households in the control group and 2,459 in the treatment group. The randomization process was built into the CAPI (Computer Assisted Personal Interview) questionnaires administered in the surveys. As our research hypothesis suggests stronger effects of nudges for more vulnerable populations, we focus on IDPs for the main analysis. The availability of the HFS sample provides a comparison group of non-IDP households for the experiment, which will be used for robustness checks. The treatment and control groups are relatively balanced. There is a higher share of male headed households in the treatment group, which have also more members, though in practical terms these differences are relatively small. As gender of the household head and household size are potentially correlated with poverty, these variables are included in the regression models and interacted with the treatment to control for potential impacts (Lanjouw and Ravallion 1995).

Table B5-1: Balance across treatment and control arms (IDP sample).

	Control	Treatment	Difference, p-value
Household size	4.835	5.098	0.003***
	(0.060)	(0.064)	
Gender of household head	0.492	0.448	0.005***
	(0.011)	(0.011)	
Literacy of household head	0.507	0.529	0.155
	(0.011)	(0.011)	
Household head completed some primary school	0.540	0.563	0.133
	(0.011)	(0.011)	
Is the household head employed	0.328	0.319	0.555
	(0.010)	(0.010)	
Share of children in household	0.364	0.373	0.309
	(0.006)	(0.006)	
Share of elderly in household	0.011	0.010	0.582
	(0.002)	(0.001)	
First Component of Asset Principal Component Analysis	-0.126	-0.194	0.162
	(0.037)	(0.032)	
N	2079	2066	
Proportion	0.502	0.498	

*Standard errors in parentheses; **p<0.05, ***p<0.01*

[67] The methodological appendix provides an overview of the relevant questions in the food consumption module.

Empirical Strategy

To assess the effect of our prime on reporting behavior, we can formulate following simple regression equation.

$$Y_i = \beta_0 + \beta_1 T_i + \boldsymbol{T_i} * \boldsymbol{X_i}\beta_2 + X_i\beta_3 + \gamma_s + \alpha_t + \varepsilon_i, \tag{3}$$

where Y_i is the log of the outcome variable. Across different models we estimate the effect for (i) the number of consumption items consumed [referred to in the regression equation as Cons. Num.], (ii) consumption quantity per adult equivalent (in kilograms) [Cons. Quant.], (iii) monetary consumption value per adult equivalent [Cons. Val.] and (iv) daily caloric intake per adult equivalent [Cons. Cal.].

Our main treatment variable T_i is a dummy variable which takes the value of 1 if the household i was assigned to the treatment group. γ_s indicates a set of camp fixed effects, α_t are month fixed effects, and ε_i is the idiosyncratic error term. X_i denotes a vector of control variables generally associated with consumption, including household size, the gender of the household head, and the proportion of children (under 18) in the household. Moreover, we add an asset index based on the first component of a principal component analysis (Filmer and Pritchett 2001, McKenzie 2005).[68] The model will be estimated with and without controls to check the impact they may have. As the treatment might interact with the unbalanced covariates, it makes sense to add to the regression $T_i * X_i$, the interaction of the unbalanced controls with the treatment variable (Lin and Green 2016, Baranov, Bhalotra et al. 2017).

It is expected that the respondents who will be affected by the treatment are respondents that would otherwise misreport and, hence, a more likely to be at the extremes of the distribution.[69] Therefore, we complement our analysis with a quantile regression approach. The idea of the quantile regression framework, which was introduced by Koenker and Bassett Jr (1978), is to take the entire distribution of the dependent variable into account by estimating several regressions, which put more weight to the quantile of interest. The underlying minimization problem can be stated as follows:

$$Q(\theta) = arg \min_{\tau} \sum_{i:y_i>\tau} \theta|y_i - \tau| + \sum_{i:y_i\leq\tau} (1-\theta)|y_i - \tau|, \tag{4}$$

where θi is the quantile of interest and the weighted sums of deviations $|y_i - \tau|$ of the outcome per quantile. Minimizing the latter, differential effects conditional on the quantile of the dependent variable are obtained. Further, it has the advantage of being less prone to outliers and non-normality of the error term. For our purpose, quantile regressions offer the advantage that they are more flexible than simple interactions with poverty lines, which would be endogenous to consumption levels.

Results

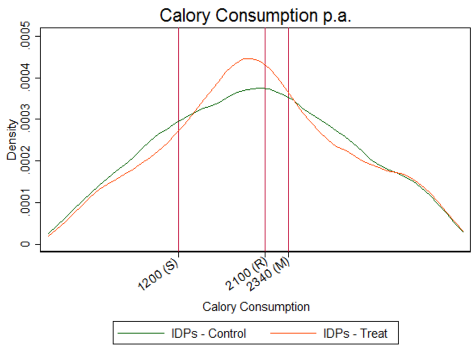
There is a slight indication that the treatment may have worked, based on consumption distributions across treatment and control group. The consumption distribution shown in Figure B5-3 shows a slight difference in caloric consumption between IDP households in the treatment group and the control group, though this is apparent only at lower levels of consumption, i.e. below the subsistence level of 1,200 kcal. The median of calorie consumption is well above the recommended daily intake. However, still a substantial part of the distribution of 16 percent reports below the subsistence level and 40

[68] As assets (bikes, fans, rickshaws etc.) can be more easily surveyed by enumerators, those are likely to capture parts of the household wealth.

[69] Although these hypotheses were not pre-registered, they are based on theoretical considerations about the mechanisms of the underlying behavioral primes.

percent below the recommended daily intake.[70] Hence, the prime would also be relevant in the adult equivalent setting to achieve more precise reporting, which is analyzed in our regression framework subsequently. Taking into account the finding that consumption levels are lower than to be expected, the most relevant treatment effects can be expected at the left tail of the distribution.

Figure B5-3: Caloric consumption p.c. (adult equivalents).



Note: The underlying data is based on per adult equivalents. Caloric consumption levels are labeled in the following graph as S subsistence equivalent (1200 kcal p.c.), R recommended daily intake (2100 kcal p.c.) and M the median (2340 kcal p.c.).

Regression results

In order to test for the influence of control variables, the regressions are estimated with and without control variables. When not conditioning on control variables, the results indicate only a significant treatment effects for the number of consumption items in Column (1). This outcome measure would be easiest to falsify as it does not undergo further cleaning, e.g., in terms of deflation or calorie scaling. When adding further controls, coefficients turn larger and imply treatment effects of 6-14 percent. The interactions of the treatment and the asset index as well as household size have negative and significant coefficients in line previous work. For example, larger households are on average more prone to consumption poverty and might react differentially (Lanjouw and Ravallion 1995). The simple treatment indicators also turn significant for the kilogram consumption quantities in Column (4) and the monetary consumption value in Column (6).[71] Yet, our main indicator of interest, the caloric food consumption remains unaffected. This is in line with our hypothesis that the average treatment effect should be limited and rather uninformative as the primes are expected to particularly affect misreporting at the tails. For this purpose, a quantile regression analysis is taken out to provide more nuanced estimates, subsequently.

To capture this heterogeneity across consumption levels, quantile regressions are applied. Results are shown in Figure B5-4.

[70] Compared to the monetary consumption levels, the calory consumption p.a. seems rather high. This is partly attributable to the fact that IDP's consumption focuses on energy intensive products, where cooking oil and sorghum constitute 45 percent of food expenditure. If we contrast the consumption shares with non-IDPs, we find that although the diet of non-IDPs is less energy intensive, it comprises a higher variety.

[71] Unintuitively, with regard to the monetary consumption values in column (5), negative coefficients are estimated, contradicting a higher consumption quantity. In line with other studies, this could be explained by larger households buying larger quantities and, hence, consuming more while paying lower bulk purchasing prices.

Table B5-2: Results from quantile regressions of different outcome variables.

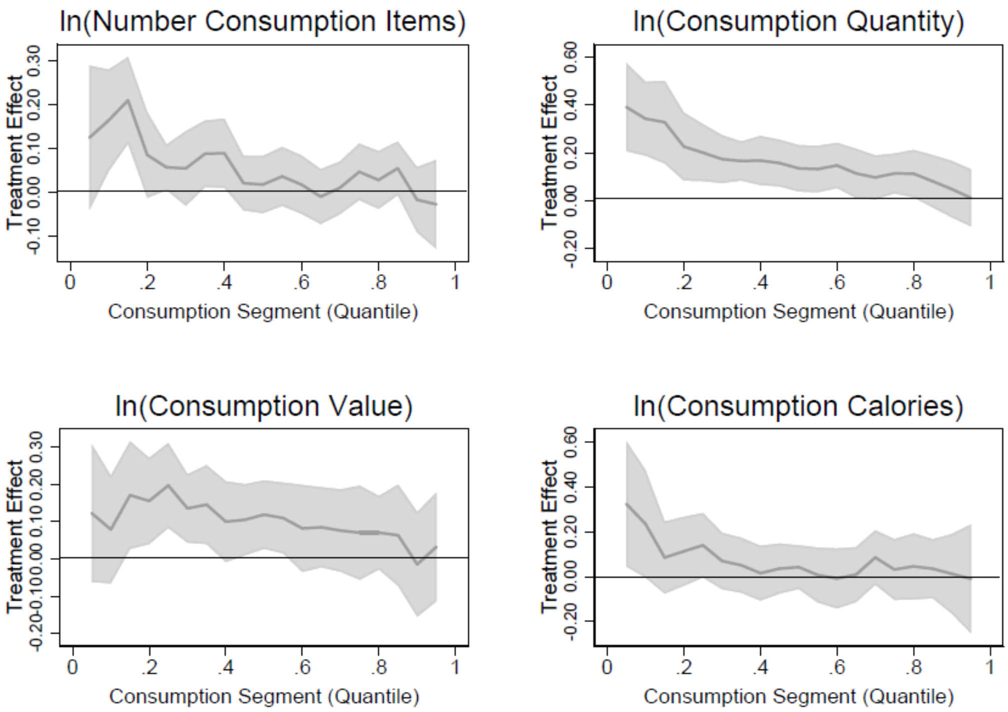
	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	0.165**	0.342***	0.079	0.235*
	(0.064)	(0.079)	(0.068)	(0.127)
Q0.25	0.058**	0.201***	0.198***	0.140*
	(0.028)	(0.067)	(0.053)	(0.080)
Q0.5	0.018	0.136**	0.119**	0.042
	(0.032)	(0.056)	(0.050)	(0.062)
Q0.75	0.047	0.114**	0.071	0.032
	(0.034)	(0.050)	(0.051)	(0.067)
Q0.9	-0.016	0.049	-0.015	0.013
	(0.028)	(0.050)	(0.054)	(0.064)
Observations	3,955	3,955	3,955	3,955
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

The priming significantly increases reported consumption in lower quintiles. Significant treatment effects occur mainly for the number of consumption items and the quantities in kilogram. Monetary and caloric consumption measures are less strongly affected (Figure B5-6). The latter might also be less susceptible to deliberate misreporting as they depend in part on variables over which the respondent has no control as the pure consumption quantities are scaled by calorie levels or deflated.[72]

[72] Conditional quantile regressions are sometimes considered as uninformative as they describe the effect on the distribution rather than on the individual. Hence, we also consider unconditional quantile regressions in the appendix. Results are robust and support an upward shift in lower quantiles of the outcome variables.

Figure B5-4: Treatment effects across quantiles.



Note: Treatment effects and confidence intervals plotted for different quantiles.

Ultimately, we are interested in the question if the prime is sufficiently strong to shift a significant share of the distribution to more credible consumption levels both in terms of monetary and caloric food consumption. For this purpose, we construct four dichotomous indicators. Those are equal to one if (i) respondent households surpass the caloric subsistence level of 1200 kcal or (ii) the recommended level of caloric intake of 2100 kcal. Two further dummies are created at (iii) 66.66 percent and (iv) 100 percent of a normalized poverty line, which is scaled by the fact that only core consumption items were assessed consistently across all surveys. Table B5-3 depicts results for the three threshold using model (3). Although the coefficients are mostly positive, only two coefficients turn significant in Column (2) and (3). Therefore, the results stress the nuanced effect of the prime, which only affects certain strata of the population.

Table B5-3: Results using poverty thresholds, model (2) and (3).

VARIABLES	(1) **>1200kcal**	(2) **>2100kcal**	(3) $>\left(\frac{2}{3}\right) \boldsymbol{Poverty\ Line}$	(4) **>Poverty Line**
Treatment	0.010 (0.027)	0.069* (0.037)	0.063* (0.037)	0.029 (0.036)
Observations	3,955	3,955	3,955	3,955
R-squared	0.067	0.098	0.118	0.135
State FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Controls Interacted	YES	YES	YES	YES

Treatment heterogeneity and robustness

Heterogenous effects:

If the primes would reduce misreporting, stronger effects are to be expected among subpopulations that have higher incentives to misreport, e.g., aid-reliant IDPs. In order to assess this channel more thoroughly, (i) heterogenous effects are estimated contingent on aid reliance and (ii) the sample is compared to a non-IDP comparison group.

Parts of the respondents from the CRS and HFS were also interviewed with regard to their previous support through UN agencies. This dummy indicator can be used for an assessment of heterogenous effects.[73] The model is analogous to equation (3), where we add UN assistance as a further control variable as well as an interaction term of UN assistance with the behavioral treatment. The results indicate no clear pattern (Table B5-2). Only for the number of consumption items a positive significant coefficient is found. The significant positive interaction of the treatment and previous aid exposure could be treated as some weak evidence that the prime is more effective for aid exposed IDPs, but should not be overstated due to the non-significance for the other three outcomes of interest.

[73] The results can only be interpreted as an explorative analysis as UN assistance was not balanced across treatment and control groups, where treatment households have a higher probability of being previously exposed to aid.

Table B5-4: Channel – UN assistance.

	(1) ln(Cons. Num.)	(2) ln(Cons. Quant.)	(3) ln(Cons. Val)	(4) ln(Cons. Cal.)
Treatment	0.100	0.195**	0.171*	0.105
	(0.066)	(0.080)	(0.081)	(0.087)
UN Assistance	-0.028	-0.065	-0.152***	-0.143***
	(0.038)	(0.045)	(0.043)	(0.046)
Treatment*UN Assistance	**0.104****	**-0.059**	**0.016**	**0.011**
	(0.051)	**(0.060)**	**(0.061)**	**(0.064)**
Observations	2,204	2,204	2,204	2,204
R-squared	0.38	0.086	0.098	0.108
State FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

*Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.*

The non-IDP subsample offers an interesting opportunity to assess the robustness of the results. Constraining the sample only on non-IDPs, the pattern of positive and significant treatment coefficients in the lower quantiles vanishes, except for Column (1). This could be interpreted as evidence that the light-touch method applied are more efficient for the vulnerable IDP population, which has higher incentives to indicate need than the non-IDPs. This would be in line with previous studies (e.g. Cilliers, Dube et al. 2015) suggesting a high degree of social desirability bias in the setting of foreign assistance. Specifically, the populations exposed to development aid, in our setting the IDPs, would be more likely to provide socially desirable answers to signal their “worthiness” for assistance. This corresponds to Table B5-2, providing some weak evidence that the primes are more effective for respondents relying on UN aid. It would be of particular interest to examine those heterogenous effects based on more fine-grained data on neediness and degree of aid reliance of recipients. For this purpose, however, a “true” benchmark would be needed. As administrative data is non-existent or of poor quality, an alternative for future research might be to build on measures from qualitative work as suggested by Blattman, Jamison et al. (2016). Moreover, one should be careful to draw too strong conclusions from these results as the number of observations is limited in this comparatively small sub-sample.

Table B5-5: Quantile Regressions – reduced sample (only non-IDPs).

	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	-0.027	-0.069	-0.026	0.032
	(0.079)	(0.102)	(0.110)	(0.113)
Q0.25	0.148**	-0.052	0.012	-0.057
	(0.073)	(0.095)	(0.107)	(0.122)
Q0.5	0.067	-0.041	-0.032	0.044
	(0.067)	(0.081)	(0.100)	(0.100)
Q0.75	-0.071	-0.072	-0.015	-0.052
	(0.054)	(0.080)	(0.092)	(0.080)
Q0.9	-0.041	0.157	0.074	0.119
	(0.047)	(0.105)	(0.144)	(0.127)
Observations	780	780	780	770
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

*Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.*

Robustness:

In line with hardly credible low consumption levels, misreporting could be considered to be more prevalent at the tails of the distribution, hence, among the extreme values. On the one hand, it makes, thus, sense to consider those outliers. On the other hand, it is problematic to base the inference mainly on those extreme values. Ideally, one would know how to distinguish the intentionally misreported outliers and the ones that are caused by errors in reporting or data entry. The log normalization in the main analysis is chosen as a compromise of keeping most data possible, but making estimates less susceptible to outliers. This suggests two natural robustness checks: (i) in a more liberal setting, the outcomes in levels are used and (ii) in a more conservative setting, the outliers at the 5th and 95th percentile are discarded. Regression results using the levels are depicted in Table B5-6.[74]

[74] As scaling of the outcome variables is different – e.g., the outliers with regard to consumption quantity in kilograms might not correspond to the consumption quantity in calories – the outliers for one measure do not always correspond to outliers in the other measure. In order to guarantee that we still base the inference on the same observations, outliers from all corresponding variables are dropped, which explains that the resulting sample is smaller than 90 percent of the full sample.

Table B5-6: Quantile Regressions – outcomes in levels.

	(1)	(2)	(3)	(4)
Outcome Variables	Cons. Num.	Cons. Quant.	Cons. Val.	Cons. Cal.
Q0.1	0.544**	0.741***	9.440	229.126*
	(0.254)	(0.173)	(13.305)	(136.013)
Q0.25	0.298*	0.675**	60.585***	179.447
	(0.156)	(0.224)	(16.261)	(157.584)
Q0.5	0.151	0.638*	49.404**	197.589
	(0.194)	(0.280)	(20.477)	(192.043)
Q0.75	0.341	0.700**	19.499	281.317
	(0.246)	(0.339)	(30.762)	(286.250)
Q0.9	-0.077	0.609	-30.117	-279.159
	(0.333)	(0.540)	(41.581)	(284.569)
Observations	3,955	3,955	3,955	3,955
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

*Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.*

Table B5-7 depicts the results without outliers and indicates a slightly less nuanced pattern. In line with our hypothesis of stronger misreporting tendencies on the extremes, Column (1) indicates significant treatment effects at the 10th and 25th percentile. Although significant treatment effects among higher quintiles can be found in Column (2) and (3), the coefficients for the 25th percentile are quantitatively larger. Finally, with regard to caloric consumption in Column (4) statistical significance vanishes, but the largest coefficient is to be found in the 10th percentile. Hence, although the pattern gets weakened when excluding outliers, the prime still significantly affects the reported consumption quantities with stronger effects in the lower quantiles.

Table B5-7: Quantile Regressions – without outliers.

	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	0.124**	0.106	0.085	0.058
	(0.049)	(0.067)	(0.064)	(0.091)
Q0.25	0.045*	0.139**	0.162***	0.042
	(0.027)	(0.055)	(0.044)	(0.077)
Q0.5	0.000	0.065	0.119**	0.037
	(0.032)	(0.050)	(0.046)	(0.059)

Q0.75	0.028	0.077*	0.086*	0.049
	(0.032)	(0.043)	(0.048)	(0.063)
Q0.9	-0.027	0.064	0.027	0.039
	(0.023)	(0.039)	(0.049)	(0.051)
Observations	3,711	3,605	3,576	3,500
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

*Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.*

Regression techniques, which are based on assumptions for large samples drawn from finite populations, are often not suitable in the context of randomized experiments (Heß 2017). The uncertainty is in this case not coming from the sampled units observed, but from the fact that we can only observe one of the potential outcomes, which is due to the treatment applied to the different units (Athey and Imbens 2017). One approach would be to take the randomization explicitly into account and follow R.A. Fisher's idea of statistical inference via permutation tests of treatment allocation (Young 2016). The idea is to assume uncertainty about the treatment allocation and compare the actual treatment allocation to re-randomizations. The results of this exercise are depicted in Table 11, underscoring the robustness of the main results.

Table B5-8: Results from baseline estimation, model (2), with random-inference based p-values.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	ln(Cons. Num.)	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Val.)	ln(Cons. Cal.)	ln(Cons. cal.)
Treatment	0.0348** (0.0200)	0.0614** (0.0560)	0.0281 (0.1300)	0.1374*** (0.0020)	-0.0178 (0.2820)	0.0812** (0.0340)	0.0189 (0.4980)	0.0007 (0.9940)
Observations	3,995	3,995	3,995	3,995	3,995	3,995	3,995	3,995
R-squared	0.0012	0.2744	0.0003	0.0805	0.0006	0.0725	0.0001	0.1232
State FE	NO	YES	NO	YES	NO	YES	NO	YES
Month FE	NO	YES	NO	YES	NO	YES	NO	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
Controls Interacted	YES	YES	YES	YES	YES	YES	YES	YES

*Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1.*

Discussion and Conclusion

The conflict in South Sudan displaced circa two million persons, constituting more than 15 percent of the country's population (UN OCHA 2017). Moreover, the majority of population is living in extreme poverty. Humanitarian crises like the one in South Sudan ask for well targeted policy responses, which address the population strata with the highest need first. This, however, is no arbitrary task as aid allocation mechanisms might set adverse incentives to underreport. Even given the extreme context, surveyed consumption levels indicate an unusually high share below subsistence levels.

For this purpose, this study assesses the effectiveness of a bundle of light-touch measures. In line with our hypothesis we find significant treatment effects, which cluster in lower (potentially underreported) consumption quintiles. Moreover, effects are stronger for the number of consumption items than for monetary consumption quantities, where former are more susceptible to deliberate misreporting. Furthermore, the significant treatment effects are driven mainly by the vulnerable IDP subpopulation, which are more likely to be in need for foreign aid. Primes can, hence, help to improve data accuracy and inform policy to develop durable solutions. However, results should be taken with a grain of salt as it is not possible to compare the reported consumption outcomes to more objective consumption data. Although the mortality rates among IDPs suggest that starvation is not happening systematically across the country, the precarious situation calls for further scrutiny. Before adjusting poverty estimates a thorough comparison with more “objective” data from administrative, anthropometric or observational sources is needed. While this type of data was not available in IDP camps due to the fragile context, future research could validate this finding in other settings.

Moreover, unbundling the primes in different treatment arms could help to shed light on the underlying causal mechanisms. The underlying design of one treatment and control arm does not allow for further disentangling the results. However, if classical measurement error would be affected only, treatment effects of the primes should be uniform. In contrast, heterogenous effects across quantiles suggest that the targeting of intentional misreporting via the appeal to honesty and prime to report more accurately would be the driver of our results. In order to design more effective primes, disentangling the pathways and trying different combinations could be a beneficial way forward. Our research can be considered as an early step to employ priming for better targeted policy responses in challenging contexts, which might not only be applicable in South Sudan, but also in other contexts facing humanitarian crises.

Appendix

<u>Construction of the caloric food intake measure:</u>

While monetary poverty lines are a key metric, when identifying the poor, caloric food poverty headcounts are of equal relevance in our context. We create a food intake approximation by multiplying the quantities of food items from the core consumption survey with average caloric values of these products. The caloric intake c_i of household i is estimated as follows: $\mathrm{caloric\ intake}_i = \frac{1}{\mathrm{hhsize}_i} \sum \mathrm{item}_j * \mathrm{calories}_j * \mathrm{quantity}_{ij}$.

Forty-three percent of household members are children, who naturally have lower consumption levels than adults. We can account for this by using adult equivalents (AE) and rely on OECD scales, which scales consumption of additional adults per household by factor 0.7 and of children by factor 0.5 (Haughton and Khandker 2009).

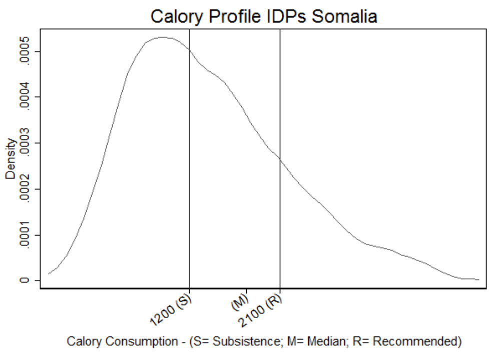
$$AE = 1 + 0.7(N_{adult} - 1) + 0.5N_{children}$$

$$AE = 1 + 0.7(N_{adult} - 1) + 0.5N_{children}$$

Caloric food poverty in Somalia:

Using the same approach, we derive caloric food intake measures, which motivated the notion that misreporting might be prevalent.

Figure B5-5: Calorie consumption - IDPs Somalia.)



Balance across survey strata:

Table B5-9: Treatment distribution by survey strata.

		Treatment with light-touch measures		
	State/Camp	Control	Treatment	Total
		No.	No.	No.
CRS	JubaPOC	223	263	486
	Wau	294	284	578
	Bor	292	257	549
	Bentiu	294	297	591
IDPCSS	Juba POC1 – IDPCSS	976	965	1,941
HFS - Wave 4	Warrap	60	60	120
	Northern Bahr el Ghazal	50	61	111
	Western Bahr el Ghazal	62	58	120
	Lakes	50	54	104
	Western Equatoria	54	50	104
	Central Equatoria	38	40	78
	Eastern Equatoria	74	70	144
	Total	2467	2459	4926

Reaction to the light-touch method:

An overwhelming majority of respondents answered in a positive manner to the fictional scenario. Less than 10 percent of respondents answered that it is ok for the character in the fictional scenario to lie to his friend. [75]

Prime to encourage more accurate reporting: I will give you a little scenario and would like to know what you think: John asks his good friend Deng if he has some money that he can lend him to help him pay for medicine for his sick son. Deng has money but was planning to buy cigarettes with it. He lies and tells John that he has none. Is it okay for Deng to lie to John?

[75] *The respondents, who find a lie inappropriate, have a higher share of male and unemployed household heads. Moreover, IDPs have a significantly lower probability to find a lie acceptable.*

	Percent	N
Yes, it is okay for Deng to lie to John.	8.8	217
No, it is not okay for Deng to lie to John.	91.2	2,240
Total	100	2,457

This might be interpreted in two ways. First, it might point to a low fraction of respondents, who would be willing to lie, which would reduce the potential of finding significant treatment effects. Second, it could indicate that the prime would increase the propensity to report truthfully. However, as studies suggest a high social desirability bias in the aid allocation setting (Cilliers, Dube et al. 2015, Stecklov, Weinreb et al. 2017), implications should not be drawn too early and will be discussed in subsequent sections.

Appropriateness of lying:

It is puzzling that IDPs have on average a lower probability to report that they would find a lie appropriate when compared to non-IDPs (see Table B5-10). This is in line with more pro-social preferences of conflict affected populations found by Voors, Nillesen et al. (2012). However, this might be misleading, as the analysis of channels indicates that the significant treatment effects are attributable to the IDP subsample, which seem to be more likely to misreport.

Table B5-10: Distribution of respondents, who would find a lie (in-)appropriate.

	Yes, it is okay for Deng to lie to John.	No, it is not okay for Deng to lie to John.	Overall	(1) vs. (2), p-value
Household size	5.041	5.123	5.119	0.696
	(0.228)	(0.061)	(0.059)	
Gender of household head	0.327	0.456	0.445	0.000***
	(0.032)	(0.011)	(0.010)	
Literacy of household head	0.544	0.532	0.533	0.734
	(0.034)	(0.011)	(0.010)	
Household head completed some primary school	0.565	0.568	0.568	0.919
	(0.034)	(0.010)	(0.010)	
Is the household head employed	0.184	0.279	0.270	0.003***
	(0.026)	(0.009)	(0.009)	
Share of children in household	0.315	0.356	0.353	0.042*
	(0.019)	(0.006)	(0.006)	
Share of elderly in household	0.014	0.015	0.015	0.890
	(0.007)	(0.002)	(0.002)	
Level of Education of Household Head	2.060	1.967	1.975	0.205
	(0.075)	(0.022)	(0.021)	
non-IDP Population	0.212	0.155	0.160	0.029*
	(0.028)	(0.008)	(0.007)	
N	217	2238	2455	
Proportion	0.088	0.912	1.000	

*** p<0.01, ** p<0.05, * p<0.1.

Robustness of results using an unconditional quantile regression:

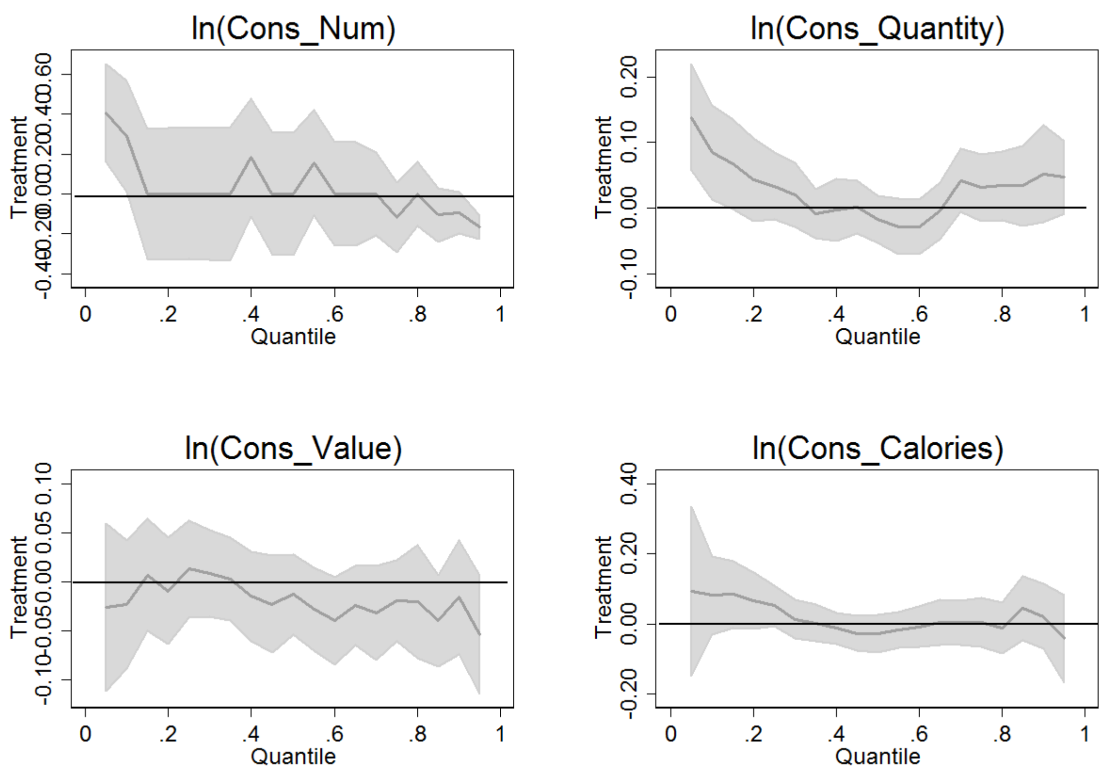
Conditional quantile regressions are sometimes critiqued on the ground that they would consider the treatment effect conditional on the distribution and not on the individual ranking. Therefore, we also replicate the main regressions within an unconditional quantile regression framework (Firpo, Fortin et al. 2009). Especially, the quantities of consumption items and kilograms experience positive treatment effects in lower quantiles. Although higher quantiles are affected as well in Column (2), the largest effects can be found in the 10% quantile, which would be consistent with the hypothesis of more accurate answers among potentially under reporting households.

Table B5-11: Results from unconditional quantile regressions of different outcome variables.

Outcome Variables	(1) ln(Cons. Num.)	(2) ln(Cons. Quant.)	(3) ln(Cons. Val.)	(4) ln(Cons. Cal.)
Q0.1	0.105**	0.259***	0.076	0.134
	(0.046)	(0.090)	(0.079)	(0.145)
Q0.25	0.078**	0.210***	0.169***	0.075
	(0.032)	(0.067)	(0.062)	(0.077)
Q0.5	0.004	0.104**	0.118**	0.071
	(0.035)	(0.053)	(0.056)	(0.063)
Q0.75	-0.012	0.132**	0.067	0.025
	(0.040)	(0.066)	(0.059)	(0.089)
Q0.9	0.024	0.075	-0.003	0.062
	(0.044)	(0.087)	(0.077)	(0.119)
Observations	3,955	3,955	3,955	3,955
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

*Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.*

Figure B5-6: Treatment effects across quantiles (unconditional quantile regressions).



Note: Treatment effects and confidence intervals plotted for different quantiles.

Robustness of results in extended IDP and non-IDP subsample

Results from a quantile regression for an extended sample of IDPs and Non-IDPs correspond to the previously found larger coefficients in the lower quintiles. Coefficients are of similar size and the pattern remains qualitatively similar. However, statistical significance is reduced in column (3) and (4) with regard to the indicators that are measured with more noise (e.g., monetary consumption values and caloric consumption).

Table B5-12: Quantile Regressions – extended sample IDPs and Non-IDPs.

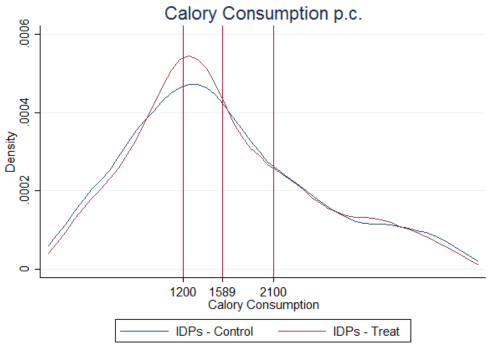
	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	0.136***	0.254***	0.072	0.153
	(0.049)	(0.058)	(0.067)	(0.094)
Q0.25	0.085***	0.123**	0.085	0.044
	(0.031)	(0.049)	(0.052)	(0.064)
Q0.5	0.024	0.088*	0.092**	0.037
	(0.029)	(0.049)	(0.043)	(0.053)
Q0.75	0.018	0.094**	0.052	0.028
	(0.031)	(0.042)	(0.044)	(0.052)
Q0.9	-0.019	0.058	-0.026	0.035
	(0.024)	(0.050)	(0.048)	(0.050)
Observations	4,735	4,735	4,735	4,735
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

*Robust standard errors in parentheses (White 1980): *** p<0.01, ** p<0.05, * p<0.1.*
Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

Robustness to per capita instead of per adult equivalents:

There is some uncertainty about the per adult equivalent scaling in the data. Ideally the distribution might be estimated from more fine-grained data on the intra-household consumption distribution. This is often not available, and, as Deaton and Zaidi (2002) summarize, “no satisfactory” scaling method is identified so far. Therefore, the OECD scaling methodology is still frequently used (e.g. Euler, Krishna et al. 2017, Van Den Broeck and Maertens 2017). Yet, one might be concerned that the main results are not robust to different scaling. Therefore, we construct our outcome measure alternatively using agnostic per capita scales. In line with the low consumption levels, the median of per capita calorie intake (1,589 kcal. per day) is well below the recommended daily intake of 2,100 kcal (Ravallion and Benu 1994). Almost one third of respondents (30.1 percent) report a calorie intake below the daily subsistence level of 1,200 kcal per day. In contrast, several respondents report overly high consumption levels, which surpass conventional consumption levels by far (> 4,000 kcal. per day). This supports previous evidence that misreporting is prevalent. As with the number of consumption items, the graph indicates that there is a slight shift in reported consumption among the treated regarding very low consumption levels.

Figure B5-7: Calory Consumption p.c.



The results of a quantile regression using agnostic per capita scales indicate that the treatment effects remain stable and respondents would report statistically significantly higher quantities in Column (1) and Column (2) if treated. Hence, scaling does not explain our results, but is a factor to take into account, when interpreting the outcomes.

Table B5-13: Results from quantile regressions of different outcome variables (pc scales).

Outcome Variables	(1) ln(Cons. Quant. p.c..)	(2) ln(Cons. Val. p.c.)	(3) ln(Cons. Cal. p.c.)
Q0.1	0.358***	0.040	0.207
	(0.087)	(0.068)	(0.135)
Q0.25	0.161***	0.160**	0.076
	(0.059)	(0.053)	(0.081)
Q0.5	0.124***	0.079	0.073
	(0.057)	(0.054)	(0.066)
Q0.75	0.050	0.055	0.021
	(0.049)	(0.054)	(0.071)
Q0.9	0.057	-0.003	0.027
	(0.063)	(0.051)	(0.081)
Observations	3,955	3,955	3,955
Month FE	YES	YES	YES
State FE	YES	YES	YES
Controls	YES	YES	YES
Interacted Controls	YES	YES	YES

Table B5-14: Results – full set of (interacted) controls.

VARIABLES	(1) ln(Cons. Num.)	(2) ln(Cons. Quant.)	(3) ln(Cons. Val.)	(4) ln(Cons. Cal.)
Treatment	0.061*	0.137***	0.081**	0.001
	(0.033)	(0.042)	(0.039)	(0.067)
Household size	0.033***	-0.024***	-0.044***	-0.105***
	(0.005)	(0.005)	(0.005)	(0.008)
Female Gender of household head	0.009	0.043*	0.022	-0.047
	(0.018)	(0.026)	(0.025)	(0.039)
Share of children in household	0.106**	0.243***	0.190***	0.027
	(0.046)	(0.053)	(0.054)	(0.085)
1st component of asset PCA	0.026***	0.013	0.022***	0.039***
	(0.005)	(0.008)	(0.008)	(0.012)
Treatment * Household Size	-0.008	-0.015**	-0.009	0.006
	(0.006)	(0.007)	(0.007)	(0.011)
Female Gender of household head #0b.treat	-0.020	-0.007	-0.008	-0.059
	(0.028)	(0.036)	(0.036)	(0.054)
1.treat# Share of children in household	-0.001	-0.066	-0.107	-0.016
	(0.059)	(0.074)	(0.073)	(0.116)
1.treat#1st component of asset PCA	-0.017**	-0.003	-0.007	-0.003
	(0.007)	(0.011)	(0.010)	(0.016)
Observations	3,955	3,955	3,955	3,955
R-squared	0.274	0.073	0.080	0.123
State FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES

*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.*

Household Size and purchasing prices per kilo price:

In order to find out if the data bores out the pattern that larger households pay lower prices, e.g., due to bulk purchasing, we regress the log of the reported price on household size, state, month and consumption good specific fixed effects.

$$\ln(\text{price}_i) = \propto + \beta_1 \text{hhsize}_i + \gamma_s + \delta_t + \theta_g + \varepsilon_i$$

The results are depicted in Table B5-15 and indicate a negative average correlation. This supports the choice of interacting unbalanced controls with the treatment indicator.

Table B5-15: Correlation of household size and purchasing prices per kilo.

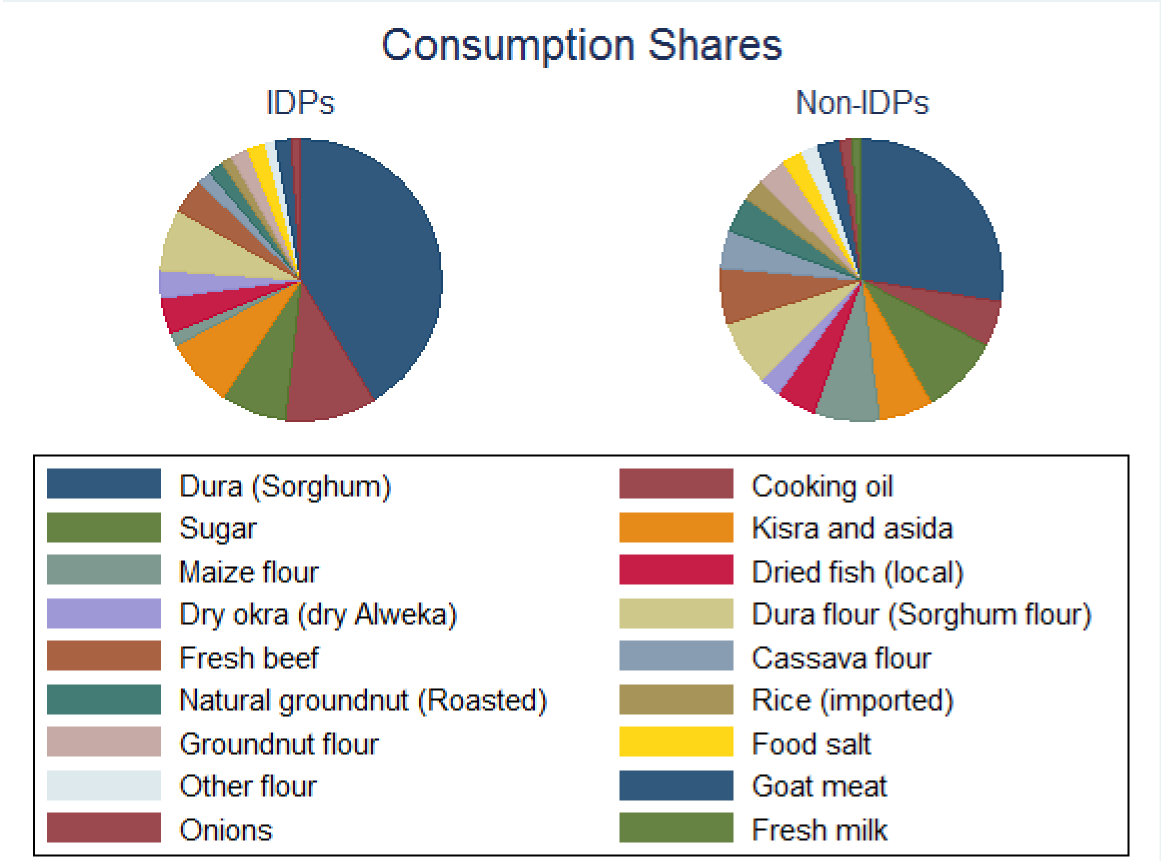
VARIABLES	(1) lnprice
Household size	-0.003** (0.001)
Observations	24,409
R-squared	0.548
State FE	YES
Month FE	YES
Item FE	Yes

Robust standard errors in parentheses (White 1980):
*** p<0.01, ** p<0.05, * p<0.1

Consumption shares of IDP and non-IDP populations:

Figure B5-8 describes the consumption shares of IDPs and non-IDPs. While the figure shows that the diet of IDPs is slightly less diverse than the diet of non-IDPs, it is also revealed that large shares of IDP budget are spent on goods, which offer a high caloric intake per SSP spent, e.g., sorghum and cooking oil. The high energy content of IDP's food consumption also corresponds to the counter intuitive pattern found in the data, where IDPs consume less than non-IDPs in terms of monetary value, but more in terms of caloric food intake.

Figure B5-8: Consumption Shares (SSP values).



Note: The figure lists the consumption shares of items, which constitute at least 1% of household consumption.

6. Estimating Poverty in a Fragile Context - The High Frequency Survey in South Sudan[76]

Utz Pape and Luca Parisotto[77]

Introduction

Civil war broke out across The Republic of South Sudan in December 2013 only two years after gaining independence on the 9th of July 2011. The South Sudanese conflict has since continued to escalate, resulting in a large-scale humanitarian crisis where more than a third of the population has been forcibly displaced (Pape, Parisotto et al. 2018).[78] Given the extremely difficult context, very little was known about welfare and livelihoods during the early years of the country's independence in 2011.[79] The last nationally representative household survey measuring consumption and poverty was conducted as far back as 2009. To fill this data gap, the High Frequency South Sudan Survey (HFS), implemented by the National Bureau of Statistics (NBS) in collaboration with the World Bank and funded by the U.K. Department for International Development, conducted several waves of representative surveys across seven of the ten former states between 2015 and 2017 (Appendix A). In the period prior to and during the first wave of the HFS in 2015, conflict had primarily been concentrated in the Greater Upper Nile region (Figure B6-12 in Appendix D).[80] This period of relative stability across the remaining Greater Equatoria and Greater Bah El-Ghazal regions allowed the preparation and relatively calm implementation of Waves 1 and 2 of the country in 2015 and early 2016.

In summer 2016, clashes broke out in Juba. The escalation of the conflict coincided with the beginning of the implementation of Wave 3 of the HFS, a second urban-rural representative wave measuring consumption and poverty. The third wave of the HFS provides a relatively rare and extremely valuable glimpse of trends in welfare, consumption, and poverty in a country going through a period of upheaval. Indeed, the South Sudanese economy has since displayed all the characteristics of a war economy, including severe output contraction, rapid currency devaluation, and soaring inflation (IMF 2016, FAO, IFAD et al. 2017). Unsurprisingly, driven by these powerful shocks the incidence of poverty has risen to extremely high levels. In 2016, the HFS estimated that more than 4 in 5 people across seven of the ten former states were living under the international poverty line of US$ 1.90 PPP 2011 (82 percent). Such high levels of deprivation are not merely a direct result of the crisis but also reflect a history of instability, characterized by a poorly functioning state and a lack of institutional services provision (de Waal 2014, de Vries and Schomerus 2017, World Bank 2017). In 2017 South Sudan

---

[76] UP developed the research question and designed as well as supervised the field work. LP and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

[77] Authors in alphabetically order. Corresponding author: Utz Pape (upape@worldbank.org). The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. The authors would like to thank Kristen Himelein, Syedah Iqbal and Ambika Sharma for discussions. In addition, the authors thank Véronique Lefebvre, Sarchil Qadar, Amy Nineman and Tom Bird from Flowminder and WorldPop for modelling and imputing poverty from spatial data, in collaboration with the authors. This work is part of the background papers produced in the support to the South Sudan Country Economic Memorandum (P169121). Support from the State and Peacebuilding Fund (Grant # TF0A9011, The Dynamics of South Sudan's Conflict Economy) is gratefully acknowledged.

[78] See, UNOCHA: https://www.unocha.org/south-sudan & UNHCR: http://data.unhcr.org/SouthSudan/regional.php

[79] Not only has insecurity made fieldwork dangerous, but much of the South Sudanese population lives in isolated and hard to access areas. More than 85 percent of the 12 million South Sudanese reside in sparsely populated rural areas connected by a mere 200 km of paved roadways – about 2 percent of all roads – spanning an area of 650,000 square kilometers, approximately the size of France. The poor state of infrastructure combined with the size of the country means nationally representative surveys are expensive and time-consuming.

[80] The Greater Upper Nile region was where the opposition forces, the SPLM-IO, kept their stronghold and were thus contested in the fighting. In Appendix Figure 11 this region corresponds to the non-HFS states, where the number of conflict events in non-HFS covered states is much greater throughout 2014. During the year 2015 the conflict lost some of its intensity. Especially in the HFS states, where although conflict events continued to be recorded most of the violence remained concentrated, particularly in a few select areas which were relatively close to the border with the non-HFS states.

ranked 187 of 189 countries in the Human Development Index, with a life expectancy of merely 57 years.[81]

The HFS was designed with the expectation of potential instability and thus capitalized on recent technological and methodological innovations to obtain reliable national poverty statistics in difficult contexts.[82] Closely monitoring fieldwork is key to implementing such a large project in a risky context. The HFS leveraged the expansion of cellular networks across South Sudan to build a near real-time monitoring system, whereby the data could be uploaded daily to a dedicated server and checked for consistency. Computer Assisted Personal Interviewing (CAPI) also allowed built-in consistency checks, eliminating the need for expensive and potentially dangerous re-visits. Adherence to the sample design can be closely monitored with GPS software, tracking enumerators inside and outside areas with mobile phone coverage. The HFS also leveraged innovations in questionnaire design which permitted reducing the number of consumption items asked to the respondents while still obtaining unbiased poverty estimates through within-survey multiple imputation (Pape 2015, Pape and Mistiaen 2018). The lower amount of time spent collecting consumption data allowed the HFS to devote more time to collecting complementary data. Indeed, the HFS questionnaires contained additional modules covering asset ownership, education, labor market outcomes, perceptions of government performance and provision of public goods and services, psychological well-being, perceptions of violence and safety, allowing a well-rounded depiction of welfare and livelihoods.

The rapid escalation of the conflict in the summer of 2016, including several violent incidents affecting international humanitarian and development staff, led to the closure of the World Bank Office in South Sudan, disrupting the implementation of the third wave of the HFS. Therefore, the NBS implemented the third wave of the survey more independently relying mainly on remote support. A multitude of challenges had to be met, including large inflation, fuel unavailability, electricity shutdowns, insecurity, delay in payment of staff salaries, high NBS staff volatility, and cash flow limitations. Even though the NBS and the World Bank project team managed to mitigate a number of those challenges, the final sample reached only about 50 percent of the intended sample size. Nevertheless, this paper will argue that despite the enormity of challenges faced during fieldwork and the slight methodological departures from established approaches to poverty estimation (e.g. Deaton and Zaidi 2002), the data collected by the HFS provide the best-possible insights on welfare and livelihoods during a critical period of the country's history.

The data from the HFS are complemented by video testimonials providing a glimpse of the lives of the people of South Sudan. At the end of the interviews, respondents are offered to provide a short video testimonial where they can share their views and give a sense of their lives. The testimonials capture the dire situation on the ground and provide a much richer qualitative picture that accompanies and complements the quantitative data. While the data may help the government fine tune its policies, the videos may reach a broader audience and depict the sense of powerlessness, the pain of hunger, the stress of hopelessness and the feelings of disappointment that express people's experiences. Overall, this helps to create a more rounded perception of the situation on the ground in South Sudan.[83]

The levels of deprivation documented by the HFS are staggering. As mentioned above, more than 4 in 5 people across the seven states covered in 2016 were living under the international poverty line of $1.90 USD PPP (83 percent). Such breadth of poverty places South Sudan among some of the poorest

[81] UNDP Human development index, available at: http://hdr.undp.org/en/composite/HDI; and World Development Indicators.

[82] For a comprehensive review of issues in data collection in fragile and conflict situations see Mneimneh, Z. N., B.-E. Pennell, J. Kelley and K. C. Hibben (2016). Surveys in societies in turmoil. The SAGE Handbook of Survey Methodology 178.

[83] The translated testimonials are available at: http://www.thepulseofsouthsudan.com.

countries in the world. The depth of poverty is just as important as its breadth, with the average poor household consuming about one-half of the international poverty line (a poverty gap index of 47 percent). The incidence of poverty is much more widespread in rural areas compared to urban areas, with the rural poverty headcount reaching up to 86 percent compared to 65 percent in urban areas ($p<0.001$). The rural poor also to experience a deeper poverty than urban residents, with a poverty gap equal to 50 percent compared to 31 percent in urban areas ($p<0.001$). Widespread fighting and large-scale displacement over several consecutive planting seasons have disrupted many households' normal agricultural activities, resulting in increasingly large production deficits each year and widespread food insecurity. This has had a devastating effect on livelihoods, given that except for a few oil enclaves the productive structure of South Sudan is one of a rural pastoralist society where more than 4 in 5 people practice subsistence agriculture (Pape, Parisotto et al. 2018).

Despite initial intentions to expand the HFS across the entire country, continued insecurity made it impossible to reach the former states of Jonglei, Unity, and Upper Nile. To account for this gap in coverage and obtain countrywide poverty rates, a statistical model imputes poverty in inaccessible areas. The resulting poverty predictions are intended as supplemental to the survey estimates and serve as a proof-of-concept for using geo-spatial information alongside on-the-ground data collection. A growing body of research has emerged leveraging the increasing availability of alternative data sources such as satellite imagery and other geo-spatial characteristics. The estimates are derived by exploring the potential correlations between existing spatial data sets as well as custom-derived spatial data with geo-referenced poverty estimates obtained in the HFS. Once a set of spatial correlates were selected several models were trained and evaluated using a cross-validation approach. The final model was used to predict poverty rates at the 100m*100m level into all settled areas of the country including where survey data were not available. To aggregate the estimates at the state and county level, the 100m*100m level are weighted using a newly developed data set of human settlements across South Sudan constructed by combining a variety of publicly available data sources.

This paper describes the design and analysis of the third wave of the HFS in 2016.[84] The paper is focused on Wave 3 of the HFS, conducted between mid-2016 and early 2017, representing the most recent wave covering both urban and rural areas. Furthermore, the period between late 2016 and early 2017 was a critical period in South Sudan's history, when the conflict and refugee crises were reaching their peak. In Section 2, the paper describes the survey design and implementation, including the deviations from the original sample frame presenting consistency-checks used to evaluate potential selection issues that affect representativeness. Section 3 will detail the process of calculating consumption aggregates and estimating poverty using within-survey multiple imputations, including calculating durables consumption flow and spatial-time deflators. Section 4 gives a brief overview of the results of the poverty estimation, while a comprehensive assessment of poverty trends is available elsewhere (Pape, Parisotto et al. 2018). Section 5 describes the estimation of poverty rates using satellite data as a proof-of-concept while Section 6 concludes the paper with a short discussion of main limitations.

Survey Design and Implementation

Sample Design

The 2016 Wave of the HFS was conducted between mid-2016 to early 2017 and consisted of the second nationally representative survey wave of the HFS. The survey covered rural and urban areas

[84] The data from Wave 3 (2016) of the HFS and the code used to process these data can be downloaded from the World Bank MicroData Library at the following link: http://microdatalib.worldbank.org/index.php/catalog/9584/

across 7 of the 10 former states of South Sudan. The regions covered include Greater Equatoria, Greater Bahr el Ghazal, and Lakes. The 10 former states are used in planning for the HFS instead of the 28 more recent ones because the sample was constructed based on the sampling frame derived from the 5th Sudan Population and Housing Census from 2008.[85] The survey was designed to be representative at the state level and employs a stratified two-stage clustered sample design. Within each state the primary sampling units are enumeration areas (EAs) that were drawn randomly proportional to size. The EAs were drawn by the NBS for the 2008 Census (Southern Sudan Center for Census 2010).[86] The number of EAs and households was equalized across states in order to balance the fieldwork across teams. Within the EAs, 12 households were drawn randomly as the unit of observation based on a listing exercise.[87]

The EAs were allocated across urban and rural areas within each state to minimize the variance of indicators of interest across the strata while explicitly taking into consideration the design effect. The data used for the sample size calculations came from the NBHS 2009, and the indicator used for the sample size calculations was the real total per capita household expenditure.[88] While this variable is one of several that are of interest in the HFSSS, consumption/expenditure is generally strongly positively correlated with other indicators of interest. For the purposes of comparison, the relative standard error (complex standard error / mean) is used. The allocation was done so as to ensure a minimum of 10 EAs per combination of urban-rural and state distinction, according to the following rule:

$$n_u = \begin{cases} if\ n_u \geq 10 & n_u \quad n_r \\ if\ n_u < 10 & 10 \quad 40 \end{cases}, n\left(\frac{N_u S_u * deff}{N_u S_{u*deff} + N_r S_r * deff}\right), n_r = n\left(\frac{N_r S_r * deff}{N_u S_u * deff + N_r S_r * deff}\right), n = n_u + n_r = 50$$

where n_h is the sample size in stratum *h, n* is the total sample size, *H* is the total number of strata, N_h is the total population of stratum *h*, *N* is the total overall population, and S_h is the standard deviation in stratum *h.* The results from the sample size calculations are shown in Appendix B, Table B6-3. The chosen sample allocation provides estimates that are representative at the national, urban/rural, and state level. Sampling weights were calculated on the basis of the 5th Sudan Population and Housing Census from 2008 (Appendix B). In cases where fewer than 12 households were interviewed in an EA, the sampling weights were adjusted at the EA level to reflect this.

Data collection was intended to be implemented in two phases, by randomly splitting each stratum into two equal-sized parts, where each phase of data collection would cover half of the sample. The advantage of a two-phased approach was early availability of representative data after half of the survey was implemented. The two-phased approach reduces the risk that an eruption of violence during field work invalidates the representativeness of the survey. However, such an approach is not guaranteed to maintain representativeness if some areas remain inaccessible throughout the entirety of fieldwork. It also comes at the cost of optimizing the organization of fieldwork by reducing the enumerators' ability to sweep over their intended area.

[85] The more recent states have largely been drawn based on the counties subdivision of the former states, the geographical boundaries have therefore largely remained intact.
[86] Urban EAs were drawn to contain approximately 100 to 150 households, while urban EAs would generally contain between 200 to 300 households.
[87] The number of households per EA was determined to be 12 to allow an equal split into 4 groups per EA to facilitate the implementation of the Rapid Consumption Methodology. The specific options of 8, 12, and 16 were considered. Eight households per cluster was deemed as too small as the number of EAs necessary and the associated travel time could not be done within the fieldwork calendar. Sixteen resulted in very high design effects, over 3 in most cases and as high as 5 for some strata, and was therefore deemed too large. Twelve households per EA was therefore selected as the ideal cluster size.
[88] The top and bottom 1 percent of outlier observations were trimmed for the sample size calculations.

Survey Implementation

The survey was implemented using tablets as survey devices. The data collection system consisted of Samsung Galaxy Tablet computers equipped with SIM cards, mobile data plans, microSD cards (16 GB capacity), and external battery packs.[89] Teams of four enumerators and one supervisor were provided with a mobile generator using fuel to ensure that tablets could be charged overnight. Computer Assisted Personal Interviewing (CAPI) data collection can be used to improve data quality by imposing sophisticated systems of constraints on the enumerators' entries. This was particularly relevant for consumption and price data, which need to be measured precisely as a prerequisite for a reliable poverty analysis. Indeed, CAPI has been experimentally shown to improve data collection while minimizing the potential for enumerator error (Fafchamps, Mckenzie et al. 2010, Caeyers, Chalmers et al. 2012). Furthermore, it can be used to create more sophisticated questionnaires, with elaborate conditional skipping patterns that are much easier to implement (De Leeuw, Hox et al. 1995).

The rapidly expanding cellular network in South Sudan meant that the data could be transmitted via mobile networks and made available quickly to data analysts (Pape and Mistiaen 2014). The near real-time transmission of data to a cloud enabled the implementation of a monitoring system including a dashboard tracking the cumulative number of interviews, the fraction of missing variables, as well as additional quality indicators at any level of disaggregation.[90] This helped to identify challenges in the field work as well as weak enumerators early on and mitigate their impact on data quality, e.g. by providing individually tailored extra trainings for selected enumerators. In addition, the real-time analysis code calculates core indicators of the survey, e.g. consumption, educational attainment, and unemployment, to check incoming data while field work is still ongoing. This head-start on building the analysis code ensures that swiftly after the end of data collection the cleaned data can be made available, which considerably accelerated the process from data collection to the publication of results.

The availability of real-time data facilitated monitoring by allowing much closer tracking of the geographic progression of fieldwork. The GPS coordinates for each interview were recorded and uploaded along with the data, allowing tracking enumerators and ensuring the sampling design was implemented. Furthermore, GPS tracking software helped to track devices at all times using a web interface (www.gps-server.net), the exact path of the devices was recorded even retrospectively and uploaded to the server once they entered areas with 3G/WIFI connection. Given the frequent disruptions and slow rate of data collection their combination provided a useful reference to understand where field teams were at any time, and could be cross-checked with reports of conflict activity etc. Overall, this system allowed close supervision of the implementation of the sampling design (Pape and Mistiaen 2014).

Fieldwork and Insecurity

Sporadic eruptions of fighting meant that teams of enumerators were at times forced to remain idle and wait for the situation to deescalate before reaching certain areas. A few areas that had been subjected to heavy fighting and that may have experienced mass displacement could not be reached at all. Therefore, fieldwork was delayed and the quality of documentation was negatively affected. In the end, despite the relatively long duration of data collection, the final sample fell short of the intended sample. Fortunately, the two-phased approach described above implies that representative data are already available after the first half of the survey implementation. Indeed, the final sample that was collected

[89] The Android application AirDroid was used to remotely manage devices, this remote management software meant that errors in the tablet configuration were detected and could be solved by updating the tablets remotely in cases where enumerators may have needed help from the survey analysts.

[90] In areas without 3G activities, enumerators saved conducted interviews on the tablet and submitted data once they had 3G connectivity.

during Wave 3 only reaches only about 50 percent of the intended sample size, i.e. the first of the two phases. This was true across all states (Table B6-4).

Nevertheless, many of the selected EAs had to be replaced when security rendered field work unfeasible.[91] One hundred EAs were surveyed of the 350 EAs in the original sample, the rest of the 64 EAs were replacement EAs. Replacements were done in three batches where each time new enumeration areas had to be drawn from the master sample frame. The replacement sequence was defined by assigning enumeration areas randomly to the original enumeration areas, maintaining the order of the original enumeration areas as in the original sample. The large number of replacements was concerning given fear of selection bias. Therefore, the team ran checks to ensure that the set of EAs surveyed do not systematically differ from a random sample as best as it could. It is important to keep in mind that assessing representativeness is a difficult task, generally due to the lack of a counterfactual or a point of reference to compare estimates. Despite these checks, it is plausible that selection bias in favor of less conflict-affected areas leads to an under-estimation of poverty. The resulting estimates are therefore interpreted as lower-bound estimates.

The checks are based on comparisons of Wave 3 data from 2016 with the nearest available reference point, Wave 1 data from 2015. Specific outcomes were compared across the two waves as well as at lower levels of aggregation and within specific regions (Table B6-8 shows an example). This process was severely complicated by the magnitude of the South Sudanese crisis. The conflict, displacement crisis, and near-hyperinflationary price increases are powerful shocks, which are expected to cause severe disruption even in a relatively short amount of time.[92] The checks therefore concentrated on outcomes that are less likely to be affected by the crises and are relatively time-invariant.

Adults' educational outcomes is one such indicator which is expected to remain relatively stable from one year to the next assuming only small demographic changes. In South Sudan, the adult literacy rate (18+), the proportion of adults with no education, and the proportion of adults with only primary education were comparable between 2015 and 2016 (Table B6-8). Similarly, cultural norms should be expected to remain stable, such as the prevalence of polygamy and the gender of the household head, both of which are again seemingly unchanged. Some types of infrastructure can provide good indicators if they are not susceptible to be destroyed in the fighting. Mobile telephone networks are a good example, since they generally comprise relatively heavy infrastructure that is not easily destroyed through the type of warfare occurring in South Sudan. This is also a good indicator of sample selection favoring wealthier areas, especially in the context of South Sudan where only one in four households is covered. Access to electricity is a similar indicator given that it is exclusive to a few selected areas of South Sudan. Again, the latter two indicators do not seem different from 2015 and 2016. Finally, the share of households living far from schools, health centers, and markets, did not change significantly – this generally holds for various thresholds.

More importantly, the path of enumerators and geographic coverage of Wave 3 data was closely inspected to ensure that it remained broadly comparable to that of previous HFS waves and other sources of population data. This helped to control that entire areas were not systematically excluded. As an exception, the city of Yei was not surveyed at all in Wave 3 because it was the site of several large battles during fieldwork and subsequently experienced a massive wave of displacement. This was likely the most severe case, and in many other instances where fighting affected specific areas enumerators

[91] Replacement EAs were approved by the project manager. Replacement of households were approved by the supervisor after a total of three unsuccessful visits of the household.

[92] At the very start of Wave 3 data collection year-on-year inflation was equal to almost 650 percent. The CPI between the start of Wave 1 and the end of Wave 3 had increased by almost 1,600 percent. Similarly, more than a third of the population was displaced by mid-2018.

simply delayed fieldwork until it was safe to continue. This explains to some extent the prolonged duration of fieldwork relative to the low number of interviews conducted in total.

Measuring Poverty in a Fragile Context

Calculating Consumption Aggregates

Poverty in the HFS was measured according to a standardized methodology best described in the seminal contribution by Deaton and Zaidi (2002). Poverty analysis consists of comparing a welfare measure to a predetermined poverty line. Therefore, the first step is to calculate a measure of welfare. The measure chosen for the HFS is the households' consumption expenditure per capita.[93] The nominal household consumption aggregate consists of the sum of consumption expenditure per person on three primary components, i) total expenditures on food items, ii) total expenditures on non-food items, and iii) the value of the consumption flow from the durable goods owned by the household.[94] The consumption aggregate is then deflated to reflect spatial and temporal cost of living differences.

(1) $$y_i = y_i^f + y_i^n + y_i^d$$

Accurately measuring consumption in highly volatile environments is a complex task, primarily because insecurity and uncertainty severely restrict the time that can safely be spent by enumerators in certain areas and the time spent conducting each interview. Consumption modules tend to be bulky and take a long time to administer. At the very least, it requires asking information on quantities consumed, quantities purchased, and prices of purchase –including additional information on home production in a context such as South Sudan – for what is often upwards of 300 to 400 consumption items (Beegle, De Weerdt et al. 2012). Reducing the length of the questionnaire is therefore a key strategy when designing surveys for fragile contexts. For example, it is common to remove rarely consumed items or to combine categories of items (e.g. vegetables). However, Beegle et al. (2012) and Olson Lanjouw and Lanjouw (2001) show that such approaches tend to result in underestimated consumption levels, and hence overestimate the poverty rate.

Another set of approaches for obtaining poverty estimates in a fragile context consists of modeling consumption, or poverty, based on a set of observable covariates and then projecting estimates using cross-survey imputation. In this manner, infrequent bulky consumption surveys can be combined with more frequent surveys that collect information on the covariates necessary for imputing poverty (for example labor force surveys as in Douidich et al. (2013) or SWIFT[95]). However, this methodology is problematic in contexts where there is no consumption survey to underlie the estimation, or where there may have been deep structural change that changes the relationship between covariates and poverty across time (Christiaensen, Lanjouw et al. 2010, Beegle, Christiaensen et al. 2016). This is most likely the case in South Sudan, where the last full consumption survey was conducted in 2009 and which had experienced a period of rapid development leading up to independence in 2011 and until the breakout of the current conflict in 2013.

Within-survey imputation can alleviate some of these concerns because the assumption of similar covariate distribution between the data used to estimate poverty and that used to project is more likely

---

[93] In the context of South Sudan using consumption as a measure of welfare is preferable to a measure of income for two main reasons: (i) there exists no real reliable information on income given poor administrative record keeping, and (ii) employment is primarily irregular and informal in nature, with subsistence agriculture accounting for about two-thirds of employment, non-farm business ownership for one-eighth, and salaried labor also only about one-eighth.

94 In some cases, housing is included in the consumption aggregate. However, calculating the consumption flow obtained from housing requires estimating rental values from the open market. Unfortunately, the housing market in South Sudan is highly underdeveloped, making such estimations impossible in any sort of accurate manner. Indeed, in the 2016 HFS, 91 percent of households were owned by the residents and fewer than 4 percent were rented. Thus, housing was excluded from the consumption aggregate.

[95] Survey of Well-being via Instant and Frequent Tracking.

to hold, or the differences may not be as great. One approach consists of administering a full consumption module to a subset of respondents, generally in more secure areas where time-constraints are not binding, and then impute consumption for less secure areas based on a smaller set of covariates (Fujii and Van der Weide 2013). However, safer areas where the full consumption module can be administered may still systematically differ from insecure areas where only the covariates are collected, thus violating the assumption of equally distributed covariates.

The HFS employed a method of within-survey imputation, but instead of imputing the totality of consumption in certain areas based on data from other areas it imputed a randomly different fraction of consumption across all enumeration areas covered in the survey (Pape 2015, Pape and Mistiaen 2018). Food and non-food consumption items were first into a core and multiple optional modules. Each household was then asked only about the core items and those items in one of the optional modules, and consumption of items in the remaining optional modules was estimated through multiple imputation. The imputation does not suffer from bias caused by different covariate distributions, since data on every one of the optional consumption modules are collected within each enumeration area. Furthermore, because a majority of consumption is accounted for by a relatively small set of items collected for each household, additional variance introduced by the imputation is minimized.

This section will describe the rapid survey consumption methodology, a more detailed treatment and simulations can be found in (Pape 2015, Pape and Mistiaen 2018). First, food and non-food consumption for household *i* are estimated by the sum of expenditures for a set of items

$$(1) \qquad y_i^f = \sum_{j=1}^{m} y_{ij}^f \text{ and } y_i^n = \sum_{j=1}^{m} y_{ij}^n$$

where y_i^f and y_i^f denote the food and non-food consumption of item *j* in household *i*.[96] Previous consumption surveys in the same country or consumption surveys in neighboring / similar countries can be used to estimate food shares.[97] In South Sudan, the item assignment could draw from the NBHS 2009 survey.[98] The list of items was partitioned into 1 core and 4-optional modules each with m_k items:

$$(2) \qquad y_i = \sum_{k=0}^{4} y_i^{(k)} \text{ with } y_i^{(k)} = \sum_{j=1}^{m_k} y_{ikj}$$

The core module was designed to maximize its consumption share based on NBHS 2009 consumption, and therefore contains all the most commonly consumed items. This includes staple foods such as dura, maize, onions, okra, common types of flour (e.g. millet, maize, cassava, and groundnut flour), common types of meat (e.g. goat, sheep, poultry, beef), and some fruits. The nonfood core module similarly captures common expenditures including fees for education, common types of transportation, common medicines and health related expenditures, and clothing. Optional modules were constructed using an

[96] As the estimation for food and non-food consumption follows the same principles, we neglect the upper index *f* and *n* in the remainder of this section.

[97] In a case where a previous survey is not available the items can be randomly assigned to the module. This would result in larger standard errors but would not introduce bias.

[98] With manual modifications to treat 'other' items correctly. Items 'other' are often found to capture remaining items for a food category. Using the Rapid Consumption Methodology, this creates problems as 'other' will include different items depending on which optional module is administered. This can lead to double-counting after the imputation. Therefore, 'other' items are re-formulated and carefully assigned so that double counting cannot occur.

algorithm to assign items iteratively to optional modules so that items are orthogonal within modules and correlated between modules.[99]

This step is followed by the actual data collection. Conceptual division into core and optional items is not reflected in the layout of the questionnaire. Rather, all items per household are grouped into categories of consumption items (like cereals, meats, etc.). Using CAPI, it is straight-forward to hide the modular structure from the enumerator. For each household, only the core module $y_i^{(0)}$ and one additional optional module $y_i^{(k^*)}$ are collected. In each enumeration area, 12 households were interviewed with an ideal partition of three households per optional module.[100] The assignment of optional modules was stratified per EA to ensure that an equal number of households are assigned to each optional module. This served to minimize potential EA effects during the imputation process.

Household consumption was then estimated using the core module, the assigned module and estimates for the remaining optional modules:

$$\hat{y}_i = y_i^{(0)} + y_i^{(k^*)} + \sum_{k \in K^*} \hat{y}_i^{(k)} \tag{3}$$

where $K^* := \{1, \dots, k^* - 1, k^* + 1, \dots, M\}$ denotes the set of non-assigned optional modules. Consumption of non-assigned optional modules is estimated using multiple imputation techniques taking into account the variation absorbed in the residual term.

Multiple imputation was implemented using multi-variate normal regression based on an EM-like algorithm to iteratively estimate model parameters and missing data. 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.[101] Iteratively, the coefficients are updated based on re-estimation using imputed values for missing data drawn from the posterior distribution of the model. The implemented technique 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:

$$\hat{y}_i^{(k)} = \beta_0^{(k)} y_i^{(0)} + x_i^T \beta^{(k)} + u_i^{(k)} \tag{4}$$

[99] In each step, an unassigned item with the highest consumption share was selected. For each module, total per capita consumption was regressed on household size, the consumption of all assigned items to this module as well as the new unassigned item. The item in questions was then 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. Increasing d puts increasing weight on orthogonality within and correlation between modules. The parameter was set to d=3 balancing the two objectives.

[100] Field work implementation aimed to achieve a balanced partition among optional modules but due to challenges in following the protocol exactly some enumeration areas are not completely balanced.

[101] The model employed in the HFS was constructed using the following indicators: demographics variables including household size, the fraction of children, the fraction of elderly persons, the sex of the household head, and the employment status of the household head; indicators of access to amenities including the water source, whether the household had electricity to power its lighting, the number of sleeping rooms, and whether the household had access to a toilet; geographic indicators including an urban-rural dummy and state fixed effects; finally, the model included dummies for each quartile of consumption of food and non-food per capita. One hundred imputations were run for the consumption imputation process to maximize the accuracy of results.

The performance of the estimation technique was assessed based on an *ex post* simulation using the NBHS 2009 data and mimicking the Rapid Consumption methodology by masking consumption of items that were not administered to households. The results of the simulation were compared with the estimates using the full consumption from NBHS 2009 as reference. The simulation results distinguish between different levels of aggregation to estimate consumption.[102] The methodology generally does not perform well at the household level (HH) but improves considerably already at the enumeration area level (EA) where the average of 12 households is estimated. At the national aggregation level, the Rapid Consumption methodology slightly over-estimates poverty by 1.6 percent. Assessing the standard poverty measures including poverty headcount (FGT0), poverty depth (FGT1) and poverty severity (FGT2), the simulation results show that the Rapid Consumption methodology retrieves almost unbiased estimates. Generally, the estimates are robust as suggested by the low standard errors.[103]

The assumption that the imputed components of consumption follow a joint normal distribution might provide an explanation as to why poverty is slightly overestimated. This would be due to the imputed means of consumption of the imputed items being slightly lower than the actual means since their true distributions are generally skewed to the right. This possibility was explored by assuming a non-parametric error term in the imputation procedure through the use of chained equations, which performed almost indistinguishably as well as the multivariate-normal approximation.

Figure B6-1: Relative bias of simulation results using Rapid Consumption estimation.

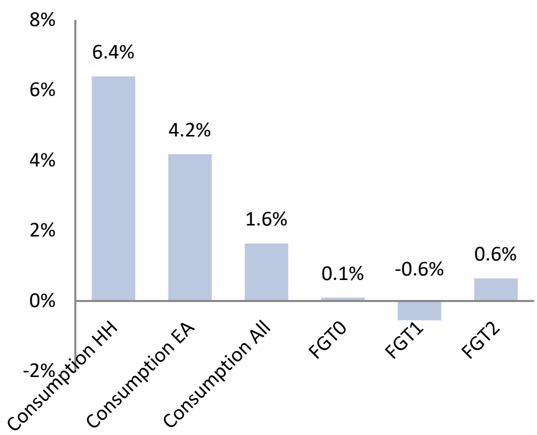
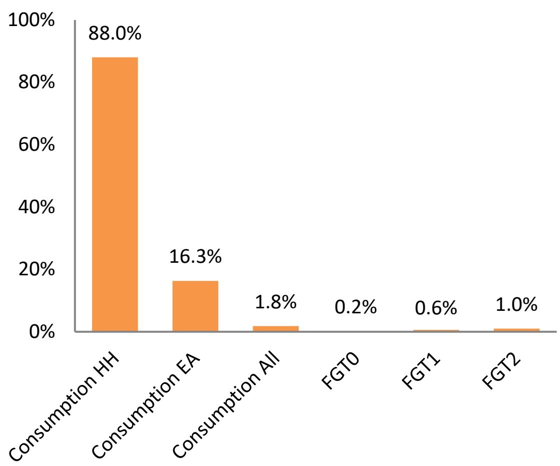


Figure B6-2: Relative standard error of simulation results using Rapid Consumption estimation.



Durable Consumption Flow

The consumption aggregate includes the consumption flow of durables calculated based on the user-cost approach, which distributes the consumption value of the durable over multiple years (Amendola and Vecchi 2014). The user-cost principle defines the consumption flow of an item as the difference of selling the asset at the beginning and the end of the year as, this is the opportunity cost of the household for keeping the item. The opportunity cost is composed of the difference in the sales price and the forgone earnings on interest if the asset is sold at the beginning of the year. The

---

[102] The performance of the estimation techniques is presented using the relative bias (mean of the error distribution) and the relative standard error. The relative error is defined as the percentage difference of the estimated consumption and the reference consumption (based on the full consumption module, averaged over all imputations). The relative bias is the average of the relative error. The relative standard error is the standard deviation of the relative error. The simulation is run over different household-module assignments while ensuring 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.

[103] These standard errors are estimated empirically using a bootstrap approach taking into account intra-cluster correlation within enumeration areas.

current price of the durable is p_t. If the durable item would have been sold one year ago, the household would have received the market price for the item twelve months ago plus the interest on the revenue for one year. The market price from 12 months ago is calculated by adjusting for inflation π_t and annual physical or technological depreciation rate δ arriving at[104]

$$\frac{p_t(1+i_t)}{(1+\pi_t)(1-\delta)} \tag{2}$$

with the nominal interest rate denoted as i_t. Alternatively, the household can use the durable and sell it after one year of usage for the current market price p_t. The difference between these two values is the cost that the household is willing to pay for using the durable good for one year. Hence, the consumption flow is:

$$\mathrm{y}^d = \frac{p_t(1+i_t)}{(1+\pi_t)(1-\delta)} - p_t \tag{3}$$

By assuming that $\delta \times \pi_t \cong 0$, the equation simplifies to

$$\mathrm{y}^d = \frac{p_t(r_t+\delta)}{(1+\pi_t-\delta)} \tag{4}$$

where r_t is the real market interest rate $i_t - \pi_t$ in period t. Therefore, the consumption flow of an item can be estimated by the current market value p_t, the current real interest rate r_t, the inflation rate π_t and the depreciation rate δ. Assuming an average annual inflation rate π, the depreciation rates δ can be estimated utilizing its relationship to the market price[105]:

$$p_t = p_{t-k}(1+\pi)^k(1-\delta)^k \tag{5}$$

The equation can be solved for δ obtaining:

$$\delta = 1 - \left(\frac{p_t}{p_{t-k}}\right)^{\frac{1}{k}} \frac{1}{(1+\pi)} \tag{6}$$

The depreciation rates estimated the 2015 HFS wave were used to calculate the consumption flow in the 2016 wave. The reason being that estimating depreciation rates is much more prone to errors in a context of high and unstable inflation such as that observed in South Sudan in 2016.[106] Furthermore, there are few reasons to expect depreciation rates to drastically change over such a short period of time. In 2015, based on equation (6), item-specific median depreciation rates are estimated assuming an inflation rate of 0.5 percent, a nominal interest rate of 5.5 percent and, thus, a real interest rate of 5 percent (Table B6-6). For all households owning a durable but that did not report the current value of the durable, the item-specific median consumption flow is used. For households that own more than one durable, the consumption flow of the newest item is added to the item-specific median of the consumption flow times the number of those items without counting the newest item.[107]

[104] Assuming a constant depreciation rate is equivalent to assuming a "radioactive decay" of durable goods (see Deaton and Zaidi, 2002).
[105] In particular π solves the equation $\prod_{i=t-k}^{t}(1+\pi_i) = (1+\pi)^k$.
[106] One potential source of bias being that the value placed by respondents on durable goods may be inflated given high levels of uncertainty regarding the future of the currency. Another is that the volatility of inflation across time periods is problematic given the formula assuming one inflation rate prevailing across the different years.
[107] The 2016 HFSSS questionnaire provides information on a) the year of purchase and b) the purchasing price only for the most recent durable owned by the household.

Spatial and Temporal Price Deflators

Prices fluctuated considerably in South Sudan in 2016 (Pape and Dihel 2017, Pape, Parisotto et al. 2018). Prices therefore need to be adjusted to make consumption comparable across the several months of fieldwork. Furthermore, there are important differences in the cost of living between urban and rural areas. This is particularly marked in South Sudan given the sheer isolation of rural areas and state of poor market linkages across the country (African Development Bank 2013, Pape, Benson et al. 2017). A Laspeyres deflator was chosen to calculate price differences across urban and rural areas and months of data collection, due to its relatively light data requirements. The base period for deflating prices was chosen as October 2016 in urban areas. Urban areas were chosen as a reference because the national CPI calculated by the NBS is based on prices in urban markets across some of the largest cities in South Sudan, and hence would facilitate deflating consumption across the frequent waves of data collection in the HFS.

The Laspeyres index reflects the item-weighted relative price differences across products. Item weights are estimated as household-weighted average consumption share across all households before imputation. Based on the democratic approach, consumption shares are calculated at the household level. Core items use total household core consumption as reference while items from optional modules use the total assigned optional module household consumption as reference. The shares are aggregated at the national level (using household weights) and then calibrated by average consumption per module to arrive at item-weights summing to 1. The item-weights are applied to the relative differences of median item prices for each urban/rural and month pair. Missing prices are replaced by the item-specific median over all households. The reference strata was chosen as the urban strata for one specific month of data collection. The month with the most data points was generally chosen for the reference time period. The Laspeyres deflator can be expressed as such:

(7)
$$L_{i,t} = \sum_{k=1}^{k} w_{i,k,m} \left(\frac{p_{i,k}}{p_{0,k}} \right)$$

The Laspeyres $L_{i,t}$ for strata i and month t is equal to the sum of, over all items k: $w_{i,k,m}$, the national budget share of item k in optional module m, times the ratio of $p_{i,k,m,t}$, the median price of item k in strata i at month t, and $p_{0,k,m,0}$, the median price of item k in the reference strata in the reference month. Two sets of price deflators were calculated, one for food and another for nonfood items, the nonfood price deflator was used to deflate the consumption flow of durable goods.

Poverty Line

Determining a household's poverty status requires a poverty line against which to compare consumption. A poverty line serves as a reference point for what might be an acceptable minimum standard of well-being, below which one could be considered deprived, or living in poverty (Ravallion 1998, Ravallion 2017). The choice of the poverty line considers what might constitute an acceptable minimum standard of living and the potential impact of resulting poverty estimates on policy decisions. Once a poverty line has been chosen, poverty analysis is then typically based on comparing the first three poverty measures of the Foster-Green-Thorbecke (FGT) class of poverty indicators. FGT measures consist essentially of variations of specification 0, where the parameter α takes the value of 0 for the poverty headcount, 1 for the poverty gap, and 2 for poverty severity (Foster, Greer et al. 1984).

$$FGT(\alpha) = \frac{1}{n} \sum_{i=1}^{p} \left[\frac{z - y_i}{z} \right]^{\alpha}$$

Where y_i denotes the consumption y of individual i, n denotes the total population, and z the poverty line.

Theoretically, a national poverty line could have been estimated for South Sudan in the year 2016 using the survey data. However, the international poverty line of US$1.90 PPP was chosen.[108] Given that the international poverty line was based on the predicted poverty line for the world's 15 poorest countries, combined with the expectation that poverty in South Sudan was to be relatively high, the international poverty line was considered an appropriate metric, also offering the ability to make international comparisons. Hence, the $1.90 USD PPP (2011) poverty line was first converted into current SSP and adjusted to reflect South Sudanese purchasing power using the South Sudan PPP conversion factor for 2011. It was then adjusted for inflation up to October 2016 using the national CPI calculated by the National Bureau of Statistics, resulting in a value of approximately 65 SSP (October 2016).

Results from the HFS

In 2016, more than 4 in 5 South Sudanese people in the seven states covered in the HFS lived under the international poverty line of US$1.90 PPP (2011) per capita per day. The poverty headcount was equal to 83 percent in 2016, with a 95 percent confidence interval from 81 to 85 percent. These levels of poverty place South Sudan among some of the poorest countries in the world. South Sudan's poverty headcount ratio is much higher than the average estimates of other countries at similar levels of development (Figure B6-3). The estimated poverty headcount ratio is not particularly sensitive to the choice of poverty line, since average consumption levels are so low that the poverty line lies at a point where the slope of the cumulative distribution of consumption tapers off (Figure B6-5). The deterioration of economic conditions has driven many poor households down to hardship conditions. The poverty gap, which measures poor households' average deficit in consumption relative to the poverty line, is equal to 47 percent in 2016. The average poor household is therefore consuming about one-half of the poverty line in 2016 (US$ 1.00 2011 PPP). The poverty severity index, which is the square of the poverty gap and thus places more weight on people with consumption levels further below the poverty line, was equal to 0.31 ($p<0.001$).

[108] The international poverty line was first introduced in the 1990 World Bank World Development Report with the intent of measuring poverty across countries in a consistent manner. This international poverty line used data on 33 national poverty lines for the 1970s and 1980s and represented the predicted poverty line for the poorest country in the sample, equal to about $0.76 USD PPP (1985). The international poverty line was subsequently adjusted for inflation as new sets of PPP were made available through the International Comparison Program. The computation of the current international poverty line of $1.90 USD PPP per day was obtained as the unweighted average of the poverty line for the 15 poorest countries, as such: i) by adjusting the national poverty lines of the 15 poorest countries for inflation up to 2011; ii) then converting the national poverty lines to real USD using the 2011 PPPs; and iii) then computing the simple average of the 15 national poverty lines. The resulting average poverty line is equal to $1.88 USD PPP (2011) per person per day, which was rounded up to $1.90 USD PPP (2011).

Figure B6-3: Poverty headcount in low and lower middle-income countries.[109]

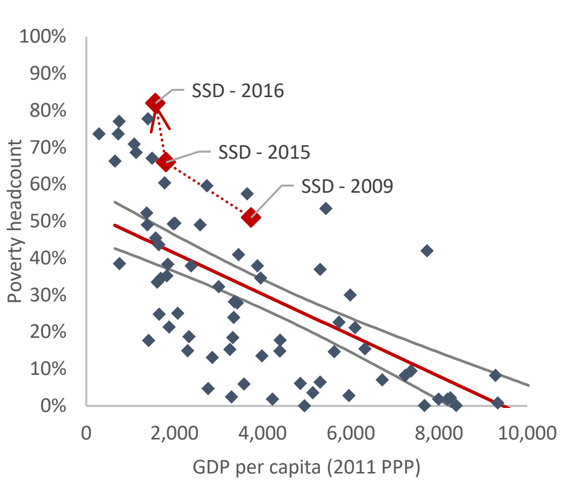


Figure B6-4: Gini index in SSA countries.

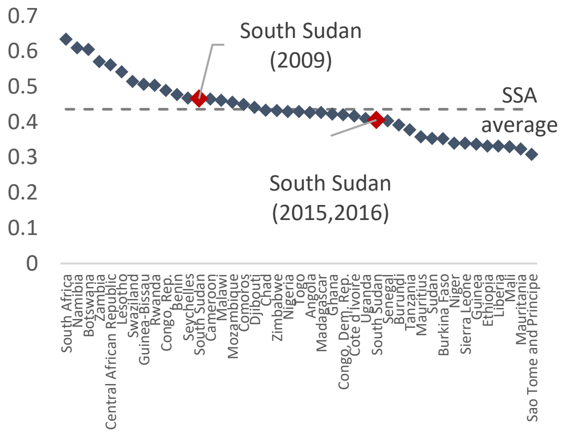


Figure B6-5: Cumulative consumption distribution.

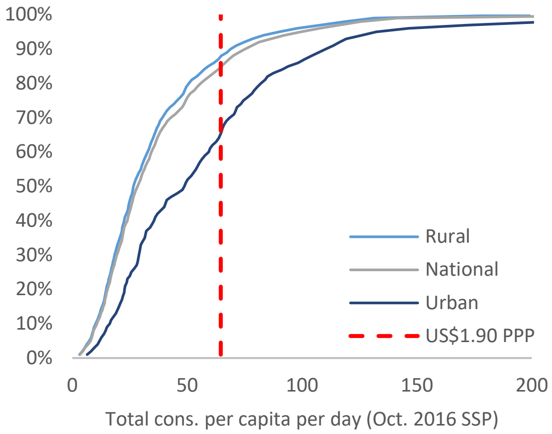
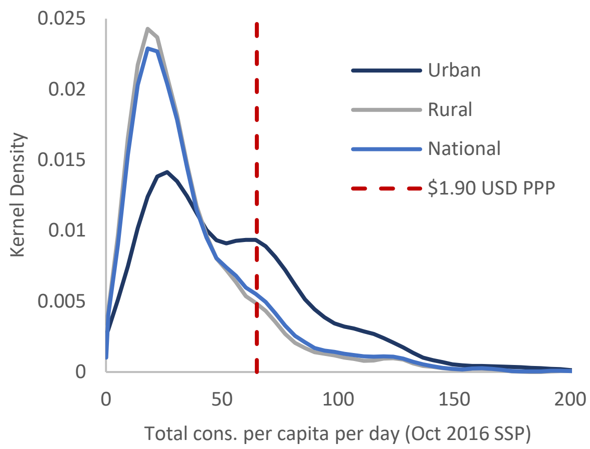


Figure B6-6: Consumption distribution, 2016.



Note: Figure B6-3 includes low income and lower middle-income countries with poverty data post-2008. All data for South Sudan refers to the seven states covered by the HFS.

Such high levels of deprivation translate into widespread hunger and food insecurity. Disruptions to agricultural production and the near hyperinflationary increases in prices of most staple foodstuffs have left most households struggling to find enough food to sustain themselves (Pape, Parisotto et al. 2018). Widespread fighting and large-scale displacement over several consecutive planting seasons have disrupted many households' normal agricultural activities, resulting in increasingly large production deficits each year and widespread food insecurity (FAO, IFAD et al. 2017). This has had a devastating effect on livelihoods, given that except for a few oil enclaves the productive structure of South Sudan is one of a rural pastoralist society where more than 4 in 5 people practice subsistence agriculture (World Bank 2016, Pape, Parisotto et al. 2018). Food security has continuously deteriorated since late 2012, sometimes even reaching famine conditions in certain vulnerable counties. During the most recent harvest season in 2017, a time when food should be abundant, as many as 4.8 million people were severely food insecure (FAO, IFAD et al. 2017). By mid-2018, the number of severely food insecure people is expected to rise to 6.2 million, reaching more than half of the total population.[110]

[109] Data for real GDP per capita in 2011 PPP for South Sudan were obtained from the IMF World Development Outlook Database.
[110] FEWSNET Food Security Outlook, February to September 2018.

Table B6-1: Poverty headcount and average consumption per strata for covered states, 2016.

	Poverty headcount ratio				Mean consumption				
	Mean	Standard Error	[95% CI]		Mean	Standard Error	[95% CI]		N
National	0.83	0.01	0.80	0.86	73.30	2.68	67.99	78.60	1,848
Rural	0.86	0.02	0.83	0.89	67.36	2.70	62.03	72.70	1,281
Urban	0.65	0.02	0.60	0.70	113.99	5.59	102.94	125.05	567
Warrap	0.86	0.05	0.77	0.95	63.98	7.13	49.88	78.08	135
Northern Bahr El Ghazal	0.90	0.03	0.84	0.95	62.63	5.64	51.49	73.77	299
Western Bahr El Ghazal	0.90	0.02	0.87	0.94	60.17	6.33	47.66	72.68	310
Lakes	0.84	0.02	0.80	0.88	71.22	3.46	64.38	78.06	232
Western Equatoria	0.53	0.04	0.46	0.61	130.51	7.45	115.79	145.23	300
Central Equatoria	0.80	0.05	0.70	0.90	86.53	8.27	70.18	102.88	311
Eastern Equatoria	0.95	0.01	0.93	0.98	43.88	3.58	36.80	50.96	261

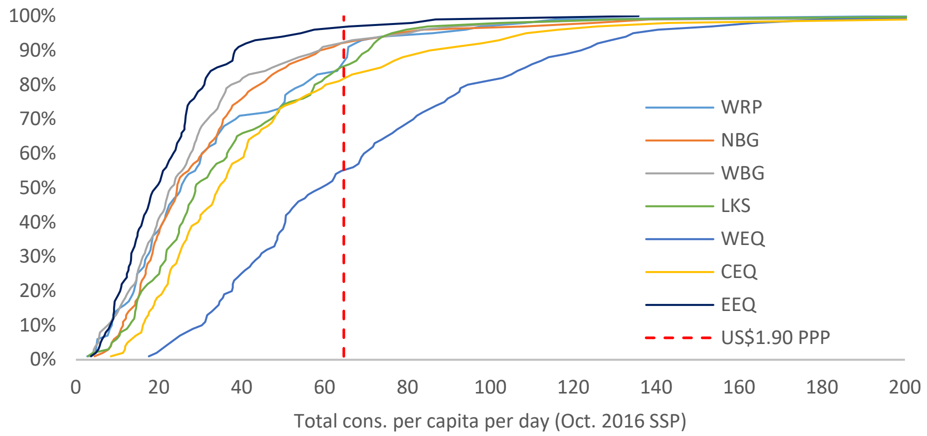
Note: Standard errors estimated through linear regressions; all estimates weighted using population weights.

The incidence of poverty is much more widespread in rural areas compared to urban areas. Rural poverty was equal to 86 percent in 2016 compared to 65 percent in urban areas ($p<0.001$, Figure B6-5). The rural poor also experience deeper poverty than urban residents, with a higher poverty gap and poverty severity. In 2016, the urban poverty gap was equal to 31 percent compared to 50 percent for the rural poverty gap ($p<0.001$, Figure B6-5). A similar pattern can be observed for poverty severity, the urban severity index was equal to 19 percent and the rural index equal to 33 percent ($p<0.001$). A stochastic dominance analysis based on a comparison of the cumulative consumption expenditure distribution across rural and urban areas reveals that this is not due to the chosen poverty line but that at any point along the distribution the urban consumption expenditure curve lies consistently below the rural curve (Figure B6-5). The isolated nature of many rural areas contributes to these observed poverty rates, given that they are often cut off from public services as well as humanitarian assistance.

Measuring inequality, the Gini index in South Sudan declined from 2009 to 2016, from about 0.47 in 2009 to 0.41 in 2016 (Figure B6-4).[111] The average Gini index for countries in Sub-Saharan Africa is approximately 0.44, with South Sudan at 0.41 indicating slightly lower inequality but higher inequality compared to the global average Gini index of 0.38. While all households suffered consumption losses because of the conflict and macroeconomic crises, the consumption losses experienced by better off households were larger than those of the poorer households (Pape, Parisotto et al. 2018). Thus, the driver of the reduction in inequality was not pro-poor growth but rather a greater decline in welfare for wealthier households relative to poorer households. This is not entirely unexpected since the poorer households already experienced extreme deprivation, and thus could not fall much further even as the crisis worsened. Inequality remains nevertheless greater in urban areas than in rural areas though only slightly, at 0.41 and 0.39 respectively. Indeed, many of the households with the highest consumption levels reside in urban areas, with better access to markets and opportunities.

[111] The Gini index is calculated from the area under the Lorenz curve, which plots the cumulative percentage of consumption expenditure against the cumulative percentage of the population, with perfect equality lying along the 45-degree line.

Figure B6-7: Cumulative consumption distribution by state.



Poverty in 2016 is generally high but it is higher in former states that were more exposed to the conflict. The incidence of poverty reached extremely high levels in the former states of Eastern Equatoria, Northern Bahr el Ghazal, and Western Bahr el Ghazal, where about 9 in 10 people live under the international poverty line (95, 90, and 90 percent, respectively). In the former states of Lakes and Central Equatoria, the poverty headcount is slightly lower at about 8 in 10 people, though still extremely high by international standards (84 and 80 percent, respectively). One notable exception is the former state of Western Equatoria, as it was less affected by the conflict compared to the other states and has benefitted from high fertility and favorable weather conditions. Indeed, Western Equatoria, in the "green belt" of South Sudan, was the only state to record a consistent cereal production surplus in the years from 2014 to 2016 (FAO, IFAD et al. 2017). Accordingly, the residents of Western Equatoria were much more likely to be able to sustain their livelihoods through own production compared to those in other states and thus maintain better standards of living (Pape, Parisotto et al. 2018).

Imputing Poverty Using Geo-Spatial Data

Extending Poverty Estimates to Non-Covered Areas

Despite initial intentions to expand the HFS across the entire country, continued insecurity made it impossible to extend the survey to the former North-Eastern states of Jonglei, Unity, and Upper Nile. To account for this gap in coverage and obtain countrywide poverty rates, a statistical model was developed to impute poverty in non-covered areas leveraging the growing availability of satellite imagery and geo-spatial data. Recent advances in the processing and availability of satellite imagery and geo-spatial data have led to a growing field of research on predicting a range of outcomes based on diverse such data sources.[112] Indeed, there is a growing body of evidence indicating that household-survey derived indices of poverty correlate strongly with many geographic features that can be observed from space or derived from ground-based data (Krizhevsky, Sutskever et al. 2012, Sedda, Tatem et al. 2015, Neal, Burke et al. 2016, Engstrom, Hersh et al. 2017).

One of the earlier applications of the use of satellite and geo-spatial data to predict outcomes was the use of night-time lights to predict GDP. Night-time lights are well-suited to predicting cross-country levels of GDP (Henderson, Storeygard et al. 2012, Pinkovskiy and Sala-i-Martin 2016). However, at the within-country level they are much better suited to predicting population density than welfare, and

[112] An organization called Planet currently operates more satellites than even the U.S. and Russian governments. Planet recently launched 88 additional satellites, allowing almost daily coverage of the entire globe with a resolution of 3 to 5 meters per pixel.

the correlation of night-time lights with local wages and local poverty rates has typically been found to be weak (Mellander, Lobo et al. 2015, Engstrom, Hersh et al. 2017). Night-time lights may therefore not be very well suited to uses akin to small-area estimation, particularly in a place such as South Sudan where only about 3 percent of households have access to electricity (Pape, Parisotto et al. 2018). More recent research has focused on training deep-learning algorithms to extract a diverse range of features from high resolution satellite imagery, for example counting the number of cars on a street, distinguishing road types, recognizing materials roofs are made of, tree coverage, the contrast and number of jagged edges, etc. (Engstrom, Hersh et al. 2017). This allows making poverty predictions at a much higher level of disaggregation (Krizhevsky, Sutskever et al. 2012, Sedda, Tatem et al. 2015, Neal, Burke et al. 2016). Engstrom et al. (2017) provide a useful overview of the current state of the literature and show the predictive power of a range of indicators constructed from satellite data in estimating poverty at the village-level.

In the case of the HFS in South Sudan, predictions from a set of linear models were used to project poverty estimates to inaccessible areas based on already extracted satellite features and geo-spatial data, given the objective of creating reliable and transparent poverty measures. The poverty imputation follows a process that is relatively similar to small area estimation, though only the point estimates were estimated and not higher moments of the outcome distribution (see for example: Elbers, Lanjouw et al. 2003, Guadarrama, Molina et al. 2016, Haslett 2016). Poverty as measured in the 2016 wave of the HFS is regressed on a range of geo-spatial characteristics such as distance to urban centers, distance to the electricity grid, annual rainfall, annual temperatures, urban-rural status, IPC phase classification, and others. The estimated model is then used to calculate expected poverty rates across regions where the household survey data are not available, but where the geo-spatial data are available. Poverty rates are predicted at the 100m*100m level across South Sudan. The poverty estimates then need to be weighted by local population counts to eliminate potential bias caused by vast uninhabited areas. Given the lack of reliable administrative data on settlements or population counts, local populations were in turn estimated using a set of covariates derived from geo-referenced data such as urbanicity, roads, clinics, and buildings.

Estimating Settlements Data

The aggregation of poverty estimates to the county and state levels needs to be calibrated against suitable population estimates. Naively aggregating poverty rates across broad geographic regions would result in extremely high poverty rates given the vast uninhabited expanses isolated from the rest of the country, in which a model would likely predict high poverty rates. Indeed, South Sudan is sparsely populated relative even to most other large African countries, in 2008 South Sudan had a population density of approximately 13 persons per kilometer squared compared to the Sub-Saharan Africa average of 35.[113] Because an accurate high-resolution map of population density is not available for South Sudan, the spatial distribution of settlements was used as a proxy for population density in order to calculate weights with which to weight poverty estimates.

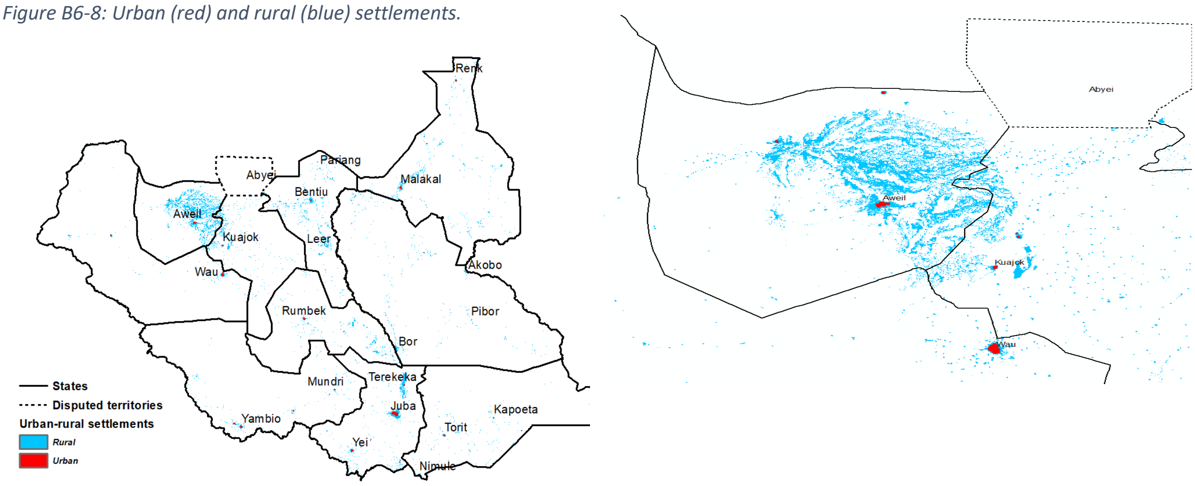
The methodology had to employ a novel process to generate estimates of settlements given the absence of more recent and up to date population data since the 2008 Census. This process was based on a wide variety of data sources and variables associated with population density, leveraging varied sources of data such as open source data from Open Street Maps on residential areas, roads, health facilities, schools, data from the Global Urban Footprint project, as well as data form the survey itself. The map of settled areas in South Sudan was built by processing and regrouping the data sets (Table B6-9). The map of settled areas was created as a binary map (1=settled, 0=not settled) at 100m

[113] World Development Indicators.

resolution. While drawing the map, the data sets were manually checked against Google Satellite imagery for the presence of settlements. One advantage of this system of estimation for settlements is that each component can be updated independently as new data become available or the situation within the country changes. Finally, the map of settlements was adjusted for displacement and for the locations of IDP camps, given extreme rates of displacement in South Sudan.

Other variables were tested but not used for the creation of the map of settled areas (Table B6-9). This includes night-time lights, which are commonly used in studies predicting outcomes from satellite data (Mellander, Lobo et al. 2015, Pinkovskiy and Sala-i-Martin 2016). However, given that only about 3 percent of households in South Sudan have access to a stable source of electricity, there is very little variation to exploit in trying to identify within-country correlations between deprivation and electric light (Pape, Parisotto et al. 2018). Indeed, night time lights would only really predict small industrial enclaves such as oil fields and did not accurately capture where the population actually lives.

Figure B6-8: Urban (red) and rural (blue) settlements.



An 'urban gradient' variable was also derived from the map of settled areas. This estimation was based in large part on the distance to major roads and the wave 1 and wave 3 survey points labelled as 'urban', i.e. the urban classification of enumeration areas based on the 2008 Census exercise. Each 100x100m pixel was classified as a city, city extent, town, town extent, large village, small village, villages far from major roads and unsettled. Distinction between villages and towns was primarily based on the presence of major road intersection and settlement size. A simpler urban/rural settlements map was also produced with only 3 classes: unsettled, rural, urban (towns and cities). All HFS survey points labelled as 'urban' fall in the urban category. Finally, a map of 'distance to urban centers' was created based on the generated urban/rural settlements map.

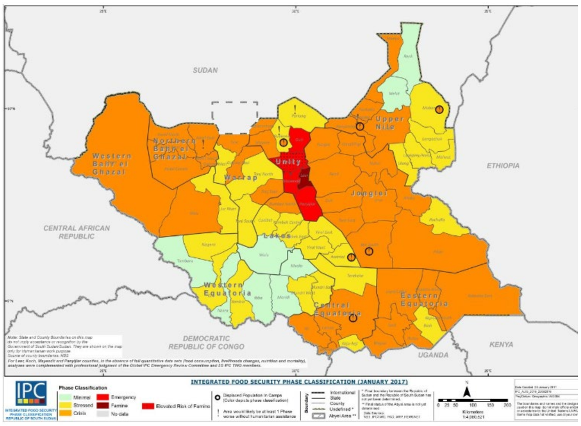
Variable Selection and Model Estimation

Many variables were tested for correlation against each household's probability of being poor averaged per EA. Given that the variance of the probability of being poor was greater across EAs than within EAs, the choice was made to average the probability of poverty per EA. In this manner, a greater degree of spatial variation could be observed, thus increasing the potential to observe meaningful correlations between the probability of poverty and the predictors, i.e. the geo-spatial variables. The variables tested included more traditional geo-spatial characteristics that are commonly used in such applications, such as average temperatures, average rainfall, annual cloud cover variation and annual cloud cover (Table B6-11). It also tested determinants of public services provision and proxies for distance to economic activity, such as distances to different types of roads, urban centers, the electricity grid cultivated areas, schools, and water bodies. Finally, a set of variables indicative of the crisis were used, such as the number of people in need as calculated by OCHA, the IPC phase classification, and the number of conflict fatalities as collected by the Armed Conflict Location Events Data between 2011-16 and between 2014-2016. Finally, the various urban gradients calculated in the previous step were also tested for correlation with poverty rates.

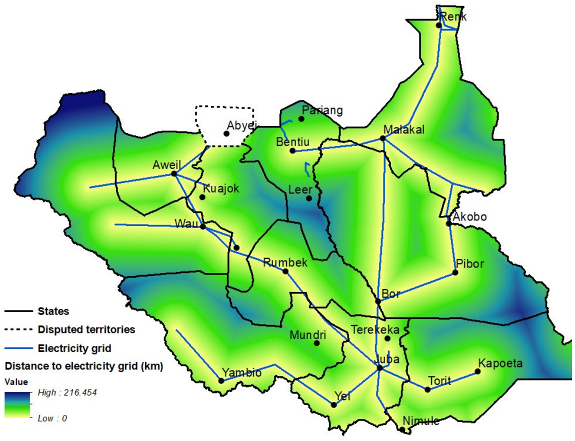
A dummy variable was added for the capital city, Juba, and the former state of Western Equatoria because no variable tested alone could explain the lower levels of poverty observed in Western Equatoria or Juba. The urban gradient alone provided little predictive power as other large towns such as Wau had very high average poverty rates. Therefore, a spatial variable indicating Western Equatoria and Juba was created, with its values smoothed for 200km across the WEQ border and smoothed 2km around the city center of Juba. The resulting map takes the value of 1 in Western Equatoria and in the Juba center, the value of 0 outside these two regions, and a gradient of values between 0 and 1 across its border. This variable does not help explain variation in poverty, but merely reflects observations from the survey and helps to account for chance correlations in the prediction. In other words, this avoids predicting low poverty in the entire western part of the country based on the low poverty rates observed around Western Equatoria and Juba. Of the variables having a relatively large correlation with poverty, some are redundant, some are due to 'chance' as explained above – and some show a trend both within Western Equatoria / Juba and in the rest of the country and hence are deemed as reliable correlations.

Figure B6-9: Example maps of variables used in the estimation.

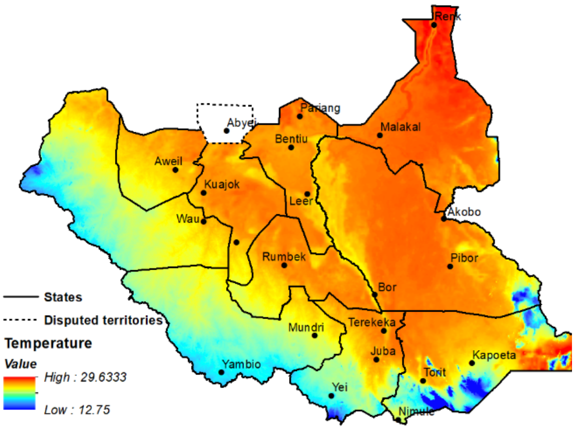
IPC phase classification in January 2017.



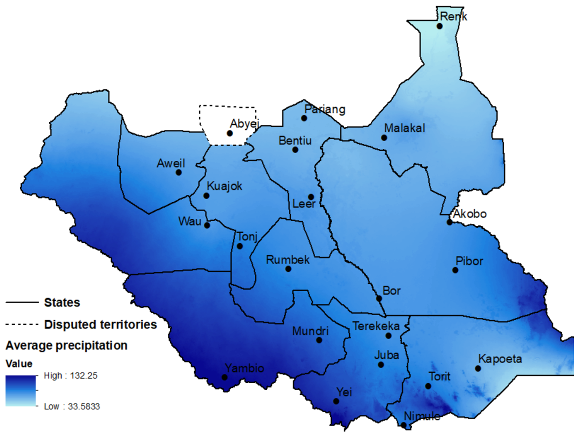
Distance to electricity grid.



Annual average temperatures.



Annual average precipitation.



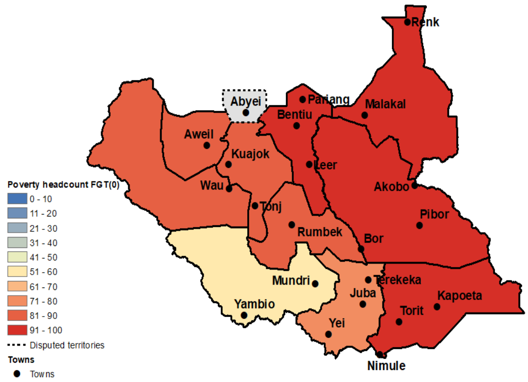
While each of the covariates described above provides some level of predictive power for poverty, a combination of non-orthogonal variables is more likely to better predict poverty. Because of the relatively small number of enumeration areas used in this study (156), focus was placed on a simple linear model. Furthermore, comparisons against polynomial and more complex models indicated that a linear model retained the largest R^2 (=0.7). The level of predictive power was confirmed using an out-of-sample cross validation. In the cross-validation exercise the model was first built using 75 percent of the survey data. Then, the remaining 25 percent was used to predict EA-level poverty values and check the predictive power of the model, therefore confirming the efficiency and validity of the results. The cross-validation approach was performed 10 times and the average predictive power was used. The following variables were selected in the final model: the IPC phase classification, distance to urban centers, annual average temperature, distance to the electricity grid, annual average precipitation, an urban/rural/unsettled dummy, and a dummy for Juba and Western Equatoria (Table B6-12).

Results

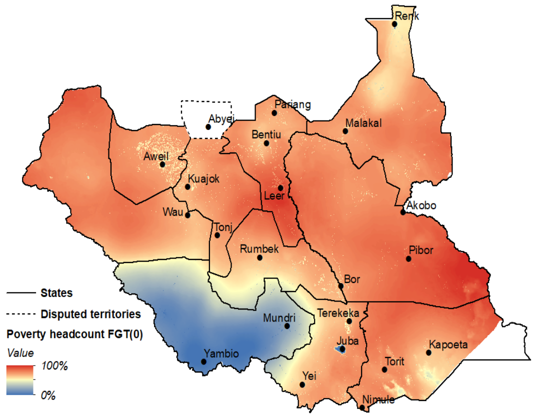
Imputing poverty headcount ratios in the states not covered by the HFS based on satellite and geo-spatial data indicate potentially extremely high levels of poverty in those regions as well. Estimating poverty for every kilometer squared across South Sudan results in the map shown in Figure 10. The poverty map obtained reflects the variations of the in WEQ or Juba variable (lower poverty in WEQ and Juba), and variations of the IPC phase (e.g. North East). The influence of the Distance to urban centers can be seen e.g. around Raga (North West town), and the distance to the electricity grid can also be seen but to a lesser extent. Influence of temperature and precipitations can be seen along the Nile and in the South East. At a smaller geographic scale predicted poverty follows the urban/rural/unsettled classification (Figure 8). The weighted poverty rates indicate extremely high poverty rates in the Greater Upper Nile regions, which is expected given the predominantly rural nature of the region and its state of instability. The poverty headcount across almost all the non-covered states reaches upwards of 9 in 10. Therefore, based on the trends depicted in Table B6-13, the extent of deprivation has reached extremely high levels throughout almost the entire country except for Western Equatoria.

Figure B6-10: Poverty maps, headcount FGT(0) in 2016.

Aggregate per state – imputation in non-HFS states



Poverty predictions at 100mx100m level



Limitations

The results presented here are an attempt to make the best use of available data given a number of limitations. Firstly, no spatial random effect was used in the present model largely due to the fact that EAs were mostly sampling in a North-West / South-East gradient, with little information available on the East-West spatial structure. In the present case, geographic covariates have provided sufficient predictive power that this lack of spatial autocorrelation is not necessarily an issue. However, further data from other regions in the country would provide significant advantages for defining this spatial random component. A related issue is the use of spatially smooth predictors, for example the distance to urban centers and the distance to the electricity grid. Such variables are informative especially with respect to their impact on poverty and can be better predictors than binary variables indicating access based on a cutoff might be. However, they also can have difficulty predicting "pockets of poverty" sitting in otherwise wealthier areas, for example slums in urban areas, or the converse. This could exacerbate the spatially smooth predictions already introduced by the assumption of constant coefficients from the linear regression. Unfortunately, the impact this may have had on the estimation is difficult to test using cross-validation with survey data that was designed to be widely distributed geographically. Therefore, it is impossible to test what the share of variation in welfare across EAs is dampened by the use of these spatially smooth predictors. This is an area which warrants future research given the predictive power of such variables and could be better tested using data collected more finely over large areas such as a Census.

Secondly, there is a very poor understanding of the population distribution in South Sudan and no reliable sampling frame against which to extrapolate our predictions. The implications of this are that while the model can predict into geographic pixels based on the existing data, it is difficult to aggregate by county without knowing how to weight each pixel according to the population present within it.

Thus, poverty maps aggregated by area are likely to over-estimate poverty rates as most areas within each county are likely to have lower population density and high poverty. The solution to this problem is to define a new sampling frame for the country, then re-calculate county-level predictions based on this sampling frame. This was attempted in this study by estimating an urban gradient based on multiple data sources and their relationship with urbanicity. However, some of these data are likely to be out of date for many of the same reasons that a traditional Census exercise is complicated. The rapid and enormous movement of people caused by the conflict is likely to have compounded this problem. Building newer and more up to date population sample frames should be a priority for researchers interested in South Sudan. This could be achieved either by conducting a traditional census, or by leveraging the recently available satellite imagery using and machine-learning based methods to extract features. Such extract could be used to help define settled areas and their associated population density to create a predictive population surface (Neal, Burke et al. 2016, Engstrom, Hersh et al. 2017, Pasquale, McCann et al. 2017). Based on this, new sample frames can be built to use for future data collection work, which is badly needed in the context of South Sudan.

The model structure was voluntarily kept simple (linear combination) to ease its interpretation given that it was constructed as a proof of concept to show the potential of spatial data for imputing poverty to supplement poverty survey estimates. Furthermore, although where these techniques may have the most value, which is where there might have been a crisis or emergency or where safety is a concern, these techniques are also the most difficult to apply. Indeed, the link between poverty and such variables is much more likely to be structural than transient across much of South Sudan. Indeed, a set of issues that arise in this estimation method is the difficulty of modeling the dynamics of poverty and shocks. Many of the areas where the enumerators could not go were inaccessible because of recent conflict and it is difficult to account for this in a cross-sectional model as such, given the potentially endogenous nature of conflict and poverty whereby some conflict events are concentrated around wealthier areas. One of the areas for future research might be to leverage the time series that area available for various types of geo-spatial data to try to account for some of these dynamics relating poverty rates to shocks and imbalances.

Conclusion

The HFS conducted several rounds of data collection at a time of upheaval in the short history of South Sudan. In particular, Wave 3 of the HFS consisted of a major data collection effort during what effectively became one of the deepest humanitarian crises in recent history. The HFS was conceived within the context of the crisis and was therefore designed to leverage new technologies for monitoring and implementation as well as methodological innovations in survey design. This allowed the HFS team to monitor closely the survey and facilitate the implementation while facing a multitude of challenges induced by the escalating crisis. Unfortunately, the growing intensity of the conflict eventually led to a shortened survey with deprived sample size. In the end, after almost 9 months of fieldwork, only about one-half of the intended sample of households was interviewed. While the disruptions caused by the conflict have had impact on the data collected, consistency checks suggest that this impact was relatively small. In addition, any introduced sample selection bias due to the conflict is likely to be a downward bias leading to under-estimation of poverty.

The HFS presents a rare data point in a fragile setting. Only very few similar surveys have managed to collect comprehensive data on welfare and livelihoods in such a complicated and volatile context. Indeed, the HFS documents some staggering levels of deprivation, which are also corroborated by accounts from a multitude of organizations operating in the country. The methodology employed to estimate poverty in the HFS is based on the best available methodologies specifically adapted to the context of fragility. The estimation is also entirely reproducible through the publicly available code and data published in the World Bank MicroData Library.[114] Overall, the HFS provides an extremely detailed picture of welfare and livelihoods for the South Sudanese population between 2015 and 2017. This is especially true when combined with the other three waves conducted between 2015 and 2017, as in the South Sudan Poverty Assessment (Pape, Parisotto et al. 2018).

Finally, the satellite imputation, although limited in scope and means, provides an additional glimpse of livelihoods across the country. Although the results are only a proof-of-concept, it remains a useful exercise to complement the survey-based data rather than assuming a national average for inaccessible areas. Much research has already gone into the field of small area estimation, which is likely to benefit enormously from the recent availability of cheaper and more encompassing – geospatial – data sets. Although such models are not likely to replace survey data, as these are needed to train the models, they can be used to supplement data collection and provide information either at more frequent intervals or for hard-to-reach areas. One particular area for future research that might be especially relevant would be to explore how such sources of data can be leveraged to estimate outcomes during rapidly evolving and dynamic events, exactly when representative surveys and other traditional data collection exercises are especially difficult to implement.

Appendix

APPENDIX A

The High Frequency Survey in South Sudan

The High Frequency Survey conducted waves of almost nationally representative surveys across South Sudan between 2015 and 2017. The HFS was based on a pilot which collected six waves of panel data across 4 of the largest urban centers between 2012 and 2014. The pilot was then scaled up in 2015 to a representative wave covering 6 of the 10 former states of South Sudan. Between 2015 and 2017, the HFS was expanded to a seventh state and conducted three more waves. Waves 2 and 4 were limited to urban areas but included a panel component. The HFS was accompanied by market price surveys which collected weekly price data and daily exchange rate data in 17 locations across the entire country.

Table B6-2: Dates and sample for data collection for all four waves of the HFS.

	Wave 1 Feb.-Oct.2015			**Wave 2 Feb.-Apr.2016**	**Wave 3 Sep.2016-Feb.2017**			**Wave 4 May-Jul.2017**
EAs/HH	**Rural**	**Urban**	**Total**	**Urban**	**Rural**	**Urban**	**Total**	**Urban**
Warrap	-			**15/173**	8/95	5/40	**13/135**	**15/144**
Northern Bahr El Ghazal	40/480	10/120	**50/600**	**15/177**	20/239	5/60	**25/299**	**15/126**
Western Bahr El Ghazal	20/225	30/360	**50/585**	**11/126**	14/166	12/144	**26/310**	**15/137**
Lakes	40/478	10/120	**50/598**	**15/180**	19/172	5/60	**24/232**	**15/133**
Western Equatoria	34/406	16/192	**50/598**	**15/176**	18/216	7/84	**25/300**	**15/156**
Central Equatoria	16/192	34/408	**50/600**	**15/177**	16/192	10/119	**26/311**	**15/95**
Eastern Equatoria	40/453	10/116	**50/569**	**15/180**	20/201	5/60	**25/261**	**15/153**
Total	190/2,234	110/1,316	**300/3,550**	**101/1,189**	115/1,281	49/567	**164/1,848**	**105/944**

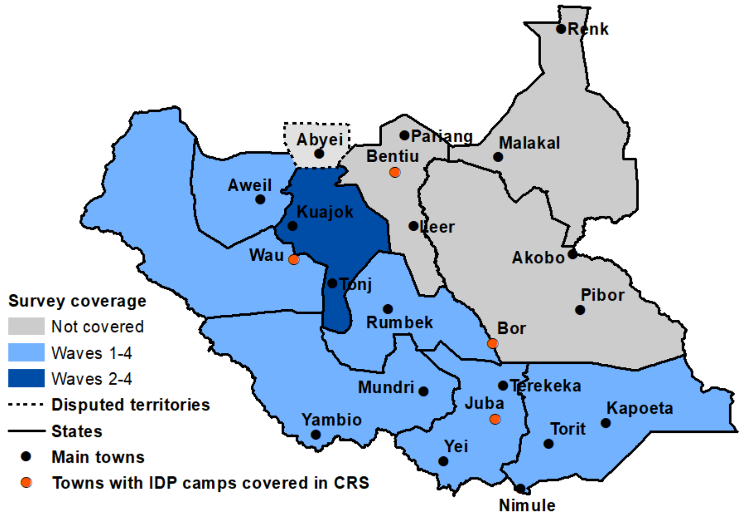
The fourth wave of the HFS was accompanied by the Crisis Recovery Survey (CRS), a representative survey of four of the largest IDP camps in South Sudan. The CRS was conducted simultaneously to Wave 4 of the HFS in mid-2017. It covered the four largest protection of civilian (PoC) camps with well-defined boundaries accessible to enumerators. The camps include Bentiu PoC, Bor PoC, Juba PoC1 and 3, and Wau PoC. Although the CRS covers PoCs, where only 12 percent of South Sudan's IDPs are

[114] See: http://microdata.worldbank.org/index.php/catalog/2914

located, the detailed microdata fill important information and knowledge gaps for IDP-focused programming.

The HFS and CRS questionnaires cover a large range of topics and draw a well-rounded picture of socio-economic livelihoods of people in South Sudan. The HFS questionnaire covers topics including demographics, employment, education, consumption, as well as perceptions of well-being and of the effectiveness of public institutions. Consumption is measured using the newly developed rapid consumption methodology. The CRS and Wave 4 HFS questionnaires, designed to be exactly comparable, also collected details on displacement-specific outcomes guided by the IASC framework.[115] These were developed to understand the motivations for displacement, return intentions, sense of security, relations with the surrounding community, social capital, and pre-displacement outcomes in the standard of living, education and labor.

Figure B6-11: High Frequency Survey coverage, 2015-2017.



APPENDIX B

Table B6-3: Sample design calculations.

	No. HH (Census)	Urban (%)	Mean (Cons.)	std dev	equal/optimal (10 min) Urban EAs	Rural EAs	rel. err.
Central Equatoria	175,962	31.2%	133.0	90.0	13	37	0.032
Eastern Equatoria	151,199	9.9%	107.3	80.2	10	40	0.045
Western Equatoria	115,595	17.1%	126.1	99.9	13	37	0.028
Warrap	167,654	7.6%	73.3	49.8	12	38	0.043
Western Bahr El Ghazal	57,487	44.7%	122.1	144.6	33	17	0.029
Northern Bahr El Ghazal	130,832	6.3%	61.1	52.1	10	40	0.049
Lakes	90,315	7.2%	119.3	119.0	10	40	0.019
Rural	746,136	--	94.3	74.0	--	249	0.003
Urban	142,908	--	152.4	155.1	101	--	0.098
Total	889,044	16.1%	103.5	90.1	101	249	0.026

Sampling weights

[115] The Inter Agency Standing Committee Framework on Durable Solutions for Internally Displaced Persons aims to provide guidance for achieving durable solutions following internal displacement in the context of armed conflict, situations of generalized violence, violations of human rights and natural or human-made disasters. The Framework primarily aims to help international and non-governmental actors to better assist governments dealing with humanitarian and development challenges resulting from internal displacement. The Framework is also designed so that it can be used to assist those in the field in determining whether a durable solution to internal displacement has been found, depending on the context of the local environment.

Sampling weights are used to make survey observations representative for the sample. The sampling weight is the inverse probability of selection. The selection probability *P* for a household can be decomposed into the selection probability P_1 of the EA and the selection probability P_2 of the household within the EA:

(1) $$P = P_1 P_2$$

The selection probability P_1 of an EA *k* is calculated as the number of households within the EA divided by the number of households within the stratum multiplied by the number of selected EAs in the stratum:

(2) $$P_1 = \frac{|K|\hat{n}_k}{\sum_{k' \in K} \hat{n}_{k'}}$$

where $\hat{n}_k$ denotes the number of households in EA *k* estimated using the Census 2008 data and K is the set of EAs selected in the corresponding stratum. Replacement enumeration areas were assigned the sampling weight of the enumeration area that they were replacing. In Wave 3, the number of enumeration areas surveyed in each stratum differed from the original sample. The weights were therefore scaled to correct for the change in the value of K.

The selection probability P_2 for a household within an EA *k* is constant across households and can be expressed as:

(3) $$P_2 = \frac{|H|}{n_k}$$

where $|H|$ is the number of households selected in the EA and n_k denoting the number of listed households in EA *k*. Usually, the number of households per EA is 12 while a few exceptions exist due to invalid interviews.

Sampling weights were scaled to equal the number of households per strata using the Census 2008 data. Thus, the sampling weight *W* can be written as:

(4) $$W = \frac{c}{P} \text{ with } c = \frac{\sum_{k \in K} \hat{n}_k}{\sum_{k \in K} n_k}$$

Table B6-4: No. of enumeration areas per strata, 2016.[116]

	Intended			Actual		
	Rural	**Urban**	**Total**	**Rural**	**Urban**	**Total**
Warrap	37	13	50	8	5	13
Northern Bahr El Ghazal	40	10	50	20	5	25
Western Bahr El Ghazal	37	13	50	14	12	26
Lakes	38	12	50	19	5	24
Western Equatoria	17	33	50	18	7	25
Central Equatoria	40	10	50	16	10	26
Eastern Equatoria	40	10	50	20	5	25
Total	**37**	**13**	**350**	**115**	**49**	**164**

APPENDIX C

Cleaning consumption data

[116] Note that the date of data collection refers to the period when most of the interviews were collected. In some cases, a few interviews were conducted in the month after the end of fieldwork as part of follow-ups to improve data quality.

Food expenditure data are cleaned in a three-step process. First, units for reported quantities of consumption and purchase are corrected. Second, quantities consumed and purchased converted into kilograms are cleaned, where potential data entry errors and outliers are detected and corrected. Third, prices per kilogram calculated using the cleaned quantities are corrected in a similar manner. The cleaning rules were maintained across the 4 survey waves to ensure comparability. More details on the specific cleaning rules are provided below:

Rule 1 (data entry errors for units): For records that have the same figure in quantity purchased and consumed but have different units, it is assumed that the correct unit is the one that takes the quantity (consumed or purchased, converted into kilograms) closer to the weighted median value for the same item.

	N	%
Not-tagged	14,818	99.5
Tagged	70	0.5
Total	14,888	100

Rule 2 (mistakes in reported units): Items that are likely to be reported in the wrong unit are corrected following generic rules. An example of a typical mistake is to report consumption of 100 kilograms of a product (like salt) where the supposed correct unit is grams. In this case, all quantities given in kilograms that exceed 10s0 would be corrected so as to be given in grams instead.

Cons. Q.	N	%
Not-tagged	14,871	99.9
Tagged	17	0.1
Total	14,888	100

Purc. Q.	N	%
Not-tagged	14,507	97.4
Tagged	381	2.6
Total	14,888	100

Rule 3 (missing quantities): Items that were consumed but have a missing quantity, consumed or purchased, are replaced with the item-specific median quantity.

Cons. Q	N	%
Not-tagged	12,851	86.3
Tagged	2,037	13.7
Total	14,888	100

Purc. Q.	N	%
Not-tagged	13,211	88.7
Tagged	1,677	11.3
Total	14,888	100

Rule 4: (quantities beyond 'hard' constraints): Quantities consumed and purchased that are below or above the item-unit quantity constraints are replaced with the item-specific median.

NONE

Rule 5 (data entry errors for quantities or prices): Records with the same value for quantity consumed or quantity purchased and price, or with the same value for all three, are assumed to have a data entry error in the price or quantity. They are replaced with the item-specific medians.

	N	%
Not-tagged	14,859	99.8
Tagged	29	0.2
Total	14,888	100

Rule 6 (quantities per capita too high): For items consumed by more than 300 households, quantities that were 3 standard deviations above the mean value per capita were replaced with item-specific medians.

Cons. Q	N	%
Not-tagged	14,757	99.1
Tagged	131	0.9
Total	14,888	100

Purc. Q.	N	%
Not-tagged	14,780	99.3
Tagged	108	0.7
Total	14,888	100

Rule 7 (missing prices): Items that were consumed but have zero or missing prices are replaced with the item-specific median prices. The reason why this is so high is because many households obtained much of the food consumed from home production, and thus could not answer when asked the price at which they purchased these goods.

	N	%
Not-tagged	11,715	78.7
Tagged	3,173	21.3
Total	14,888	100

Rule 7 (price outliers): Prices in the item-specific price per kilogram distribution that lie above the 95th percentile are replaced with item-specific medians, so are prices for items consumed by more than 300 households that lie above 3 standard deviations above the mean.

Hard constraints	N	%
Not-tagged	14,531	97.6
Tagged	357	2.4
Total	14,888	100

3 sd	N	%
Not-tagged	13,885	93.3
Tagged	1,003	6.7
Total	14,888	100

All medians are estimated at the EA level if a minimum of 5 observations are available. If the minimum number of observations is not met, weighted medians are estimated at the strata-level requiring a minimum number of 10 observations before proceeding to the item level. Medians are estimated excluding zero values and tagged values so as not to replace reported values with zeroes or invalid values.

The non-food data set only contains price values without quantities and units, the cleaning process was therefore much simpler. Two cleaning rules are applied and tagged observations are replaced with item-specific medians at the EA, state, and survey level as is done for food consumption. The cleaning rules are the following:

Rule 1 (price outliers): Prices that are beyond the hard constraints, above or below, are replaced with item-specific medians. Given the high inflation over the subsequent HFS waves, the value of the hard constraints used in Wave 1 were adjusted for inflation using the national NBS CPI.

Max	N	%
Not-tagged	10,864	94
Tagged	689	6
Total	11,553	100

Min	N	%
Not-tagged	10,969	94.9
Tagged	584	5.1
Total	11,553	100

Rule 2 (zero or missing prices): Zero and missing prices for consumed items are replaced with item-specific medians.

Zero	N	%
Not-tagged	11,310	97.9
Tagged	243	2.1
Total	11,553	100

Missing	N	%
Not-tagged	10,862	94
Tagged	691	6
Total	11,553	100

The medians are calculated following exactly the same process as in food cleaning. All medians are estimated at the EA level if a minimum of 5 observations are available. If the minimum number of observations is not met, weighted medians are estimated at the strata-level requiring a minimum number of 10 observations before proceeding to the item level. Medians are calculated excluding zero values and tagged values so as not to replace reported values with zeroes or invalid values.

For durables, the cleaning process involved cleaning ownership statistics as well as the calculated depreciation rates. The quantity of an item is replaced by the item-specific survey median (due to paucity of data) if the reported quantity is unrealistically high assessed by manual inspection. The purchase value of durables is recorded in the year and currency of purchase. Outliers of purchase values in the reported currency are identified by hard constraints and replaced by the item-specific survey median. Items with at least 3 observations purchased in the same year are replaced by the respective item-year specific median. Alternatively, the item-state-level median prices are used if at least 5 observations are given. Hypothetical selling prices are replaced by the item-state level median if at least 5 observations are available. Without the minimum number of observations available, the item-specific median is used. All prices reported in foreign currencies are converted into SSP through conversion to USD.

Rule 1 (quantity outliers): Quantities above 100 units of an asset are replaced with the item-specific median.

	N	%
Not-tagged	5,007	99.9
Tagged	5	0.1
Total	5,012	100

Rule 2 (price outliers): (i) Prices above hard constraints are replaced with the item-specific median. (ii) For specific assets where outliers are identified that fall below the hard constraints and for which we have enough observations to estimate a distribution, the top 5 percent of observations are replaced with item-specific medians.

Selling Above	N	%
Not-tagged	5,004	99.8
Tagged	8	0.2
Total	5,012	100

Purchase Above	N	%
Not-tagged	4,759	95
Tagged	253	5
Total	5,012	100

Selling Below	N	%
Not-tagged	4,851	96.8
Tagged	161	3.2
Total	5,012	100

Purchase Below	N	%
Not-tagged	4,654	92.9
Tagged	358	7.1
Total	5,012	100

Rule 3 (missing prices): Missing prices are replaced with the item-specific median.

Missing Purchase	N	%
Not-tagged	3,713	74.1
Tagged	1299	25.9
Total	5,012	100

Missing Selling	N	%
Not-tagged	2,569	51.3
Tagged	2443	48.7
Total	5,012	100

Rule 4 (missing vintages): Items with missing vintages are replaced with the item-specific median.

	N	%
Not-tagged	4,602	91.8
Tagged	410	8.2
Total	5,012	100

Table B6-5: Core vs. module shares117

	Food Consumption				Non-Food Consumption			
	Number of items	Share NBHS 2009	Share HFS 2016 (collected)	Share HFS 2016 (imputed)	Number of items	Share NBHS 2009	Share HFS 2016 (collected)	Share HFS 2016 (imputed)
Core	33	80%	92%	73%	26	65%	89%	61%
Module 1	27	5%	3%	12%	21	8%	2%	8%
Module 2	26	5%	2%	6%	20	9%	4%	14%
Module 3	26	5%	2%	6%	18	7%	3%	10%
Module 4	28	5%	1%	3%	25	11%	2%	7%
Total	**140**	**100**	**100**	**100**	**110**	**100**	**100**	**100**

Table B6-6: Estimated median depreciation rates.[118]

Asset	Depreciation rate	Asset	Depreciation rate
Cars	0.05	Radio or transistor	0.17
Trucks	0.02	Mobile phone	0.21
Motorcycle/motor	0.12	Computer or laptop	0.03
Rickshaw	0.12	Refrigerator	0.05
Bicycle	0.04	Fan	0.16
Canoe or boat	0.04	Mattress or bed	0.10
Plough	0.21	Mosquito net	0.11
Television	0.04	Electric ironer	0.07
Satellite dish	0.12	Hoe, spade or axe	0.12
DVD or CD player	0.16		

Table B6-7: Urban and rural Laspeyres deflators, 2016.

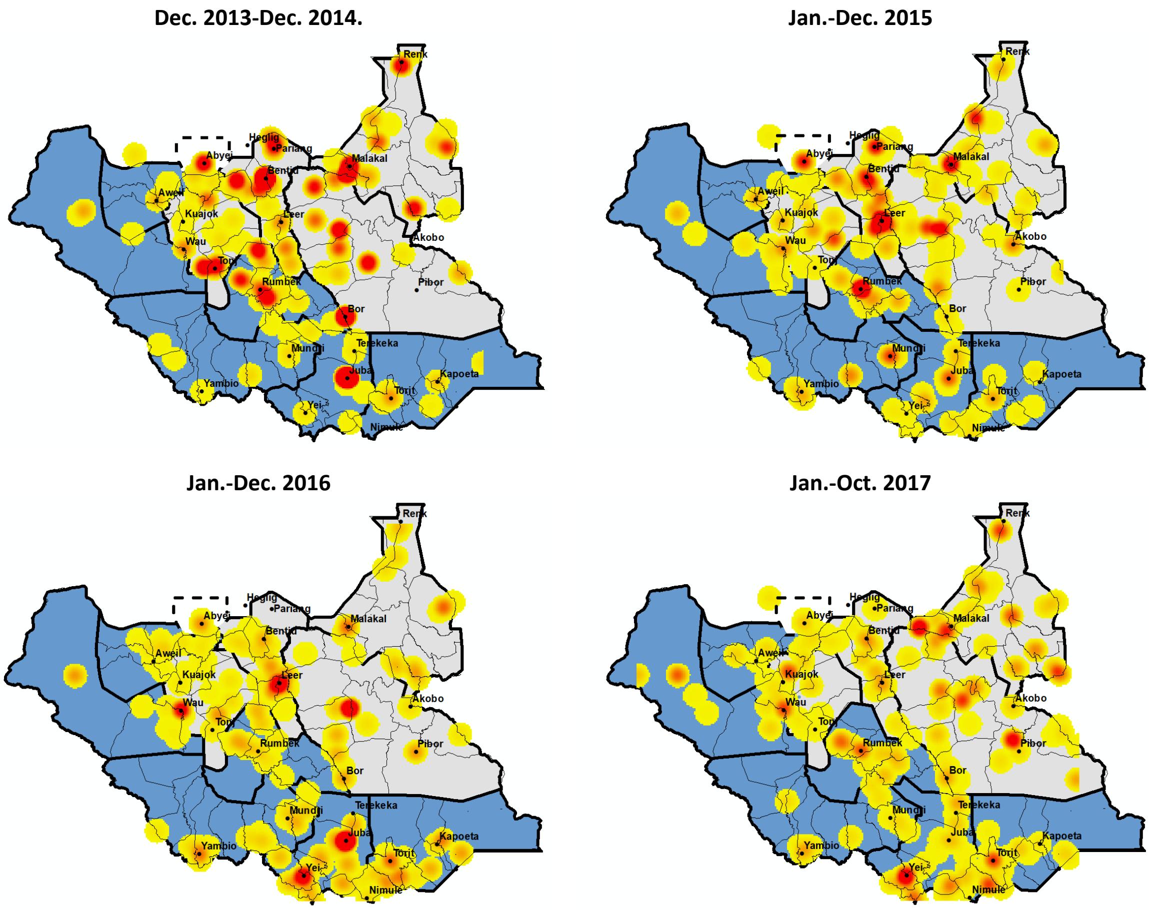
	Food		Non-Food	
	Rural	Urban	Rural	Urban
Sep-16	1.09	1.15	0.83	1.08
Oct-16	1.18	1.00	0.88	1.00
Nov-16	1.21	1.23	1.08	1.67
Dec-16	1.05	1.23	0.86	1.67
Jan-17	1.11	0.99	0.95	1.25
Feb-17	1.07	1.37	1.07	1.43
Mar-17	1.25		1.54	
Apr-17	1.46		1.43	

Refence strata and period is Urban areas in October 2016.

[117] The share of module 4 is missing in the HFS 2015 data due to a technical glitch. See footnote 21.
[118] Washing machines and Air conditioners were not bought.

APPENDIX D

Figure B6-12: Heatmap of conflict fatalities, Dec. 2013-Oct. 2017.



Note: all densities in maps above are color-labelled on the same scale; counties lying outside of the state boundaries are disputed territories.

Table B6-8: Difference in means between selected variables in Wave 1 and Wave 3.

	(1)	**(2)**	**t-test**
	Wave 3 - 2016	**Wave 1 - 2015**	**(1)-(2)**
	Mean/SE	**Mean/SE**	**Difference**
Household owns its property	0.905	0.916	-0.011
	[0.009]	[0.009]	
Phone network available at household	0.236	0.244	-0.008
	[0.012]	[0.013]	
Household is more than two hours walking from a health center	0.299	0.27	0.029
	[0.016]	[0.012]	
Household is more than two hours walking from a school	0.103	0.124	-0.021
	[0.011]	[0.008]	
Household is more than two hours walking from a market	0.327	0.32	0.008
	[0.016]	[0.012]	
Household has access to electricity	0.014	0.028	-0.014
	[0.003]	[0.008]	
Adult literacy rate (18+)	0.376	0.353	0.023
	[0.009]	[0.008]	
Adults with no education (18+)	0.564	0.55	0.014

	[0.009]	[0.008]	
Adults with only primary education (18+)	0.242	0.233	0.009
	[0.008]	[0.007]	
Household head practices polygamy	0.333	0.338	-0.005
	[0.015]	[0.012]	
Household head is male	0.579	0.603	-0.024
	[0.016]	[0.013]	
Household head is employed	0.732	0.715	0.018
	[0.014]	[0.012]	
Average age	19.311	19.026	0.285
	[0.217]	[0.186]	

*** and * indicate significance at the 1 and 5 percent level.*

APPENDIX E

Table B6-9: Variables used to create a map of settled areas.

Variable name	Description	Processing step
Variables used		
Global Urban Footprint	Infrared-based raster of predicted presence or absence of buildings.	Dilated 100m
OSM residential areas	Volunteer-reported residential locations	Rasterized
OSM buildings	Volunteer-reported point locations of buildings	Rasterized, dilated 100m
OSM residential roads	Volunteer-reported vector of road locations	Rasterized, dilated 100m, then eroded to identify blob-like structures (residential areas)
OSM road intersection	Volunteer-reported point locations of road intersections	Rasterized, Dilated 100m
OSM health sites	Volunteer-reported locations of health facilities	Rasterized, Dilated 200m
WB health facilities	Point locations of health facilities reported in WB Points of interest database.	Rasterized, Dilated 200m
Schools	Point locations of schools reported in WB Points of interest database.	Rasterized, Dilated 200m
Household survey interviews, HFS Wave 3.	Data points from the HFS Wave 3 survey	Rasterized, Dilated 100m
Variables rejected		
Variable name	**Description**	**Reason not used**
Night time lights DMSP	Satellite-detected intensity of night-time visible light radiance.	Brightest for power plants and oil fields in the north. These data do not bring more information on settled areas
Night time lights VIIRS	Satellite-detected intensity of night-time visible light radiance.	Often high level in areas that do not appear to be settled on satellite imagery
Waterpoints	GPS coordinates of reported water points	Many water points were not in settlements - perhaps because dataset is dated (<2012)

Note: 'rasterized' means that point or vector data were converted to gridded data at100m. 'Dilated' means that pixels were added around 'on' pixels, expanding shapes or points by a constant radius. 'Eroded' means that outer pixels of shapes were removed, suppressing linear structures and keeping only the core of blob like structures.

Table B6-10: Summary Statistics of Geo-Spatial variables

	Mean country	Min country	Max country	Mean settle	Min settle	Max settle	Mean sample	Min sample	Max sample
Distance to electricity grid (km)	114.01	0.00	505.08	27.67	0.00	213.40	12.80	0.00	72.65
Distance to schools (km)	86.42	0.00	459.24	6.33	0.00	200.77	3.59	0.00	24.83
Distance to waterpoints (km)	91.96	0.00	470.21	7.18	0.00	119.63	1.78	0.00	21.99
Distance to national roads (km)	127.22	0.00	491.87	17.03	0.00	235.94	6.96	0.00	66.82
In WEQ or Juba	0.06	0.00	1.05	0.09	0.00	1.05	0.26	0.00	1.04
IPC phase Jan. 2017 smoothed 50km	2.54	0.98	4.58	2.70	0.99	4.57	2.36	1.01	3.00
MODCF intra annual SD 100mres	1763.08	449.00	3005.00	1906.11	650.00	2933.0	1622.45	980.00	2249.00
MODCF mean annual 100mres	5338.89	3014.00	9199.00	5115.11	3084.0	8195.0	5600.55	4316.00	7945.00
SSD conflicts 2011 2016	21.46	0.00	4130.17	239.32	0.00	4130.1	378.19	0.04	3051.54
SSD conflicts 2014 2016	12.34	0.00	1608.22	128.57	0.00	1608.2	181.64	0.04	960.01
Distance to major roads (km) 100mres	83.87	0.00	470.11	5.56	0.00	131.34	2.40	0.00	26.39
Distance to plantations in 2014 100mres	71.66	0.00	458.39	2.22	0.00	96.44	2.13	0.00	23.62
Distance to urban centres (km) 100mres	158.95	0.00	543.57	38.77	0.00	256.28	14.74	0.00	74.80
Precipitations 100mres	959.98	405.54	1586.70	892.39	411.00	1584.6	1015.16	752.43	1538.01
OCHA percent people in need, 2016	24.63	0.00	252.70	51.67	0.00	252.70	36.70	33.01	115.18
Temperature 100mres	26.90	12.81	28.63	27.24	18.13	28.59	26.85	23.53	27.86
Urban gradient	0.03	0.00	8.00	3.55	1.00	8.00	4.92	2.00	8.00
Urban-rural settlements	0.01	0.00	2.00	1.04	1.00	2.00	1.35	1.00	2.00

Table B6-11: Variables tested for correlation with poverty.

Variable	Correlation with poverty
IPC phase (01/2017)	0.34
Seasonal cloud cover variations	0.28
Annual cloud cover	-0.37
OCHA nb people in need	0.02
Mean conflict fatalities 2011-2016	-0.49
Mean conflict fatalities 2014-2016	-0.51
Distance to 1,2,3 roads	0.02
Distance to cultivated areas 2014	0.17
Distance to urban centres	0.5
Annual temperature	0.41
Distance to electricity grid	0.36
Distance to schools	0.25
Distance to water bodies	0.10
Distance to national roads	0.25
Annual precipitation	-0.61
Urban gradient	-0.41
Urban-rural-unsettled	-0.45
In WEQ	-0.62
In Juba	-0.44
In WEQ or Juba	-0.81

Table B6-12: Estimated coefficients for best-fit linear model.

Variable name	Coefficient Estimate
(Intercept)	0
IPC phase	0.04
Distance to urban centers	4.7e-4
Annual temperature	0.03
Distance to electricity grid	3.6e-4
Annual precipitation	2.0e-4
urban/rural/unsettled	-0.13
In WEQ or Juba	-0.46

Table B6-13: State-level predictions of poverty headcount (percent).

	Poverty (survey)	***Poverty (predicted)***	***Poverty Rural (survey)***	***Poverty Rural (predicted)***	***Poverty Urban (survey)***	***Poverty Urban (predicted)***
Central Equatoria	*80*	*76*	*84*	*84*	*17*	*63*
Eastern Equatoria	*95*	*91*	*97*	*94*	*28*	*42*
Jonglei		*92*		*95*		*17*
Lakes	*84*	*86*	*86*	*89*	*29*	*47*
Northern Bahr el Ghazal	*90*	*90*	*91*	*93*	*12*	*68*
Unity		*92*		*95*		*17*
Upper Nile		*92*		*95*		*36*
Warrap	*86*	*89*	*90*	*92*	*43*	*65*
Western Bahr el Ghazal	*90*	*88*	*53*	*92*	*38*	*70*
Western Equatoria	*53*	*68*	*61*	*74*	*39*	*31*
Total	***83***	***92***	***86***	***92***	***66***	***77***

7. Estimation of Poverty in Somalia Using Innovative Methodologies[119]

Utz Pape and Philip Wollburg[120]

Introduction and related literature

Somalia gained independence in 1960. The collapse of Siad Barre's post-independence regime in 1991 led to civil war between local power factions and dismantled the central state completely. Between 1995 and 2000, regional administrations emerged across the country, as security improved and economic development accelerated.[121] The formation of the Transitional Federal Government in 2004 and of its successor, the Federal Government of Somalia, in 2012 marked the return of a significant central state institution. After peaceful elections in 2016, a new government was formed in 2017 committed to embark on a development trajectory (World Bank 2017).

Though Somalia remains one of the world's poorest countries (World Bank 2015, World Bank 2016), a vibrant but largely informal private sector sprouted in the absence of government, drove growth in the Somali economy, and took on the provision of services. Several economic activities including telecommunications, money transfer businesses, livestock exports, and localized electricity services grew well during this period (World Bank 2017). Large-scale out-migration of skilled Somalis who sent back part of their earnings made diaspora remittances essential to the Somali economy, equivalent to between 23 and 38 percent of GDP and outweighing both international aid flows and foreign direct investment (World Bank 2015).

Despite improvements in political stability, Somalia remains fragile. Parts of southern Somalia are inaccessible due to the presence of Al-Shabaab, which also repeatedly carried out terroristic attacks, and violent clashes between various power factions continue to occur throughout the territory.[122] In addition to conflict, the cyclical El Nino phenomenon caused severe droughts in 1991/92, 2011/12, and 2016/17 which exacerbated preexisting vulnerabilities in the Somali population. Both conflict and drought have led to large-scale internal displacement (World Bank 2018). The recent 2016/17 drought led to the displacement of approximately one million Somalis, adding to an existing population of internally displaced persons of 1.1 million (UNHCR 2018).

As is typical for fragile states, Somalia is highly data-deprived, leaving policy makers to operate in a statistical vacuum (Beegle, Christiaensen et al. 2016). Specifically, years of civil war and ongoing conflict have eroded Somalia's statistical infrastructure and capacity, leading to the lack of key macro- and micro-economic indicators, including the poverty rate (Hoogeveen and Nguyen 2017). The government conducted and published the last full population census in 1975, while Somalia Socioeconomic Survey of 2002 was the last country-wide household survey (UNFPA 2014). Most recent existing data sources are local FSNAU and FAO food and nutrition surveys, while organizations operating within Somalia implemented a range of smaller surveys. In 2014, UNFPA implemented the first nationwide Population Estimation Survey (PESS) in preparation for a national census, finding the

---

[119] UP developed the research question and designed as well as supervised the field work. PW and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

[120] Authors in alphabetically order. Corresponding author: Utz Pape (upape@worldbank.org). The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. Gonzalo Nunez contributed to the survey design and poverty analysis and provided inputs to this manuscript. The authors would like to thank Kristen Himelein and Wendy Karamba for discussions. In addition, the authors thank Véronique Lefebvre, Sarchil Qader, Amy Ninneman, Dana Thomson and Tom Bird from Flowminder and WorldPop for designing the population sampling frame using quadtrees and producing fieldwork maps, and for modelling and imputing poverty from spatial data, in collaboration with the authors.

[121] Somaliland self-declared independence in 1991.

[122] See the Armed Conflict Location and Events Database (ACLED), Somalia, for a disaggregated overview.

total population to be 12.3 million, of which 42 percent are urban, 23 percent rural, 26 percent nomadic, and 9 percent are internally displaced (UNFPA 2014).

Funded by the World Bank, Somaliland carried out a household budget survey (SLHS) in 2013, which generated much-needed indicators, including poverty estimates, but the sample was not representative especially for the rural population and did not cover the nomadic and displaced populations. The World Bank conducted the first wave of the Somali High Frequency Survey (SHFS) in the spring of 2016, representative of the accessible urban, rural, and IDP population in 9 of 18 prewar regions as well as Mogadishu, providing a baseline dataset for monitoring poverty and contributing to other key statistical indicators. However, in addition to large inaccessible areas, the sample excluded nomadic population and households in insecure areas. Furthermore, the rural sampling frame had to be derived ad-hoc with only limited representativeness. Wave 2 of the SHFS, implemented in December of 2017, significantly expanded coverage to urban and rural areas in central and southern Somalia and included the nomadic population for the first time, while a newly derived sampling frame enhanced overall representativeness.

The specific context of insecurity and lack of statistical infrastructure in Somalia posed a number of challenges for implementing a household survey and measuring poverty. First, in the absence of a recent census, no exhaustive lists of census enumeration areas along with population estimates existed, creating challenges to derive a probability-based representative sample. Second, while some areas remained completely inaccessible due to insecurity, even most accessible areas held potential risks to the safety of field staff and survey respondents, so that time spent in these areas had to be minimized. Third, poverty in completely inaccessible areas had to be estimated by other means. Finally, the non-stationary nature of the nomadic population required special sampling strategies. This paper outlines how these challenges were overcome in wave 2 of the SHFS through methodological and technological adaptations in four areas: sampling strategy, survey design, fieldwork implementation, and poverty measurement. In line with the challenges outlined above, this paper contributes to several themes in the literature on poverty measurement and data collection in the context of conflict and fragility, involving hard-to-survey populations.

First, geospatial techniques and high-resolution imagery were used in the SHFS to model the spatial population distribution, build a probability-based population sampling frame, and generate enumeration areas in an effort to overcome the lack of a recent population census. The SHFS sampling strategy bears resemblance to the strategy proposed by Muñoz and Langeraar (2013), which relies satellite imagery and grid cells to build a sampling frame in Myanmar. Wardrop, Jochem et al. (2018) review various efforts to produce spatially disaggregated population estimates based on satellite imagery, in contexts where census data is absent or inaccurate. Barry and Rüther (2005) and Turkstra and Raithelhuber (2004) employ satellite imagery to study informal urban settlements in South Africa and Kenya, respectively, while Aminipouri, Sliuzas et al. (2009) estimates various slum populations in Dar-es-Salaam, Tanzania. Himelein, Eckman et al. (2016) compare the viability of various satellite and area-based sampling methods in second-stage sample selection in Mogadishu, Somalia.

Second, risks to the safety of field staff required spending as little time in enumeration areas as possible. One strategy to address this issue is to call or message respondents on their mobile phones and not visit dangerous areas at all. A growing body of literature explores the use of mobile technology in this context (e.g. Dillon 2012, Demobynes and Sofia 2016, Firchow and Mac Ginty 2016). However, administration of necessary consumption modules to estimate poverty is not feasible via phone surveys.

To address security concerns, the SHFS adapted logistical arrangements, sampling strategy, and questionnaire design to limit time on the ground. In logistical arrangements, a detailed and timely security assessment ensured that the enumeration areas to-be-visited were safe on the day of fieldwork. The fieldwork protocol was designed such that teams would spend as little time as possible in any given region and draw little attention, ensuring enumerator and respondent safety. Concerning sampling strategy, it was not feasible to conduct a full listing of all households in an enumeration area, as this was too time-intensive and may have raised suspicion. Instead, a micro-listing approach was used, which required enumeration areas to be segmented into smaller enumeration blocks using satellite imagery. Enumeration blocks are small enough for enumerators to list and select households immediately before conducting the interview. Himelein, Eckman et al. (2016) compare this methodology with other second-stage sampling strategies designed for use in fragile and time-sensitive settings.

Complete food and nonfood consumption modules result in an overall questionnaire length that is prohibitive in areas with high insecurity. The length of consumption modules can be reduced by removing rarely consumed items from the module or to combine categories of items (e.g. vegetables) and ask aggregates rather than individual items. Beegle, De Weerdt et al. (2012) and Olson-Lanjouw and Lanjouw (2001) provide evidence that both approaches lead to an underestimation of consumption and hence an overestimation of poverty. Fujii and Van der Weide (2013) propose an alternative approach which could be adapted for use in fragile settings, by assigning a full consumption module to households in areas without a binding security and time constraint, with only the covariates of consumption administered to households in insecure areas. Consumption and poverty could then be imputed based on those covariates. This approach, however, potentially leads to biases as the assignment of the two different modules depends on security and is not necessarily random. Instead, the Rapid Consumption Methodology (Pape and Mistiaen 2018) was used to significantly reduce the length of the survey's consumption modules. The Rapid Consumption Methodology used in the SHFS relies on a set of core consumption items administered to all households. The remaining items are algorithmically partitioned into optional modules distributed systematically across households, with multiple imputation techniques used to impute total consumption and poverty. Pape and Mistiaen (2018) show that this design yields reliable poverty estimates.

Third, the SHFS relies on correlates derived from satellite imagery and other geo-spatial data to estimate poverty in areas that remained completely inaccessible as a result mainly of insecurity. A growing field of research is dedicated to predicting a range of outcomes based on a diverse set of such data sources. Early applications use night-time lights data to predict economic activity. These data are particularly successful at predicting GDP at the country-level (Henderson, Storeygard et al. 2012, Pinkovskiy and Sala-i-Martin 2016), but appear less well-suited for measuring income and when variation in welfare is desired at a highly disaggregated level (Mellander, Lobo et al. 2015, Engstrom, Hersh et al. 2017). More recently, deep learning techniques applied to daytime imagery in order to classify such objects as roof types, roads, tree coverage, and crops has led to advances in measuring welfare at more disaggregated levels (Krizhevsky, Sutskever et al. 2012). Neal, Burke et al. (2016) use a convolutional neural network based on daytime satellite features to predict per capita consumption at the level of the enumeration area from living standards measurement surveys. Their model is successful in predicting consumption and explains 46 percent of variation on average across four countries and out-of-sample. Engstrom, Hersh et al. (2021) provide a recent overview of the state of the literature and use high-resolution satellite features to estimate poverty at the village-level. In the SHFS, estimating poverty in inaccessible areas relied on a linear model with the objective of creating reliable and transparent poverty measures.

The remainder of this paper proceeds as follows. The next section discusses the sampling strategy. The following section provides an overview of the data collection process, with the subsequent section describing the derivation of the consumption aggregate, including the Rapid Consumption Methodology. The paper closes by presenting the imputation of poverty in inaccessible areas, and an overview of poverty in Somalia.

Sampling strategy

Wave 2 of the SHFS employed a multi-stage stratified random sample, ensuring a sample representative of all sub-populations of interest, while optimally balancing cost and precision of estimates. Strata were defined along two dimensions – administrative location (pre-war regions and emerging states) and population type (urban areas, rural settlements, IDP settlements, and nomadic population), leading to a total of 57 strata (Table B7-12). Sub-populations in the urban centers of Mogadishu, Baidoa, and Kismaayo, in fisheries livelihood zones in coastal areas (Figure B7-13), and IDP host communities were of particular interest and therefore deliberately oversampled.

The total planned sample size was 6,384 interviews, allowing for high-precision consumption estimates with less than 10 percent relative standard errors for key sub-populations and overall. The sample was allocated across strata following optimal (Neyman) allocation, minimizing the global sampling error of the consumption estimates (Neyman 1934). Optimal allocation is given by

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h},$$

Error! Bookmark not defined.(1)

where n_h is the sample size in stratum *h, n* is the total sample size, *H* is the total number of strata, N_h is the total population of stratum *h*, *N* is the total overall population, and S_h is the standard deviation in stratum *h.* Hence, the number of households to be interviewed per stratum is mainly determined by the variability of consumption within the stratum (S_h). S_h was derived from the results of the SHFS Wave 1. The population size only matters for practical purposes in very small strata below 10,000 households. In the absence of a recent population census, the population of each stratum was derived from UNFPA's 2014 Population Estimation Survey (PESS), which contains detailed estimates for each population type and administrative unit of interest.

The optimal allocation of interviews was subject to the following requirements:

(i) 500 expected interviews in IDP settlements and 500 in nomadic populations;
(ii) At least 600 interviews expected per administrative unit;
(iii) Oversampled populations with
- Mogadishu (urban): 900 interviews, including IDPs;
- Kismaayo (urban) and Baidoa (urban): at least 500 interviews each;
- Coastal fisheries livelihood zones: at least 300 interviews;
- IDP host communities: 500 expected interviews.

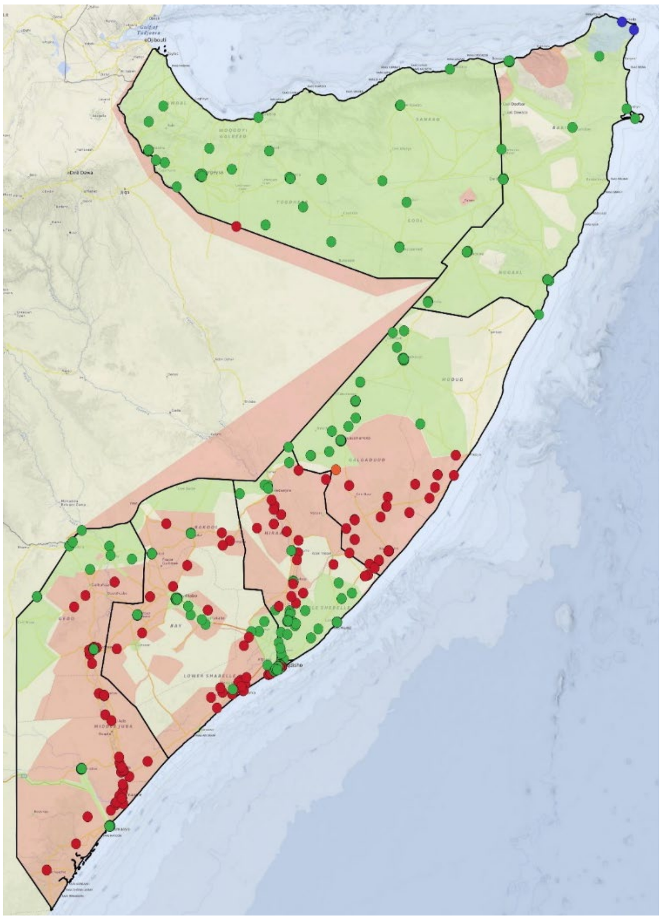
Households are clustered into enumeration areas (EAs), with 12 interviews expected for each selected EA. A larger number of households per enumeration area would only marginally benefit the statistical estimation of indicators because of potential homogeneity among households in geographic proximity. A smaller number of households would result in less than 3 observations for each of the four optional modules capturing household consumption based on the Rapid Consumption Methodology, and thus affect the reliability of poverty estimates.

The sampling design addressed the challenging security situation on the ground in two ways. First, a security assessment was conducted to exclude areas too dangerous for field teams to visit. Second, a micro-listing approach was used in second-stage sample selection to allow field teams to spend limited time on the ground. Replacement of sampling units during fieldwork followed a transparent and predefined replacement schedule, which was necessary to correctly calculate sampling weights.

Incorporating inaccessibility into the sampling frame

A geo-spatial access map depicting accessibility was created through key informant interviews with security experts and regional fieldwork coordinators based in the field. Publicly available information and incident reports provided by a local security company were used as auxiliary inputs. Finally, the information in the access map was triangulated with security analysts from a security NGO and private security company.

Figure B7-1: Security assessment access map.



Note: Red color indicates inaccessibility, green color indicates accessibility. Circles represent urban centers.

The security assessment led to the complete exclusion of pre-war region Middle Juba. Several other pre-war regions in south and central Somalia were only partially accessible. The security situation differed substantially between different cities, with some completely inaccessible and some at least partially accessible even though they were located in insecure regions. The sampled IDP and nomadic populations fell within safe areas. Survey estimates for these populations were thus considered to be representative.

Table B7-1: Accessibility rates by pre-war region.

Pre-war region	Percentage of population in accessible areas	
	Urban areas	Rural areas
Awdal	100%	94%
Bakool	35%	21%
Banadir	87%	96%
Bari	99%	92%
Bay	86%	46%
Galgaduud	88%	50%
Gedo	100%	43%
Hiraan	44%	28%
Lower Juba	92%	9%
Lower Shabelle	28%	33%
Middle Juba	0%	0%
Middle Shabelle	98%	77%
Mudug	100%	76%
Nugaal	100%	100%
Sanaag	100%	100%
Sool	89%	98%
Togdheer	100%	98%
Woqooyi Galbeed	100%	96%
Overall	89%	48%

Low accessibility in south and central Somalia motivated the imputation of poverty in inaccessible areas using geo-spatial information. The accessibility map was incorporated into the sampling frame to draw EAs only from accessible areas. The resulting sample was thus representative of the entire Somali population within secure areas.

Sampling frame and sample selection

The sampling frame for wave 2 of the SHFS is the exhaustive list of sampling units for every stage in the multi-stage selection process (denominated according to the stage of selection, i.e. primary sampling units (PSUs) in the first stage, secondary sampling units (SSUs) in the second stage, and so on) employed in the survey's sampling strategy. Sampling units are listed separately by stratum. Each sampling unit must have information concerning the population residing in it to allow for selection proportional to size (United Nations Statistical Division 2005). In the absence of a recent population census, no readily useable enumeration areas and population estimates existed. To overcome these challenges the SHFS drew from a variety of data sources and GIS techniques to create a population sampling frame, strata boundaries, and a comprehensive list of enumeration areas.

Strata boundaries

In line with stratification at the intersection of administrative regions and population type, the following GIS datasets were combined to spatially demarcate strata boundaries:

(i) Pre-war region boundaries;
(ii) IDP settlement boundaries;
(iii) Urban area boundaries;
(iv) Rural settlement boundaries;
(v) Security assessment access map.

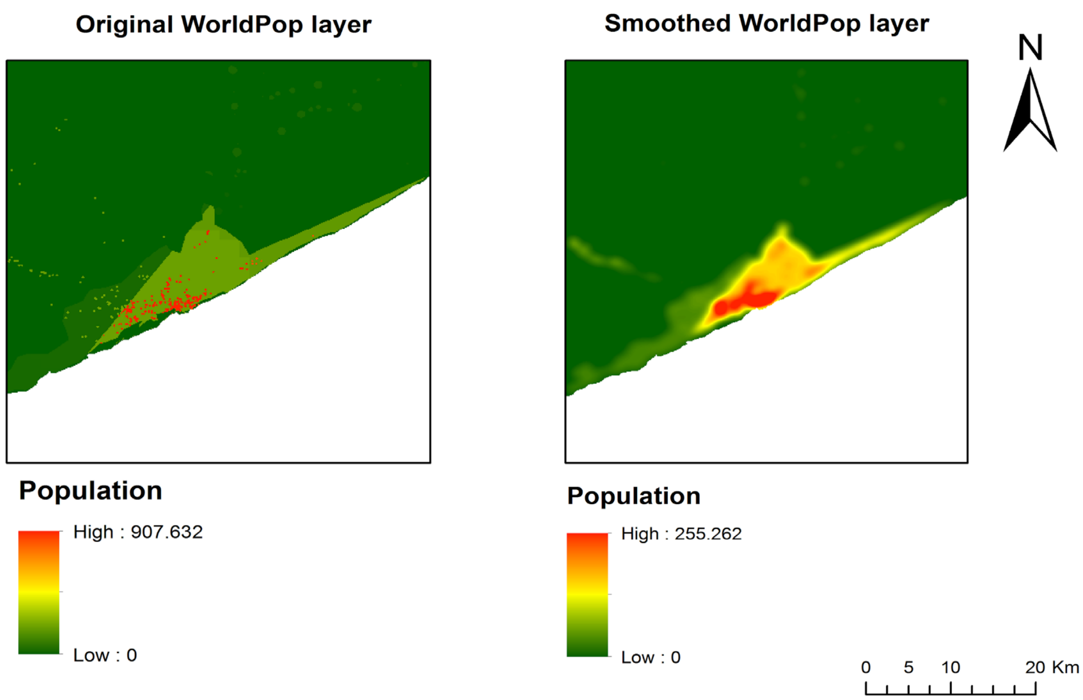
Pre-war region boundaries are available as shapefiles from UNDP. The boundaries of urban areas were defined by the urban enumeration areas previously used in UNFPA's Population Estimation Survey

2014 (PESS). Boundaries of IDP settlements were provided by UNHCR's Shelter Cluster and PESS. The IDP strata boundaries were subtracted from the urban and rural strata to prevent duplicate sampling. The remaining areas outside of the urban and IDP strata were considered as rural strata. Areas determined too dangerous through the security assessment were removed from the sampling frame.

Population sampling frame

In urban and rural strata, population estimates were derived from the 2015 WorldPop dataset, detailed in Linard, Alegana et al. (2010). This dataset uses a combination of data sources and methods, including satellite imagery, to derive highly spatially disaggregated population estimates. First, the starting point are 2005 population estimates at the district-level from the UN Office for the Coordination of Humanitarian Affairs (OCHA) for 74 districts. Second, an Africover GIS dataset depicting 22 landcover classes was combined with 2005 Landsat satellite imagery depicting settlement outlines. Third, settlement point location data, based on the efforts of various NGOs and UN agencies, with more than 11,000 settlement points along with some population estimates, including urban and rural areas, and IDP settlements. To achieve higher spatial resolution, the OCHA estimates were disaggregated using the information contained in the settlement points data and the landcover class data. The result is a gridded population dataset at 100m-by-100m spatial resolution. For each 100m-by-100m cell, the dataset contains a population estimate, which, aggregated within the PSU, provides a population estimate for each primary sampling unit (PSU) in urban and rural strata, which was later used for sample selection proportional to size. Due to inadequacies of the population density map for the purpose of creating a sampling frame, a set of corrections had to be made to this dataset. In the original WorldPop layer, the population values were not always distributed smoothly. For instance, a village might have only one pixel with a high population number creating a sharp contrast, although its coverage area is larger and the transition from sparse to dense population is more progressive. Hence, a Gaussian smoothing kernel technique with standard deviation of 500m was applied. This distributed higher values smoothly in areas surrounding a high-density pixel while preserving close to the total population count in the area (Figure B7-2).

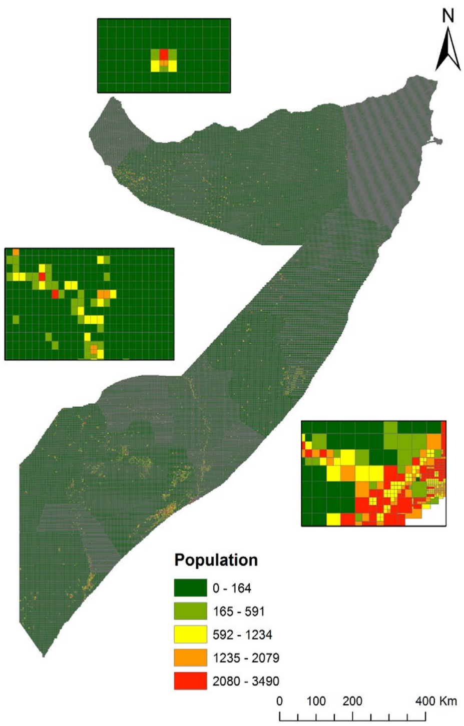
Figure B7-2: Gaussian smoothing of the WorldPop population density layer.



Primary Sampling Units (PSUs) and first-stage sample selection

PSUs were generated using a variety of techniques depending on the population type. The primary sampling unit (PSU) in urban as well as rural strata was the enumeration area (EA). The boundaries for urban EAs were derived from the enumeration areas used in UNFPA's 2014 Population Estimation Survey (PESS). Overlaps with IDP settlements were removed. The EAs thus obtained were combined with the corresponding population estimates from the 100m-by-100m WorldPop dataset to form the sampling frame for urban strata in which each PSU has a positive and known probability of selection. In case a strata boundary cut through any grid cell in the WorldPop dataset, the grid cell was split and the population estimates re-calculated weighted by geographical area. In rural strata – defined as those permanently settled areas outside of urban areas and IDP settlements – no list of enumeration areas comparable to the PESS EAs exists. The entire area of rural strata assessed as secure was divided into rectangular grid cells of different sizes using a quadtree algorithm. The approach splits an area into successively smaller quadtratures by checking to see whether the content of each split is greater or less than a prescribed value. In this case, the population map was used as the unit of measure, and was split successively until each square had a population of less than a target population of 3500. This approach also allowed the definition of each grid cell per a set of combined parameters, specifically geographic extent and population size (Minasny, McBratney et al. 2007). Thus, each cell has a minimum estimated population size of 3500 and a maximum geographical area of 3 km x 3 km. to keep enumeration areas manageable in size for field teams.

Figure B7-3: Quadtree grids



For IDP strata, primary sampling units were IDP settlements as defined by UNCHR's Shelter Cluster. PSU boundaries, which, given the choice of PSU, are equivalent IDP settlement boundaries, were derived from UNHCR's GIS shapefiles. In several cases information on settlement boundaries was missing. In these cases, the missing information was drawn from PESS IDP enumeration areas. PESS population data, available at the pre-war region level, was used to obtain population estimates for each IDP settlement. To match the pre-war region IDP population to each IDP settlement in the sampling frame, the following protocol was applied: Whenever there was exactly one IDP settlement per pre-war region, the PESS IDP population for that pre-war region was used as the population estimate for the respective settlement, thus taking the settlement population as representative of the IDP population in its pre-war region. In cases where there is more than one settlement per pre-war region, the PESS IDP population was assigned to each settlement proportional to the geographical area each settlement covers relative to the total area of all settlements in the given pre-war region.

Across all strata, PSUs were selected using a systematic random sampling approach with selection probability proportional to size (PPS), where size is given by the estimated population in each PSU. In PPS sampling, PSUs are selected into the sample based on their size so that large PSUs have a greater chance of being part of the final sample. In urban and rural areas, the EA served was the primary sampling unit. In IDP strata, PPS sampling is applied at the IDP settlement level, determining how many enumeration areas are to be selected in each settlement. PSUs were drawn separately for each stratum, with at least 20 percent additional PSUs selected to serve as replacements in case one of the main PSUs needed to be replaced.

Secondary sampling units (SSUs) and second-stage sample selection

Even in areas deemed accessible per the security assessment, it was critical to the safety of field staff and respondents that teams would spend as little time as possible in each EA. Himelein, Eckman et al. (2016) discuss and compare several second-stage sample selection strategies for use in contexts such as this one. In wave 2 of the SHFS, a micro-listing approach is used in second- and final-stage sample selection. In micro-listing, enumeration areas are divided into smaller enumeration blocks. Rather than performing a time-consuming full listing of all households in the EA, enumerators list only households in one enumeration block, then select the household to be interviewed, and immediately conduct the interview, greatly reducing the time required in the EA.

Enumeration blocks were generated through different means, depending on the population type. In urban and rural strata, the EAs selected in the first stage were manually segmented into enumeration blocks (EBs) using satellite imagery from Google Earth or Bing, counting the number of structures visible in each. Enumeration blocks served as secondary sampling units (SSUs) in the sampling design. Enumeration blocks were created as per the following general criteria:

- Each selected EA would be comprehensively covered by enumeration blocks.
- Each EA would be delineated into 12 enumeration blocks, expecting one interview per block.
- Each enumeration block would contain at least 1 and at most 12 structures.
- Enumeration blocks in the same EA should have roughly the same number of visible structures.
- Blocks would be drawn to take account of natural boundaries.
- Each block should have a central point from which all structures in the block can be seen.

The general criteria for block delineation allow for several special cases:

(i) If any PSU contained less than 12 structures, it would not be possible to delineate 12 blocks of the same size.

(ii) If any PSU contained more than 150 structures, more than 12 blocks were delineated, following the above criteria.

(iii) Given the design features of the sample, a fraction of PSUs was selected more than once. This occurred in two instances: First, given the nature of the first-stage sample selection with PPS, very large PSUs were selected twice or three times. This was especially likely in strata with a relatively short list of PSUs and a relatively large number of required interviews in the stratum. Second, as outlined in the previous section, PSUs were selected more than once if they formed part of one of the oversamples. The number of required interviews and consequently the required number of enumeration blocks was scaled up proportionately in these cases. For instance, if a PSU was selected twice, 12*2=24 interviews and blocks were required, and if a PSU was selected three times, 12*3=36 interviews and blocks were required. All other criteria for block delineation remained in place (Figure B7-4).

Figure B7-4: Example of EA delineated into blocks

Enumeration blocks were selected with equal probability. In the general case of 12 blocks per enumeration area, every single block was selected as 12 interviews per EA were required (and equivalently for PSUs with 24 or 36 required interviews in special case (iii)). In PSUs where more than 12 (or 24, or 36) blocks had been delineated due to the high number of visible structures (special case (ii)), selection of 12 (or 24, or 36) blocks with equal probability was implemented using equal probability random sampling. In PSUs with less than 12 (or 24, or 36) visible structures (special case (i)), two selection mechanisms were possible: First, if field teams found that there were indeed less than 12 structures in the PSU (as the satellite imagery suggested), all structures were interviewed. Second, when field teams found that the number of structures was higher than the satellite imagery suggested, enumerators counted the number of structures and randomly selected 12 (or 24, or 36) households to be interviewed with equal probability.

A similar second-stage sampling strategy was employed for IDP strata. Each IDP settlement was segmented manually into enumeration blocks with approximately 10 structures each. Where sensible, 12 enumeration blocks were combined into one enumeration area. In some cases, however, IDP settlements consisted of geographically dispersed pockets within urban areas, each far away from the next. To keep enumerator travel time in check, facilitate supervision, and ensure safety, the construction of IDP EAs followed these geographical contingencies to some extent. Hence, some EAs were created to contain more than 12 blocks and others contained less than 12.

Several of the most recent IDP settlement boundaries provided by UNHCR were a few years old, while the recent drought caused perturbations to the size, composition, and localization of the IDP population. Thus, each selected IDP enumeration area was inspected to ensure that it was still inhabited by displaced communities. This led to several IDP EAs being dropped and replaced by backup EAs. Enumeration areas served as secondary sampling units and were selected with probability proportional to size, with size given by the number of blocks per EA. The required number of EAs in each IDP settlements was fixed through first-stage sample selection. Then, where there were more or fewer than 12 blocks per IDP EA, blocks were selected with equal probability.

Final-stage sample selection: households

Except for the special cases discussed in the previous sections, enumerators were expected to interview one household per block in all selected blocks within the enumeration area. The household was selected randomly with equal probability in two stages, following the micro-listing protocol: From a central point in the block, the enumerator listed all residential structures within the current block into the tablet. The enumerator's tablet then randomly selected a residential structure for the enumerator to visit. At the structure, the enumerator recorded the number of households residing in the structure, and the tablet again randomly selected a household to be interviewed.

Oversamples

For Baidoa, Kismaayo, and fisheries areas, a second-stage oversampling strategy was used. In second-stage oversampling, PSUs selected in the first stage and falling into the specified urban centers or coastal areas were selected again to reach the minimum sample size for each oversample. Through this process, PSUs in Kismaayo were selected twice, and PSUs in Baidoa and in fisheries areas were selected a total of three times. Fisheries livelihood zones in coastal areas were defined by FEWSNET and FSNAU (Figure B7-13, zones SO7 and SO8). For the host communities oversample, all urban enumeration areas adjacent to IDP settlements were pre-selected as a separate sampling frame. The resulting list was stratified implicitly by pre-war region. 42 enumeration areas were selected with probability proportional to size to reach the desired oversample.

Sampling of the nomadic population

Nomadic households, who make up around a quarter of the Somali population according to UNFPA's Population Estimation Survey (PESS) of 2014, are inherently difficult to sample because, by definition, they have no permanent place of residence (Kalsbeek 1986, Soumare, Tempia et al. 2007). Himelein, Eckman et al. (2014) use a random geographic cluster sample approach, in which points are randomly selected from a map and all nomadic households within a radius around the point are interviewed. The SHFS followed a different approach. The strategy for sampling nomadic households relied on lists of water points used by nomadic households to water their livestock, which served as the primary sampling units. UNFPA's 2014 PESS took a similar approach to estimate the nomadic population (UNFPA 2014). The SHFS project deployed 200 purpose-designed tracking devices to nomadic households who gave consent, which track their movements for two years. This will improve the understanding of the patterns of movement the nomadic population in Somalia, which will facilitate sampling this population in the future.

Nomadic sampling frame

Nomadic strata were defined at the federated member state level, with the population count for each stratum provided by PESS. The list of water points was divided up by stratum. The list was put together from a combination of two sources. First, the list of water points used in PESS. Second, a regularly updated list of water points kept by the UN Food and Agriculture Organization (FAO). Given this combination of sources, the resulting list of water points used as sampling frame was viewed to be close to or completely exhaustive. The list contained the GPS location and information on type of water point (Berkad, Borehole, Dam, Dug Well, Spring, Other). Other water point characteristics such as the number of households using the water point and the predominant type of cattle watered were available only for an incomplete subset of water points. The list was stratified implicitly by pre-war region (each federated member state encompasses several pre-war regions) and type of water point.

First-stage sample selection

Water points from this list served as primary sampling units. In the absence of reliable estimates of the population size of water points, 42 water points were selected in the first stage with equal probability, with 12 interviews to be conducted at each selected water point. A further challenge in sampling nomadic household peculiar to the timing of SHFS wave 2 was the ongoing drought, which led to many water points having run dry. Therefore, a series of Key Informant Interviews (KIIs) and Focus Group Discussions (FGDs) in each federated member state verified whether each selected water point was currently frequented by nomadic households. In case a selected water point was not currently frequented by nomadic households, it was replaced.

Selection of nomadic households at water points

Selection of nomadic households to interview relied on a listing process at each water point whose aim was to compile an exhaustive list of all nomadic households at the water point. However, the total number of nomadic households at a given water point is not static as nomadic households are not resident at water points, but only stay there for a limited time, and arrive and leave at various times during the day. It was determined in KIIs that nomadic households need to spend a very minimum of two hours at a given water point to water their cattle and that cattle watering would occur during daylight hours. To allow for a complete listing, daylight hours were segmented into two-hour time slots, during each of which enumeration team leaders completed a full listing of all nomadic households at the water point at that time. As not all persons present at water points were members of nomadic households, but may instead be from close-by rural settlements, the listing form contained a number of questions identifying nomadic households. The form also asked for informed consent to be interviewed. Upon completing a two-hour listing period, up to three households were randomly selected from the list of consenting nomadic households gathered during this time slot. Interviews were then scheduled with the selected households at a time and place convenient for the household respondent.[123] Based on this sampling design, sampling weights were calculated after the completion of data collection.

Data collection

Wave 2 of the Somali High Frequency Survey was implemented using computer assisted personal interviewing (CAPI), whereby enumerators were equipped with tablet computers which contained the

[123] Additional rules applied in special cases: (i) If no nomadic households were found or arrived at a given water point, enumeration teams remained at the water point for three days. If no nomadic household had arrived, the water point was replaced. (ii) If nomadic households were present but arrived at very low frequencies, so that teams struggled to reach the required number of interviews, they would stay for a maximum of 12 days. Then teams would leave whether or not they have reached the required number of interviews. If 3 or less (but at least 1) nomadic households arrived during any two-hour listing period, all listed households were interviewed during that period.

survey questionnaire and would upload completed interviews to the project's Survey Solutions cloud servers daily. The choice of CAPI was guided, on the one hand, by the finding that this technology greatly reduces the number of errors relative to pen-and-paper interviewing (e.g. Caeyers, Chalmers et al. 2012). On the other hand, this technology was essential to the near real-time monitoring data of collection and quality control, which were deemed necessary in the Somali context where insecurity and remoteness make close supervision challenging and follow-up visits costly.

Survey instrument

The consumption modules were the central components of the SHFS wave 2 survey questionnaire. The questionnaire also contained other key components of a multi-topic household survey, particularly those relevant to the Somali context. These included an individual-level module with information on education, employment, and health, household characteristics, remittances, displacement, perceptions and subjective welfare, and shocks. The questionnaire was designed in line with best practices (Deaton and Grosh 2000) and went through several iterations of internal and external expert revision.

The food consumption module consisted of 114 food items drawn from a list of CPI items provided by statistical authorities. To meet the requirements of the Rapid Consumption Methodology (section 0), items were divided in one core and four optional modules, with most commonly consumed items assigned to the core module. The list of items was highly specific (e.g. apples, pears rather than fruits) and selected to cover the basic food categories and adequately reflect the local diet (Smith, Dupriez et al. 2014, Zezza, Carletto et al. 2017). The list of food items contained various items for food away from home, accounting for both food bought away from home and consumed at home and food consumed outside of the home. Further, to facilitate food quantity reporting for respondents, a list of non-standard units, along with their conversion to kilograms, was developed for each item, with inputs from regional experts and experience from the accompanying market price survey (Oseni, Durazo et al. 2017).[124] The questionnaire was designed to capture purchased food, home production, and gifts.

The nonfood consumption module consisted of 90 items, which were assigned to core and optional modules in the same manner as the food items. The choice of nonfood items followed the COICOP classification system, with all relevant COICOP categories represented in the list of nonfood items.

Fieldwork and monitoring

The fieldwork strategy was designed to facilitate high-quality data collection and safety of field teams.[125] All enumerators and team leaders attended rigorous training sessions and had to sit a final exam to be hired. 45 teams were assembled for fieldwork, staffed each with one team leader, three regular enumerators, and two reserve enumerators. The large number of teams was essential, on the one hand, for security reasons. It allowed teams to enter and exit an area swiftly before their presence would draw too much suspicion and endanger their safety and that of survey respondents. On the other hand, this arrangement allowed teams to be composed of enumerators native to the areas which they covered.

The survey was piloted in each region before the beginning of fieldwork. Fieldwork was monitored in near real-time to verify data collection progress, data quality, and enumerator performance. To implement near real-time monitoring, field teams uploaded interviews onto the project's Survey Solutions server at the end of each day. An automated pipeline of Stata code downloaded and

[124] The Market Price Survey (MPS) is a component of the SHFS. The MPS collects weekly exchange rates and prices of a broad range of 91 products and services as well as exchange rates from 14 key markets across all Somali regions.

[125] The Somali High Frequency Survey was implemented by Altai Consulting in coordination with the respective statistical authorities. The team worked closely with the Directorate of National Statistics, Ministry of Planning, Investment and Economic Development of the Federal Government of Somalia.

processed the data, creating a detailed monitoring dashboard in Microsoft Excel, which headquarters reviewed daily. The dashboard tracked the number of submissions meeting the quality standards to be considered acceptable, interview duration, and unit non-response rates separately by EA, enumerator, team, and strata. It further assessed item non-response by listing the number of household members, proportion of missing values, 'No', and 'Don't know' or 'Refused to respond' entries in all modules and several other key questions, which would trigger follow-up questions. Unusually high proportions of missing values, 'No', 'Don't know', or 'Refused to respond' entries indicated possible enumerator shirking, as this behavior would reduce enumerators' workload. For example, entering that a household self-identifies as displaced would trigger an entire module on displacement. Enumerators returning low-quality or displaying suspicious behavior received warnings and follow-up training. If the issues could not be resolved in this way, enumerators were replaced by reserve enumerators from their team. Overall, however, enumerator performance was high, requiring few replacements, while the unit non-response rate was very low at 0.16 percent among urban, rural, and IDP households, and 0.50 percent among nomadic households.

Submissions quality standards

Each enumerator submission was subject to a set of minimum standards to ensure data quality. Interviews were classified as valid or invalid based on the criteria listed in the following.

- *Valid EAs*. If the EA was not part of the final sample (i.e. it was replaced), the interview was classified as invalid and thus excluded from the final dataset.

- *Valid EBs.* If the EB was not part of the final sample (i.e. it was replaced), the interview was excluded.

- *Duration.* If the duration of the interview did not exceed the minimum threshold of 30 minutes, the interview was excluded.

- *Location.* If the interview did not have GPS coordinates associated with it, the interview was considered invalid. If the GPS coordinates fell outside buffer zone of a 50m+accuracy of GPS (based on the minimum latitude-longitude formula + 50m buffer) around the EA, the interview was excluded.

- *Follow up visits*. If the interview was not conducted in the first visit, the interview for the first visit must be valid except for the minimum duration, and both records must contain matching GPS positions (with a 10m + precision maximum distance), otherwise the interview completed in the follow-up visit was excluded.

- *Replacement interview*. If the interview was from a replaced household, the record of the original household must be valid except for the minimum duration and the reason for no interview must also be valid, otherwise the interview was excluded.

Beyond these criteria, the Survey Solutions CAPI platform allowed to 'reject' submissions on a case-by-case basis and send them back to enumerators to correct whenever headquarters found problems with a submission.

Consumption aggregate

The main welfare measure used in this and other analyses using SHFS data is per-capita consumption, rather than income (Deaton and Zaidi 2002). The SHFS collected data realized consumption rather than the total money spent on consumption items, as this measured actual realized welfare in a utility-consistent way (Ravallion and Benu 1994). This section discusses the various adjustments made to the SHFS data to construct the consumption aggregate using the Rapid Consumption Methodology.

Cleaning of consumption data

Before deriving the consumption aggregate, the components of consumption –data on food consumption, nonfood consumption, and durable assets– must undergo a cleaning process to correct outliers and other mistakes (Deaton and Zaidi 2002).

Food expenditure data is cleaned in a four-step process. First, units for reported quantities of consumption and purchase are corrected. Typical mistakes include recorded consumption of 100 kg of a product (like salt) where the correct quantity is grams. These mistakes are corrected using generic rules (Table B7-14). Then, a conversion factor to kg for all units is introduced. For example, a small piece of bread will likely have a different weight than a small piece of garlic. To avoid mistakes, enumerator trainings focused on units and introduced a common understanding of what each unit means for each food item. In addition, the conversion to kilograms was made explicit on the enumerators' tablets (Table B7-15). The third step consisted of correcting issues with the exchange rate selected (

Table B7-16). Finally, outliers in each component of consumption are detected using a set of cleaning to correct quantities and prices (see Appendix). The non-food dataset only contains values without quantities and units. First, the same cleaning rules for currencies are applied (

Table B7-16), followed by a set of specialized cleaning rules (see Appendix). Likewise, for durables, the same cleaning rules for currencies are applied (

Table B7-16), and then a set of durables-specific cleaning rules (see Appendix).

Consumption aggregate using the Rapid Consumption Methodology

The nominal household consumption aggregate is the sum of four components, namely expenditures on food items, expenditures on non-food items, the value of the consumption flow from durable goods, and housing (Deaton and Zaidi 2002). Without a housing market functioning well enough to derive credible estimates for the cost of housing, the SHFS consumption aggregate is based on the first three components: food consumption, nonfood consumption, and consumption of durable assets.

Food and nonfood consumption in the Rapid Consumption Methodology

The SHFS used the Rapid Consumption Methodology to estimate the consumption aggregate. Pape and Mistiaen (2018) provide a detailed and general exposition of the Rapid Consumption Methodology including an ex-post assessment of the methodology. The methodology is based on dividing food and nonfood consumption items in one core and several optional modules. With each household assigned the core module and one optional module, this methodology reduces the time spent on enumerating the consumption modules. Deriving the consumption aggregate with this methodology is a two-step process. First, core and optional modules are constructed. Core items are selected based on their importance for consumption. The remaining items are partitioned into optional modules. Optional modules are assigned to groups of households. Second, after data collection, consumption of optional modules is imputed for all households. Then, the resulting consumption aggregate is used to estimate poverty indicators.

Module construction

Food and non-food consumption for household *i* are estimated by the sum of expenditures for full list of consumption items[126]

(2) $$y_i^f = \sum_{j=1}^{m} y_{ij}^f \text{ and } y_i^n = \sum_{j=1}^{m} y_{ij}^n$$

where y_i^f and y_i^f denote the food and non-food consumption of item *j* in household *i*. As the estimation for food and non-food consumption follows the same principles, the upper indices *f* and *n* are neglected in the remainder of this section. The list of items can be partitioned into M+1 modules each with m_k items:

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

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

Item assignment to the core module was designed to maximize the core module's share of total consumption, so that a large share of consumption would be enumerated from each household. Important items were identified by their average food share across households from wave 1 of the SHFS.[127] This strategy relies on the fact that, in Somalia, a few dozen items capture the majority of consumption. The core modules captured 94 percent of food consumption and 79 percent of nonfood consumption, respectively (Table B7-2). Optional modules were constructed such that items are orthogonal within modules and correlated between modules, using an iterative algorithm (Pape and Mistiaen 2018).

Table B7-2: Item partitions and consumption shares in SHFS wave 2.

	Food Items			**Non-food Items**		
	Number of items	Share Wave 2	Share Wave 2 Imputed	Number of items	Share Wave 2	Share Wave 2 Imputed
Core	38	94%	79%	29	79%	47%
Module 1	21	2%	8%	14	5%	14%
Module 2	18	2%	6%	15	6%	15%
Module 3	19	1%	5%	16	7%	18%
Module 4	18	1%	4%	15	5%	11%

In fieldwork, a sufficient number of households must be assigned each optional module to obtain a reliable total consumption estimate. In wave 2 of the SHFS, this was ensured by interviewing 12 households per EA allowing for the ideal partition of three items per optional module.

Consumption estimation

Household consumption was then estimated using the core module, the assigned optional module, and estimates for the remaining optional modules

[126] The list of consumption items used in wave 2 of the SHFS is discussed in section 0.

[127] Generally, previous consumption surveys in the same country or consumption shares of neighboring or similar countries can be used to estimate food shares. In the worst case, a random assignment results in a larger standard error but does not introduce a bias. 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.

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

where $K^* := \{1, \dots, k^* - 1, k^* + 1, \dots, M\}$ denotes the set of non-assigned optional modules. Consumption of non-assigned optional modules was estimated using multiple imputation techniques taking into account the variation absorbed in the residual term (Pape and Mistiaen 2018). Multiple imputation was implemented using multivariate normal regression based on an EM-akin algorithm to iteratively estimate model parameters and missing data.[128] The standard errors capture the error distribution of the multiple imputation process. The underlying model is a welfare model relating consumption to key household characteristics thus explaining 71 percent of variation in food consumption and 64 percent in nonfood consumption. The model parameters were household size, share of children in household, share of seniors in household; household head gender, employment, and education; dwelling type, dwelling drinking water access, dwelling floor, and dwelling ownership status; household experience of hunger; receipt of remittances; population type (urban, rural, IDP, nomadic) and a region-population type interaction, as well as each household's core consumption quartile.[129] Pape and Mistiaen (2018) demonstrate that the Rapid Consumption Methodology yields reliable estimates of poverty using an ex-post assessment with household budget data from Hargeisa and mimicking the Rapid Consumption methodology by masking consumption of items that were not administered to households.[130]

Durable consumption flow

The consumption aggregate includes the consumption flow of durables calculated based on the user-cost approach. The consumption flow distributes the consumption value of the durable over multiple years. The user-cost principle defines the consumption flow of an item as the difference of selling the asset at the beginning and the end of the year as this is the opportunity cost of the household for keeping the item. The opportunity cost is the difference in the sales price and the forgone earnings on interest if the asset is sold at the beginning of the year.

If the durable item is sold at the beginning of the year, the household would receive the market price p_t for the item and the interest on the revenue for one year. With i_t denoting the interest rate, the value of the item thus is $p_t(1 + i_t)$. If the item is sold at the end of the year, the household will receive the depreciated value of the item while considering inflation. With π_t being the inflation rate during the year t, the household would obtain $p_t(1 + \pi_t)(1 - \delta)$ with the annual physical or technological depreciation rate denoted as δ assumed constant over time.[131] The difference between these two values is the cost that the household is willing to pay for using the durable good for one year. Hence, the consumption flow is:

(5) $$\mathrm{y}^d = p_t(1 + i_t) - p_t(1 + \pi_t)(1 - \delta)$$

[128] Various other techniques for imputing total consumption were tested, including OLS and tobit module-wise regression and multiple imputation chained equations, concluding that multivariate normal regression is the preferred technique.

[129] Negative imputed values are corrected by scaling all associated imputed values to an average of zero without affecting the variance.

[130] The imputation results are compared with consumption estimates from the full consumption modules of the 2013 Somaliland Household Survey. The authors present the performance of the estimation techniques in terms of the relative bias (mean of the error distribution) and the relative standard error. The methodology generally does not perform well at the household level (HH) but improves considerably already at the enumeration area level (EA) where the average of 12 households is estimated. At the national aggregation level, the Rapid Consumption methodology slightly over-estimates consumption by 0.3 percent. Assessing the three standard poverty measures including poverty headcount (FGT0), poverty depth (FGT1) and poverty severity (FGT2), the simulation results show that the Rapid Consumption methodology retrieves estimates within 1.5 percent of the reference measure. Generally, the estimates are robust as suggested by the low standard errors. Simulations were also run for the complete dataset from the Somaliland 2012 household budget survey producing comparable results.

[131] Assuming a constant depreciation rate is equivalent to assuming a "radioactive decay" of durable goods.

By assuming that $\delta \times \pi_t \cong 0$, the equation simplifies to

(6) $$\text{y}^d = p_t(i_t - \pi_t + \delta) = p_t(r_t + \delta)$$

where r_t is the real market interest rate in period t. Therefore, the consumption flow of an item can be estimated by the current market value p_t, the current real interest rate r_t, and the depreciation rate δ. Assuming an average annual inflation rate π, the depreciation rates δ can be estimated utilizing its relationship to the market price[132]:

(7) $$p_t = p_{t-k}(1 + \pi)^k(1 - \delta)^k$$

The equation can be solved for δ obtaining:

(8) $$\delta = 1 - \left(\frac{p_t}{p_{t-k}}\right)^{\frac{1}{k}} \frac{1}{(1 + \pi)}$$

Based on this equation, item-specific median depreciation rates are estimated assuming an inflation rate of 0.5 percent, a nominal interest rate of 2.0 percent and, thus, a real interest rate of 1.5 percent (Table B7-18).

For all households owning a durable but did not report the current value of the durable, the item-specific median consumption flow is used. For households that own more than one of the durable, the consumption flow of the newest item is added to the item-specific median of the consumption flow times the number of those items without counting the newest item.[133]

Deflators

Spatial price indices were calculated using a common food basket and spatial prices to make consumption comparable across regions. The Laspeyres index is chosen as a deflator due to its moderate data requirements. The deflator is calculated by analytical strata areas based on the price data collected in wave 2 of the SHFS. The Laspeyres index (Table B7-3) reflects the item-weighted relative price differences across products. Item weights are estimated as household-weighted average consumption share across all households before imputation. Based on the democratic approach, consumption shares are calculated at the household level. Core items use total household core consumption as reference while items from optional modules use the total assigned optional module household consumption as reference. The shares are aggregated at the national level (using household weights) and then calibrated by average consumption per module to arrive at item-weights summing to 1. The item-weights are applied to the relative differences of median item prices for each analytical stratum. Missing prices are replaced by the item-specific median over all households.

[132] In particular, π solves the equation $\prod_{i=t-k}^{t}(1 + \pi_i) = (1 + \pi)^k$

[133] The SHFS wave 2 questionnaire provides information on a) the year of purchase and b) the purchasing price only for the most recent durable owned by the household.

Table B7-3: Spatial Laspeyres index

Analytical strata	Foo deflator
IDPs	0.856
Nomads	1.030
Banadir (Urban)	0.910
Nugaal (Urban)	1.058
Bari and Mudug (Urban)	0.976
Woqooyi Galbeed (Urban)	1.181
Awdal, Sanaag, Sool and Togdheer (Urban)	1.181
Hiraan, Middle Shabelle and Galgaduud (Urban)	1.119
Gedo, Lower and Middle Juba (Urban)	0.960
Bay, Bakool and Lower Shabelle (Urban)	0.931
Bari, Mudug and Nugaal (Rural)	0.960
Awdal, Togdheer and Woqooyi (Rural)	0.887
Hiraan, Middle Shabelle and Galgaduud (Rural)	0.925
Bay, Bakool and Lower Shabelle (Rural)	0.945

To obtain the US$1.90 PPP (2011) poverty line and correct for price differences over time, a price index was created –in the absence of a national CPI– using consumption shares from the survey and prices collected by the Market Price Survey (MPS) and by the Food Security and Nutrition Analysis Unit, Somalia (FSNAU).[134] Inflation between 2011 and December 2017 was obtained from the growth in the price index, which was estimated in two steps. First, the price index was calculated from 2011 to February 2016 using data from Wave 1 of the SHFS and prices from FSNAU, and then from February 2016 to December 2017 with data from Wave 2 of the SHFS and prices from the MPS.[135]

In the first step, consumption shares of 109 food and 68 nonfood items were aggregated according to their Classification of Individual Consumption by Purpose (COICOP) code, and then combined with monthly prices from FSNAU for 51 products. As a result, 32 matched COICOP codes were used to calculate the price index between 2011 and February 2016. In the second step, consumption shares of 114 food and 89 nonfood items were aggregated by COICOP code, in combination with weekly price series from the MPS for 109 products. This resulted in 49 matched COICOP codes that were then used to estimate the price index until December 2017.

Imputing consumption data in North-East and Jubbaland regions

Despite methodological innovations, field team training, and a stringent security protocol (section 0), some challenges with data collection persisted in certain geographic areas. These were mainly related to human resource capacity constraints and remote monitoring to ensure the quality of the data. Specifically, in the Jubbaland and rural North-East regions,[136] the information collected turned out to be only representative of a very small, idiosyncratic part of the population or did not consistently meet the survey's high-quality standards.

Jubbaland

The implementation of Wave 2 of the SHSF required some concessions to the local authorities in terms of the recruitment of field teams. Some enumerators who performed sub-optimally during training and the pilot were recruited as agreed with local authorities in Jubbaland. Likewise, there were some constraints to replace enumerators during the data collection if they were found to underperform. Based on internal discussions and consultations with trusted team leaders, the SHFS team judged that

[134] FSNAU collects monthly prices of commodities in 50 markets across all regions. The MPS collects weekly prices of a broad range of products and services as well as exchange rates from 14 key markets across all Somali regions.
[135] The products and services in the MPS are a close match with the food and nonfood items that form part of the consumption module of the Somali High Frequency household survey component. The price survey is implemented using a stringent set of quality standards.
[136] Jubbaland region consists of pre-war regions Gedo, Middle Juba, and Lower Juba (Middle Juba was completely inaccessible). North-East region consists of pre-war regions Nugaal, Bari, and Mudug (Table B7-22).

this affected the quality of the data collected in Jubbaland, particularly of the more demanding consumption modules, compared to other regions. Furthermore, insecurity remained widespread in Jubbaland, mainly due to a strong presence of Al-Shabaab. The entire region of Middle Juba was excluded from wave 2 of the SHFS due to security reasons. Likewise, large parts of Lower Juba, and to a lesser extent Gedo were also excluded.

In rural Jubbaland, field teams only collected data in areas that were relatively close from main cities (e.g. within a 10-km radius around Kismayo, Afmadow and Dhobley in Lower Juba). This was due to insecurity and because many rural EAs considered in the sampling frame were found to be empty after reviewing the satellite imagery. The EAs sampled for rural Jubbaland were peri-urban areas that correspond to large villages or small cities and thus the information was not representative of the rural population there. In addition, data from teams surveying rural Jubbaland showed signs of inconsistency and relatively low quality (highest percentage of invalid submissions compared to other urban and rural areas (Table B7-4); largest number of flags in the cleaning process of the consumption modules (Table B7-6), and large differences in the consumption of many food items relative to other rural areas (Table B7-7). Interviews with rural households from this region were therefore entirely excluded from the final dataset, and poverty estimated from satellite imagery and other geo-spatial data.

In urban areas, data collection lasted longer than in any other area covered due to over-sampling. Insecurity also made it more difficult collecting interviews and thus required more time. Team leaders reported that these issues contributed to fatigue on the part of enumerators, presumably impacting the quality of the data collected in urban Jubbaland.

Table B7-4: Percentage of valid submissions for urban and rural areas

Region	%
Mogadishu (Urban)	99.9
North-east Urban	99.6
North-east Rural	100.0
North-west Urban	99.2
North-west Rural	100.0
Central regions Urban	99.0
Central regions Rural	97.0
Jubbaland Urban	99.3
Jubbaland Rural	94.6
South West Urban	98.6
South West Rural	98.1

Table B7-5: Percentage of missing values for food items in urban and rural areas

Region	Percentage
Mogadishu (Urban)	54.8
North-east Urban	58.4
North-east Rural	61.2
North-west Urban	58.2
North-west Rural	61.2
Central regions Urban	57.9
Central regions Rural	58.3
Jubbaland Urban	49.8
Jubbaland Rural	49.1
South West Urban	57.5
South West Rural	56.5

Table B7-6: Number of flags in the cleaning of food items for urban and rural areas

Region	Average number per household
Mogadishu (Urban)	1.0
North-east Urban	0.8
North-east Rural	0.8
North-west Urban	0.9
North-west Rural	0.8
Central regions Urban	1.8
Central regions Rural	1.1
Jubbaland Urban	2.1
Jubbaland Rural	2.6
South West Urban	1.0
South West Rural	0.9

Table B7-7: Items consumed by 10% more/less households relative to strata averages

Region	Number of core food items
Mogadishu (Urban)	5
North-east Urban	6
North-east Rural	20
North-west Urban	7
North-west Rural	17
Central regions Urban	1
Central regions Rural	13
Jubbaland Urban	20
Jubbaland Rural	20
South West Urban	10
South West Rural	8

While the validity rate of submissions was in line with other regions (Table B7-4), the consumption data were flagged as outliers more often than in other regions during the review and cleaning process (Table B7-6). Further, the profile of food consumption for households in urban Jubbaland was different than in other urban areas for 20 of 38 core food items (Table B7-7). These issues in the consumption modules led to inconsistent poverty rates. Therefore, the information on the consumption modules (food, non-food and assets) was discarded and poverty estimated based on sociodemographic and other household characteristics in a multiple imputation process routine.

Rural North-East

The implementation of the survey also experienced some constraints in the recruitment of field teams in the rural North-East regions. The access of some areas in this region is possible only for team members from certain clans. Thus, enumerators had to be selected and replaced based on this criterion. Some of these candidates might not otherwise have been selected given their performance during training, the pilot, and data collection. This was judged to have affected the quality especially of the consumption data collected.

Moreover, the EAs sampled were spread across a vast territory and mostly in remote areas. They were far from each other, and far from urban centers. NE teams who covered rural areas had to travel up to two days to reach some EAs, longer than teams in any other region. Team leader reports from the field indicate that these large distances and conditions created fatigue among enumerators. Further, direct monitoring of field teams by supervisors was limited due to poor connectivity, and thus sending frequent and timely feedback more challenging that for other teams. As a result, the performance of teams did not improve as in other regions.

Finally, the consumption profile of most core food items was different to other rural areas, including nearby and ostensibly similar areas covered by other teams (Table B7-7). Hence, the consumption data (food, non-food and assets) was discarded and poverty estimated from a multiple imputation process

Consumption imputation process

Consumption data in North-East rural and Jubbaland urban was imputed in Stata with Multiple Imputation (MI) techniques. The same multiple imputation process and model described to estimate the consumption of non-assigned optional modules from equation (4) was used to obtain the four consumption components, and thus the total consumption expenditure for households in these regions.

The dependent variable of the model is total consumption expenditure per capita with data from North-East rural and Jubbaland urban set as missing and to be imputed.[137] The independent variables were chosen based on explanatory power with respect to household consumption: household size, share of children in household, share of seniors in household; household head gender, employment, and education; dwelling type, dwelling drinking water access, dwelling floor, and dwelling ownership; household experience of hunger and receipt of remittances; population type (urban, rural, IDP, nomadic) and a region-population type interaction, as well as consumption quartiles. With an R-Squared of 71 percent, this model had high explanatory power.

The model for imputing consumption had two caveats: first, each value or category of the right-hand-side variables of the model must overlap with some non-missing values of the dependent variable. Otherwise, there is no basis for simulating the relationship between consumption and these explanatory variables. This means that the region-population type interaction variable must be modified, as North-East rural and Jubbaland urban are two categories of that variable without overlap with any non-missing consumption values. To do this, the North-East rural category was combined with North-East urban to form a general North-East category. Jubbaland urban was combined in the final specification with adjacent South-West urban (Table B7-8, column I). Various other specifications were tested in which Jubbaland urban was combined with Central Regions urban, as an assessment of the sensitivity of the final estimates to this choice (Table B7-8, column III and IV). Second, the model contains consumption quartiles as a key right-hand-side variable. Since the consumption data for North-East rural and Jubbaland urban was inconsistent, consumption quartiles were calculated for North-East rural and Jubbaland urban separately, to include this variable in the final specification. Other specifications excluding the quartile variable were assessed as a sensitivity test as well (Table B7-8, column II and IV).

Table B7-8: Multiple Imputation results.

	(I)	(II)	(III)	(IV)
Region	**Poverty rate**			
Mogadishu (Urban)	73.67% (69.45%, 77.9%)	72.25% (67.83%, 76.64%)	73.74% (69.54%, 77.94%)	72.28% (67.81%, 76.58%)
North-east Urban	58.78% (43.17%, 74.38%)	56.93% (40.45%, 73.68%)	59.01% (43.54%, 74.49%)	57.21% (40.44%, 73.72%)
North-east Rural	62.46% (62.1%, 62.81%)	64.97% (64.3%, 64.88%)	63.59% (52.36%, 75.21%)	65.01% (64.3%, 64.88%)
North-west Urban	62.71% (51.81%, 73.62%)	61.5% (50.93%, 72.24%)	62.7% (51.83%, 73.68%)	61.48% (50.97%, 72.27%)
North-west Rural	77.3% (67.07%, 87.53%)	75.29% (64.66%, 86.04%)	76.48% (65.52%, 87.4%)	75.41% (64.75%, 86.48%)
IDP Settlements	75.62% (62.35%, 88.88%)	74.55% (61.43%, 88.1%)	75.62% (62.31%, 88.86%)	74.45% (61.4%, 88.03%)
Central regions Urban	59.18% (47.46%, 70.9%)	58.21% (46.2%, 70.24%)	59.18% (47.42%, 70.85%)	58.24% (46.25%, 70.32%)
Central regions Rural	65.06% (27.44%, 102.7%)	64.77% (27.28%, 102.6%)	65.01% (27.41%, 102.5%)	64.81% (27.28%, 102.5%)
Jubbaland Urban	53.34% (42.4%, 64.29%)	59.33% (54.81%, 63.53%)	53.85% (42.51%, 64.31%)	48.81% (44.01%, 54.32%)
South West Urban	62.72% (43.1%, 82.35%)	60.8% (40.43%, 80.96%)	62.39% (42.62%, 82.22%)	60.91% (40.57%, 80.88%)
South West Rural	74.94% (61.43%, 88.44%)	73.61% (59.25%, 88%)	75% (61.52%, 88.45%)	73.53% (59.17%, 88.02%)
Nomadic population	71.61% (63.1%, 80.12%)	70.86% (62.27%, 79.54%)	71.71% (63.18%, 80.22%)	70.87% (62.28%, 79.53%)

Note: (I) final model used to impute consumption and poverty; (II) sensitivity test without income quartiles in imputation model; (III) sensitivity test with Jubbaland urban combined with Central regions instead of South-West; (IV) as (III) but without income quartiles. 95% confidence interval in parentheses

The results from the imputation process are stable and robust considering these different specifications. The imputation process and these results were judged the best alternative to overcome the issues experienced in data collection.

[137] A logarithmic transformation is not feasible in this case due to its singularity at zero. As the core module was constructed to capture maximum consumption shares, many optional modules – almost by definition – obtained zero consumption especially among the poorer households, which have a less diversified diet.

Imputing poverty in inaccessible areas using geo-spatial data

Prevalent insecurity and conflict meant that parts of Somalia remained inaccessible for the SHFS field teams. In the ten least accessible urban and rural strata, less than 50 percent of the population could safely be reached.[138] The survey poverty estimates in these regions are therefore insufficiently representative of the regions' entire urban and rural populations. Hence, poverty in each region was predicted making use of correlations between geo-spatial information and survey estimates. The resulting poverty predictions are supplemental to survey estimates and serve as a proof-of-concept for using geo-spatial information alongside on-the-ground data collection. This section describes selection of geo-spatial variables and the model used to impute poverty.

Selection of variables for poverty predictions

Spatial variables expected to predict poverty well were drawn from three types of sources. First, a custom-derived global database of over 300 spatial covariates from the WorldPop research group at the University of Southampton (see Stevens, Gaughan et al. 2015).[139] Second, spatial variables were computed from geo-tagged data from publicly available sources such as ACLED conflict data or FEWSNET food security data, and OpenStreetMap. Third, population and population type data drawn from a novel population density map using recent data from OpenSteetMap, BMGF / Digital Globe spatial data, UNFPA survey and SHFS data.

From these sources, 15 variables were selected based on their correlation with survey poverty estimates at the EA-level. These contained information on the type of land cover (distance to bare land cover, distance to cultivated areas)[140], climate (temperature, precipitation, distance to drought-affected areas), population characteristics (population density, distance to urban areas), infrastructure (distance to major roads, medical sites, schools, water sources, and waterways), conflict and insecurity (distance to conflict incidents, distance to insecure areas), and food security (distance to food insecure areas). A detailed list of the selected variables, their sources, preparation for analysis, illustration (Table B7-18), summary statistics (Table B7-19), and linear correlations with survey poverty estimates (Table B7-21) are available in the Appendix.

Model selection

The final model to predict poverty was selected in two steps. First, a range of model types was compared based on a five-fold cross validation scheme.[141] The data was randomly partitioned in five folds, four of which made up the training set and one served as the validation set, ensuring that each model was trained and validated on identical data. Models' prediction success in the validation set determined which models were selected, with R-squared and Root mean squared error (RSME) as goodness-of-fit measures. The models were fitted separately for each population type.[142] The survey poverty estimates aggregated at the EA-level served as the response variable.[143] Linear models yielded the best results. Second, the selection of covariates, from the 15 spatial variables presented in Table A.8, was refined using stepwise regression to minimize the RMSE of the linear models and maximize their predictive power. In this process, a sequence of linear combinations of up to 15 covariates, as well as covariate interactions, was iteratively fitted to the response variable with different starting

[138] Of these 10 strata, 4 were urban and 6 were rural. The survey of IDP and nomadic populations was not subject to similar accessibility problems so that survey results are considered representative for these populations.

[139] WorldPop "Global high resolution population denominators", project funded by the Bill & Melinda Gates Foundation (OPP1134076)

[140] Variables produced by D. Kerr, H. Chamberlain and M. Bondarenko (WorldPop) in the framework of the WorldPop "Global high resolution population denominators"

[141] Several models in each of the following categories were tested: linear models, random forest models, Support Vector Machine models, and Gaussian Process Regressions.

[142] IDP and nomadic households did not suffer from accessibility problems.

[143] EA-level poverty estimates are preferable to household-level estimates as EAs cover larger areas and contain fewer binary values. Thus, the model was trained and tested at the EA-level and within EA variability was not considered.

points and criterion for selecting the covariates, using the full data set of survey poverty estimates at the EA-level.[144] The final model for each population type was the one with the lowest AIC and RSME value.[145] Furthermore, the residuals did not present any patterns and, therefore, were treated as random.

Final models for predicting urban and rural poverty

The final model for predicting poverty in urban areas contained 12 covariates and various covariate interactions (Table B7-9; Figure B7-14). Most variables individually, and all variables collectively, are statistically significant in explaining variation in poverty. However, the model's overall explanatory power is limited, with an adjusted R-squared of 52 percent. To check for potential issues with over-fitting, 10 percent of the sample was randomly excluded and the model from Table 10 estimated. The process was repeated 1,000 times and Figure B7-16 shows the results for the in-sample and out of sample R^2 of this validation.

Table B7-9: Final model to predict urban poverty.

Coefficients	Coefficient estimate	Standard error	p-value
(Intercept)	-1.946	0.355	0.000
Distance to bare areas	0.165	0.028	0.000
Distance to cultivated areas	0.000	0.008	0.969
Distance to dry areas	0.001	0.000	0.017
Distance to major roads	0.084	0.026	0.002
Distance to medical sites	0.062	0.024	0.011
Distance to schools	0.083	0.028	0.003
Distance to unsafe areas	0.002	0.001	0.001
Distance to urban areas	-0.057	0.021	0.009
Distance to water sources	0.001	0.001	0.147
Distance to waterways	0.001	0.001	0.437
Population density	0.000	0.000	0.001
Temperature	0.084	0.013	0.000
Distance to bare areas x Distance to waterways	0.000	0.000	0.000
Distance to bare areas x Temperature	-0.006	0.001	0.000
Distance to cultivated areas x Distance to waterways	-0.001	0.000	0.012
Distance to dry areas x Distance to schools	0.000	0.000	0.001
Distance to major roads x Distance to schools	-0.010	0.003	0.001
Distance to major roads x Distance to unsafe areas	-0.003	0.001	0.000
Distance to major roads x Distance to urban areas	-0.090	0.024	0.000
Distance to medical sites x Distance to water sources	-0.001	0.000	0.015
Distance to schools x Temperature	-0.002	0.001	0.025
Distance to urban areas x Distance to water sources	0.002	0.000	0.000
Model statistics			
Unit of observation	Enumeration areas		
Observations	252		
Degrees of freedom	229		
R-squared	0.56		

[144] The MATLAB function 'fitlm' was used to obtain a first model for EA-level poverty. The MATLAB 'step' function, which implements stepwise regression, was then used to select model terms, including interactions of terms. See https://uk.mathworks.com/help/stats/fitlm.html and https://uk.mathworks.com/help/stats/linearmodel.step.html for the MATLAB documentation of these functions. Goodall, C. R. (1993). "13 Computation using the QR decomposition." provides the basis for the 'fitlm' fitting algorithm and Draper, N. R. and H. Smith (2014). "Applied regression analysis." give an overview of stepwise regression on which 'step' is based.

[145] Minimizing the Akaike information criterion (AIC) of a linear model is equivalent to minimizing the cross-validation error. See Shao, J. (1997). An Asymptotic Theory for Linear Model Selection. Statistica Sinica. **7:** 221 - 264. and Stone, M. (1977). "An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion." Journal of the Royal Statistical Society **39**(Series B): 44 - 47..

Adjusted R-squared	0.518
Root mean squared error	19.8
F-Statistic	13.3

The model's relatively low predictive power is likely because the explanatory variables do not vary at a high enough spatial frequency relative to urban poverty estimates which can vary significantly across a small space. Furthermore, distance explanatory variables could result in relatively smooth predictions across space and not accurately capture small geographical clusters of low/high consumption.[146] For example, in urban settings, poverty levels may be quite different in two EAs which are only several hundred apart. In contrast, the same two EAs will have very similar levels of precipitation or, depending on spatial resolution, may indeed be covered by the same precipitation data point. Predictors such as the density of buildings or building patterns would likely improve the model.

Further, two different sets of night-time lights data were used to improve the predictive power of the urban model, but these turned out to be poorly correlated with survey poverty estimates and did not improve the urban model's predictive power.[147] This failure to improve the model is likely due to the night-time lights data's coarse resolution of 1km and 500m, respectively.

In rural areas, EAs are highly dispersed and poverty levels somewhat more spatially homogenous. Hence, the rural model was more successful at explaining variation in poverty in rural areas, with an adjusted R-squared of 94 percent (Table B1-1).

The uncertainty from using spatial covariates as explanatory variables was not considered in the estimation of standard errors. However, the data points used in the model were randomly selected to ensure they were taken from places far from each other. The resulting weighted average coefficient of variation (CV) from estimates for urban districts is 0.19 and 0.73 for rural districts. Moreover, EAs were randomly selected for the survey with the multi-stage stratified process described above, which combined with a random selection of data points to estimate the model, aims to derive a sample of EAs with different values within the range of each explanatory variable, similar to the range from the overall EA population.

Table B7-10: Final model to predict rural poverty

Coefficients	**Coefficient estimate**	**Standard error**	**p-value**
(Intercept)	2.075	0.320	0.000
Conflicts density	0.000	0.000	0.003
Distance to cultivated areas	0.040	0.008	0.000
Distance to food insecure areas	0.020	0.008	0.018
Distance to major roads	-0.019	0.005	0.001
Distance to medical sites	-0.026	0.004	0.000
Distance to schools	0.034	0.005	0.000
Distance to unsafe areas	-0.009	0.004	0.027
Distance to urban areas	0.011	0.002	0.000
Distance to water sources	0.007	0.002	0.000
Distance to waterways	0.000	0.001	0.830
Precipitations	0.001	0.000	0.001
Temperature	-0.089	0.012	0.000
Conflicts density x Distance to cultivated areas	0.000	0.000	0.005
Distance to cultivated areas x Distance to major roads	0.001	0.000	0.002
Distance to cultivated areas x Distance to medical sites	-0.001	0.000	0.001

[146] The same issue can arise with non-distance variables computed within a buffer, such as precipitation, temperature and conflict density.
[147] The two datasets are from 'Visible Infrared Imaging Radiometer Suite' (VIIRS) and Defense Meteorological Satellite Program (DMSP).

Distance to cultivated areas x Distance to schools	-0.002	0.000	0.000
Distance to food insecure areas x Distance to schools	0.000	0.000	0.002
Distance to food insecure areasx Distance to urban areas	0.000	0.000	0.001
Distance to food insecure areas x Distance to water sources	0.000	0.000	0.000
Distance to food insecure areas x Temperature	-0.001	0.000	0.000
Distance to medical sites x Distance to urban areas	0.001	0.000	0.000
Distance to unsafe areas x Distance to water sources	0.000	0.000	0.025
Distance to urban areas x Distance to water sources	0.000	0.000	0.000
Distance to urban areas x Distance to waterways	0.000	0.000	0.027
Model statistics			
Unit of observation	Enumeration areas		
Observations	92		
Degrees of freedom	67		
R-squared	0.953		
Adjusted R-squared	0.937		
Root mean squared error	11.2		
F-Statistic	56.9		

Both the urban and the rural model were used to predict poverty at the 100m-by-100m pixel-level for all urban and inhabited rural areas. In order to derive imputed poverty estimates at the pre-war region and district levels, pixels were aggregated using as population weights an updated version of the WorldPop population layer.[148]

Poverty in Somalia

Poverty is a complex phenomenon that refers to the deprivation of a person, household, or community in multiple dimensions (Deaton and Zaidi 2002). In general, it considers whether individuals or households have enough resources to meet their needs. Identifying the poor population or those living below a minimum threshold is a first crucial step for evidence-based planning aimed at alleviating poverty in any country. Profiling the poor and vulnerable is crucial to inform policies, design targeted interventions, as well as to monitor and evaluate the evolution of living standards and poverty reduction efforts (Baker 2000). This section presents an overview of quantitative measures used to assess poverty and inequality in Somalia using SHFS wave 2 data. The analysis focuses on the monetary dimensions of poverty. The World Bank's forthcoming Somali Poverty Assessment, and therein especially the first chapter, provides a more detailed analysis of poverty and deprivation, including non-monetary dimensions of deprivation.

Measuring poverty

Three components are required for poverty analysis. First, a measure of welfare. Second, a poverty line that defines a level of welfare at which individuals are either considered poor or not poor. Third, an aggregate poverty measure (Coudouel, Hentschel et al. 2002, Ravallion 2008, Haughton and Khandker 2009).

Poverty line

There are two types of poverty lines: relative to the overall distribution of consumption in a country, or anchored in an absolute level of what a household should consume to meet basic needs (Beegle, Christiaensen et al. 2016). Many countries define a national poverty line based on the cost of essential food items or a minimum calorie intake in that country, along with an allowance for non-food products. While a national poverty line allows for a precise measure of poverty according to national standards and circumstances, it is not comparable with other countries. Thus, absolute poverty lines are preferred to measure poverty across countries.

[148] The poverty estimates were obtained using an updated WorldPop population density map of Somalia with the latest data from wave 2 of the SHFS and DigitalGlobe.

This analysis uses the international poverty line which was introduced in the 1990 World Development Report with the aim of measuring poverty consistently across countries (Ravallion, Chen et al. 2009). To be representative of poverty in the poorest countries, it was computed using data from national poverty lines of 33 of the poorest countries. The international poverty line is expressed in terms of purchasing power parity (PPP) rather than traditional currency exchange rates to compare both poverty and GDP across countries (Beegle, Christiaensen et al. 2016).[149] The value of the poverty line has been revised through the years and adjusted to reflect welfare conditions of low-income countries. In 2008 this international line was estimated at \$1.25 per capita per day at 2005 prices. In 2015 the line was updated to its current level at a daily value of US\$ 1.90 (2011 PPP) per person (World Bank 2016).

Poverty and inequality measures

The poverty measure is primarily based on the three standard poverty measures following Foster, Greer et al. (1984). These measures are derived from the following general function:

$$F(\alpha) = \frac{1}{n}\sum_{i=1}^{p}\left[\frac{z - y_i}{z}\right]^{\alpha} \tag{9}$$

Here y_i denotes the consumption of individual i, n the total population, p the poor population and z the poverty line. The poverty headcount ratio is obtained when the parameter α takes the value of 0, the poverty gap and severity when this parameter is set to 1 and 2 respectively. The poverty headcount ratio or poverty incidence is the most common poverty measure. It is the share of population in a given region that is poor by virtue of having a total consumption lower than the poverty line. With $\alpha = 0$, the poverty headcount ratio can be expressed as the sum of poor individuals (p) over the total population (n), such that

$$F(0) = \frac{p}{n} \tag{10}$$

The poverty gap, obtained when α takes the value of 1, measures how far households or individuals are from overcoming poverty, by measuring the distance poor households are from the poverty line. It captures the difference between poor households' current consumption and the poverty line as a proportion of the poverty line. It can be interpreted as the minimum amount of resources that would have to be transferred to the poor, under a perfect targeting scheme, to eradicate poverty (Deaton 2006). This measure is obtained by adding up all the shortfalls of the poor relative to the poverty line and dividing the total by the population:

$$F(1) = \frac{1}{n}\sum_{i=1}^{p}\left[\frac{z - y_i}{z}\right] \tag{11}$$

The poverty severity index measures the level of inequality among the poor. This measure is estimated as the square of the poverty gap. It attributes a larger weight to the poorest among the poor, with the formula given by:

$$F(2) = \frac{1}{n}\sum_{i=1}^{p}\left[\frac{z - y_i}{z}\right]^{2} \tag{12}$$

In the context of monetary poverty, equality can be defined as an equal distribution of consumption across the population, with inequality being the departure from that equal distribution. Measures of

[149] The poverty line was derived considering the regression-based PPP estimate for Somalia, which corresponds to a private consumption conversion factor of US\$1 PPP (2011) worth 10,731 SSh.

inequality are thus defined over the entire population, aiming instead to capture the full consumption distribution without depending on the mean of the consumption distribution. It is important to note that measuring inequality with consumption, instead of income, tends to underestimate inequality in the population as consumption-based measures do not consider savings or wealth (Beegle, Christiaensen et al. 2016).

The Gini index or coefficient is the primary measure of inequality presented in this analysis. It ranges between 0 and 1, such that a coefficient equal to 0 indicates perfect equality and equal to 1 complete inequality. The Gini index is graphically represented by the Lorenz curve, a visual representation of the distribution of consumption across the population. It plots the cumulative population distribution by consumption percentile against the cumulative consumption distribution. The Gini index is the area between perfect equality, as represented by the 45-degree line, and the Lorenz curve observed from the data, relative to the maximum area that would be attained given perfect inequality (Figure B7-19). Formally,

$$Gini = 1 - \sum_{i=1}^{N} (x_i - x_{i-1})(y_i + y_{i-1}) \tag{13}$$

where y denotes the cumulative proportion of the total country-wide consumption expenditure for the *ith* person and x the cumulative proportion of the total population for the *ith* person. An alternative measure of inequality presented below is the Theil index. It is part of a larger family of measures referred to as the general entropy class (Coudouel, Hentschel et al. 2002), with the general formula given by:

$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N} - \sum_{i=1}^{N} (\frac{y_i}{\bar{y}})^{\alpha} - 1\right] \tag{14}$$

where y_i denotes the total consumption for individual *i,* $\bar{y}$ the mean expenditure per capita and *N* is the total population. The parameter α regulates the emphasis placed on higher or lower incomes. As with the Gini index, higher values of the Theil index represent higher levels of inequality, but unlike the Gini coefficient, this measure is not bounded between 0 and 1. Moreover, the Theil index is sensitive to inequality among the poor, and has the advantage of being additive across different subgroups in the country, allowing to decompose inequality into how much of it is explained by differences within groups and how much by differences between groups.

Results

As data collection in wave 2 of the SHFS was restricted to accessible areas, survey poverty headcount estimates are representative of only of the population living in these areas. The SHFS filled this critical gap by imputing poverty based on data extracted from satellite images for inaccessible areas. Section 0 describes the imputation methodology in detail. Survey and satellite imputation estimates for all population types were combined to compute a poverty headcount rate representing the entire Somali population (Table B7-23; Figure B7-7).[150] Overall, 77 percent of the Somali population lived below the poverty line in December 2017. This poverty incidence was 26 percentage points higher than unweighted average of low-income countries in Sub-Saharan Africa (51 percent) in 2017. The country has the third-highest poverty rate in the region, after Burundi and South Sudan (Figure B7-5).[151] The

[150] To derive a nation-wide poverty rate, survey and satellite estimates were combined in the following way. For each pre-war region and population type, the satellite prediction was considered if the accessibility rate in wave 2 was 90 percent or less, and the survey estimate was used if accessibility exceeded this threshold.

[151] The countries used for regional comparison are all the African low-income countries as defined by the World Bank: Benin, Burkina Faso, Burundi, Central African Republic, Chad, Comoros, Democratic Republic of Congo, Eritrea, Ethiopia, Guinea, Guinea-Bissau, Liberia, Madagascar, Malawi, Mali, Mozambique, Niger, Rwanda, Senegal, Sierra Leone, South Sudan, Tanzania, Togo, Uganda, and Zimbabwe. For each country, we include the most recent available year for each indicator.

high poverty incidence of Somalia is in line with its low levels of Gross Domestic Production (GDP) per capita, which was estimated at US$450 in 2017 (Figure B7-5).[152]

Figure B7-5: Cross-country comparison of poverty and GDP

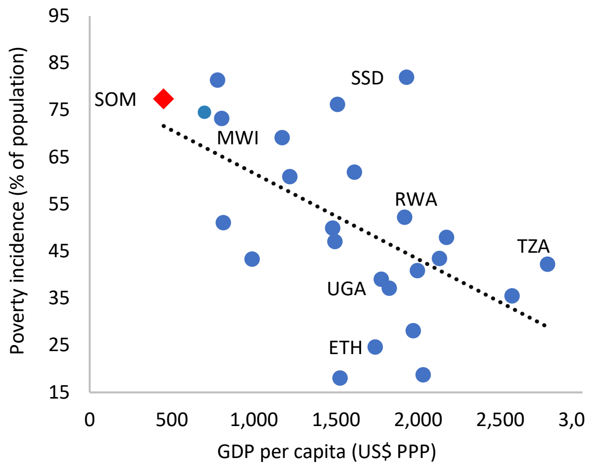
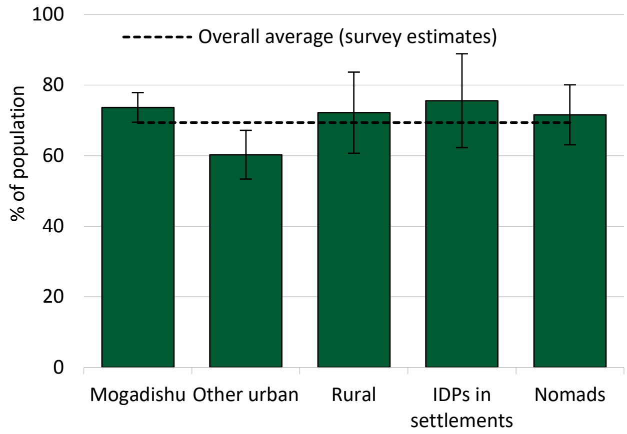


Figure B7-6: Poverty incidence



Poverty is somewhat heterogeneous between different population types and regions. Urban areas have a lower poverty headcount rate (60 percent), than the rest of the Somali population (Figure B7-6; $p<0.01$ vs. Mogadishu, $p<0.05$ vs. IDPs in settlements and nomads, and $p<0.10$ vs. rural areas).[153] This comparison excludes the capital, Mogadishu, whose residents are poorer than in other urban areas (between 72 and 76 percent). This higher poverty rate in Mogadishu compared to other urban areas is likely be the result of a larger concentration of the displaced population and the challenges associated with the displacement crisis, which the 2016/17 drought recently exacerbated.[154]

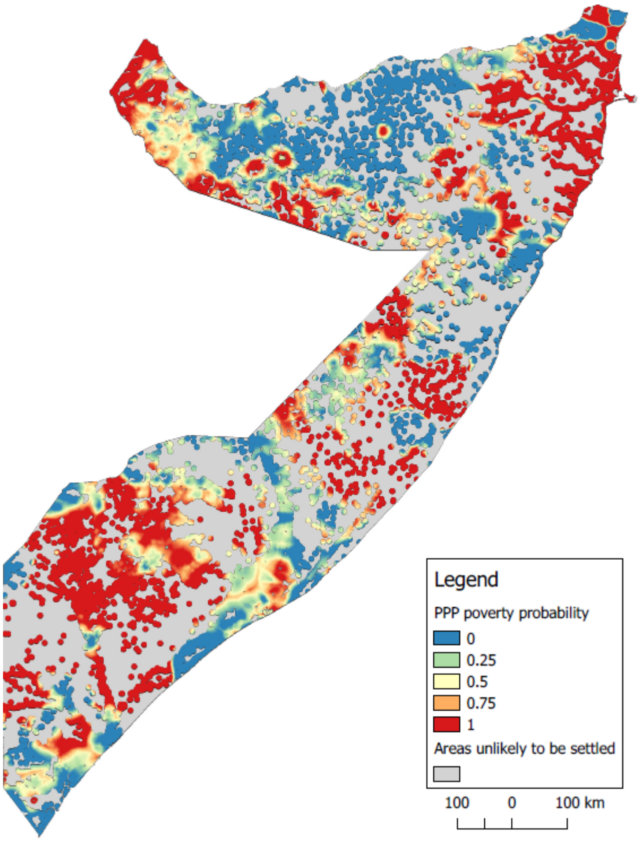
Poverty is also heterogeneous across space. Based on estimates from satellite imputation, the highest levels of poverty are clustered in south-western Somalia, and several districts in northern Somalia (Figure B7-7).

[152] For international comparisons, the poverty rate for Somalia was derived from satellite estimates. In the rest of the section, the figures refer to survey estimates unless explicitly noted.

[153] Urban areas usually benefit from agglomeration effects that result in more economic opportunities and access to services, relative to rural areas (Lall et al., 2017).

[154] Banadir/Mogadishu concentrates 41 percent of IDPs in settlements and 28 percent of the overall displaced population according to the second wave of the SHFS. The share is similar (22 percent) for the overall displaced population with data from Protection & Return Monitoring Network of the United Nations High Commissioner for Refugees (UNHCR).

Figure B7-7: Map of poverty incidence at the district-level based on satellite imputation[155]



Note: The poverty incidence of each region does not include IDPs in settlements.

Figure B7-8: Poverty gap

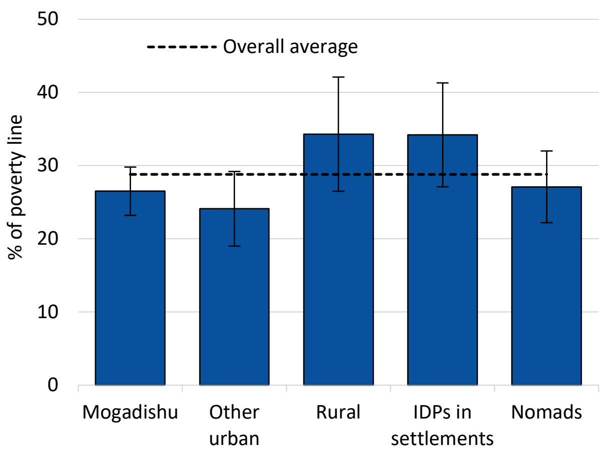
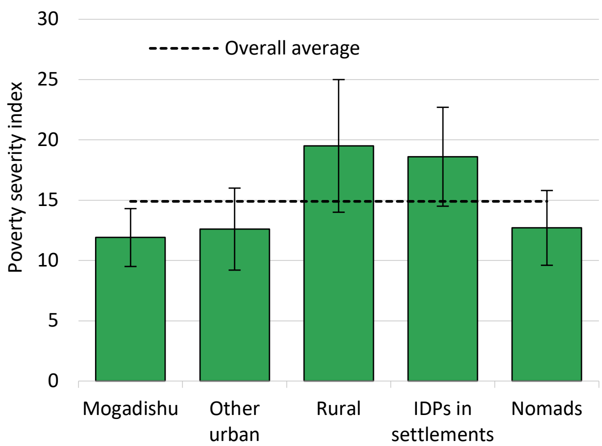


Figure B7-9: Poverty severity



The average poverty gap in Somalia was estimated at 29 percent (Figure B7-8), implying that the average consumption level of a poor Somali is about 71 percent of the international poverty line. Poverty was deeper in rural areas and IDP settlements (34 percent for both), compared to Mogadishu (27 percent, $p<0.1$) and other urban areas (24 percent, $p<0.05$). A large share of Somalis living in poverty, together with a considerable shortfall in their consumption expenditure relative to the poverty line means that a substantial boost in consumption would be necessary to overcome poverty.

[155] The boundaries on the map show approximate borders of Somali pre-war regions and do not necessarily reflect official borders, nor imply the expression of any opinion on the part of the World Bank concerning the status of any territory or the delimitation of its boundaries.

A transfer of around US$ 1.64 billion per year would lift the poor population out of poverty, assuming a perfect targeting scheme and ignoring administrative and logistical costs.[156] In line with these results, the average poverty severity index is 15 percent pointing to inequalities among the poor. These inequalities were concentrated in rural areas and IDP settlements (Figure B7-9).

Consumption was relatively homogenous due to the high levels of monetary deprivation shared by most households. Hence, inequality was relatively low with a Gini index of 34 percent in 2017. Other low-income countries in Sub-Saharan Africa with similar levels of poverty tend to have higher levels of inequality. For example, Malawi and South Sudan which have a poverty incidence of 69 and 82 percent respectively, have around a 12 percentage points higher Gini index than Somalia (Figure B7-10). The Gini index is 41 percent in rural areas, 34 percent in other urban areas and 26 percent in Mogadishu (Figure B7-11). Donor support concentrated in urban areas due to insecurity and accessibility constraints may help in levelling the consumption of the urban population, leading to lower levels of inequality.

Overall inequality stems largely from differences within regions and population groups, rather than from differences between them. The Theil index indicates that between 98 and 99 percent of total inequality are the result of inequality within groups (Table B7-11). Differences between households from within the same region or population group (Mogadishu, other urban, IDPs in settlements and nomads) largely explain inequality in consumption.

Figure B7-10: Cross-country comparison of poverty and inequality.

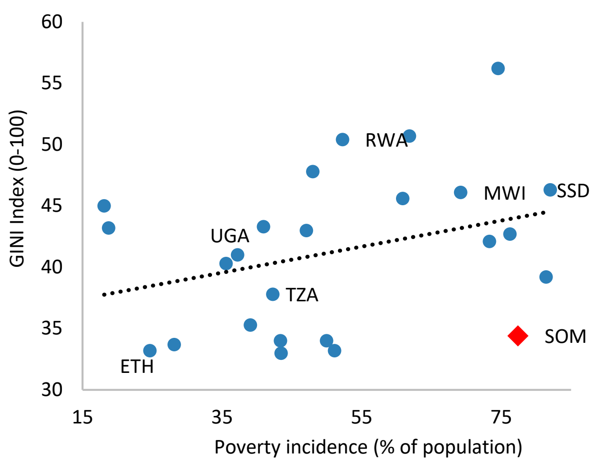
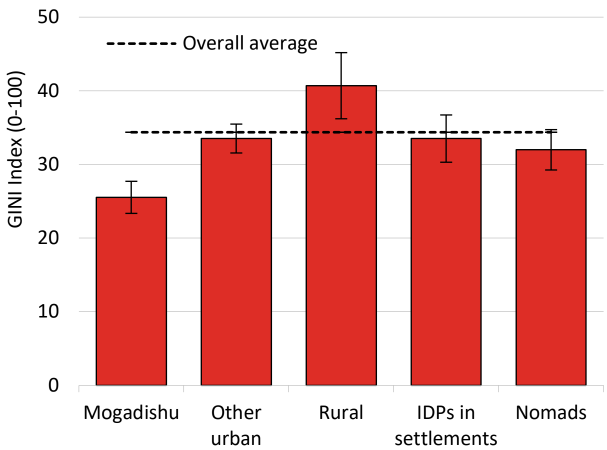


Figure B7-11: Inequality



[156] Corresponds to an annual value for all the regions, including areas not covered in wave 2 of the SHFS. For these, the same poverty incidence and gap was assumed as in regions covered by the survey.

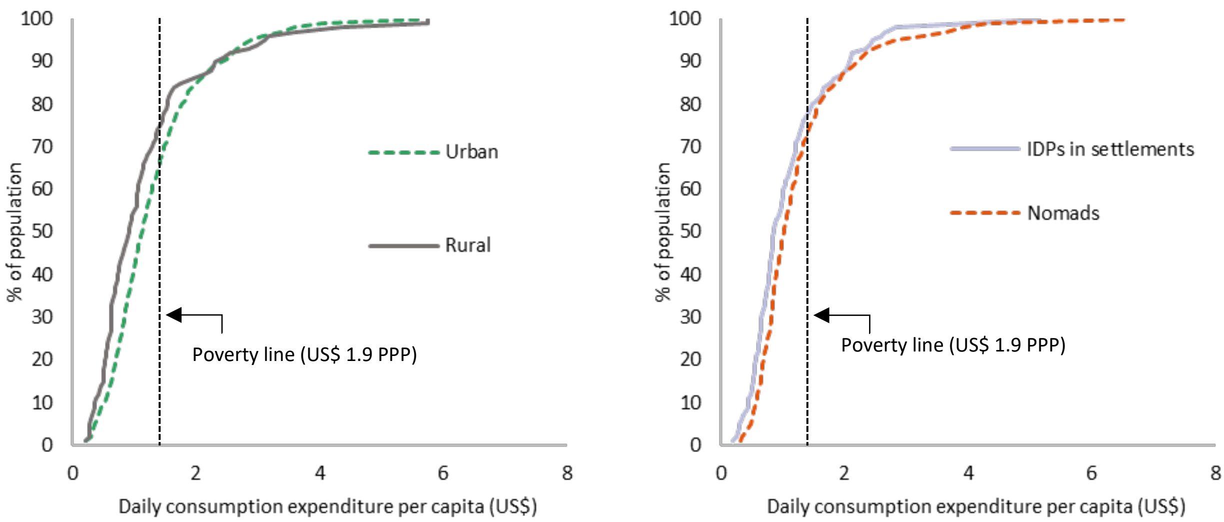
Table B7-11: Inequality decomposition

Theil GE(1) inequality index

Decomposition	By population type	By region
Between group	0.002	0.005
Within group	0.208	0.205
Total	0.210	0.210

The consumption distributions of the different population groups are relatively similar. The largest differences between rural and urban areas, as well as between IDPs in settlements and nomads, are found below the poverty line (Figure B7-12). A considerable share of 10 percent of non-poor population is clustered within 20 percent of the poverty line. This population is susceptible to fall into poverty in case of an unexpected decrease in their consumption levels.

Figure B7-12: Consumption distribution



Conclusions

The lack of data in Somalia poses a risk to evidence-based interventions aimed at alleviating poverty and inequality. To mitigate this risk, the World Bank implemented Wave 2 of the Somali High Frequency Survey to better understand welfare conditions of the population and to estimate the incidence of poverty. An analysis of the dataset has been published as the Somali Poverty and Vulnerability Assessment (World Bank 2018).

This paper contributes to several themes in the literature on poverty measurement and data collection in the context of conflict and fragility, involving hard-to-survey populations. It outlines how challenges associated to the context of insecurity and lack of statistical infrastructure in Somalia were overcome through four methodological and technological adaptations: i) building a probability-based population sampling frame; ii) minimizing the time spent in the field using the Rapid Consumption Methodology; iii) estimating poverty in completely inaccessible areas with correlates derived from satellite imagery and other geo-spatial data; and iv) employing a special sampling strategy for the nomadic population.

Further improvements in terms of human resource capacity should be considered to minimize disruptions to the quality of the data, besides field team training and stringent security protocols. Also, future applications should consider refining the model to predict poverty from satellite imagery by incorporating predictors with higher spatial frequencies, as well as data on building footprint which are likely improve the estimates. Other alternatives are thresholding some of the distance variables or applying a sigmoid transformation to capture variations in small areas. Furthermore, the accuracy of satellite-based imputations should be assessed based on a reference dataset, ideally in a more stable environment.

Appendix

Table B7-12: Sample overview.

Strata ID	Administrative unit	Population type	Total Interviews	Total EAs
1	Central Regions	IDP	36	3
2	Galmudug	IDP	0	0
3	Jubaland	IDP	84	7
4	Mogadishu	IDP	108	9
5	North East	IDP	192	16
6	North West	IDP	24	2
7	South West	IDP	24	2
8	Central Regions	nomadic	60	
9	Galmudug	nomadic	36	
10	Jubaland	nomadic	84	
12	North East	nomadic	96	
13	North West	nomadic	144	
13	South West	nomadic	84	
25	Hiraan	rural	144	12
26	Hiraan	urban	48	4
27	Middle Shabelle	rural	264	22
28	Middle Shabelle	urban	48	4
29	Galgaduud	rural	144	12
30	Galgaduud	urban	396	33
31	Lower Juba	urban	804	67
32	Gedo	rural	108	9
33	Gedo	urban	48	4
34	Lower Juba	rural	108	9
35	Middle Juba	rural	0	0
36	Middle Juba	urban	0	0
37	Banadir	urban	792	66
38	Bari	rural	48	4
39	Bari	urban	264	22
40	Mudug	rural	24	2
41	Mudug	urban	96	8
42	Nugaal	rural	12	1
43	Nugaal	urban	36	3
44	Awdal	rural	24	2
45	Awdal	urban	36	3
46	Sanaag	urban+rural	72	6
47	Sool	urban+rural	24	2
48	Toghdeer	rural	12	1
49	Toghdeer	urban	108	9
50	Woqooyi Galbeed	rural	36	3
51	Woqooyi Galbeed	urban	156	13
52	Bay	urban	540	45
53	Bakool	rural	48	4
54	Bakool	urban	12	1
55	Bay	rural	180	15
56	Lower Shabelle	rural	204	17
57	Lower Shabelle	urban	48	4
N/A	Host community sample	urban (IDP adjacent)	504	42
	Total		**6,384**	

Emerging state	Total interviews
Central Regions	684
Galmudug	576
Jubaland	1,248
Banadir	984
North East	840
North West	732
South West	1,296

Pre-war region	
Hiraan	264
Middle Shabelle	420
Galgaduud	576
Gedo	228
Lower Juba	996
Middle Juba	24
Bari	420
Mudug	324
Nugaal	96
Awdal	84
Sanaag	108
Sool	48
Toghdeer	192
Woqooyi Galbeed	300
Bakool	84
Bay	900
Lower Shabelle	312
Banadir	984

Urban / rural / IDP / nomad	
urban	3,936
rural	1,356
IDP	468
nomad	504

Oversampled populations	
Fisheries	324
Baidoa	540
Kismaayo	612
Mogadishu	900
Host communities	504

Figure B7-13: Fishery livelihood zones Somalia.

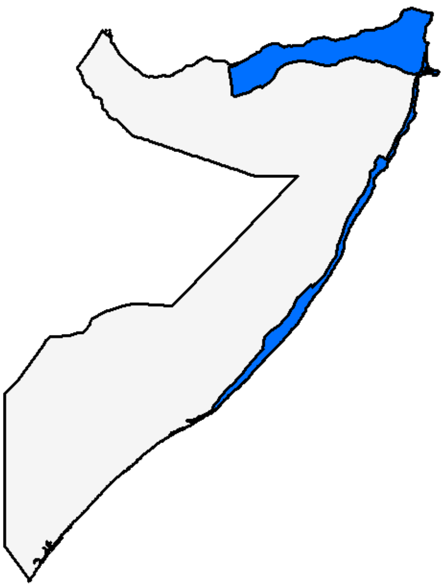


Table B7-13: Source of IDP settlement boundaries

#	Pre-war region	IDP name	Sources	Year
1	Bay	Baidoa	PESS	2016
2	Hiraan	Beletweyne	UN Shelter Cluster	2016
3	Nugaal	Garowe	UN Shelter Cluster	2016
4	Lower Juba	Kismayo	UN Shelter Cluster	2016
5	Bari	Qardho	UN Shelter Cluster	2016
6	Hiraan	Buloburto	UN Shelter Cluster	2015
7	Hiraan	Maxaas	UN Shelter Cluster	2015
8	Lower Juba	Afmadow, Diff and Dhobley	UN Shelter Cluster	2014
9	Togdheer	Burao	UN Shelter Cluster	2014
10	Mudug	Gaalkacyo North	UN Shelter Cluster	2014
11	Mudug	Gaalkacyo South	PESS	2014
12	Woqooyi Galbeed	Hargeisa	UN Shelter Cluster	2014
13	Middle Shabelle	Jowhar	UN Shelter Cluster	2014
14	Lower Juba	Kismayo	UN Shelter Cluster	2014
15	Gedo	Luuq	PESS	2014
16	Lower Shabelle	Marca	PESS	2014
17	Banadir	Mogadishu	PESS	2014

Replacement of sampling units

Sampling units (EAs, EBs, structures, households) may need to be replaced for a variety of reasons, but their replacement must follow a predetermined schedule that allows each interviewed household to be assigned a sampling weight and to preserve the sample's representativeness.

Replacement of enumeration areas (EAs)

An enumeration area (EA) was replaced only in one of the following scenarios:

(i) The EA was insecure for field teams to conduct interviews.
(ii) The EA could not be accessed for logistical reasons

(iii) The EA did not contain any residential structures.
(iv) All residential structures in the EA were visited unsuccessfully.

Main EAs were replaced from the pool of replacement EAs drawn for the same stratum during sample selection. All replacement EAs had a replacement rank thus setting the order of replacement in a replicable way. Replacement occurred both before fieldwork and during fieldwork. All selected EAs (including replacements) were manually checked prior to fieldwork to establish whether they were empty of structures (scenario (iii) above). If an EA was found to be empty, it was replaced with the highest-ranked replacement EA within its stratum. If the replacement EA was also empty, the next highest-ranked replacement EA was used to replace it, and so on (Table A.3). Prior to fieldwork 3 percent of selected urban EAs and 53 percent of selected rural EAs were found to be empty and thus replaced. If an EA needed to be replaced during fieldwork in any of the four scenarios listed above, the same schedule for replacement applied.

Replacements of enumeration blocks (EBs)

An entire EB was replaced in the following scenarios:

(i) The EB was insecure for field teams to conduct interviews.
(ii) The EB was empty or not comprised of inhabited dwellings (e.g. market).
(iii) All residential structures in the EB were visited unsuccessfully.

If an EB needed to be replaced in any of the three scenarios, the enumerator responsible for the EB randomly drew a replacement EB from the list of EBs in the current enumeration area using his/her tablet. Since, in most cases, there were exactly 12 EBs per EA and one interview had to be completed in each EB, EB replacement thus led to two or more households interviewed in the same EB.

Replacement of Households

Once the enumerator randomly selected a household, he/she made contact, trying to find a knowledgeable person in the household (an adult of 15 years or older with good knowledge of the household and its members). Where no knowledgeable person was currently present, enumerators scheduled follow-up visits before replacing the household.

Once contact was made during the first visit, it was possible to arrange a meeting at another time of the day or following day if more convenient for the respondent. However, if no knowledgeable person was at home and no later appointment was scheduled, the enumerator had to go back to the same household a second and a third time. At least 5 hours separated these consecutive visits. A household was replaced in any of the following scenarios:

(i) If the household was deemed unsafe by the enumerator, and this was confirmed by the team leader.
(ii) If someone in the household said that no knowledgeable person was around nor would be in the next 2 days. In this case, the household was replaced, without a second and third visit.
(iii) The household was found to be empty even after three visits to the household.
(iv) The head of household or a person 15 or above who was sufficiently knowledgeable to respond the survey was not available after three visits.
(v) The respondent refused to give his/her consent to continue the interview.

(vi) The interview that was conducted with that household is incomplete (the respondent stopped the interview in the middle or some required fields were not filled in) without the possibility to return to the household to complete the interview.

Sampling weights

The sampling weight of each household is the inverse of its probability of selection. Its probability of selection is the combination of selection probabilities at each stage of sample selection, in line with SHFS wave 2 sampling design discussed in section 0. A household's probability of selection is the probability of selection of the primary sampling unit in which it is located, multiplied by the probability of selection of the secondary sampling unit in which it is located, and so on.

Urban (non-host communities) and rural households

In urban and rural households, the EA was the primary sampling unit and the enumeration block (EB) was the secondary sampling unit. Enumerators followed a micro-listing protocol on the ground, in which they first listed all the structures in the EB, selected a structure, and then listed all the households in the selected structure. Thus, the probability of selection for urban and rural households is the following:

$$P_{hij} = P_1 P_2 P_3 P_4 = \frac{EA_j \, H_i}{H_j} \, \frac{BS_i}{B_i} \, \frac{SS_k}{S_k} \, \frac{HS_m}{H_m},$$

where

P_1: Probability of selecting the EA, given by $\frac{EA_j \, H_i}{H_j}$.

P_2: Probability of selecting the enumeration block, given by $\frac{BS_i}{B_i}$.

P_3: Probability of selecting the structure, given by $\frac{SS_k}{S_k}$.

P_4: Probability of selecting the household, given by $\frac{HS_m}{H_m}$.

EA_j: Number of EAs selected in strata *j.*

H_i: Number of households in the sample frame for the original EA *i.*

H_j: Number of households in the sample frame in strata *j.*

BS_i: Number of blocks selected in EA *i.*

B_i: Total number of blocks in EA *i.*

SS_k: Number of selected structures in block *k.*

S_k: Total structures in block *k.*

HS_m: Number of households selected in structure *m.*

HL_m: Total number of households in structure *m.*

Urban and host communities

Since the host community sample was drawn from a subset of urban enumeration areas, urban households selected in the host communities sample were part of two separate sampling processes. They thus had two positive probabilities to be selected into the final sample. To reflect this, the probability of selection for this groups is the following:

$$P_{hij} = P_1 P_2 P_3 P_4 = P_1 \frac{BS_i}{B_i} \frac{SS_k}{S_k} \frac{HS_m}{H_m},$$

where

P_1: Probability of selecting the EA given by two successive sampling processes

$$P_1 = (P_{1\text{a}} + P_{1\text{b}} - P_{1\text{a}} * P_{1\text{b}}),$$

such that,

$P_{1\text{a}} = \frac{EA_j H_i}{H_j}$ and $P_{1\text{b}} = \frac{EA_{host} H_{host}}{H_{host}}$.

EA_j: Number of EAs selected in urban strata *j.*

H_i: Number of households in the sample frame for the original urban EA *i.*

H_j: Number of households in the sample frame in urban strata *j.*

EA_{host}: Number of EAs selected in the host community sample.

H_{host}: Number of households in the sample frame for the original host community EA.

H_{host}: Number of households estimated in the host community sample.

P_2: Probability of selecting the enumeration block, given by $\frac{BS_i}{B_i}$.

P_3: Probability of selecting the structure, given by $\frac{SS_k}{S_k}$.

P_4: Probability of selecting the household, given by $\frac{HS_m}{H_m}$.

BS_i: Number of blocks selected in EA *i.*

B_i: Total number of blocks in EA *i.*

SS_k: Number of selected structures in block *k.*

S_k: Total structures in block *k.*

HS_m: Number of households selected in structure *m.*

HL_m: Total number of households in structure *m.*

IDP households

For IDP settlements, wave 2 of the SHFS employed a slightly different sampling strategy. IDP settlements were first sampled with probability proportional to size to determine the number of enumeration areas to be selected in each settlement. Then, the probability of selection follows the same schema as for urban and rural households, multiplying the first-stage probability of selection with the probability of selecting EA, which is selected with probability proportional to size, with the size given by the number of equal size enumeration blocks (EBs) in the EA. This is then multiplied by the probability of selection of the EB, the structure, and the household, all of which are selected with equal probability. Thus, the probability of selection of IDP households is given by:

$$P_{hijc} = P_1 P_2 P_3 P_4 P_5 = \frac{C_j H_c}{H_j} \frac{EA_\text{c} B_i}{B_c} \frac{BS_i}{B_i} \frac{SS_k}{S_k} \frac{HS_m}{H_m},$$

with

P_1: Probability of selecting the IDP settlement, given by $\frac{C_j H_c}{H_j}$.

P_2: Probability of selecting the EA, given by $\frac{EA_c B_i}{B_c}$.

P_3: Probability of selecting the enumeration block, given by $\frac{BS_i}{B_i}$.

P_4: Probability of selecting the structure, given by $\frac{SS_k}{S_k}$.

P_5: Probability of selecting the household, given by $\frac{HS_m}{H_m}$.

CA_j: Number of camps selected in strata *j.*

H_c: Number of households in the sample frame for camp *c.*

H_j: Number of households in the sample frame in strata *j.*

EA_c: Number of EAs selected in camp *c.*

B_i: Number of blocks in the original EA *i.*

B_j: Number of blocks in camp *c.*

BS_i: Number of blocks selected in the EA *i.*

B_i: Total number of blocks in EA *I.*

SS_k: Number of selected structures in block *k.*

S_k: Total structures in block *k.*

HS_m: Number of households selected in structure *m.*

HL_m: Total number of households in structure *m.*

Nomadic households

The sampling strategy for nomadic households was based on water points and a listing exercise of households in each of the water points (see above). First, the water point is selected with equal probability. Then, households are selected with equal probability in each listing round. Thus, the probability of selecting a nomadic household is given by:

$$P_{hij} = P_1 P_2 P_3 = \frac{WS_j}{W_j} \ \frac{LS_r}{L_r} \ \frac{HS_r}{H_r},$$

with

P_1: Probability of selecting the water point, given by $\frac{WS_j}{W_j}$.

P_2: Probability of selection for the listing round, given by $\frac{LS_r}{L_r}$.

P_3: Probability of selecting the household, given by $\frac{HS_r}{H_r}$.

WS_j: Number of selected water points in strata *j.*

W_j: Total number of water points in strata *j.*

LS_r: Selected number of listing rounds *r.*

L_r: Total number of listing rounds *r.*

HS_r: Number of households selected in listing round *r.*

H_r: Total households listed in listing round *r.*

Of note, since all the listing rounds were always selected, $\frac{LS_r}{L_r}$ = 1 and the probability of selection becomes:

$$P_{hij} = P_1 P_3 = \frac{WS_j}{W_j} \frac{HS_r}{H_r}.$$

Table B7-14: Summary of unit cleaning rules for food items.

Unit	Condition	Correction
250 ml/gr units	<=.03	multiply by 4
animal back, ribs, shoulder, thigh, head or leg	>=10 kg	divide by 10
basket or dengu (2 kg)	>=10	divide by 10
kilogram (1 kg)	>=100	divide by 1,000
kilogram (1 kg)	>20	divide by 10
spoonfull (200g)	>=2	divide by 2
faraasilad (12kg)	>12	divide by 12
Gram	<=0.001 (<1 gram) & item is a spice	multiply by 100
Gram	<=0.001 (<1 gram) & item is not a spice	multiply by 1,000
haaf (25 kg)	>=25	divide by 25
heap (750g)	>=7.5	divide by 10
large bag (50 kg)	>=50	divide by 50
spoonfull (4 g)	<0.004	multiply by 25
piece (30 g)	<=0.02	multiply by 3.334
piece (40 g)	<=0.03	multiply by 2.5
piece (50 g)	<=0.04	multiply by 2
piece (60 g)	<=0.05	multiply by 1.6667
piece (75 g)	<=0.065	multiply by 1.3334
piece (100g)	>=10	divide by 100
piece (110g)	>=11	divide by 110
piece (120g)	>=12	divide by 120
piece (125g)	>=12.5	divide by 125
piece (130g)	>=13	divide by 130
piece (150g)	>=15	divide by 150
piece (300g)	>=30	divide by 300
piece (350g)	>=35	divide by 350
piece (400g)	>=40	divide by 400
piece (500g)	>=50	divide by 500
piece (600g)	>=60	divide by 600
piece (800g)	>=80	divide by 800
rufuc/Jodha (12.5kg)	>12.5	divide by 12.5
large bag (10 kg)	>10	divide by 10
large bag (8 kg)	>8	divide by 8
large bag (7 kg)	>7	divide by 7
large bag (6 kg)	>6	divide by 6
large bag (5 kg)	>5	divide by 5
large bag (4 kg)	>=16	divide by 4
large bag (3 kg)	>=9	divide by 3
large bag (2.5 kg)	>=6.25	divide by 6.25
large bag (1.5 kg)	>=2.25	divide by 1.5

large bag (15kg)	>=15	divide by 10
saxarad (20kg)	>=20	divide by 20
large bag (30kg)	>=30	divide by 30
large bag (100kg)	>=100	divide by 100

Table B7-15: Conversion factor to Kg for units of food items.

Unit	Conversion factor to 1kg
1 liter tin (about 1 kg)	1
1 meal portion (about 300g)	0.3
250 ml tin (250g)	0.25
250gr tin (250g)	0.25
500 gr tin (500g)	0.5
500 ml tin (500g)	0.5
Animal Back (around 1.5kg)	1.5
Animal leg (around 1.5kg)	1.5
Animal Ribs (around 2kg)	2
Animal Shoulder (around 1kg)	1
Animal Thigh (around 1 kg)	1
Basket (dengu, around 4kg)	4
Bottle (1l)	1
Bottle (2.5l)	2.5
Bottle (350g)	0.35
Bottle (400g)	0.4
Bottle (500g)	0.5
Bottle (600g)	0.6
Bottle (750g)	0.75
Bottle (750ml)	0.75
Bottle (800g)	0.8
Bottle (800ml)	0.8
Breast (130g)	0.13
Cup (100g)	0.1
Cup (125g)	0.125
Cup (1l)	1
Cup (250g)	0.25
Cup (250ml)	0.25
Cup (400g)	0.4
Cup (400ml)	0.4
Cup (500g)	0.5
Cup (500ml)	0.5
Cup (750g)	0.75
Faraasilad (12kg)	12
Gram	0.001
Haaf (25kg)	25
Heap (125g)	0.125
Heap (25kg)	25
Heap (2kg)	2
Heap (300g)	0.3
Heap (350g)	0.35
Heap (500g)	0.5
Heap (5kg)	5
Heap (750g)	0.75
Kilogram	1
Large bag (100kg)	100
Large bag (10kg)	10
Large bag (12kg)	12
Large bag (15kg)	15
Large bag (1kg)	1
Large bag (25kg)	25
Large bag (2kg)	2
Large bag (30kg)	30

Large bag (3kg)	3
Large bag (4kg)	4
Large bag (50kg)	5
Large bag (5kg)	5
Large bag (6kg)	6
Large bag (7kg)	7
Large bag (8kg)	8
Leg (250g)	0.25
Liter	1
Loaf (200g)	0.2
Madal/Nus kilo ruba (0.75kg)	0.75
Mass (1.5kg)	1.5
Packet (1kg)	6
Packet (3kg)	3
Packet (sealed box/container, 1.5kg)	1.5
Packet (sealed box/container, 10kg)	10
Packet (sealed box/container, 12.5kg)	12.5
Packet (sealed box/container, 120g)	0.12
Packet (sealed box/container, 150g)	0.15
Packet (sealed box/container, 15kg)	15
Packet (sealed box/container, 1kg)	1
Packet (sealed box/container, 20kg)	20
Packet (sealed box/container, 250g)	0.25
Packet (sealed box/container, 2kg)	2
Packet (sealed box/container, 300g)	0.3
Packet (sealed box/container, 350g)	0.35
Packet (sealed box/container, 3kg)	3
Packet (sealed box/container, 500g)	0.5
Packet (sealed box/container, 5kg)	5
Packet (sealed box/container, 6kg)	6
Piece (1.5kg)	1.5
Piece (100g)	0.1
Piece (110g)	0.11
Piece (120g)	0.12
Piece (125g)	0.125
Piece (150g)	0.15
Piece (1kg)	1
Piece (200g)	0.2
Piece (250g)	0.25
Piece (2kg)	2
Piece (300g)	0.3
Piece (30g)	0.03
Piece (350g)	0.35
Piece (400g)	0.4
Piece (500g)	0.5
Piece (50g)	0.05
Piece (600g)	0.6
Piece (60g)	0.06
Piece (750g)	0.75
Piece (75g)	0.075
Piece (large)	0.9
Rufuc/Jodha (12.5kg)	12.5
Saxarad (20kg)	20
Small bag (150g)	0.15
Small bag (15kg)	15
Small bag (1kg)	1
Small bag (2.5kg)	2.5
Small bag (250g)	0.25
Small bag (2kg)	2
Small bag (3kg)	3
Small bag (4kg)	4
Small bag (500g)	0.5
Small bag (5kg)	5

Small bag (6kg)	6
Small bag (750g)	0.75
Spoonful (125g)	0.125
Spoonful (200g)	0.2
Spoonful (40g)	0.04
Spoonful (4g)	0.004
Tray (1kg)	1
Tumin (125g)	0.125

Table B7-16: Summary of cleaning rules for currency.

Currency	Condition	Correction
Somaliland shillings thousands	Price>1,000 for food and nonfood item Price>10,000 for durable goods	Divide by 1,000 because respondents meant units, not thousands.
Somali shillings thousands	Price>1,000 for food and nonfood items Price>10,000 for durable goods	Divide by 1,000 because respondents meant units, not thousands.
US$	Price >1,000	Replace currency to Somali(land) shillings.

Cleaning rules for food consumption data

- *Rule 1.*
    - Consumption quantities with missing values for items reported as consumed were replaced with item-specific median consumption quantities.
    - Missing purchase quantities and missing prices for items consumed were replaced with item-specific median purchase quantity and item-specific median purchase price.
- *Rule 2.* Records where the respondent did not know or refused to respond if the household had consumed the item, were replaced with the mean value, including non-consumed records.
- *Rule 3.* Records with the same value for quantity consumed or quantity purchased and price are assumed to have a data entry error in the price or quantity and are replaced with the item-specific medians.
- *Rule 4.* Records that have the same value in quantity consumed and quantity purchased but different units are assumed to have a wrong unit either for consumption or purchase. For both quantities, the item-specific distribution of quantities in kg is calculated to determine the deviation of the entered figure from the median of the distribution. The unit of the quantity that is further away from the median is corrected with the unit of the quantity closer to the median.
- *Rule 5.*
    - Missing and zero prices are replaced with item-specific medians
    - Outliers for unit prices were identified and replaced with the item-specific median. This includes unit prices in the top 10 percent of the overall cumulative distribution (considering all items), and unit prices below 0.07 US$.
- *Rule 6.* the consumption value in US$ was truncated to the mean plus 3 times the standard deviation of the cumulative distribution for each item, if the record exceeded this threshold.

All medians are estimated at the EA level if a minimum of 5 observations are available excluding previously tagged records. If the minimum number of observations is not met, medians are estimated at the strata-level before proceeding to the survey level. In addition, medians greater than 20 kg and smaller than 0.02 kg were not considered for quantities, while medians greater than 20 US$ and smaller than 0.005 US$ were also excluded for unit prices.

Cleaning rules for nonfood consumption data

- *Rule 1.* Zero, missing prices and missing currency for purchased items are replaced with item-specific medians.
- *Rule 2.* Records where the respondent did not know or refused to respond if the household had purchased the item, were replaced with the mean value, including non-consumed records.
- *Rule 3.* Prices that are beyond a specific threshold for each recall period (Table B7-17) are replaced with item-specific medians.
- *Rule 4.* Prices below the 1 percent and above the 95 percent of the cumulative distribution for each item are replaced with item-specific medians
- *Rule 5.* the purchase value in US$ was truncated to the mean plus 3 times the standard deviation of the cumulative distribution for each item, if the record exceeded this threshold.

The item-specific medians were applied at the EA, strata and survey level as described above.

Table B7-17: Threshold for non-food item expenditure (US$)

Recall period	Min	Max
1 Week	0.05	30
1 Month	0.20	95
3 Months	0.45	200
1 Year	0.80	1,200

Cleaning rules for durable assets

- *Rule 1.* Vintages with missing values and greater than 10 years are replaced with item-specific medians.
- *Rule 2.* Current and purchase prices equal to zero are replaced with item-specific medians.
- *Rule 3.* Records that have the same figure in current value and purchase price are incorrect. For both, the item-vintage-specific distribution is calculated to determine the deviation of the entered figure from the median. The one that is further away from that median is corrected with the item-year-specific median value.
- *Rule 4.* Depreciation rates are replaced by the item-specific medians in the following cases:
    - Negative records
    - Depreciation rates in the top 10 percent and vintage of one year
    - Depreciation rates in the bottom 10 percent and a vintage greater or equal to 3 years
- *Rule 5.* Records with 100 items or more, and those that reported to own a durable good but did not report the number were replaced with the item-specific medians of consumption in US$

- *Rule 6.* Consumption in the top and bottom 1 percent of the overall distribution were replaced with item-specific medians
- *Rule 7.* Records where the respondent did not know or refused to respond if the household owned the asset, were replaced with the mean consumption value, including non-consumed records.
- *Rule 8.* the consumption value in US$ was truncated to the mean plus 3 times the standard deviation of the cumulative distribution for each item, if the record exceeded this threshold.

All medians are estimated at the EA level if a minimum of 3 observations are available excluding previously tagged records. If the minimum number of observations is not met, medians are estimated at the strata-level before proceeding to the survey level. Table B7-18 presents median expenditure and median depreciation rates for each durable item.

Table B7-18: Median consumption and depreciation rate of durable assets.

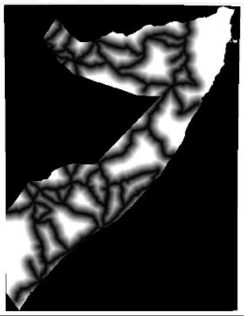
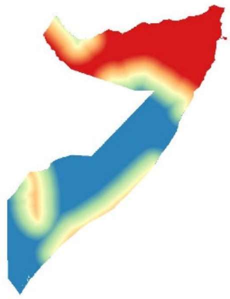
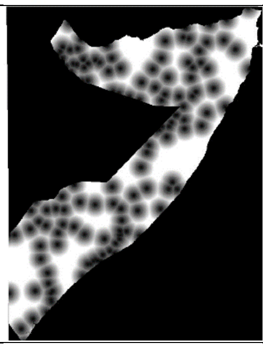
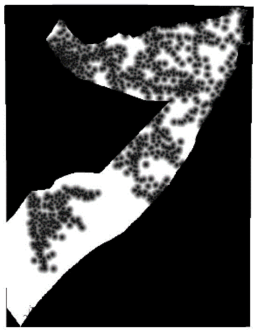
Item	Median consumption (current US$/week)	Median depreciation rate
Air conditioner	0.002	0.264
Bed with mattress	0.092	0.229
Car	0.013	0.159
Cell phone	0.085	0.245
Chair	0.015	0.242
Clock	0.007	0.267
Coffee table (for sitting room)	0.002	0.209
Computer equipment & accessories	0.010	0.182
Cupboard, drawers, bureau	0.019	0.213
Desk	0.001	0.349
Electric stove or hot plate	0.000	0.204
Fan	0.006	0.188
Gas stove	0.005	0.159
Generator	0.017	0.333
Iron	0.007	0.229
Kerosene/paraffin stove	0.000	0.248
Kitchen furniture	0.006	0.296
Lantern (paraffin)	0.000	0.092
Lorry	0.001	0.209
Mattress without bed	0.041	0.267
Mini-bus	0.002	0.248
Mortar/pestle	0.005	0.244
Motorcycle/scooter	0.004	0.229
Photo camera	0.000	0.005
Radio ('wireless')	0.005	0.276
Refrigerator	0.007	0.210
Satellite dish	0.004	0.213
Sewing machine	0.001	0.229
Small solar light	0.002	0.195
Solar panel	0.002	0.188
Stove for charcoal	0.002	0.296
Table	0.014	0.229
Tape or CD/DVD player; HiFi	0.001	0.337
Television	0.056	0.201

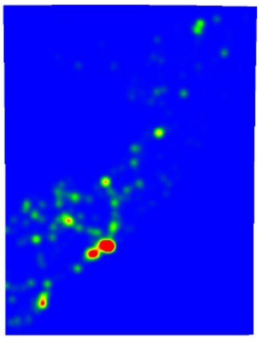
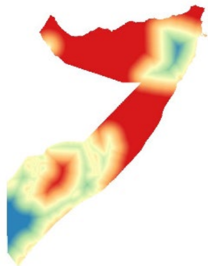
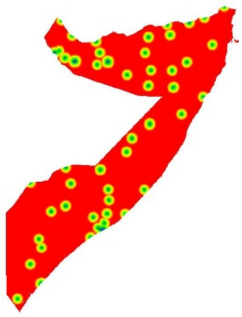
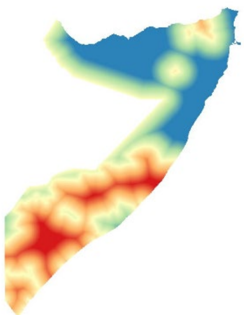
Upholstered chair, sofa set	0.021	0.267
VCR	0.000	0.161
Washing machine	0.013	0.210

Table B7-19: Overview of spatial variables used in poverty imputation.

Variable	Source	Description	Illustration
Distance to bare areas	WorldPop Global covariate dataset[157]. The ESA-CCI 300m annual global landcover dataset was used to produce this layer.	Distance to borders of areas of which land cover was classified as bare. The distance is positive outside of the areas and negative inside.	
Distance to cultivated areas	WorldPop Global covariate dataset[157]. The ESA-CCI 300m annual global landcover dataset was used to produce this layer.	Distance to borders of areas of which land cover was classified as cultivated. The distance is positive outside of the areas and negative inside.	
Temperature	WorldClim v2	Average annual temperature. Original layer from World Clim.	
Precipitations	WorldClim v2	Average annual precipitations. Original layer from World Clim.	

[157] The WorldPop "dist-to" datasets have been produced by D. Kerr, H. Chamberlain and M. Bondarenko in the framework of the WorldPop “Global high resolution population denominators”, project funded by the Bill & Melinda Gates Foundation (OPP1134076).

Distance to major roads	WFP	Distance to primary and secondary roads in km. FM/WP rasterized the original shapefile to 100m and computed the distance transform.	
Distance to drought areas	FAO SWALIM	Distance in km to borders of areas labelled as 'moderate drought' and 'severe drought' (computed by FM/WP). The distance is positive outside of the areas and negative inside.	
Distance to medical sites	UNICEF, FAO SWALIM	Distance to medical sites. FM/WP computed the distance to the points given in the source dataset. 2005	
Distance to schools	UNICEF, FAO SWALIM	Distance to schools. FM/WP computed the distance to the points given in the source dataset. 2004	
Distance to water sources	FAO SWALIM	Distance to strategic water points or sources. FM/WP computed the distance to the points given in the source dataset. 2008	

Distance to waterways	OSM extract	Volunteer-reported vector data of waterway locations. FM/WP rasterized the vectors at 100m and computed the distance transform.	
Conflict density	ACLED	Reports on violent events (e.g. battles, riots) from news outlets. FM/WP computed the spatial average of the number of fatalities from January 2014 to May 2018, within a 25 km radius.	
Distance to food insecure areas	FEWS NET	Food security outcomes for October 2017. FM/WP computed the distance to borders of areas with an IPC phase of 3 or more. The distance is positive outside of the areas and negative inside.	
Distance to urban areas	UNFPA / PESS urban EAs	Distance to borders of urban areas. FM/WP used the UNFPA urban EAs and filled in gaps within urban areas. Then we computed the distance to the urban borders. The distance is positive outside of the areas and negative inside.	
Distance to unsafe areas	World Bank	Distance to areas labelled as unsafe by the World Bank. FM/WP rasterized the shapefile provided at 100m then computed the distance to the unsafe areas border. The distance is positive outside of the areas and negative inside.	

<table>
<tr><td>Population density</td><td>World Bank / Flowminder</td><td>Population density inferred at 100m as part of the work on: Defining a new Somali national sampling frame.</td><td></td></tr>
</table>

Table B7-20: Summary statistics of collected spatial variables.

	Urban mean	Urban median	Urban min	Urban max	Rural mean	Rural median	Rural min	Rural max
Conflicts density	1302.29	134	0.11	5697.2	61.24	9.15	0	5688.03
Distance to bare areas	8.89	2.95	-2.72	101.48	22.92	12.27	-10.78	132.95
Distance to cultivated areas	2.95	0.7	-2.69	57.76	10.18	5.19	-12.29	93.15
Distance to dry areas	-15.07	-18.03	-484.27	239.91	-7.02	5.78	-505.07	281.1
Distance to food insecure areas	-80.46	-54.13	-272.25	92.98	-65.2	-41.66	-329.51	171.76
Distance to major roads	1.89	0.9	0	188.74	29.96	17.93	0	237.97
Distance to medical sites	3.17	2.48	0	27.34	35.2	30.51	0	158.01
Distance to schools	8.74	3.34	0	78.34	27.01	21.92	0	165.1
Distance to unsafe areas	37.86	27.4	-107.32	186.93	27.78	16.38	-142.18	262.72
Distance to urban areas	-1.1	-0.8	-4.8	-0.1	52.88	48.91	0	200.32
Distance to water sources	45.82	5.32	0	237.8	31.57	13.26	0	295.61
Distance to waterways	13.76	3.69	0	111.89	25.49	15.34	0	141.71
Population density	78.1	17.57	0.11	978	0.18	0	0	1551.57
Precipitations	358.63	419.07	9.05	551.15	287.98	266.56	9.05	746.43
Temperature	25.04	26.5	17.24	29.88	29.81	26.77	14.57	30.51

Notes: distance to bare, cultivated, drought, food insecure areas, unsafe and urban areas is positive outside these areas and negative inside these areas (e.g. location within unsafe area has negative distance to unsafe areas).

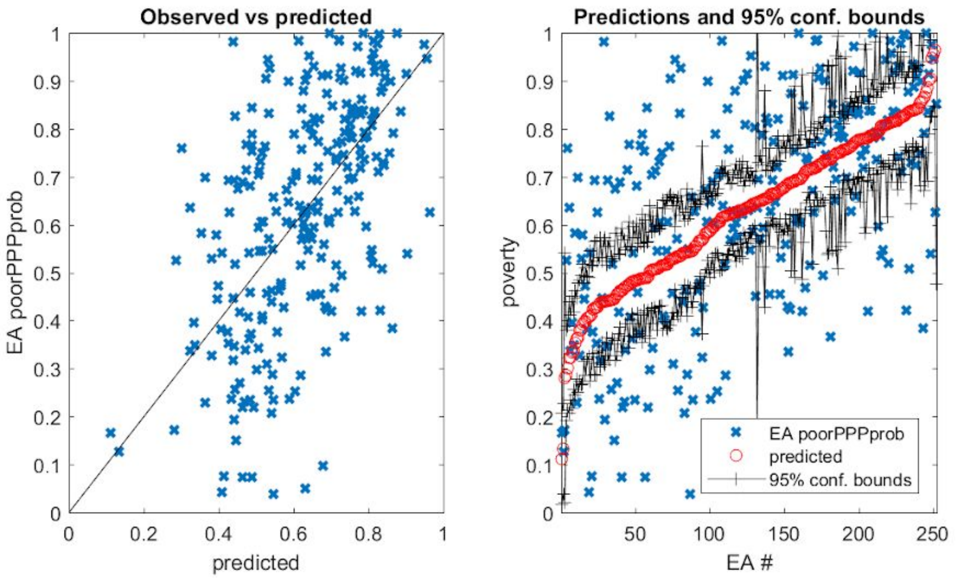
Table B7-21: Linear correlations between spatial variables and poverty.

Spatial variables	poverty	Urban poverty	Rural poverty	Nomad poverty	IDP poverty
Conflicts density	0.12	0.18	-0.1	0.2	0.37
Distance to bare areas	-0.12	0.01	0.14	-0.53	0.14
Distance to cultivated areas	-0.08	0.04	-0.08	-0.15	-0.14
Distance to dry areas	-0.21	-0.27	-0.21	-0.26	-0.36
Distance to food insecure areas	-0.1	-0.04	-0.33	-0.34	0.25
Distance to major roads	-0.01	0.02	0.02	-0.08	0.18
Distance to medical sites	-0.02	0.11	-0.22	0.01	-0.27
Distance to schools	0.16	0.29	0.11	0.08	0.34
Distance to unsafe areas	-0.04	0.06	0.29	-0.2	-0.18

Distance to urban areas	-0.22	-0.01	-0.18	-0.61	0.17
Distance to water sources	-0.1	0.13	-0.22	-0.55	0.1
Distance to waterways	-0.28	-0.11	-0.25	-0.58	0.07
Population density	-0.16	0.11	-0.03	-0.11	-0.68
Precipitations	0.06	-0.04	0.08	0.01	0.37
Temperature	-0.06	0.04	-0.19	-0.31	0.14

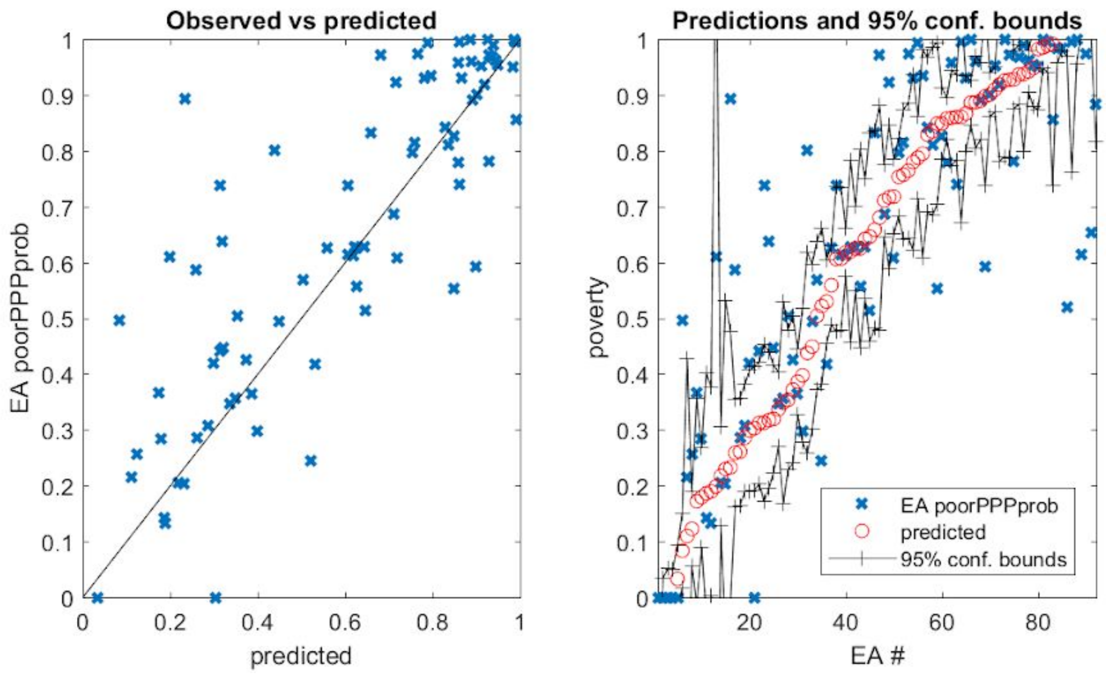
Notes: distance to bare, cultivated, drought, food insecure areas, unsafe and urban areas is positive outside these areas and negative inside these areas (e.g. location within unsafe area has negative distance to unsafe areas).

Figure B7-14: Visual representation of urban model fit.



Notes: Blue crosses are urban EAs. The left figure plots the survey estimates against the model predictions for each EA, the black line shows the perfect fit. The right figure shows the model predictions for each EA ordered by increasing poverty (red circles) and compares it to the lower and upper 95% confidence bounds on the predictions (black crosses) and to the survey estimates (blue crosses).

Figure B7-15: Visual representation of rural model fit.



Notes: Blue crosses are rural EAs. The left figure plots the survey estimates against the model predictions for each EA, the black line shows the perfect fit. The right figure shows the model predictions for each EA ordered by increasing poverty (red circles) and compares it to the lower and upper 95% confidence bounds on the predictions (black crosses) and to the survey estimates (blue crosses).

Figure B7-16: In and out-of-the sample R-squared.

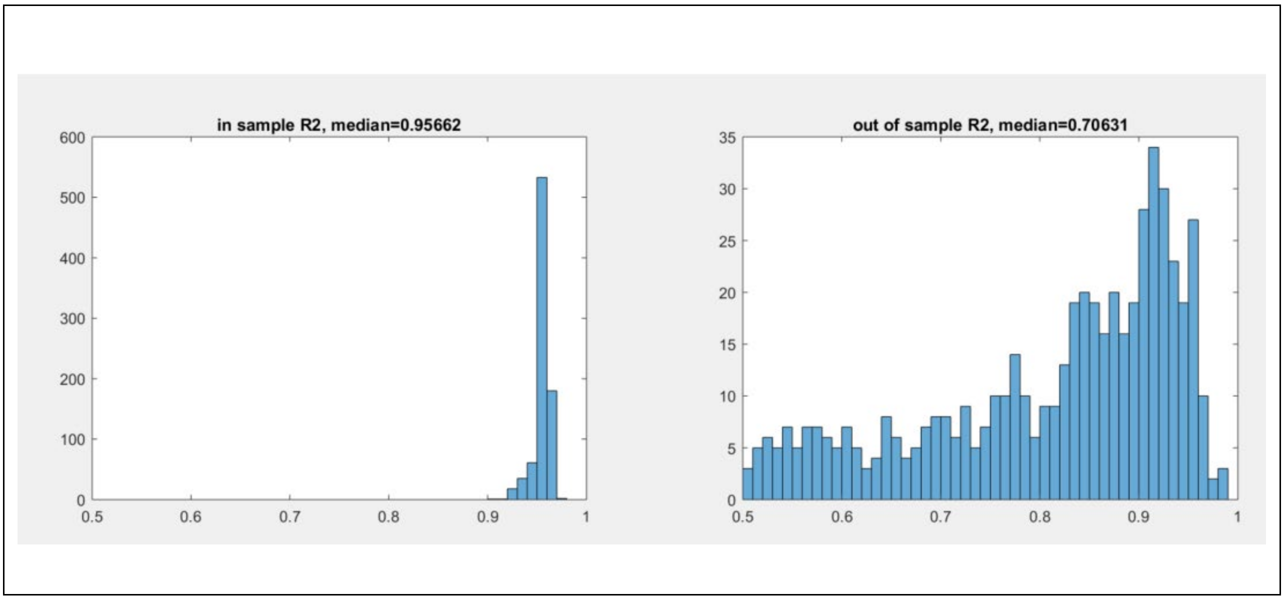


Figure B7-17: Relative bias of simulation results using the rapid consumption estimation.

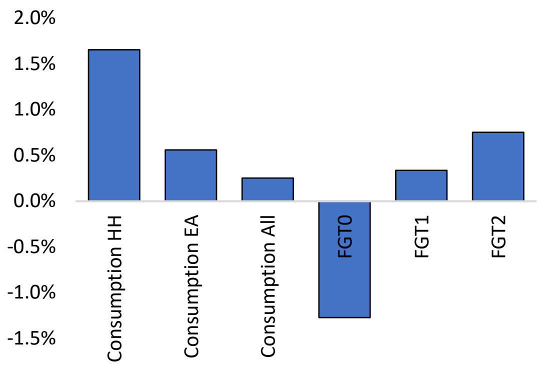


Figure B7-18: Relative standard error of simulation results using the rapid consumption estimation.

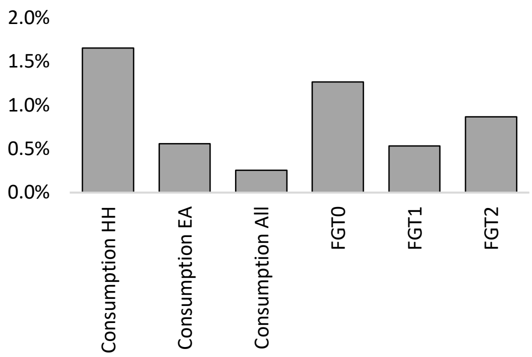


Table B7-22: Fieldwork regional breakdown

Region	Pre-war region
North-West	Awdal, Sanaag, Sool,Togdheer, Woqooyi Galbeed
North-East	Nugaal, Bari, Mudug
Central regions	Hiraan, Middle Shabelle, Galgaduud
Mogadishu	Banadir
Jubbaland	Gedo, Lower Juba
South-West	Bay, Bakool, Lower Shabelle

Figure B7-19: Lorenz curve based on SHFS data.

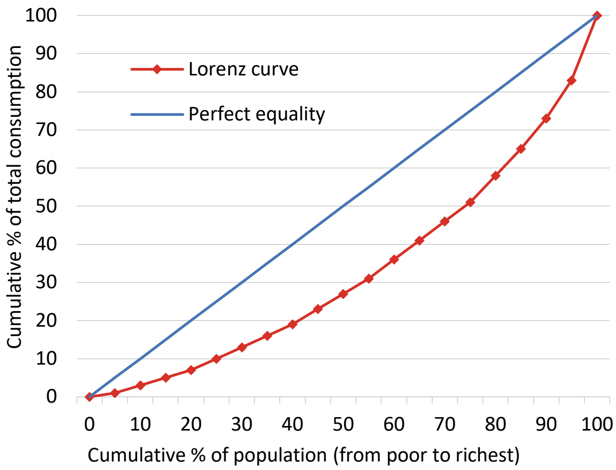


Table B7-23: Poverty incidence by pre-war region.

Pre-war region	Poverty incidence (% of population)					
	Urban areas			Rural areas and Nomads		
	Accessibility rate	Survey estimate	Satellite estimate	Accessibility rate	Survey estimate	Satellite estimate
Awdal	100%	21% (6%, 36%)	23% (14%, 33%)	94%	76% (72%, 79%)	68% (57%, 79%)
Bakool	35%	55% (55%, 5%)	16% (11%, 27%)	21%	26% (13%, 39%)	55% (33%, 78%)
Banaadir	87%	74% (69%, 78%)	69% (61%, 77%)	N/A	N/A	N/A
Bari	99%	77% (58%, 95%)	71% (60%, 81%)	92%	63% (54%, 71%)	84% (76%, 90%)
Bay	86%	83% (80%, 85%)	77% (67%, 88%)	46%	92% (87%, 97%)	73% (57%, 84%)
Galgaduud	88%	49% (42%, 55%)	42% (32%, 56%)	50%	47% (40%, 54%)	75% (61%, 85%)
Gedo	100%	58% (52%, 65%)	66% (55%, 76%)	43%	42% (0%,100%)	60% (47%, 70%)
Hiraan	44%	71% (39%, 100%)	76% (66%, 85%)	28%	18% (0%, 52%)	40% (24%, 61%)
Lower Juba	92%	50% (34%, 66%)	55% (45%, 64%)	9%	N/A	31% (9%, 59%)
Lower Shebelle	28%	50% (20%, 79%)	61% (50%, 72%)	33%	58% (46%, 70%)	41% (30%, 60%)
Middle Juba	0%	N/A	97% (74%, 100%)	9%	N/A	54% (26%, 80%)
Middle Shebelle	98%	72% (38%, 105%)	79% (64%, 92%)	77%	75% (57%, 93%)	47% (33%, 63%)
Mudug	100%	41% (34%, 48%)	45% (35%, 56%)	76%	53% (42%, 64%)	48% (36%, 60%)
Nugaal	100%	48% (35%, 61%)	56% (47%, 65%)	100%	90% (73%, 100%)	72% (52%, 86%)
Sanaag	100%	95% (94%, 95%)	77% (74%, 79%)	100%	100% (N/A)	86% (80%, 92%)
Sool	89%	85% (85%, 85%)	80% (70%, 81%)	98%	79% (61%, 97%)	56% (37%, 75%)
Togdheer	100%	69% (59%, 78%)	60% (50%, 71%)	98%	96% (90%, 100%)	83% (69%, 92%)
Woqooyi Galbeed	100%	64% (49%, 79%)	64% (54%, 74%)	96%	82% (74%, 89%)	58% (43%, 75%)

Note: N/A=not applicable. 95% confidence interval in parentheses.

C. Part II: Impacts of shocks and fragility

1. Impact of Conflict on Livelihoods in South Sudan[158]

Luca Parisotto and Utz Pape[159]

Introduction

Civil wars and violent conflict are inextricably linked with poverty. The 19 countries classified by FAO as being in a protracted food crisis in 2017 were all experiencing violent conflict. Similarly, 60 percent of the 815 million people who are undernourished and 79 percent of the 155 million stunted children worldwide live in countries affected by violent conflict (FAO, IFAD et al. 2017, Brück and d'Errico 2019). A large macro-level literature documents the empirical association between conflict and poverty (Collier and Hoeffler 2007, Blattman and Miguel 2010). Poverty is both a strong predictor for the onset of conflict and the incidence of conflict is associated with heightened deprivation and poverty, at least in the short run. More recently, the increasing availability of comprehensive micro-data from post-conflict regions has led to the emergence of a new literature examining the consequences of conflict exposure and its mechanisms at the household level (Justino 2009, Justino 2012, Martin-Shields and Stojetz 2019).

Some of the more salient findings that have emerged from this micro-level literature concerns the persistence of the impact of conflict exposure on human capital within countries. It manifests itself primarily through lower anthropometric and health outcomes, which can be observed even long after the end of the fighting (e.g., Bundervoet, Verwimp et al. 2009, Shemyakina 2011, Minoiu and Shemyakina 2014). This body of evidence rests on the well-established empirical observation that exposure to adverse events during critical developmental years can have long-lasting consequences on individuals' future outcomes (Almond and Currie 2011). For example, children born in conflict-exposed regions had significantly lower height-for-age z-scores and higher rates of stunting than those who were not (Bundervoet, Verwimp et al. 2009, Akresh, Verwimp et al. 2011). Akresh, Bhalotra et al. (2012) show that, even 40 years later, Nigerian women exposed to the Biafran civil war were shorter than their counterparts on average, especially for those exposed between 13-16 years old. Similarly, Camacho (2008) first showed that Colombian women exposed to violent conflict during the first three months of their pregnancy gave birth to children with lower birth weights. A smaller but related literature looks more specifically at the impact of conflict on poverty and food security, arguably the primary driver of these differences in human capital, and finds a strong association between conflict exposure and greater food insecurity and lower consumption levels (Dabalen and Paul 2012, D'Souza and Jolliffe 2013, Mercier, Ngenzebuke et al. 2016).

Due to the lack of micro-data collected during or shortly after conflict exposure, much of this literature is forced to rely on ex-post data typically collected several years after the end of the conflict, and as such remains relatively silent on the short-term impacts of conflict exposure. While there is an implicit link between conflict and food insecurity, strong evidence documenting this relationship is still lacking (Martin-Shields and Stojetz 2019). This paper contributes to this literature by leveraging representative consumption expenditure data collected during the most recent conflict in South Sudan.

South Sudan gained its independence in July 2011, it was only two years later, in December 2013, that clashes broke out in Juba between factions of soldiers loyal to President Salva Kiir and factions loyal to former vice-president Riek Machar. This was followed by a wave of violence sweeping throughout

---

[158] UP developed the research question and designed as well as supervised the field work. LP and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

[159] Corresponding author: Utz Pape (upape@worldbank.org). The authors thank XX for their contributions. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

the country. Although a peace agreement was signed in August 2015, a constant state of violence largely prevailed throughout the country. The conflict intensified in July 2016 after renewed clashes in Juba, and by the end of 2017 the conflict had escalated into a large-scale humanitarian crisis, with almost 4.5 million people forcibly displaced and 6 million facing heightened food insecurity – out of a population of about 12 million (UNHCR 2018). It was during this period of intense violence, between late 2016 to early 2017, that the High Frequency Survey in South Sudan (HFS) conducted a representative consumption and expenditure survey. The HFS was explicitly designed to be conducted in a context of insecurity and many households were thus be interviewed very recently after having been exposed to the conflict – up to less than a month following exposure (Pape and Parisotto 2019).

This study combines the HFS 2016-17 data with data collected before the conflict began during the National Baseline Survey in 2009 to estimate a repeated cross-section difference-in-differences model of the impact of conflict exposure on welfare, as proxied by average consumption levels, the poverty headcount, and poverty gap. By differencing across time and across groups, DID estimation nets out both group specific heterogeneities and overall time trends, which is key in dealing with the non-random incidence of exposure to conflict and the macro-economic crisis driven by the rapid devaluation of the South Sudanese Pound over the later phase of the conflict. Given the repeated cross-section empirical setup this study relies on an external measure of conflict exposure, derived from geo-coded event data collected by the Armed Conflict Event & Location Data Project (ACLED). Given this estimation strategy, results are capturing the broader impact of residing in an area exposed to conflict and insecurity. This includes households which are directly subject to conflict or violent events like looting, as well as households which are not. Our study shows that the conflict led to a large decline in consumption levels and a corresponding increase in poverty across the entire country. However, households residing in areas that were exposed to more intense violence experienced greater declines in average consumption, higher poverty incidence as well as deeper poverty. The results are driven by households residing in areas exposed to high-intensity conflict related violence, proxied by total conflict fatalities.

This paper is structured as such: Section 2, the next section, describes the sources of data and the measures of welfare used in the study; Section 3 describes the estimation strategy and Section 4 presents the results; Section 5 discusses the results and concludes.

Data

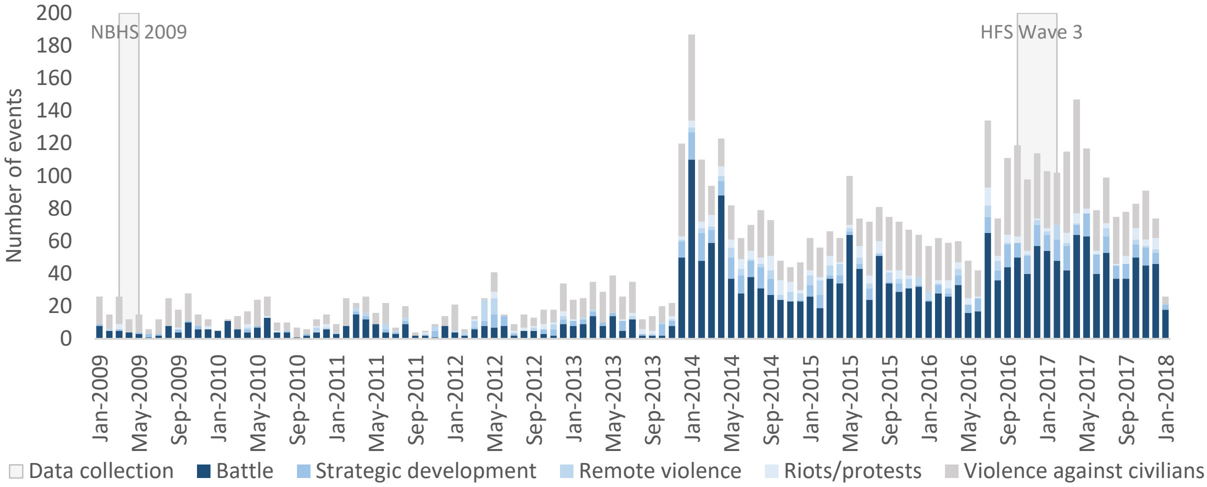
Household data

This study uses two waves of representative household surveys to build a repeated cross-section of households interviewed before and during the conflict (Figure C1-1). Pre-conflict exposure data is obtained from the National Baseline Household Survey (NBHS) in 2009 and post-conflict data is obtained from the High Frequency South Sudan Survey (HFS) in 2016-17 (Figure C1-1). Due to continued insecurity in the North-East of the country, the Greater Upper Nile region, the HFS could only be conducted across seven of the ten states of South Sudan, covering the Southern and Western regions of Greater Equatoria and greater Bahr El-Ghazal, respectively.[160] Pre-conflict exposure data is obtained from the National Baseline Household Survey (NBHS) in 2009. [161] Data from the states in Greater Upper Nile collected during the NBHS is thus excluded from the analysis. Overall, the final

[160] The HFS was designed to maintain a high degree of comparability with the NBHS. The sample frame used to select respondents is based on the same Census and most of the variables used in the analysis are directly comparable in terms of wording of answer choices.
[161] Despite the different name both surveys were implemented by the National Bureau of Statistics using virtually the same questionnaire to maximize comparability. Most relevant to this paper is the consumption module which included exactly the same set of items across the two surveys.

sample used in this study includes 5,296 households – 3,454 from the NBHS 2009 and 1,843 from the HFS 2016. Households are weighted using population weights representative at the state level.[162]

Figure C1-1: Conflict events and fatalities in South Sudan, 2011-2017.



As is South Sudan, the sample is primarily rural, almost 9 in 10 households reside in sparsely populated rural areas and about 8 in 10 rely on agricultural production as their primary source of livelihood. Poverty is high, in 2009 about 55 percent lived under the national poverty line, by 2016-17 this figure had jumped to 86 percent. A large majority of household consumption is accounted for by expenditures on food, 88 percent in 2016-17 (Table C1-1). The deprivation experienced by the South Sudanese encompasses multiple dimensions of well-being, and households are generally lacking in access to most amenities and services. Educational attainment is low, with more than two thirds of household heads having never had any education and less than 5 percent with any post-secondary education. Very few households have access to electricity, and most do not have access to adequate sanitation, a safe water supply, cooking fuels, quality housing, etc. (Table C1-1). Furthermore, in the many years between 2009 and 2016 most of these indicators of well-being hardly changed at all, highlighting the lasting impact of the protracted crises on economic and human development.

Conflict exposure

Indicators of conflict exposure used in this study are based on conflict data from the ACLED database. ACLED database records events of various types of conflict reported by different news media outlets.[163] The types of events covered range from battles between major actors, strategic developments and changes in territory, violence against civilians perpetrated by armed forces, spontaneous riots and protests, etc.. Each observation in the ACLED database consists of an event linked to a location and date. The database also contains additional information on the type of event, the actors involved, and a conservative estimate of the number of casualties. ACLED codifies the type of event for each observation but for the purposes of this study only violent events are considered, this includes battles, remote violence (i.e. IEDs, bombings, mortar attacks, etc.), and violence against civilians (all of which make up 92 percent of all events). Conflict event data is aggregated at the Payam level, equivalent to Admin level 3 (Figure C1-1).[164]

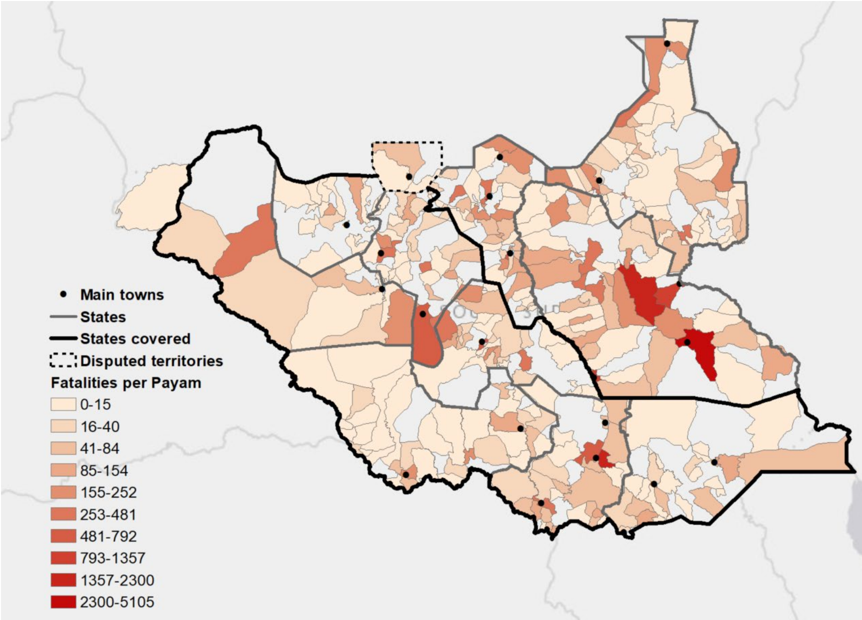
---

[162] The sampling design of the NBHS was based on the 2008 Census exercise and was stratified at the state level, which mapped exactly to the states of the future Republic of South Sudan. The HFS used the same sampling frame as the NBHS, despite the fact that the official administrative boundaries of South Sudan have gone through two phases of restructuring since then.

[163] With some emphasis placed on the outlet's reputation when compiling data.

[164] There are about 540 Payams in South Sudan and 279 in the states covered by the surveys. The reason for aggregating conflict events at the Payam administrative level is that even though each conflict event is geo-located the provided coordinates are sometimes inaccurate and would typically indicate the center of the nearest town/administrative capital. A distance-based measure of conflict exposure might therefore introduce bias. For example, presuming a violent event occurred near a certain town but the associated set of coordinates points to the center of this town, A distance-based measure of conflict exposure would assign a greater value of conflict exposure to an urban

Figure C1-2: Conflict related fatalities per Payam between Dec. 2013 and Feb. 2017.



Over the two waves of household surveys, about 58 percent of households interviewed reside in a Payam in which there had been at least one conflict event since the conflict began in December 2013. Most of the households interviewed in 2016-17 who were exposed to the conflict were exposed relatively recently, approximately 6 months on average, and about 80 percent were exposed within the last 12 months (Figure C1-4). However, the intensity of exposure varies across the regions and the distributions of both the number of events and the associated fatalities both are highly skewed with a long right tail. On average, conflict exposed households were exposed to about 8 conflict events, with this figure reaching upwards of 80 events in some of the main cities including Juba, Wau, and Yei (Figure C1-3).[165] In order to test whether low-intensity exposure has a different impact than high-intensity exposure, a categorical indicator of exposure is derived based on the median number of fatalities observed in conflict exposed Payams, equal to 26.

Measures of welfare

The primary indicator of welfare used in this analysis is households' total monetary consumption value of food and nonfood items per capita. Consumption aggregates are calculated based on standardized methodology detailed in Deaton and Zaidi (2002).[166] Given the large price variation in South Sudan between 2009 and 2016-17 the consumption aggregates are deflated spatially across urban and rural areas within each survey wave and then up to 2016 SSP using the consumer price index calculated by the National Bureau of Statistics.[167] It is not just the average consumption levels that are of interest when evaluating the consequences of conflict exposure but also the number of people who cannot achieve an adequate level of nutrition. Therefore, this study will also estimate impact of conflict exposure on the first two poverty indices from the Foster-Greer-Thorbecke (FGT) class of poverty measures (Foster, Greer et al. 1984), the poverty headcount and the poverty gap. FGT measures

household living in the center of the town relative to a nearby rural household. Therefore, although aggregating conflict exposure at the administrative level results in a general loss of accuracy, it does not introduce bias.

[165] A similar distribution is observed for conflict fatalities.

[166] There is an important caveat to consider with respect to the consumption measures used in this study. To facilitate fieldwork in the difficult context the HFS was designed to estimate consumption and poverty using the rapid consumption methodology. The rapid consumption methodology allows administering a shorter consumption module by randomly excluding some items from different households and estimate poverty and consumption through within-survey multiple imputations, thus reducing the time needed for each interview. Households are therefore not asked about the full set of 270 food and nonfood consumption items, but all households are asked about a consistent set of about 65 core food and nonfood consumption items specifically chosen to capture at least 80 percent of total consumption. In order to make consumption exactly comparable across the two surveys and abstain from any effect introduced by the imputation method only core consumption items are considered when calculating the consumption aggregate.

[167] In the HFS prices are deflated across both space and time within the wave given that fieldwork covered a relatively long time compared to the NBHS, i.e. per month and across urban and rural areas.

consist essentially of variations of the following specification, where the parameter α takes the value of 0 for the poverty headcount and 1 for the poverty gap:

$$FGT(\alpha) = \frac{1}{n}\sum_{i=1}^{p}\left[\frac{z - y_i}{z}\right]^{\alpha}$$

The poverty headcount, with $\alpha = 0$, reduces to the share of a population living under the poverty line. The poverty gap, with $\alpha = 1$, captures the average consumption deficit of the poor relative to the poverty line. Non-poor households are assigned a value of 0. The sum of the poverty gap across all individuals is a measure of the consumption deficit of the entire population relative to the poverty line.

The poverty line is based on the national poverty line derived calculated using the NBHS 2009 survey data, it is equal to 32 SSP per capita per day (2016), or approximately 2 USD PPP per capita per day (2016).[168] The poverty line is derived from a cost of basic needs approach, and is equivalent to the monetary value required to obtain a consumption basket that covers basic food and non-food consumption needs. The food component of the poverty line is equal to the monetary value require to achieve adequate nutrition, set to 2,400 calories per person per day. This value was calculated based on the average consumption bundle realized by the bottom 60 percent of the population in terms of real per capita consumption, which was then scaled proportionately to obtain the average price of consuming 2,400 calories. The non-food component of the poverty line was based on the consumption bundle of households living within 10 percent to the food poverty line. The guiding assumption is that if an individual is spending on food what has been determined as the minimum necessary to be healthy and to maintain certain activity levels, then this person is also likely to have acquired the minimum non-food goods and services to support this lifestyle.[169] We refer readers to the National Bureau of Statistics NBHS 2009 report for a more detailed treatment of the derivation of the poverty line.

Estimation

Estimation strategy

This study exploits two features of the data to estimate the impact of conflict exposure, (i) data availability before and after the conflict, and (ii) regional variation in conflict exposure. These features allow estimating a DID model of the impact of conflict exposure, which is key in dealing with non-random exposure to violent conflict, which is observed in South Sudan but that has also been documented in various contexts (Blattman and Miguel 2010). Table C1-1 provides some descriptive statistics from the sample and details differences across households within each survey wave and across conflict exposed and non-exposed regions. Conflict exposure is defined as a household residing in a Payam where there have been any conflict events during the study period. Poverty and deprivation are more widespread in non-exposed areas, non-exposed households are also more likely to reside in rural areas and obtain their livelihoods from agricultural production, and households heads are less likely to have received no education. The simple cross-sectional comparison of outcomes across exposure in 2016-17 would therefore yield counter-intuitive results, and from Table C1-1 we can see that there are no differences across conflict exposure in average consumption levels nor in the poverty headcount and poverty gap in 2016-17. Rather, although conflict exposed areas were better off prior to the conflict, the conflict would have induced a larger relative decline in outcomes.

By differencing across time and across groups DID estimates the relative changes in outcomes, thus netting out overall time trends and group specific initial levels, or group specific heterogeneity. The main specification takes the following form:

$$(1) \quad Y_{i,p,s,t} = \alpha Post_t + \gamma C_p + \beta(Conflict * Post)_{p,t} + \delta X_{i,s,p,t} + \varphi_{s,t} + \varepsilon_{i,s,p,t}$$

[168] Using the PPP conversion factor of 15.637 for 2016 obtained from the World Bank Development Indicators

[169] As per the RCS methodology the poverty line is scaled by 80 percent to reflect the lower number of consumption items considered.

Where $Y_{i,t}$ denotes the dependent variable for household i residing in Payam p, in state s, in period t.[170] $Post_t$ is a post-conflict binary variable which takes the value 1 for all households interviewed after the start of the conflict in 2016-17. $Conflict_p$ is a binary variable which takes the value of 1 for households residing in a conflict exposed Payam. Therefore, β, the coefficient on the interaction of the post-conflict and conflict-exposure indicators, is the coefficient of interest which describes the impact of conflict exposure. Standard error are estimated by bootstrap replication. Household specific control variables are denoted by $X_{i,t}$, Table C1-1 provides summary statistics for the set of control variables included. It is plausible that some of the controls may have been impacted by conflict exposure, therefore the full set of control variables is de-meaned and interacted with the conflict indicator so as not to introduce bias in the estimation of the average treatment effect on the treated when these characteristics are correlated the impact of conflict exposure. The specification also includes time trends for each state, denoted by $\varphi_{s,t}$.There are two main assumptions required by a cross-sectional DID empirical setup for the estimation to be identified: (i) that exposed and non-exposed households would have experienced the same overall trend in consumption were it not for conflict exposure, which allows attributing the differences in trends across to the two groups solely to conflict exposure; and (ii) that the make-up of the two groups did not change across time, which allows comparing outcomes across time within each group without necessarily observing the same households. Unfortunately, neither of these two assumptions can be formally tested. There is no data from South Sudan prior to the NBHS 2009 survey, rendering it impossible to check for pre-exposure trends nor estimating placebo effects. This assumption can nevertheless be made conditional on additional explanatory variables, for this reason the specifications include a comprehensive set of controls variables which can account for plausible factors expected to be correlated with any confounding factors that might drive differring trends across the groups. The estimates also include state-specific time trends that further account for unobservable time-varying characteristics across states.[171] Regarding possible changes in group composition across time, it is certain that the high levels of internal and external displacement raise some credible concerns.

[170] The natural logarithm of continuous consumption variables is taken, since we expect the estimated effect to be nonlinear in that it is likely proportional to households' level of consumption. Before taking the natural logarithm +1 is added to continuous measures in order to minimize the effect of low decimal values resulting in a long left-tail.

[171] Unfortunately time trends cannot be included at a finer level of disaggregation because of the requirement that both exposed and non-exposed households are observed within each fixed effect unit.

Table C1-1: Descriptive statistics by year and exposure status

	2009				2016-17			
Variable	**Overall**	**Control**	**Conflict**	**Difference (C-T)**	**Overall**	**Control**	**Conflict**	**Difference (C-T)**
Total weekly cons. pc	276.46	241.467	307.332	-65.866***	129.344	124.493	132.552	-8.059
(2016 SSP)	[307.854]	[256.169]	[349.154]		[133.595]	[126.591]	[138.349]	
Food weekly cons. pc	227.152	212.631	239.962	-27.331***	114.727	113.733	115.384	-1.652
(2016 SSP)	[235.704]	[219.767]	[249.099]		[122.472]	[122.637]	[122.405]	
Share of food	0.856	0.885	0.831	0.054***	0.886	0.898	0.878	0.020**
consumption	[0.235]	[0.224]	[0.241]		[0.173]	[0.138]	[0.192]	
Non-food weekly cons. pc	49.309	28.836	67.37	-38.535***	14.617	10.76	17.167	-6.407***
(2016 SSP)	[153.721]	[113.655]	[183.812]		[29.384]	[15.018]	[35.758]	
Share of non-food	0.144	0.115	0.169	-0.054***	0.114	0.102	0.122	-0.020**
consumption	[0.235]	[0.224]	[0.241]		[0.173]	[0.138]	[0.192]	
Poverty headcount	0.555	0.603	0.513	0.090***	0.862	0.883	0.848	0.035
	[0.619]	[0.609]	[0.625]		[0.495]	[0.400]	[0.544]	
Poverty gap	0.269	0.296	0.245	0.051***	0.484	0.498	0.476	0.022
	[0.423]	[0.426]	[0.416]		[0.443]	[0.368]	[0.483]	
Controls								
Urban	0.172	0.065	0.266	-0.201***	0.127	0.088	0.153	-0.065***
	[0.374]	[0.169]	[0.483]		[0.336]	[0.250]	[0.389]	
Household size	7.714	7.7	7.726	-0.027	7.287	7.163	7.368	-0.205
	[4.843]	[4.734]	[4.922]		[5.003]	[3.432]	[5.727]	
Total number of rooms	2.763	2.652	2.86	-0.207**	2.533	2.292	2.693	-0.401***
	[2.457]	[2.393]	[2.497]		[2.439]	[1.557]	[2.823]	
Livelihood: Own-account	0.779	0.834	0.514	0.320***	0.825	0.859	0.681	0.178***
agriculture	[0.483]	[0.431]	[0.655]		[0.465]	[0.408]	[0.650]	
Livelihood: Wage labour /	0.139	0.097	0.342	-0.245***	0.126	0.101	0.23	-0.129***
own business	[0.375]	[0.320]	[0.585]		[0.393]	[0.344]	[0.564]	
Livelihood: Remittances /	0.082	0.069	0.144	-0.074***	0.049	0.04	0.089	-0.049***
Aid / Other	[0.342]	[0.315]	[0.460]		[0.254]	[0.221]	[0.381]	
Toilet is a latrine	0.23	0.129	0.32	-0.192***	0.277	0.16	0.353	-0.193***
	[0.504]	[0.386]	[0.573]		[0.613]	[0.409]	[0.722]	
Toilet is a flush toilet	0.011	0.005	0.016	-0.011***	0.002	0.001	0.003	-0.002
	[0.127]	[0.063]	[0.169]		[0.038]	[0.017]	[0.046]	
No toilet	0.759	0.866	0.664	0.202***	0.721	0.839	0.644	0.195***
	[0.514]	[0.390]	[0.583]		[0.614]	[0.409]	[0.724]	
Watersource: Borehole	0.199	0.165	0.23	-0.065***	0.175	0.119	0.212	-0.093***
	[0.508]	[0.475]	[0.534]		[0.632]	[0.393]	[0.728]	
Watersource: Hand pump	0.357	0.383	0.335	0.048**	0.491	0.523	0.47	0.053
	[0.591]	[0.599]	[0.578]		[0.774]	[0.647]	[0.843]	
Watersource: Open water	0.393	0.414	0.375	0.039*	0.232	0.261	0.214	0.047*
/ none	[0.628]	[0.633]	[0.617]		[0.600]	[0.570]	[0.610]	
Has access to electricity /	0.039	0.022	0.055	-0.033***	0.028	0.007	0.042	-0.035***
solar power / gas	[0.205]	[0.131]	[0.257]		[0.194]	[0.114]	[0.231]	
Cooks using firewood	0.887	0.941	0.84	0.100***	0.869	0.923	0.833	0.089***
	[0.333]	[0.232]	[0.403]		[0.452]	[0.295]	[0.527]	
Dwelling type: Traditional	0.672	0.655	0.687	-0.032	0.662	0.621	0.689	-0.068**
mud hut	[0.604]	[0.608]	[0.595]		[0.735]	[0.643]	[0.781]	
Dwelling type:	0.266	0.306	0.231	0.075***	0.281	0.336	0.245	0.091***
Wood/straw house	[0.582]	[0.600]	[0.553]		[0.721]	[0.637]	[0.764]	
Dwelling type:	0.062	0.039	0.082	-0.043***	0.057	0.043	0.066	-0.023*
Concrete/other	[0.274]	[0.194]	[0.333]		[0.265]	[0.220]	[0.290]	
HH head education: No	0.731	0.789	0.679	0.110***	0.658	0.774	0.581	0.192***
education	[0.556]	[0.522]	[0.581]		[0.713]	[0.534]	[0.803]	
HH head education:	0.168	0.154	0.18	-0.025	0.198	0.133	0.242	-0.109***
Primary	[0.478]	[0.474]	[0.480]		[0.595]	[0.425]	[0.676]	
HH head education:	0.08	0.05	0.107	-0.058***	0.104	0.069	0.128	-0.059***
Secondary	[0.323]	[0.253]	[0.377]		[0.459]	[0.341]	[0.518]	
HH head education: Post-	0.021	0.007	0.034	-0.027***	0.04	0.025	0.049	-0.024**
secondary	[0.173]	[0.107]	[0.220]		[0.242]	[0.175]	[0.275]	
N	3454	1478	1976		1843	772	1071	

The estimation can be extended to deal with these issues. More specifically, the specifications can be estimated using DID propensity score matching estimators. Matching estimators effectively account for selection bias by estimating the likelihood of belonging to each treatment group and using the estimated propensity to match individual observations. The matching process helps to identify observations across groups that can be used to creating a more adequate counterfactual. Matching-DID estimators can help relax the common trends assumption under the assumption that a well-matched groups of exposed and non-exposed households are also more likely to follow common trends, in the absence of conflict exposure. Furthermore, this estimation goes one step further and also matches households across the two time periods, which helps to account for the possible selection effect of migration. This study implements a matching estimator, where each group of exposed and non-exposed households by time period are matched to the post-conflict exposed group. The propensity score is estimated using a logit regression of the binary variable indicating conflict exposure on the full set of control variables mentioned in the previous section, the regression results are presented in Table C1-5. Households are matched across the periods and treatment groups using an epachenikov kernel function with a bandwidth of 0.06. Kernel matching is advantageous in our context because of the relatively small sample size because it does not require searching for one to one matches and can thus make use of more information. The density of the obtained propensity scores shows that there is ample common support in terms of propensity scores between the two groups, which alleviates some concerns regarding the extent of differences between the exposed and non-exposed groups (Figure C1-6). Only observation that fall within this common support are used for the matching estimation.

Results

Baseline results

Table C1-2, reports the DID coefficients from the estimation of the full specification 1, using simple DID and matching-DID, on log total consumption, the poverty headcount, and the poverty gap. The results shown in columns 1 report the results from regressing specification 1 on log total consumption. The estimated DID coefficients indicate an additional decline of about $\left(e^{\beta_{DD}} - 1\right) = -13$ percent for households residing in conflict affected areas relative to households residing in non-exposed areas. The estimated impact is statistically significant at the 5 percent level and robust to the inclusion of control variables and their interaction with the conflict indicator as well as the state-specific time trends. The estimated impact of conflict exposure on total consumption from the Matching-DID estimator, reported in column 2, indicates a similar effect of about 12 percent. However, it is statistically significant only at the 10 percent level. Taken together, these estimates provide some evidence for an additional impact of conflict exposure on average consumption levels. The regressions results shown in columns 3 to 6 of Table C1-2, on poverty and the poverty gap, indicate that the decline in average consumption is mirrored by a corresponding increase in the poverty headcount but the results are slightly less consistent. Poverty in conflict exposed Payams was 5 to 6.5 percentage points higher than in non-exposed Payams, as per columns 3 and 4 of , although the DID estimate is only statistically significant at the 10 percent level. The impact of conflict exposure on the poverty gap is of about 3.4 to 4.6 points, as shown in columns 5 and 6 of Table C1-2, indicating that the impact of the conflict not only led to a decline in average consumption and pushed more people into poverty, but it also pushed already poor households further into poverty. Although, in this case the estimated coefficient is only statistically significant for the matching-DID estimator.

Table C1-2: Regression results, baseline estimation.

	Ln Total Cons.		Poverty headcount		Poverty gap	
	DID	*Matching-DID*	*DID*	*Matching-DID*	*DID*	*Matching-DID*
	(1)	(2)	(3)	(4)	(5)	(6)
DID coefficient	-0.140**	-0.123*	0.0505*	0.0647**	0.0341	0.0462**
	(0.0651)	(0.0645)	(0.0294)	(0.0307)	(0.0237)	(0.0225)
Observations	5,297	5,080	5,297	5,080	5,297	5,080
Controls	YES	YES	YES	YES	YES	YES
Interactions	YES	YES	YES	YES	YES	YES
State time trends	YES	YES	YES	YES	YES	YES

*** p<0.01; ** p<0.05; * p<0.1; All estimates are weighted by population weights, bootstrapped standard errors in parentheses with 250 replications; total consumption is deflated to real 2016 SSP; control variables include: whether the household resides in an urban/rural area, household size, total number of rooms in the dwelling, gender of the household head, household head's highest educational attainment, household head age, the households' main source of livelihood, toilet type, water source, and the source of energy for lighting and cooking; all controls are fully interacted with the conflict exposure dummy; Matching-DID regressions are run on the observations within the range of common support of the propensity score and thus have fewer observations.

Accounting for the intensity of exposure

The intensity of exposure differed greatly across the country and it is likely that the impact of conflict exposure will vary with the intensity of exposure. Therefore, the estimation is extended to account for the intensity of exposure. Conflict exposure is classified into a low- and high-intensity of exposure categories, defined as having there been a total number of conflict fatalities in a Payam that is higher or lower than the median sum of conflict related fatalities, respectively.[172] Conflict fatalities are used based on the logic that an event where greater fatalities were incurred is likely to be more destabilizing than several events where no fatalities are incurred. The sum of conflict fatalities is strongly correlated with the number of conflict events, with a correlation coefficient of 0.68. The differential impact of conflict exposure at each level of intensity can then be estimated as such:

$$(2) \quad Y_{i,s,p,t} = \alpha Post_t + \gamma_j \sum_{j=L,H}^{2} Conflict_p + \beta_j \sum_{j=L,H}^{2} (Conflict * Post)_{p,t} + \delta X_{i,s,p,t} + \varphi_{s,t} + \varepsilon_{i,s,p,t}$$

Accounting for the intensity of exposure in this manner allows for a non-linear marginal impact of intensity on the impact of conflict exposure. The matching-DID estimates are obtained in a similar manner but not exactly, because we are interested in the pairwise comparison of each group of households exposed at the different levels of conflict intensity with the non-exposed group. Therefore, the specification 1 from the previous section is estimate twice, each time exactly as in the previous but excluding one of the two groups of conflict exposed households. Table C1-3 reports DID coefficients from the estimation of specification 2 on log total consumption, the poverty headcount, and the poverty gap. The *Matching-DID* columns, 2, 4, and 6, report the coefficients from two separate regressions of the impact of conflict exposure for the low-intensity exposed households against the control group and the high intensity exposed households in the same column.

The results shown in Table C1-3 indicate that the impact of exposure on consumption and poverty was largely driven by areas that were affected by high intensity conflict. The DID coefficients on log total consumption, shown in columns 1 and 2, imply an additional decline of about 25 percent relative to non-exposed households. Similarly, the poverty headcount in exposed Payams was higher by 11-14 percentage points and the poverty gap by 5-9 points. All of these effects are strongly statistically significant at the 5 or 1 percent level. Meanwhile, the DID coefficients on low-intensity exposure are of a much smaller magnitude and thus almost all statistically insignificant. This provides relatively strong evidence that the differential impact of conflict exposure observed in the previous section was

[172] The median is calculated only counting Payams that experienced at least one conflict event.

almost entirely driven by high-intensity exposure. Insecurity and macroeconomic disruptions would have affected the control group, however, the finding that relatively low-level violence does is somewhat surprising. In part, these findings may be due to power concerns, given that conflict exposure is measured at a relatively high level of aggregation, Payams, so that there is relatively little variation with which to estimate effects. Furthermore, all the coefficients have the right sign, in that average consumption seems to decline and the poverty headcount and gap increase, it is just that they are statistically insignificant.

Table C1-3: Regressions results, by level of intensity of exposure.

	Ln Total Cons.		*Poverty headcount*		*Poverty gap*	
	DID	*Matching-DID*	*DID*	*Matching-DID*	*DID*	*Matching-DID*
	(1)	(2)	(3)	(4)	(5)	(6)
Low-intensity DID	-0.0726	-0.0718	0.0476	0.00780	0.0367	0.0433**
	(0.0703)	(0.0600)	(0.0320)	(0.0307)	(0.0253)	(0.0215)
High-intensity DID	-0.284***	-0.291***	0.114***	0.141***	0.0533**	0.0878***
	(0.0807)	(0.0818)	(0.0339)	(0.0325)	(0.0269)	(0.0271)
Observations	5,297	3,474/3,512	5,297	3,474/3,512	5,297	3,474/3,512
Controls	YES	YES	YES	YES	YES	YES
Interactions	YES	YES	YES	YES	YES	YES
State time trends	YES	YES	YES	YES	YES	YES

**** p<0.01; ** p<0.05; * p<0.1; All estimates are weighted by population weights, bootstrapped standard errors in parentheses with 250 replications; total consumption is deflated to real 2016 SSP; control variables include: whether the household resides in an urban/rural area, household size, total number of rooms in the dwelling, gender of the household head, household head's highest educational attainment, household head age, the households' main source of livelihood, toilet type, water source, and the source of energy for lighting and cooking; all controls are fully interacted with the conflict exposure dummy; Matching-DID regressions are run on the observations within the range of common support of the propensity score and thus have fewer observations; the matching-DID regressions are run separately for each group of conflict exposed households with different levels of intensity against the control group.*

Robustness

This section performs a few additional robustness checks. In addition to the matching estimators, another means of testing for robustness to the indentifying assumptions of DID is to estimate the specifications on specific subsamples for which the assumption is more likely to hold. This is done in several ways ways in this estimation: Firstly, the sample is split across urbanicity into urban and rural samples. One of the primary confounding factors and possible cause for diverging trends is that urban households rely on a different mix of economic activities, primarily in that they are less likely to rely on own-account agricultural production and thus purchase food in markets. The reduced ability to produce their own consumption makes these households more susceptible to the economic crisis and the sharp devaluation of the domestic currency. The DID coefficient for the rural subsample, shown in column 1 of Table C1-4, is very similar to the main estimates, largely because 85 percent of the sample resides in rural areas – after being weighted for representativeness. However, the impact is smaller and no longer significant within the urban subsample only, shown in column 1 of Table C1-4. Nevertheless, the effect remaining within the rural only sample is significant evidence for the common trends assumption holding. The same can be done to include only households who rely primarily on their own production for their livelihoods, where once again the estimates remain relatively similar, column 3 of Table C1-4.

Table C1-4: Robustness checks

		Urban	Rural	Agricultural	Overlapping Payams	Excl. IDPs	Price index
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Binary conflict indicator (at least one conflict event)							
Total consumption	DID	-0.156	-0.194***	-0.184***	-0.166**	-0.236***	-0.237***
		(0.0953)	(0.0694)	(0.0706)	(0.0695)	(0.0684)	(0.0658)
Poverty headcount	DID	0.0719	0.0774***	0.0774***	0.0605*	0.0805***	0.0893***
		(0.0509)	(0.0295)	(0.0300)	(0.0309)	(0.0285)	(0.0279)
Poverty gap	DID	0.0367	0.0596**	0.0539**	0.0391*	0.0624***	0.0633***
		(0.0329)	(0.0242)	(0.0252)	(0.0234)	(0.0233)	(0.0221)
Observations		1,751	3,546	3,633	3,388	5,039	5,297
Controls		YES	YES	YES	YES	YES	YES
Interactions		YES	YES	YES	YES	YES	YES
State-urban time trends		YES	YES	YES	YES	YES	YES
Panel B: Indicators of conflict intensity							
Total consumption	DID low-intensity	-0.0376	-0.140*	-0.159**	-0.0827	-0.141*	-0.129*
		(0.137)	(0.0732)	(0.0768)	(0.0739)	(0.0769)	(0.0751)
	DID high-intensity	-0.160*	-0.319***	-0.257***	-0.261***	-0.373***	-0.390***
		(0.0953)	(0.0888)	(0.0921)	(0.0784)	(0.0823)	(0.0781)
Poverty headcount	DID low-intensity	0.0157	0.0672**	0.0801**	0.0219	0.0495	0.0565*
		(0.0790)	(0.0312)	(0.0332)	(0.0363)	(0.0329)	(0.0317)
	DID high-intensity	0.0745	0.106***	0.0790**	0.104***	0.126***	0.137***
		(0.0505)	(0.0384)	(0.0388)	(0.0340)	(0.0343)	(0.0339)
Poverty gap	DID low-intensity	0.0123	0.0593**	0.0606**	0.0338	0.0544**	0.0504*
		(0.0508)	(0.0256)	(0.0275)	(0.0275)	(0.0268)	(0.0258)
	DID high-intensity	0.0373	0.0713**	0.0515	0.0448*	0.0764***	0.0838***
		(0.0328)	(0.0298)	(0.0330)	(0.0246)	(0.0269)	(0.0252)
Observations		1,751	3,546	3,633	3,388	5,039	5,297
Controls		YES	YES	YES	YES	YES	YES
Interactions		YES	YES	YES	YES	YES	YES
State-urban time trends		YES	YES	YES	YES	YES	YES

*Notes: *** p<0.01; ** p<0.05; * p<0.1; Each set of cells in the table presents results from a different regression where the dependent variable is indicated by the row, i.e. total consumption, the poverty headcount, and pverty gap; all estimates are weighted by population weights, bootstrapped standard errors in parentheses with 250 replications; total consumption is deflated to real 2016 SSP; control variables include: whether the household resides in an urban/rural area, household size, total number of rooms in the dwelling, gender of the household head, household head's highest educational attainment, household head age, the households' main source of livelihood, toilet type, water source, and the source of energy for lighting and cooking; matching estimator matches households based on their propensity score and a kernel matching estimator; urban and rural columns inclue only urban and rural households, respectively; agricultural livelihood columns only includes households whose primary source of livelihood is own account agricultural production; overlapping Payams column only includes households residing in a Payam that figures in both the 2009 and 2016-17 surveys; Excluding IDPs column excludes households in 2016-17 who moved into their current place of residence after the December 2013 conflict; Price index column includes the price index as an additional control variable.*

In order to further control for high inflation and growing prices driving the impact of conflict exposure, we can partly control for price dispersion by adding a Payam-specific price index, calculated as a Laspeyres price index which reflects the item-weighted relative price differences across products.

$$L_{i,t} = \sum_{k=1}^{k} w_{i,k} \left(\frac{p_{i,k}}{p_{0,k}} \right)$$

The Laspeyres $L_{i,t}$ for Boma i in period t is equal to the sum of, over all items k: $w_{i,k,m}$, the Boma average budget share of item k, times the ratio of $p_{i,k,t}$, the median price of item k in Boma i at month t, and $p_{0,k,0}$, the median price of item k in the reference strata in the reference period, which is the overall urban consumption basket and prices in 2009. Price dispersion is much greater in 2016, as is evidence by plotting the density of the price index per survey wave (Figure C1-8). However, the price index is not statistically significantly different between conflict exposed and non-conflict exposed areas in either period. The regressions including the price index as an additional control are shown in column 6 of Table C1-4, and again they de not qualitatively differ. Finally, the sample can be further trimmed down by excluding all Payams that do not appear in both waves of the sample, in case we are concerned that there may have been issues with the sampling in the 2016-17 surveys, shown in column 4 of Table C1-4. Including only these payams does not qualitatively affect the results although the impact is smaller and less strongly statistically significant, likely due to the loss of power. As noted in the previous sections there was also a significant amount of displacement in South Sudan due to the conflict. Migration and displacement can confound our estimates because they are likely to be correlated with conflict exposure. Namely, households who were exposed to the conflict would have relocated to non-conflict affected areas, which would result in an underestimation of the impact of conflict exposure given that displaced households are likely to have lower outcomes than permanent residents. In order to test the sensitivity of our results to migration we drop from the sample all households interviewed in 2016-17 that have moved into their current place of residence after the beginning of the conflict in December 2013 from outside their current county. This leads to the removal of 258 households, the majority of which are living in areas exposed to violent conflict (184, or 72 percent). There is no detectable difference in consumption levels of IDPs and permanent residents, overall nor within control or conflict affected areas. Although this might be due to the small sample size and low statistical power. Nevertheless, removing these households does not qualitatively affect the results, as shown by the results in column 5 of Table C1-4.

Discussion

The incidence of poverty in South Sudan during the conflict in 2016-17 is extremely high, where about 86 percents of households are estimated to live under the poverty line derived for this study – approximately equivalent to $2 USD PPP per capita per day. An increase in the poverty headcount of about 5 percentage points associated with conflict exposure does not therefore seem like a profound impact, and although the conflict clearly increased its' prevalence, insecurity at large already had a strong impact on deprivation. However, again the average poverty gap in South Sudan in 2016-17 was of about 48 points, meaning that the average poor household's deficit in consumption relative to the poverty line was equal to 48 percent of the poverty line.

Appendix

Figure C1-3: Number of conflict events per Payam.

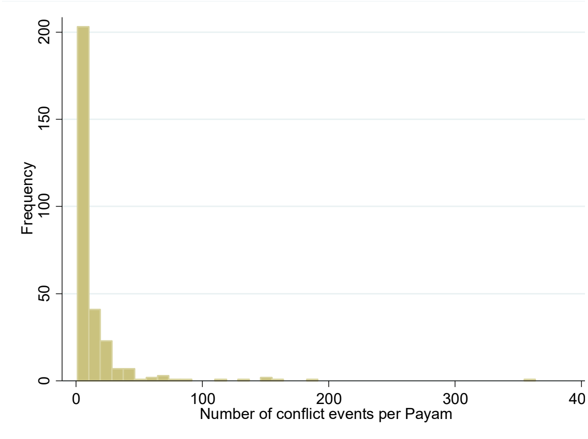


Figure C1-4: Months since the Last conflict event and date of interview.

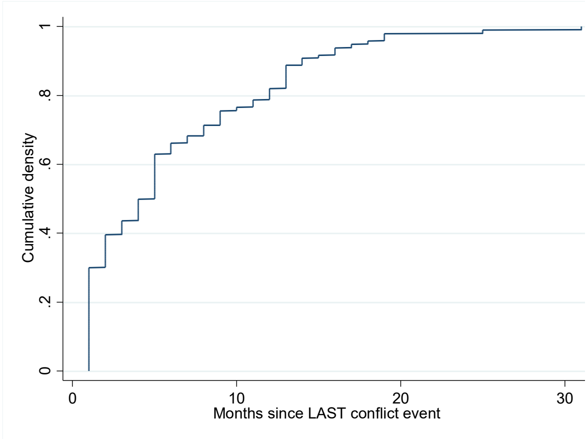


Figure C1-5: Cumulative density of the number of conflict events.

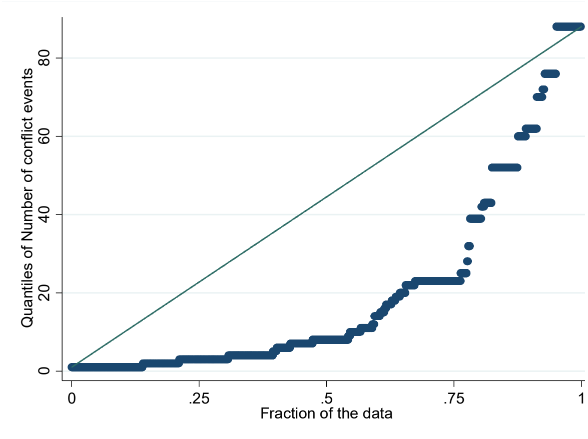


Table C1-5: Logit regression results on conflict indicator to estimate propensity score.

Conflict exposure dummy	(1) NBHS 2009	(2) HFS 2016-17
Household size	-0.0376*	-0.0105
	(0.0214)	(0.0387)
Rooms used for sleeping	0.0232	0.112
	(0.0545)	(0.0996)
HH head age	0.00213	-0.0169***
	(0.00380)	(0.00573)
HH head sex	0.326*	-0.377*
	(0.173)	(0.204)
livelihood==Remittances/Aid/Other	Base	
livelihood==Agriculture	-0.888***	-0.418
	(0.241)	(0.472)
livelihood==Wages	0.0764	-0.711
	(0.256)	(0.495)
livelihood==Own business	0.256	0.0561
	(0.283)	(0.673)
toilet==None	Base	
toilet==Latrine	0.861***	0.475
	(0.269)	(0.375)
toilet==Flush	0.639	1.185
	(0.549)	(1.282)
water source==Purchased/Other	Base	
water source ==Borehole	-0.215	0.654*
	(0.332)	(0.372)
water source ==Hand pump	-0.387	0.239
	(0.310)	(0.471)
water source ==Open water	-0.342	-0.133
	(0.282)	(0.399)
lighting==Firewood/grass/none	Base	
lighting==Electricity/solar/gas	0.0828	0.709***
	(0.310)	(0.267)
lighting==Paraffin/wax	0.0745	1.788***
	(0.193)	(0.457)
cooking==Firewood	Base	
	-	
cooking==Charcoal/other	-0.0722	0.0113
	(0.227)	(0.330)
dwelling==Concrete/other	Base	
dwelling==Mud house	-0.105	0.600
	(0.220)	(0.376)
dwelling==Wood/straw house	-0.348	0.228
	(0.266)	(0.394)
HH Head education==Post-secondary	Base	
HH Head education==No education	-0.991**	-0.0500
	(0.421)	(0.423)
HH Head education==Primary	-0.997**	0.297
	(0.431)	(0.433)
HH Head education==Secondary	-0.733*	0.182
	(0.433)	(0.412)
Observations	3,454	1,843

All estimates are weighted by population weights, standard errors estimated through linear regression in parentheses; *** p<0.01; ** p<0.05; * p<0.1.

Figure C1-6: Density of estimated propensity score by conflict exposure status.

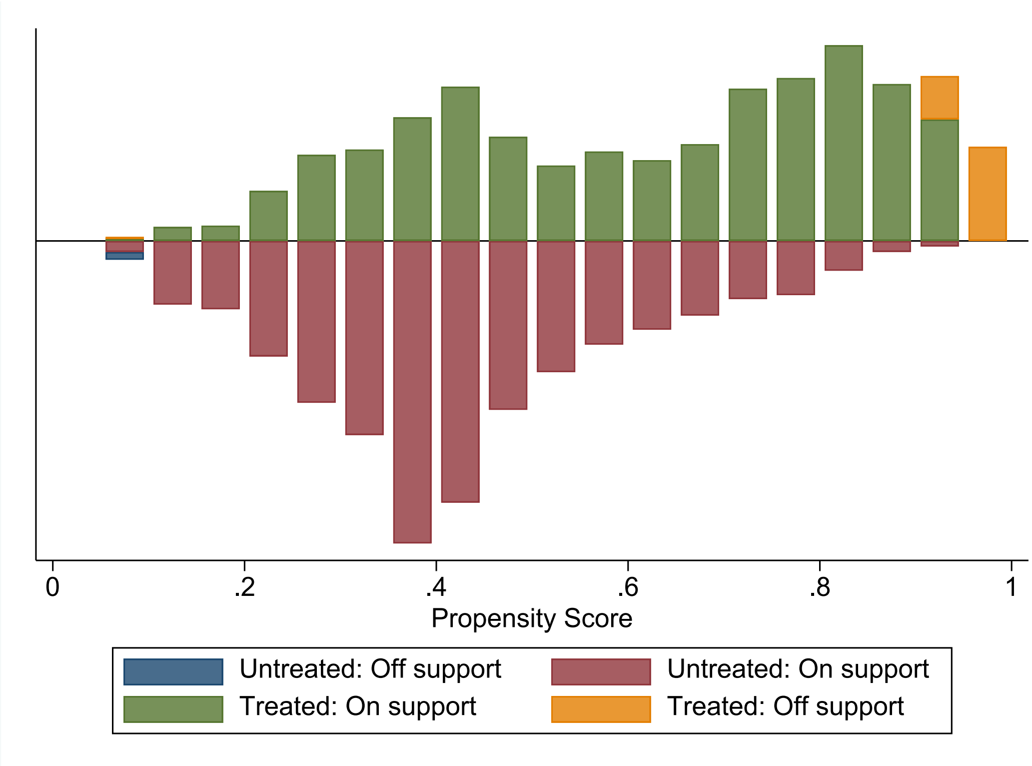


Figure C1-7: Density of estimated propensity score by conflict exposure status: Low-intensity exposure.

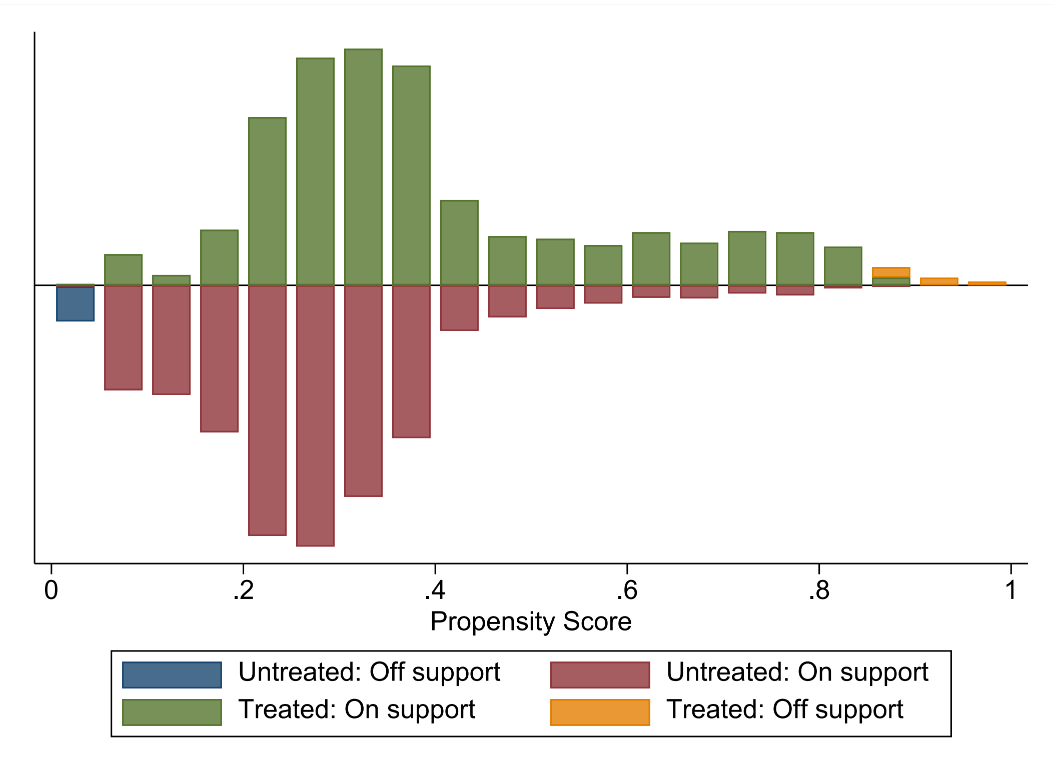


Figure C1-8: Density of estimated propensity score by conflict exposure status: High-intensity exposure.

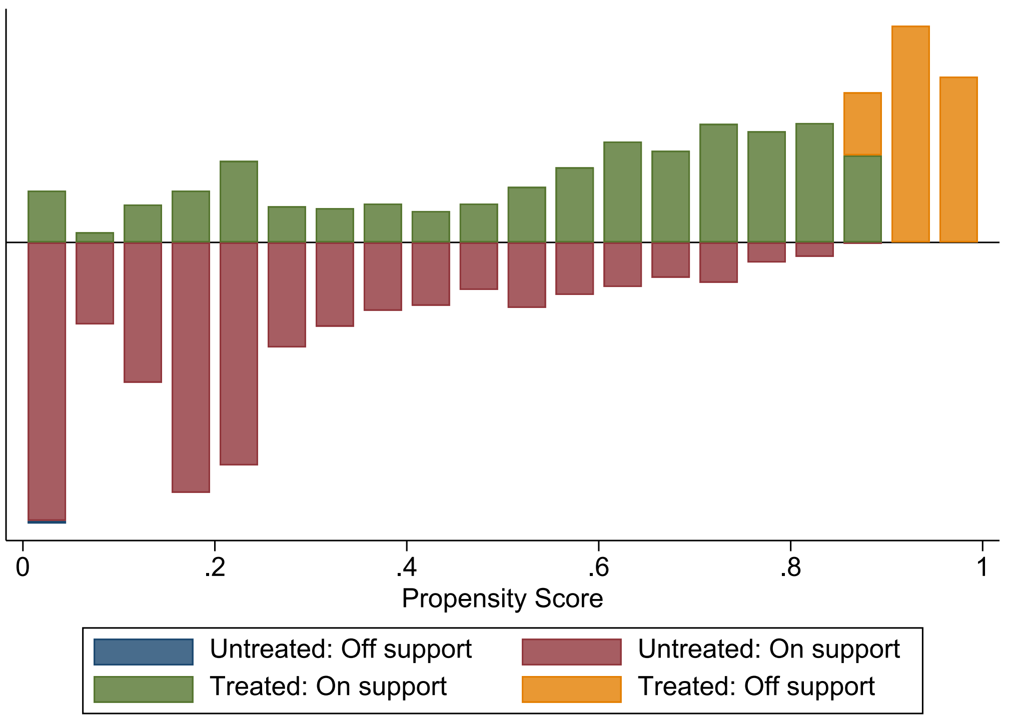
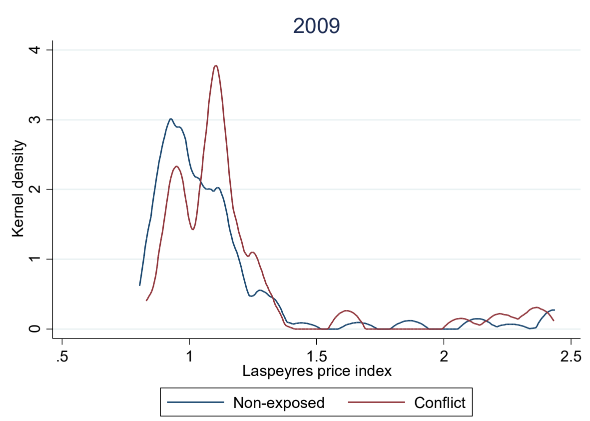
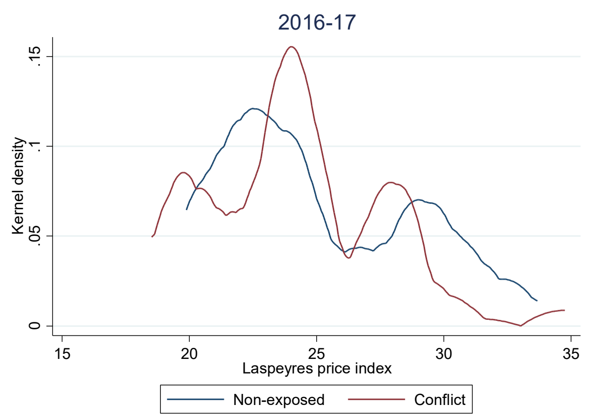


Figure C1-9: Kernel density of Laspeyres price index per survey wave.





2. Impact of conflict on adolescent girls in South Sudan[173]

Utz Pape and Verena Phipps[174]

Introduction

Conflict and displacement escalated dramatically after the civil war in South Sudan in December 2013. The December 2013 conflict between President Salva Kiir and former Vice President Riek Machar quickly became an ethnically-charged conflict particularly between the Dinka and Nuer ethnic groups. Skirmishes as well as brutal violence against civilians were reported in dozens of locations. In the days following the start of the conflict, incidences were more isolated with violence against Nuer civilians in Juba, attacks by Nuer on Dinka and other civilians in these areas as well as incidences of armed groups of different ethnic backgrounds launching revenge attacks on community members. The civil war with high rates of violence resulted in high mortality and displacement, as well as worsening livelihoods, poverty and food insecurity (Shankleman 2011, World Bank 2014, World Bank 2015, World Bank 2015).

More than 50,000 civilians have been killed since the resurgence of conflict in December 2013, in addition to various severe crimes including extrajudicial killings, abductions, rape, and torture. More than 2.2 million people have also fled the country or have been displaced internally, and it is believed that 4.8 million are at risk of famine (FAO 2017). The conflict has severely impacted welfare indicators and cost the country an estimated 6.3 percent of its GDP (World Bank 2016).

Violent conflict and instability affect men and women in heterogeneous ways, including differentiated impacts on economic, social, physical and mental well-being. Research highlights that men and boys often confront direct, first-round effects of conflict, including death and morbidity, while conflict contributes to indirect impacts on women and girls, including as related to health, e.g. to malnutrition, exposure to disease and lack of access to health services (Buvinic, Das Gupta et al. 2012). Children's health and access to education are often severely affected by exposure to conflict.

In many countries, women and children frequently account for the majority of populations displaced by conflict; in South Sudan for example, 53 percent of the 2.43 million externally displaced due to the 2013 conflict are female while 63 percent of those displaced are children under the age of 18 (UNHCR 2018). While displacement generally contributes to a critical loss in assets, including housing, land and property and other productive assets, women confront particular constraints extending from social norms that restrict women's ownership rights over land and other assets, and contributes to their exclusion from decision-making processes (Cagoco-Guiam 2013). Displacement also often gives rise to or exacerbates serious protection challenges including increased exposure to gender-based violence.

Violent conflict often changes the demographic composition of households, contributing to a rise in female-headed households due to the extended absence of males either due to conflict or abnormal migration. These shifts impact traditional gendered division of tasks through its impacts on household composition, often increases women's participation in labor markets and augmenting responsibilities of women within households (Annan, Blattman et al. 2009, Brück and Schindler 2009, Brück and Vothknecht 2011, Justino, Ivan et al. 2012, Menon and Rodgers 2013). At the same time, data on whether women's greater market participation and shifts in household responsibilities contributes to wider welfare gains and long-term social empowerment, however, is more ambiguous (Bozzoli, Brück

[173] UP and VP contributed equally to the manuscript.

[174] Corresponding author: Utz Pape (upape@worldbank.org). The authors thank Jana Bischler, Niklas Buehren, Shubha Chakravarty, Menaal Ebrahim and Rachel Firestone for their contributions. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

et al. 2011, Justino, Ivan et al. 2012). There are data to suggest that the economic and social gains women may have achieved due to the absence of men during conflict periods can erode during post-conflict periods due to a reversion in pre-conflict norms and do not always result in a comparable increase in social empowerment or improved bargaining power (Justino 2009). Non-material well-being, such as marriage outcomes and happiness, has also been negatively impacted by conflict and displacement in some cases (Wang and Weina 2016). Robust evidence also exists on the positive correlation between rates and incidence of varying forms of gender-based violence (GBV) (including sexual and physical assault, intimate partner violence, trafficking and early and forced marriage) and exposure to conflict (Annan, Blattman et al. 2009, Dijkman, Catrien et al. 2014, Ostby 2016). Lastly, studies have found that women are more vulnerable to developing anxiety disorders and struggling with psychosocial distress in conflict-affected settings (Murthy and Lakshminarayani 2006, Roberts, Ocaka et al. 2008, Farhood and Dimassi 2012, Luitel, Jordans et al. 2013, Ayazi, Lien et al. 2014).

The devastating nature of the recent conflict in South Sudan and the grim reality of its gendered effects provides the motivation for this study. The conflict has affected millions of South Sudanese people but the effects of this conflict on a particularly vulnerable group, such as adolescent girls, are worth identifying. Economic, social, and mental impacts at an early age tend to be long-lasting and should be addressed before they worsen and persist. Therefore, this paper aims to measure the impact of this conflict on adolescent girls across a set of welfare indicators to inform and guide appropriate intervention strategies.

There is growing consensus that studying conflict cannot be dissociated from how it is experienced and perceived by individuals affected by armed violence. Econometric research on the various channels through which conflict affects women, however, and the impact of conflict on gender dynamics is relatively nascent (Ibáñez, Calderón et al. 2011, Justino, Ivan et al. 2012). Within the literature on the intersection between conflict and gender dynamics, there is scant research on non-combatant adolescent girls. This study contributes to this literature by offering one of the first efforts to empirically quantify the impact of violence and conflict on educational attainment, labor market behavior, and social empowerment for non-combatant adolescent girls.[175]

This study utilizes survey data emerging from a World Bank-administered pilot project in South Sudan to contribute to the existing conflict and gender literature on several fronts. First and foremost, it uses a cluster-level difference-in-difference analysis to identify the impact of the conflict in South Sudan on girls aged 15-24. Given the high levels of mobility in South Sudan, these surveys are repeated cross-sections. Second, the study contributes new knowledge on the impact of conflict on welfare, poverty, and aspirations by offering one of the first analyses of data on adolescent girls, a generally under-researched demographic. Finally, this research contributes to a growing body of evidence examining the impacts of earlier-life environment on later life outcomes, and is closely related to a large body of literature on subjective perceptions of well-being linked to significant and potentially traumatic life events.

The analysis builds on two rounds of survey data that were collected for an impact evaluation of an adolescent girls' program. The first round of data was collected between August and October 2010 and the second round of data was collected between January and February 2015. The two surveys measure the same indicators except that the endline survey has an additional module on conflict exposure. We use the data from the conflict exposure module to obtain self-reported measures of

[175] In this conflict, adolescent girls and young women did not constitute a significant number of participating combatants.

cluster-level exposure to the conflict and examine the impact of conflict-related victimization on adolescent girls. For robustness, we also use external data on conflict events to examine the impact of the conflict exogenously. This analysis tests the hypothesis that girls exposed to the conflict had statistically different welfare outcomes than girls who were not exposed to the conflict.

The remainder of the paper is organized according to the following sections: Section 2 details literature on micro-level conflict studies and the impact of conflict on women; Section 3 provides a description of the data, followed by a description of the conflict data and how conflict treatment variables were constructed; Section 4 provides the theoretical model and Section 5 reports results from the difference-in-difference regression analysis. Finally, Section 6 provides a discussion and conclusion for the study.

Related Literature

From civil wars to riots and violent mass protests, past and present violent conflicts result in lost opportunities for human and economic development and have significant effects on the welfare, resilience and behavior of individuals, households and communities. Due in part to security studies' traditional focus on the state and state agency, research on violent conflict has until recently relied mostly on standardized macro-level measures of conflict such as the number of battle deaths per country per year. Over the last 15 years, initiatives such as the Households in Conflict Network (HiCN), housed within the Institute of Development Studies at the University of Sussex, and MicroCon (Micro Level Analysis of Violent Conflict), an EU-funded multi-institution partnership, have spearheaded micro-level conflict research. Some studies have conducted systematic empirical analysis of the mechanisms linking interactions of individual, household, and community units of analysis to processes of violent conflict (Bozzoli and Brück 2009, Bundervoet, Verwimp et al. 2009, Brück, Patricia Justino et al. 2010, Miguel and Roland 2011, Dupas 2012).

Brück and Vothknecht (2011), Justino (2012), Bozzoli, Brück et al. (2011) differentiate between two main approaches to using data in micro-level conflict research: (1) using purposively designed surveys; and (2) employing existing socio-economic data sets from conflict-affected regions. Of the two, the former aims to use specifically collected data to uncover causes and functions of conflict at the micro-level. This more uncommon approach so far, this includes ex-combatant surveys, genocide and atrocities surveys, displaced people surveys, surveys of civilian populations affected by conflict, and standardized conflict surveys (I.e. ICRC's People on War Surveys) (Blattman and Jeannie 2007, Mvukiyehe and Samii 2008/9, Ibáñez, Calderón et al. 2011, Justino, Ivan et al. 2012). The second direction uses micro-level data sets that were not explicitly collected for the analysis of conflict processes or consequences, but which can be used for that purpose when merged with conflict event data (Bozzoli and Brück 2009, Bozzoli, Brück et al. 2011, Douarin, Litchfield et al. 2011, Moya 2015, Nasir, Rockmore et al. 2015). Beyond these two directions, research on the causes and drivers of conflict at the individual and household level also includes qualitative and smaller scale quantitative analysis based on small samples and limited geographic locations (ICRC 2001, Boothby, Crawford et al. 2006, McKay, Robinson et al. 2006, Wessells 2006, Dwyer and Cagoco-Guiam 2011), which also focus on conflict processes, community structures and institutional changes at the local level.

Micro-level research has made significant contributions to measuring conflict's effect on livelihood choices and poverty dynamics. One strand of the micro-level conflict literature suggests a positive correlation between violence exposure and various measures of deprivation at the household level. Mercier, Ngenzebuke et al. (2016) compare three waves of household panel data in Burundi over 1998-2012 and deduce that violence exposure seems to trap already poor and economically vulnerable households into chronic poverty. Non-poor households exposed to violence do not exhibit the same adverse impact on welfare. Douarin, Litchfield et al. (2011) similarly find that exposure to violence has different impacts on household welfare depending on the labor and livelihood choices adopted. Households with more diverse livelihood opportunities demonstrate greater economic resilience and ease in increasing consumption levels. War-affected households in Rwanda that suffer loss of real estate or land due to conflict tend to be at greater risk of falling into chronic poverty after conflict, particularly for households accustomed to cultivation and land usage prior to the conflict (Justino and Verwimp 2013). Other studies offer evidence suggesting that the decline of infrastructure, economic opportunities, and social services due to conflict increases the likelihood of chronic poverty regardless of pre-existing assets, skills, or social capital (Bozzoli and Brück 2009, Bozzoli, Brück et al. 2015, Bratti, Mendola et al. 2016).

Analytical work linking conflict to human capital accumulation indicators finds that conflict exposure causes household trade-offs that negatively impact child schooling retention and investment in health care (Dabalen and Paul 2012, Justino, Leone et al. 2014, Minoiu and Shemyakina 2014, Brown and Velásquez 2015). These studies echo the view that conflict induces risk aversion and short-term time preferences, which, combined with real conflict-imposed economic constraints, detracts from human capital accumulation post-conflict. Micro-level studies also explore the relationship between exposure to conflict and other behaviors, such as the impact of civilian casualties on wartime informing (Shaver and Shapiro 2016) and degrees of depression (Bratti, Mendola et al. 2016).

Yet data on whether women's greater market participation and altered engagement in the domestic sphere results in welfare gains and long-term social empowerment are more ambiguous (Bozzoli, Brück et al. 2011, Justino, Ivan et al. 2012). One strand of the literature suggests that the economic and social gains women may have achieved due to the absence of men during conflict periods can erode in the post-conflict period and often do not result in a comparable increase in social empowerment (Justino 2009). Ibáñez, Calderón et al. (2011) look at displacement in Colombia as an indirect impact of conflict and find that that despite a net increase in earnings, bargaining power of displaced women is not statistically different from the control group. In contrast, domestic violence is larger for displaced women, who in turn resort to violent punishment against their children. Wang and Weina (2016) find that displacement in China during Mao's mass Send-Down Movement had a significantly negative effect on women and men's nonmaterial well-being, which they measure by marriage outcomes, social network, and happiness.

Incidence of gender-based violence, while a prevalent global challenge in many environments even before the onset of violence, often worsens in the context of conflict and instability (Buvinic, Das Gupta et al. 2012, Strachan and Haider 2015). Micro-level quantitative analysis on the impact of conflict on women and girls' vulnerability to gender-based violence (GBV) is relatively robust. Ostby (2016) explores links between armed conflict and intimate partner violence and finds a significantly damaging effect of armed conflict on rates of domestic violence. Dijkman, Catrien et al. (2014) explore the impact of conflict exposure on GBV across IDP camps, areas of return, and households of various income and education levels, where wealthier educated women had a higher likelihood of falling victim to GBV. Women in South Sudan have also experienced varying dimensions of gender-based violence (Elia 2007, Elia 2007, CARE 2014). A new study highlights that rates of varying forms of violence against women and girls in South Sudan is among the highest in the world; conducted by International Rescue Committee and the Global Women's Institute (2017), the study found that 65 percent of the women surveyed experienced some form of sexual or physical violence in their lifetime, double the global average. While intimate partner violence was most commonly reported, 33 percent of women reported experiencing sexual assault from a non-partner, frequently linked to displacement, abduction or raids.

The international donor community has traditionally come forth with descriptive reports on the effect of armed conflict on the health and well-being of both women and girls, which includes sections on GBV and security (ICRC 2001, UNFPA 2002, UNICEF 2005, Dwyer and Cagoco-Guiam 2011). These studies are mostly aimed towards informing aid programming and Disarmament, Demobilization, and Reintegration (DDR) initiatives and lack the trend analysis and controlling for bias associated with quantitative analysis. Several qualitative and quantitative studies using smaller samples of respondents focus on ex-combatant motivation and reintegration (Keairns 2003, Miranda 2003, McKay, Robinson et al. 2006), exploring both the positive and negative aspects of the combatant experience for adolescent girls. Yet while much of the literature on female combatants does address the experience of adolescent girls, there is little micro-level analysis of conflict's impact on labor market decisions, welfare, vulnerability to sexual violence, and behavior specific to civilian adolescent girls.

Quantitative research on the impact of conflict on women has focused on heads of households with little information on the experience of adolescent girls or other female members of the household. Conflict's impact on risk preferences and social capital accumulation have mostly been explored in adults. Welfare measurements related to children and youth tend to comprise of years of schooling and monetary investments in health care. A focus on adult respondents remains the norm. While girls have been included in surveys assessing the extent and impact of various forms of GBV and are the subjects of mostly qualitative research on ex-combatants, little systematic empirical research focuses specifically on non-combatant female adolescents.

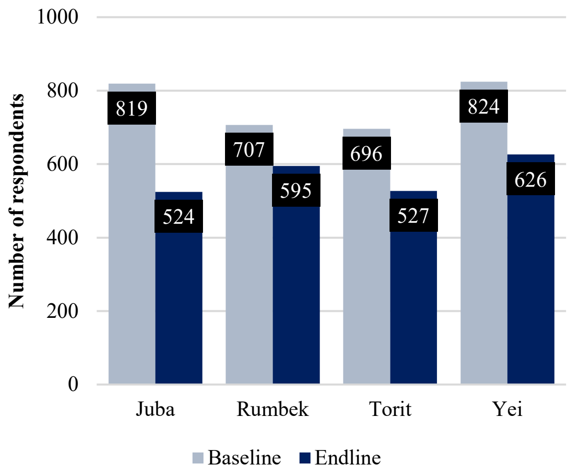
This gap can partly be explained by the fact that challenges related to collecting data in conflict-affected areas are particularly acute when targeting adolescent girls. For purposeful surveys, sampling female adolescent respondents can be more challenging compared to male or female heads of households, particularly in the context of traditional gender norms and low levels of community trust. Existing socio-economic household data also typically relate to heads of households. Lack of personal identification, which is more common among children and underage youth than adults, is often higher in conflict affected areas and can cause discrepancy in data. Attrition is a major concern due to conflict driven displacement. These challenges can be compounded by difficulties tracking younger household members who are not always registered consistently in population databases.

Data

The Adolescent Girls Initiative (AGI) was launched by the World Bank in October 2008 as a public-private partnership intended to promote the transition of adolescent girls from school to productive employment through innovative interventions that are tested, and then scaled-up or replicated if successful. The initiative was piloted in eight countries including Afghanistan, Jordan, Lao PDR, Liberia, Haiti, Nepal, Rwanda, and South Sudan. In South Sudan, the World Bank partnered with an NGO, BRAC International, to adapt and pilot its Empowerment and Livelihood for Adolescents (ELA) model which combined a range of innovative social and financial empowerment interventions targeting 3,000 girls between the ages of 15-24 in four states. Key interventions included the establishment of adolescent girls clubs to create safe spaces for social interaction and engagement, life skills and livelihoods training, financial literacy training, access to savings and credit facilities and community and parental sensitization efforts. To assess the effectiveness of the interventions in South Sudan, a rigorous impact evaluation was built into the project.

The baseline and endline surveys for this evaluation were conducted across the four target states of Juba, Rumbek, Torit and Yei in 2010 and 2015 respectively (Figure C2-1). In each state, respondents were drawn based on a two-stage random selection using clusters as the primary sampling units. Given the high levels of mobility in South Sudan, these surveys were designed as repeat cross-sections. Hence no efforts were made to re-visit baseline respondents at endline.[176]

Figure C2-1: Number of observations at baseline and endline.



Self-reported conflict exposure

The eruption of violence in 2013 impacted and delayed the implementation of the endline survey to early 2015. To measure the extent of this conflict, the endline survey incorporated an additional module on conflict exposure. Getting direct household conflict exposure measures is very meaningful. This module was developed based on similar conflict exposure questionnaires and adapted to the context in South Sudan with special consideration paid to the ethical administration of surveys in conflict-affected populations. The conflict exposure module included key questions related to looting, household damage, and physical harm (including death) to members of the household (Table C2-1). A subset of these questions has already been used in the High Frequency Pilot conducted by South Sudan's National Bureau of Statistics after comprehensive discussions of the impact of these questions on the emotions of the respondent. Understandably, several respondents chose not to answer these questions.

Table C2-1: Variables in the endline questionnaire measuring conflict exposure.

Variable	Description
Household looted	Was your household looted during the conflict?
Other household looted	Was any household in your neighborhood looted in the conflict?
Household damaged	Was your household damaged in the conflict?
Household member harmed	Was any member of your household harmed in the conflict?
Number of Household members harmed	How many members of your household were harmed in the conflict?
Household member died	Did any member of your household die due to the conflict?
Number of Household members died	How many members of your household die due to the conflict?
Member left	Did any member of your household leave due to the conflict?

Out of 3,137 respondents, 804 respondents (around 25 percent) chose not to respond to the questions in the conflict module (Figure C2-2). While it seems that respondents from Juba and Yei mostly consented, there is significant variation in the rates of consent across clusters in Rumbek and Torit

[176] Five clusters, which covered 173 households, were dropped from the analysis as these clusters were not re-visited at endline.

(Figure C2-3). Besides area, we find no bias in terms of age, household size, religion, and years of residence when exploring the characteristics of the non-consenting individuals (Table C2-2).

Figure C2-2: Non-consent to the conflict module.

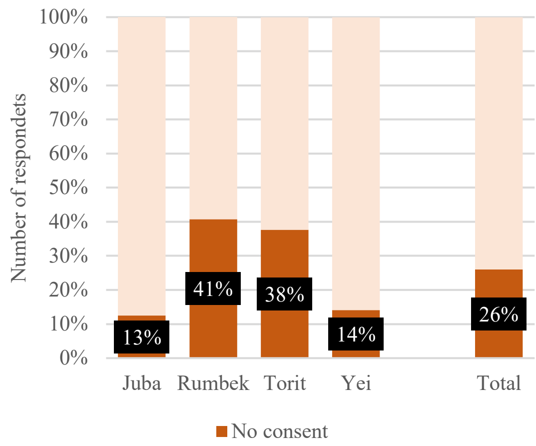


Figure C2-3: Density plot of consent by area.

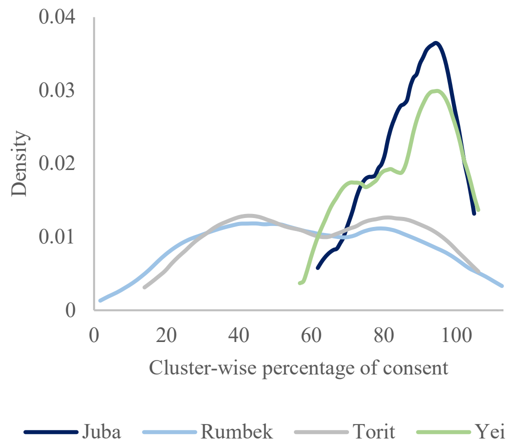


Table C2-2: Characteristics of consenting and non-consenting respondents.

Characteristics (mean)	Age	HH size	Years at residence	Years of education	Number of IGAs
Consenting	22.1	10.5	5.6	7.7	0.9
Non-consenting	21.7	9.8	5.5	7.6	1

About 40 percent of all consenting individuals experienced at least one conflict event (Figure C2-4). Additionally, about 30 percent of consenting individuals stated that a member of the household was harmed or died due to the conflict (Figure C2-5). The highest incidence is found in Rumbek, where about 67 percent of the consenting individuals experienced one or more conflict events, compared to less than 40 percent in the other three areas. Accordingly, Rumbek's residents also report the majority for most conflict events such as a member dying or being harmed, while both Rumbek and Juba's residents more often reported that members were displaced. As Rumbek also has the highest non-response rate, it is likely that the overall extent of conflict exposure is underestimated.

Figure C2-4: Respondents that experienced at least one conflict event.

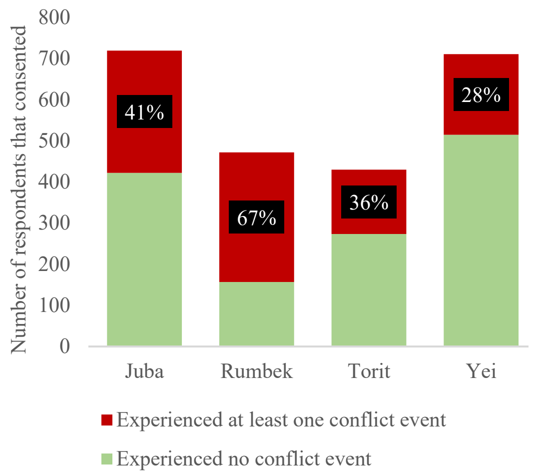
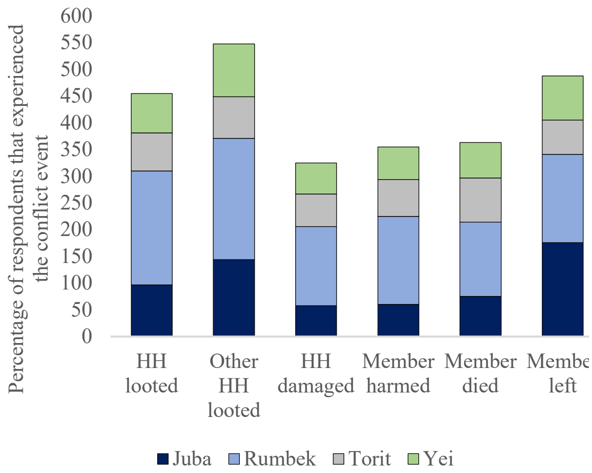
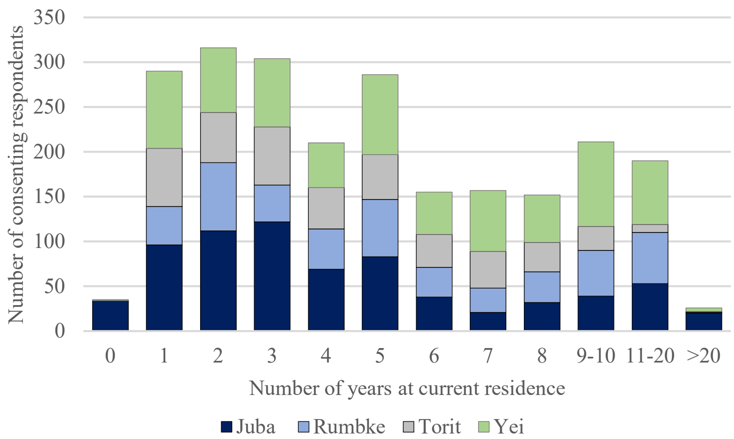


Figure C2-5: Conflict events by area.



Measuring the impact of the conflict in 2013 at the cluster level requires us to restrict the data set to respondents that spent at least three years at their current residence. Otherwise, the cluster indicators for conflict exposure and outcomes would be mixed between the population exposed to conflict at the selected cluster and the population being exposed to conflict in another cluster, who relocated to the selected cluster in the last three years. This excluded 640 consenting respondents, from which the majority (38 percent) were from Juba.

Figure C2-6: Number of years at current residence



The variables in the conflict exposure module of the questionnaire are used to construct a composite index to measure exposure to the conflict (see Appendix). According to the internal conflict indicator, 1 in 3 girls were exposed to the conflict. For ease of interpretation of the analysis results, the continuous conflict exposure indices are converted into binary values.[177] The cut-off point to identify conflict exposure is the average of the continuous conflict exposure index. Clusters above the mean index (1.93) are categorized as having been exposed to conflict, while clusters below the mean are categorized as not having been exposed to conflict. Using this cut-off, 33 percent of all clusters were exposed to conflict, most of which are from Rumbek (Figure C2-7 and Figure C2-8).

[177] The binary variable is more intuitive for a difference-in-difference approach, so results using a continuous variable are reported in the Appendix.

Figure C2-7: Density plot of the internal conflict indicator.

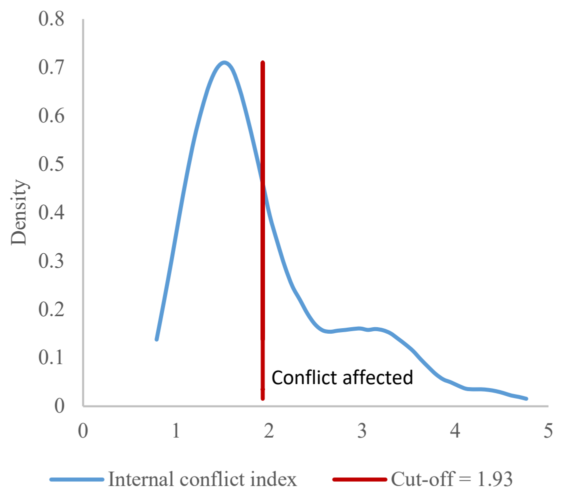
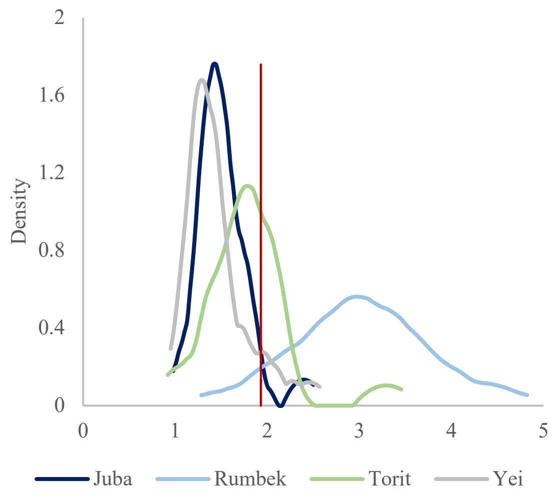


Figure C2-8: Density plot of the internal conflict indicator per area.



External conflict indicator

We construct an external conflict indicator using data from Armed Conflict Location & Event Data (ACLED) project in addition to a self-reported conflict index. There are two reasons why the self-reported conflict exposure index might be biased. First, there could have been cases of extreme under- or over-reporting induced by fear or differences in perception. This pattern of extreme reporting is evident considering the spread of the conflict index. Secondly, only 75 percent gave consent to answering the conflict questions, potentially leading to an additional source of bias through self-selection. Thus, it could bias the estimation results from a cluster-level difference-in-difference analysis. Therefore, we also construct an external conflict indicator.

Similar to the internal conflict indicator, slightly over a third of the girls were exposed to conflict according to the external indicator. Like the binary internal indicator, the average of the continuous external indicator is used to identify clusters exposed to conflict. Based on this cut-off, 34 percent of all girls were exposed to conflict (Figure C2-9). This measure only categorized clusters in Juba and Rumbek as conflict exposed (Figure C2-10).

Figure C2-9: Density plot of external conflict indicator.

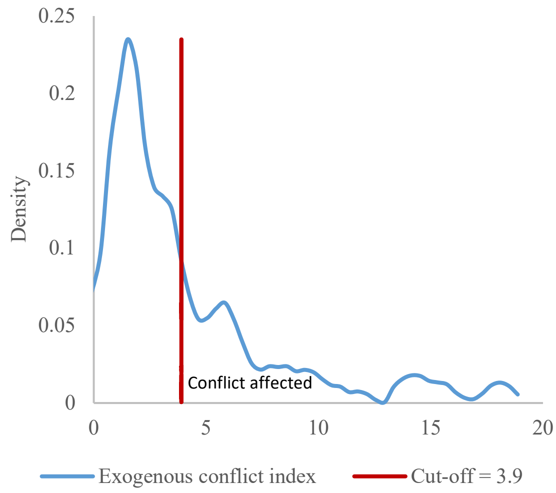
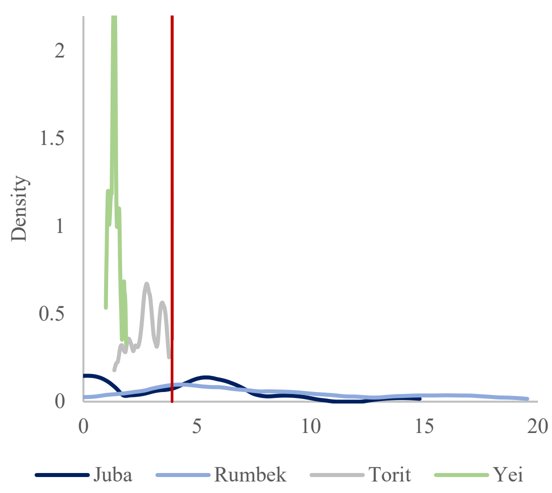


Figure C2-10: Density plot of external conflict indicator by area.



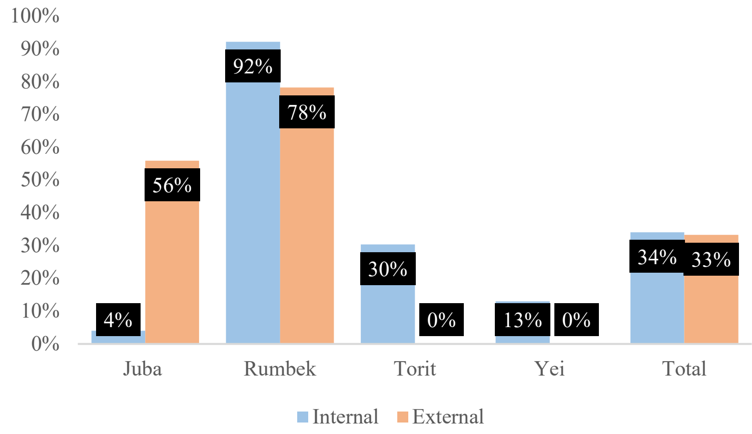
Except for education, average characteristics for girls exposed to conflict and those not exposed are similar for both the internal and external indicator. On average, girls exposed to conflict were slightly younger, had more household members, had lived in their residence longer, and participated in more Income Generating Activities (IGAs) than girls not exposed to the conflict (Table C2-3). Most of the differences are statistically significant but minor, except the household size, with girls exposed to conflict were from much larger households than girls who were not exposed to conflict. Household size is also correlated with poverty, so girls exposed to conflict may also be poorer (World Bank 2016). On average, conflict exposed girls are less educated when using the internal indicator but more educated when using the external indicator.

Table C2-3: Characteristics of girls exposed and not exposed to conflict.

	Characteristics (mean)	Age	HH size	Years at residence	Years of education	Number of IGAs
	Not exposed	22.4	8.7	5.4	7.9	0.8
Internal	Conflict exposed	21.5	12.6	5.9	7.3	1.1
	Not exposed	22.3	8.8	5.4	7.4	0.9
External	Conflict exposed	21.5	12.5	6.0	8.4	1.0

While both indicators have some caveats, they are both complementary. According to both indicators, about 1 in 3 girls were exposed to the conflict. Rumbek had the highest percentage of conflict exposed clusters (92 and 78 percent respectively) and the highest percentage of non-consent to conflict questions (Figure C2-11 and Figure C2-2). Therefore, households that were most affected may also have been unwilling to respond to conflict questions. The correlation between the internal and external indicator is significant and positive ($P<0.1$). The moderate correlation coefficients further warrant the claim that the self-reported index measured the self-perceived exposure to conflict while the external index provides a more objective but also less nuanced indication of conflict exposure.

Figure C2-11: Percentage of clusters categorized as conflict-affected.



Outcome Indicators

The dependent variables for the analysis are individual level outcome indicators. These variables cover a range of economic, social, and household condition indicators. A total of 27 outcome variables are selected from categories such as education, income generating activities, savings, marriage, aspirations, empowerment and household characteristics (Table C2-8 in the Appendix).[178]

We apply the one-way ANOVA test, to check if means for all 27 outcome variables are statistically different across clusters (see Appendix). Means being similar implies that there is low variability in the outcome variable across all clusters, which prevents us from significantly evaluating the impact of conflict on outcome variables. The test is applied to both the baseline and endline outcomes.[179] We observe that all outcome means are statistically and significantly different from each other.[180]

Methodology

We apply a difference-in-difference approach to compare outcomes for girls exposed to the conflict versus girls who were not. This method is appropriate when there are before-and-after time periods and two groups: one that is subject to the treatment, and another which is subject to all the other influences on the treatment group except the actual treatment itself (Meyer 1995). This eliminates pretreatment differences in the outcome variable and controls for anything that also changes over time and affects both groups. Therefore, the difference-in-difference estimates we report rely on the assumption that the differences in the outcomes between girls would be similar across conflict-affected and non-affected clusters had the conflict not happened.

More specifically, the difference-in-difference estimator β_3 in equation 1 is computed by comparing the first-differenced values of the outcome for the treatment and control groups. The treatment group in this case are the girls exposed to the conflict *C* while the control group are the girls who were not exposed to the conflict *NC*. The average outcome $\bar{y}$ in period 0 is subtracted from its average value in period 1 for both groups. The outcome differences for the control group are then differenced from the treatment group, which gives us the difference-in-difference estimate. The purpose of a

[178] Only those indicators were chosen which were present in both the endline and the baseline data sets. While most of the indicators were directly recorded through the questionnaire, some of the indicators have been derived through algebraic manipulations of other variables.
[179] For each variable, an analysis of variance is performed on the absolute deviations of values from the respective group means. If the P-value is less than 0.05, the hypothesis of homogeneous means is rejected. In addition to the ANOVA F statistic, we also report the Levene's test for equality of variances and the Brown-Forsyth test statistic, where the ANOVA is performed on the deviations from the group medians.
[180] The discrepancy between the Levene's statistic and Brown-Forsythe statistic can be explained by the fact that the Brown-Forsythe analysis assumes a non-normal distribution as it takes into account the cluster medians rather than the means.

difference-in-difference approach is to analyze whether the estimate β_3 is statistically and significantly different from zero.

To estimate the difference-in-difference effect of self-reported conflict exposure, we use an ordinary least squares (OLS) regression model:

$$Y_{it} = \beta_0 + \beta_1 \text{post}_t + \beta_2 \text{conflict}_i + \beta_3 \text{post}_t * \text{conflict}_i + \varepsilon_{it} \quad [1]$$

where Y_{it} is the outcome variable of adolescent girl *i* at time *t*. post_t is a binary variable indicating time period t (pre- or post-conflict) and conflict_i is the binary or continuous treatment variable, indicating conflict exposure of cluster *i*. ε_{it} is the error term.

β_1 is the expected mean change in outcome from before to after the conflict among the control group. The coefficient of the treatment variable, β_2, is the estimated mean difference in the outcome between the treatment and control groups prior to the conflict: it represents whatever baseline differences existed between the groups before the group was exposed the conflict. β_3 by itself is the difference-in difference estimator, and hence, the coefficient of interest.

However, the baseline model might still suffer from omitted variable bias as there are other confounding factors affecting the given outcome variables besides time-period and conflict exposure. Therefore, the following model is estimated:

$$Y_{it} = \beta_0 + \beta_1 \text{post}_t + \beta_2 \text{conflict}_i + \beta_3 \text{post}_t * \text{conflict}_i + \beta_4 X_{it} + \beta_5 \text{cluster} + \varepsilon_{it} \quad [2]$$

where X_{it} is a vector of control variables for girl *i* at time *t* and cluster is a cluster-level fixed effect to control for variation within clusters.

Results

The regression model formulated in equation [2] is used to analyze the effect of conflict exposure on various socio-economic outcomes at the individual and household level. We run the regression for each of the 27 outcome variables presented earlier using the self-reported and external conflict exposure indices (results from continuous variables are reported in the Appendix). An overview of the regression coefficients for the binary self-reported and external conflict variable is provided (Table C2-4).

Table C2-4: Overview of regression results for each outcome indicator and conflict variable.

Dimension	Outcome	Internal conflict indicator	External conflict indicator
Education	Enrolled	-0.03	-0.05
	Dropped out	0.01	-0.02
	Years education	0.24	1.13*
	Years before dropping out	-0.04	1.07*
Household Characteristics	Current savings	-0.08*	-0.16**
	Savings from 2 weeks	-0.02	-0.01
	Total savings	-0.23	-0.24
	People per room	0.73**	-0.133
	Food scarcity index	0.58*	-0.34
	Household asset index	-3.59***	-1.33
	Toilet	-0.16**	-0.30***
	Good walls	-0.08**	-0.10***
	Good roof	-0.01	0.04
	Household monthly income	0.26	-0.32
Income generating activities (IGAs)	Number of IGAs	0.13	-0.35***
	Individual monthly income	-0.12	0.24
	Control index	0.15	-0.03
	Entrepreneurial potential	1.01***	1.01**
	Satisfaction	0.056	0.02
Marriage	Empowerment	0.14	0.60***
	Married	0.07*	0.20***
	Pregnant	-0.09***	-0.12**
	Daughter optimist	-0.03	-0.07
	Lost pregnancy	-0.05**	-0.12***
	Children	-0.01	0.06
Aspirations	General anxiety	0.52*	0.97***
	Ladder position	-1.38***	-1.13***

*** p<0.01, ** p<0.05, * p<0.1

Conflict had a significant positive effect on the number of years in education and the number of years before dropping out. Transient education outcomes such as enrollment were not significantly impacted by the conflict, as both are often only affected in the short-term after a conflict event. However, conflict had a significant positive effect on the number of years in education and the number of years before dropping out. Specifically, girls in conflict affected areas completed an additional year of education than girls who were not in conflict affected areas. Juba is the only area for which the conflict significantly increased years of education (Table C2-5). A sorting effect is a likely explanation as most of the girls who had spent less than 3 years at the current residence were from Juba (Table C2-5). Additionally, most girls that reported a member leaving due to the conflict were also from Juba (Figure C2-5). Thus, families with higher education may have recently migrated to Juba and lower educated girls might have left due to the conflict, resulting in an overall average increase in girls' education after the conflict.

Table C2-5: Impact of the external conflict indicator on years of education by area.

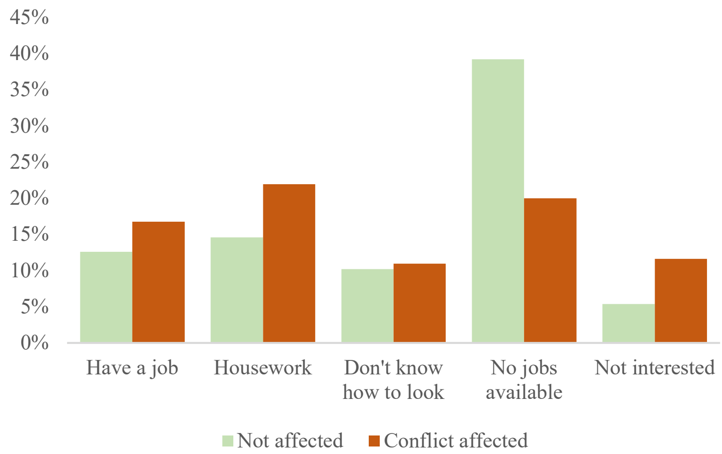
Dimension	Outcome	Rumbek	Juba
Education	Years education	0.148	1.321**
	Years before dropping out	1.324	1.491***

We find a significant negative effect of conflict on current savings. Girls exposed to the conflict were about 10 percent less likely to report any current savings compared to girls who were not exposed to conflict. This finding is consistent for both the internal and external conflict indicator. In the context of violence, looting and damage to households, savings can be used to complement consumption or repair the damage. However, the impact on total savings is not statistically significant although large and negative.

The conflict negatively affected household's socioeconomic indicators such as food security, assets and the physical condition of the house. The effect on household income is uncertain. Specifically, girls in conflict affected areas lost assets, toilets and good walls while they had to use fewer rooms for more people, and suffered from increased food scarcity after the conflict. The negative impact of conflict on food security is widely documented (Cohen and Pinstrup-Andersen 1999). The effect on toilets and walls is consistent for both the internal and external indicator. The losses of assets, the increased number of people per room and increased food scarcity are only impacted by the internal conflict indicator, potentially as it measured conflict exposure in a more nuanced way than battles but includes looting.

Engagement of girls in income generating activities (IGAs) is significantly negatively impacted by the conflict. In this case, being exposed to a conflict event resulted in girls participating in fewer IGAs. No statistically significant estimate was found for the impact on individual monthly income. Heightened insecurity might have constrained girls' mobility and ability to conduct paid work outside the home, resulting in more time spent on domestic tasks. Girls in conflict affected areas mostly reported housework as the reason for not having a job (23 percent) whereas girls in areas not affected by the conflict mostly reported the unavailability of jobs (38 percent; Figure C2-12).[181] Thus, it is likely that an increase in housework may have substituted income generating activities for girls in conflict affected areas.

Figure C2-12: Most common reasons for being unemployed.



The entrepreneurial potential index increased for girls in conflict-affected areas for both the internal and external conflict variables. The index is a score from 1 to 10 and comprises of self-perceived scores related to various future business opportunities.[182] On average, conflict increased girls' entrepreneurial potential index by about 10 percent. Conflict may lead to girls perceiving greater

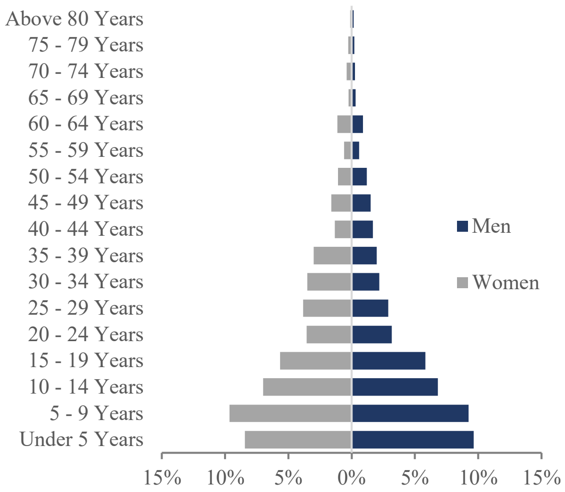
[181] An accurate comparison cannot be made as the baseline and endline surveys had different questions and answer options regarding unemployment. Furthermore, these responses are the top five most common responses from a range of many.
[182] For more details, consult 0. Entrepreneurial potential index in the Appendix.

business opportunities and to consider entrepreneurial activities as a resilience mechanism. However, the negative impact of conflict on IGAs indicates that the entrepreneurial potential is – currently – not activated. A tension between expectation and reality can explain this disconnect, such that the expectation and interest in taking up employment opportunities increase but the ability and opportunity to undertake income generating activities decrease.

Conflict increased the likelihood of girls being married. Conflict increases uncertainty and insecurity, thereby incentivizing either voluntary or forced marriage as families marry off daughters or girls engage in marriage to increase safety and economic security. This is common practice in the context of displacement. In some circumstances, women and girls who are sexually assaulted are forced to marry their perpetrators to avoid social stigma (Elia 2007). In South Sudan, sexual assault and abduction have been used as a means to initiate marriage while circumventing high bride prices. While the questionnaire does not capture indicators of gender-based violence due to ethical concerns, the conflict is like to have increased gender-based violence.

Conflict affected girls were less likely to be pregnant than girls not affected by conflict. In the context of South Sudan, high rates of male mortality or morbidity due to conflict, the general absence of men from home areas due to abnormal migration or engagement in combat are contributing factors. Population statistics indicate the absence of men in the respective age groups (Figure C2-13). Additionally, fertility rates may be impacted by additional factors, including poor nutritional status and maternal stress, which serve to lower fecundity and increase spontaneous abortions.

Figure C2-13: Population distribution, 2016.



Conflict affected girls had higher empowerment scores.[183] With the absence of men, girls might have recently assumed responsibility as head of household and responsibility for household decision making. Similarly, men may be spending most of their time outside the house fighting or looking for sources of income, which may have resulted in women taking more control of the household. This result is consistent with girls exposed to the conflict reporting higher entrepreneurial scores. Even though girls took up fewer employment activities and faced a reduction in savings and household assets, they may have felt more accountable due to the added responsibilities they face after conflict.

[183] The empowerment score considers 7 questions relating to gender roles within the household, such as 'Who should earn money for the household? – Men, Women, Both'.

Lastly, the conflict increased general anxiety and lowered the expected ladder position in 5 years by at least 1 level.[184] These results are consistent for both the internal and external conflict variable. Women are often more vulnerable than men to Post-Traumatic Stress Disorder (PTSD) and anxiety disorders when exposed to the same traumatic event (Ayazi, Lien et al. 2014). Similarly, the lowered aspirations could be driven by psychosocial impacts including trauma. As conflict leads to an increase in anxiety levels, this in turn may decrease an individual's expected ladder position standing in the next 5 years. Additionally, conflict increases uncertainty about the future and increases expectations of future conflict, which can also explain lowered aspirations.

Conclusion

This study contributes to available empirical evidence on micro-level impacts of conflict by analyzing the effects of the 2013 conflict on adolescent girls in South Sudan. Our analysis provides evidence on the negative effect of conflict exposure on various outcomes for girls such as employment opportunities, marriage-related outcomes, and the physical household condition. These results provide some perspective on both economic and social costs of the conflict, which can ideally be leveraged to inform design and evaluation of policies and programming intending to remediate the negative effects of conflict.

About half of the results were consistent when using self-reported and external conflict indicators. Here, it is important to revisit the caveats in both indicators. The self-reported indicator uses self-reported measures to assess exposure to traumatic events, where inconsistencies in recall and exaggerated responses can produce a bias (Southwick SM., Morgan CA III. et al. 1997). Additionally, the political climate may have contributed to respondents not fully trusting interviewer intentions. Given the renewed conflict in some of the border areas of South Sudan and the recent independence, it is possible that heightened caution within communities affected responses among those surveyed. This is consistent with the fact that Rumbek has the highest self-reported and external conflict exposure measure and also the lowest consent rate from the other three counties. While the external indicator is used to mitigate these biases, it may be an underestimated indicator of conflict exposure as it only includes deadly events that were reported. The self-reported conflict variable is relatively more precise as it is comprised of a wide range of micro-level conflict exposure variables and captures specific types of damage which are not reflected in the ACLED data. For these reasons, we use both external and self-reported indicators to inform our analysis.

The impacts of the ongoing conflict are overwhelming, and action must be taken immediately to prevent them from escalating. One important policy implication from this study is that adolescent girls and young women are an important resource for economic engagement and empowerment and that economic and business development initiatives should include criteria for targeting and incentivizing participation of this particular demographic in economic activities. Adolescent girls exposed to conflict reported higher empowerment and entrepreneurial index scores, indicating willingness to work and start businesses in the future. Creating such opportunities for girls would, in turn, have the potential to contribute to economic growth and poverty reduction, as well as address pervasive conditions of income inequality among the poor and among the overall population (Acharya 2008). Targeted programming to support and incentivize girls' economic engagement further improves household food security and economic welfare. Depending on the types of activities in which girls choose to engage, an integrated approach that enables a school-to-work transition through both livelihoods and

[184] Anxiety was measured by constructing an index that incorporates if a girl worries about her job, husband, money and violence. Ladder position here indicates (on a 1 to 10) the scale of how good or bad one's life is, so the 10th ladder is the best possible life scenario while the 1st one is the worst. In this case, the question asked about what the assumed ladder position would be 5 years later.

skills development, as well as with cognitive and non-cognitive skills training interventions would prove especially useful.

That said, increasing economic opportunities alone are not enough to improve the well-being of girls. The findings in this paper also help improve our understanding of the longer-term psychosocial consequences of conflict. For example, lowered aspirations and high anxiety during early years have been linked to worsening economic outcomes in adulthood (Powell and Butterfield 2003, Riegle-Crumb, Moore et al. 2011). Additionally, the issue of early and likely forced marriage is a prevalent feature of South Sudan, as are other dimensions of gender-based violence. These challenges highlight the need for interventions that focus on increasing access to education services in part to enable improved employment opportunities, building capacity for provision of psychosocial and mental health services, and wider prevention programming addressing pervasive and challenging social norms that perpetuate among other issues violence or harmful practices impacting in particular women and girls. In terms of addressing issues of trauma and PTDS, currently, despite enormous need, there are few providers for psychosocial or mental health services in South Sudan, with the exception of select services provided by non-governmental organizations. The principal delivery mechanism of health services in South Sudan is through a basic package of health services funded by the Government of South Sudan and international donors and provided by non-governmental organizations (Roberts, Guy et al. 2008). Besides scaling up these services, training of health care staff and community workers to provide basic psychosocial care or mental health support, and also to train up and enable community-based self-help support groups should also be explored (van Ommeren M., Saxena S. et al. 2005).

Without improved services and protections, it is likely that the impacts of conflict will continue to be severe particularly for vulnerable groups such as adolescent girls and young women, with dire implications for social and economic functioning of the girls themselves, as well as for their family. Immediate aid and targeting during the ongoing conflict is needed, but so is the protection of marginalized groups and long-term efforts to secure future outcomes for the people of South Sudan.

Appendix

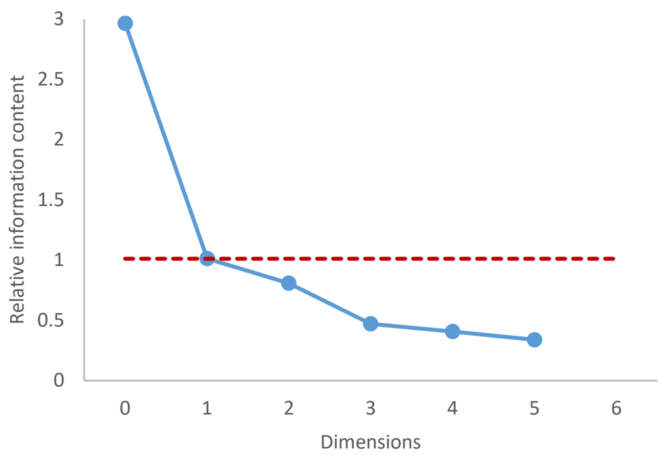
Construction of self-reported conflict indicator

The variables in the conflict exposure module of the questionnaire are used to construct a composite index to measure exposure to the conflict, using principal component analysis (PCA).[185] Constructing an index is useful as it captures key dimensions of multiple variables and makes it easy to use and interpret in regression analyses. As there are six conflict exposure variables of interest, PCA can identify key dimensions with the most variability.[186] For the PCA, the endline sample is restricted to respondents who provided consent to answer the questions in the conflict exposure module, and have stayed at their current residence for at least 3 years. The scree plot shows a break after the steepness at the second component, where it is evident that the first component captures the most variability. The first component of the PCA is chosen as it captures about half the variation (Figure C2-14). The resulting index obtained for each household is normalized, and standardized to a scale of 1 to 10.

[185] The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set..

[186] The PCA produced 6 components. The first component has an eigenvalue of close to 3, and captures 49.4% of the total variation, while the second component has an eigenvalue of approximately 1, and captures 16.9% of the total variation.

Figure C2-14: Relative Information in PCA dimensions.



Analysis of Variance (ANOVA) results for conflict variables

We use a one-way Analysis of Variance (ANOVA) to test whether conflict index means and input variable means are statistically significant across clusters.[187,188] The results suggest a significantly larger variation between clusters than within clusters for each of the measured variables.[189] Thus, the conflict exposure indicators are able to reflect the geospatial exposure of conflict where nearby households are usually co-exposed to conflict. Given that the conflict affected some areas a lot more than others, this is not surprising. In addition, this is encouraging for a cluster-level difference-in-difference approach.

A simple one-way ANOVA does not specifically indicate which clusters display significant differences with the within and between cluster variability. Post hoc tests reported at the cluster level identify the clusters with significant difference in the within and between cluster variability, and the respective levels of significance (

Table C2-7).[190] About 40 percent of all the clusters show a statistically significant difference in the within and between cluster variability and most of these clusters are in Rumbek.

[187] ANOVA uses the F-test to statistically test the equality of means. The F statistic is based on the ratio of the variation between cluster means against the variation within the clusters. In order to reject the null hypothesis that the cluster means are equal, a high F-value or a P-value below 0.05 is needed. If the cluster means do not vary, or do not vary by more than random chance allows, than we cannot be confident about the means being different.

[188] Since we restricted the sample to a set of households who consented to respond to the survey module on conflict exposure, and had stayed at the current place of residence for at least 3 years, there are unequal number of clusters in each area, and unequal number of households in each cluster.

[189] The results are confirmed by a simulation where the cluster is randomly assigned to respondents. The simulation retrieved non-significant p-values.

[190] The cluster ID's have been codified, with numerical ID's for each cluster. The cluster ID's from 10000+ are correspond to clusters in Juba province, those with ID's 20000+ are in the Rumbek province, those with ID's 30000+ are in Torit province, and those with ID's 40000+ are in the Yei province.

Table C2-6: Results of one-way ANOVA for Conflict Index and other input variables.

	W/t Group Squared Sum (SS)	W/t Group Degrees of Freedom (DOF)	B/w Group Squared Sum (SS)	B/w Group Degrees of Freedom (DOF)	F Stat	P Value
Conflict Index	840.76	90	2548	1601	5.868	<0.01
Household Looted	54.807	90	196.49	1601	4.962	<0.01
Other Household Looted	64.954	90	218.45	1601	5.289	<0.01
Household Damaged	31.553	90	156.13	1601	3.595	<0.01
Number of Members Harmed	1.562	90	10.088	1601	2.754	<0.01
Number of Members Died	1.231	90	9.13	1601	2.398	<0.01
Members Left	40.639	90	221.32	1601	3.266	<0.01

Table C2-7: Post hoc results of ANOVA for Conflict Index, grouped by clusters.

CLUSTER ID	COEFFICIENT	STANDARD ERROR
10002	-0.416	0.316
10003	0.586*	0.307
10004	-0.244	0.299
10005	-0.314	0.307
10006	-0.296	0.322
10007	-0.037	0.313
10008	-0.354	0.303
10009	-0.386	0.299
10010	-0.484	0.299
10011	-0.203	0.299
10012	-0.376	0.297
10013	-0.114	0.326
10014	-0.242	0.299
10015	0.029	0.313
10016	-0.604**	0.307
10017	-0.097	0.313
10018	-0.120	0.326
10019	-0.095	0.299
10020	0.040	0.305
10021	1.322***	0.333
10022	-0.270	0.299
10023	-0.062	0.305
10024	0.030	0.310
20001	1.541***	0.393
20002	0.915***	0.343
20003	1.383***	0.307
20004	0.894***	0.329
20005	0.553*	0.301
20006	0.933**	0.431
20007	0.117	0.297
20008	1.237***	0.338
20009	1.488***	0.568
20010	0.805**	0.319
20011	0.303	0.297
20012	1.799***	0.326
20013	1.495***	0.333
20014	0.892**	0.374
20015	0.919***	0.319
20016	0.978**	0.448
20017	1.078**	0.431
20018	1.793***	0.383
20019	1.294***	0.374
20020	0.976**	0.383
20021	0.560	0.404
20022	1.216***	0.307
20023	-0.204	0.307
30001	0.151	0.305
30002	-0.225	0.416
30003	0.212	0.303
30004	0.612*	0.354
30005	-0.034	0.374
30006	-0.119	0.301
30007	-0.356	0.305
30008	0.325	0.431
30009	0.177	0.343
30010	0.133	0.338
30011	-0.151	0.319
30012	-0.537*	0.313
30013	0.197	0.367
30014	-0.124	0.367
30015	-0.151	0.305
30016	0.026	0.448
30017	0.104	0.404
30018	-0.306	0.316
30019	0.051	0.307

40001	-0.400	0.393
40002	-0.142	0.299
40003	-0.587*	0.303
40004	-0.456	0.322
40005	0.098	0.301
40006	0.064	0.297
40007	-0.542*	0.297
40008	-0.740**	0.297
40009	-0.240	0.322
40010	-0.049	0.305
40011	-0.510*	0.305
40012	-0.502*	0.301
40013	-0.499*	0.301
40014	-0.593*	0.338
40015	-0.284	0.305
40016	0.534*	0.307
40017	-0.348	0.333
40018	-0.068	0.316
40019	0.529	0.333
40020	-0.434	0.310
40021	-0.361	0.299
40022	-0.095	0.303
40023	-0.654**	0.322
40024	-0.254	0.322
40025	-0.291	0.305
_cons	1.782***	0.221

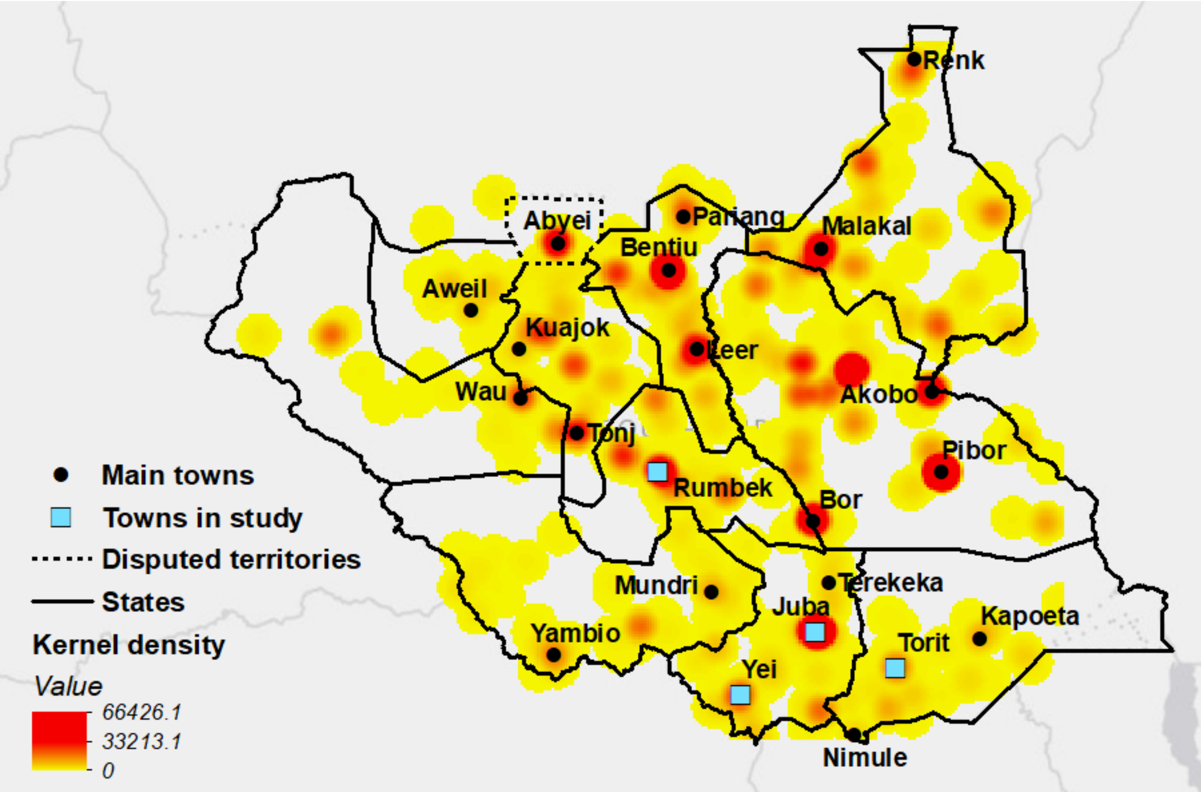
(***P < .01, **P < .05, *P<0.1)

Construction of external conflict indicator

The external indicator is based on conflict event data from the ACLED Project between December 2013 and January 2015. The data set codes the exact location of all political violence incidents that were reported during this time period.[191] For the selected time period there were 1,200 reported conflict events in South Sudan with a total of 9,209 fatalities. Most of the conflict is concentrated in the Northern part of South Sudan, particularly around Rumbek. This is consistent with Rumbek's high conflict exposure index average.

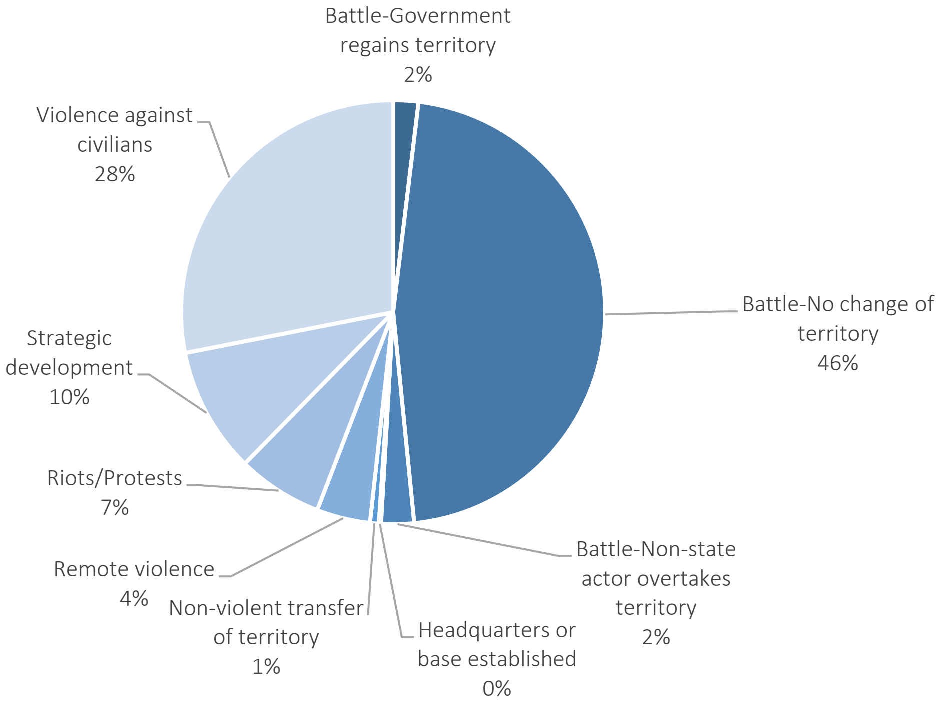
[191] Political violence is the use of force by a group with a political purpose or motivation. ACLED defines political violence through its constituent events, the intent of which is to produce a comprehensive overview of all forms of political conflict within and across states. A politically violent event is a single altercation where often force is used by one or more groups to a political end, although some instances - including protests and non-violent activity - are included in the data set to capture the potential pre-cursors or critical junctures of a conflict.

Figure C2-15: Location of conflict events in South Sudan between December 2013 and January 2015.



The ACLED data show that 465 of the reported conflict events (36 percent) were deadly and resulted in at least one fatality. Almost half of all reported conflict events (48 percent) were battles between the government and non-government forces. Violence against civilians was committed in 28 percent of all events (Figure C2-16).

Figure C2-16: Conflict events by type



Distance to a deadly conflict event is used to generate an external conflict exposure variable. The averages of latitude and longitude of all households in a cluster in the AGI survey are used to compute cluster GPS coordinates. By merging the girls' households GPS coordinates with the conflict event GPS coordinates, the distance between each cluster-conflict event pair is calculated. The continuous indicator is the normalized sum of the distances of all fatal conflict events within a radius of 5 km from the cluster.

Outcome variables

Table C2-8: Outcome variables

Variable	Description
Education	
Enrolled	Whether respondent is currently enrolled in school
Dropped out	Whether respondent dropped out from school
Years dropped out	Number of years of schooling completed by those in school
Years Education	Number of years of education completed by respondent
IGA	
Number of IGAs	Number of income generating activities currently being undertaken
Individual monthly income	Log of total income from all IGA's in the last month for the individual
Savings	
Current savings	Whether respondent has current savings
Savings from 2 weeks	Whether respondent has savings from the past 2 weeks
Total savings	Log of total savings at multiple locations
Marriage	
Empowerment	Standardized index of empowerment post marriage
Married	Whether respondent is currently married
Pregnant	Whether respondent is currently pregnant
Daughter optimist	Whether respondent sees a better future for their daughter
Lost pregnancy	Whether respondent has lost a pregnancy
Children	Whether respondent has a child
Aspirations	
General anxiety	Summative index of respondents to variables related to feelings of anxiety
Ladder position	Standardized index of difference between ladder position now vs. expected position 5 years in future
Empowerment	
Control Index	First dimension of MCA of variables relating to control over resources
Entrepreneurial potential	Summative index of binary variables relating to entrepreneurial potential
Satisfaction	Summative index of ordinal variables relating to level of satisfaction with status quo
Household Characteristics	
People per room	Number of occupants per room in household
Food scarcity index	Standardized index of food scarcity in household
Household asset index	First dimension MCA of household asset ownership variables
Toilet	Quality of toilet facilities
Good walls	Quality of walls' construction material
Good roof	Quality of roof construction material
Household monthly income	Log of total income from all IGA's in the last month for the household

Analysis of Variance (ANOVA) results for outcome variables

Table C2-9: Education outcome indicators in the baseline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Enrolled	110.081	95	654.855	3070	5.432	0.01	0.01	0.01
Dropped Out	145.844	95	617.229	3070	7.636	0.01	0.01	0.01
Years Dropped out	1345.189	94	8974.607	1050	1.674	0.01	0.01	0.66
Years Education	5176.14	95	83146.01	2526	1.655	0.01	0.01	0.01

Table C2-10: Education outcome indicators in endline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Enrolled	36.679	90	568.977	3046	2.182	0.01	0.01	0.01
Dropped Out	66.276	90	327.884	1488	3.342	0.01	0.01	0.054
Years Dropped out	2699.402	90	16236.45	1587	2.932	0.01	0.06	0.568
Years Education	2809.585	90	23010.52	2407	3.265	0.01	0.117	0.368

Table C2-11: Income generating outcome indicators in the baseline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Number of IGAs	259.78	95	1728.461	3123	4.941	0.01	0.01	0.01
Log of Last Month Income (Ind)	1775.793	95	14014.28	3115	4.155	0.01	0.01	0.01

Table C2-12: Income generating outcome indicators in the endline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Number of IGA	415.69	90	1643.674	3046	8.559	0.01	0.01	0.01
Log of Last Month Income (Ind)	4304.374	90	19612.96	3028	7.384	0.01	0.01	0.01

Table C2-13: Savings outcome indicators in the baseline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Log of Total Savings	1523.374	95	12453.6	2913	3.751	0.01	0.01	0.01
Savings	76.855	95	604.839	3035	4.059	0.01	0.01	0.01
Saved (last 2 Weeks)	55.314	95	191.983	903	2.739	0.01	0.01	0.428

Table C2-14: Savings outcome indicators in the endline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levene's P Value	Brown-Forsythe P-Value

Log of Total Savings	2094.889	90	23342.35	3046	3.037	0.01	0.01	0.126
Savings	74.481	90	700.718	3046	3.597	0.01	0.01	0.043
Saved (last 2 Weeks)	54.076	90	666.052	3046	2.748	0.01	0.01	0.001

Table C2-15: Marriage related outcome indicators in the baseline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Empowerment Index	1549.454	95	10211.77	3123	4.988	0.01	0.01	0.01
Married	96.578	95	619.706	3050	5.003	0.01	0.01	0.01
Loss of Pregnancy	5.162	95	75.652	3055	2.194	0.01	0.01	0.01
Children	90.438	95	596.643	3063	4.887	0.01	0.01	0.01
Pregnant	29.498	95	291.167	2765	2.949	0.01	0.01	0.01
Daughter's Future	50.101	95	706.791	3123	2.33	0.01	0.01	0.005

Table C2-16: Marriage related outcome indicators in the endline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levene's P Value	Brown-Forsythe P-Value
Empowerment Index	1693.787	90	12830.28	3046	4.468	0.01	0.01	0.003
Married	48.409	90	717.224	3045	2.284	0.01	0.01	0.514
Loss of Pregnancy	14.202	90	192.946	3046	2.491	0.01	0.01	0.01
Children	44.175	90	715.71	3045	2.088	0.01	0.01	0.844
Pregnant	11.423	90	268.735	3045	1.438	0.005	0.01	0.005
Daughter's Future	26.717	90	290.427	3046	3.113	0.01	0.01	0.01

Table C2-17: Aspirations outcome indicators in the baseline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Ladder position	1860.859	95	16113.65	3078	3.742	0.01	0.01	0.01
Anxiety Index	1799.872	95	8487.459	3123	6.971	0.01	0.01	0.01

Table C2-18: Aspirations outcome indicators in the endline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levene's P Value	Brown-Forsythe P-Value
Ladder position	2491.977	90	14846.88	3045	5.679	0.01	0.01	0.01
Anxiety Index	1146.942	90	8140.725	3046	4.768	0.01	0.01	0.006

Table C2-19: Empowerment outcome indicators in the baseline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
Control Index	4324.301	95	38185.11	3084	3.676	0.01	0.01	0.01
Entrepreneurial potential	1728.217	95	11119.66	3080	5.039	0.01	0.01	0.01
Satisfaction Index	4582.776	95	11335.69	3034	12.911	0.01	0.01	0.01

Table C2-20: Empowerment outcome indicators in the endline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levene's P Value	Brown-Forsythe P-Value
Control Index	3673.948	90	42138.6	3046	2.951	0.01	0.01	0.01
Entrepreneurial potential	2833.924	90	8990.101	3045	10.665	0.01	0.01	0.01
Satisfaction Index	1942.864	90	5928.676	3045	11.087	0.01	0.01	0.01

Table C2-21: Household characteristics outcome indicators in the baseline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levenes P Value	Brown-Forsythe P Value
People per room	1567.146	95	7182.065	2702	6.206	0.01	0.01	0.01
Good Roof	177.802	95	490.126	3051	11.651	0.01	0.01	0.01
Good Walls	52.282	95	370.449	3056	4.54	0.01	0.01	0.01
Toilet	83.993	95	533.509	3062	5.074	0.01	0.01	0.01
Food Scarcity Index	5724.042	95	34024.02	3041	5.385	0.01	0.01	0.01
HH Asset Index	57028.213	95	98049.49	3110	19.041	0.01	0.01	0.01
Log of Household monthly income	2784.587	95	18911.19	3109	4.819	0.01	0.01	0.01

Table C2-22: Household characteristics outcome indicators in the endline survey.

	W/t Group SS	W/t Group DOF	B/w Group SS	B/w Group DOF	F Stat	P Value	Levene's P Value	Brown-Forsythe P-Value
People per room	3827.57	90	19328.79	3040	6.689	0.01	0.01	0.01
Good Roof	228.645	90	550.473	3044	14.048	0.01	0.01	0.01
Good Walls	28.638	90	328.48	3044	2.949	0.01	0.01	0.01
Toilet	217.427	90	545.5	3044	13.481	0.01	0.01	0.01
Food Scarcity Index	7568.646	90	39438.33	3043	6.489	0.01	0.01	0.01
HH Asset Index	4996.458	90	5524.153	3044	30.591	0.01	0.01	0.01
Log of Household monthly income	4304.374	90	19612.96	3028	7.384	0.01	0.01	0.01

Entrepreneurial potential index

Read aloud: "Now we will talk about different tasks. You will rank your ability on how well you can do these activities on a scale of 0 to 10? 0 means you cannot do this activity and 10 is you definitely can"

1. Run your own business
2. Identify business opportunities to start up new business
3. Obtain credit to start up new business or expand existing business
4. Save in order to invest in future business opportunities
5. Make sure that your employees get the work done properly
6. Manage financial accounts
7. Bargain to obtain cheap prices when you are selling anything for business (outputs)
8. Bargain to obtain high prices when selling
9. Protect your business assets from harm by others
10. Collecting the money someone owes you

The index is a simple average of the answer.

Regression results

Table C2-23: Impact of conflict on education.

Variables		Enrolled	Dropped out	Years education	Years dropped out
Internal binary		-0.0259	0.0112	0.244	-0.0416
		(0.0378)	(0.0548)	(0.355)	(0.455)
		3,358	2,235	4,107	2,195
Internal continuous		-0.0410*	0.0510	0.147	-0.00931
		(0.0227)	(0.0341)	(0.287)	(0.323)
		3,358	2,235	4,107	2,195
External binary		-0.0510	-0.0225	1.127*	1.070*
		(0.0415)	(0.0605)	(0.638)	(0.573)
		2,365	1,569	1,808	1,160
External continuous		-0.00752	0.00397	0.107	0.151**
		(0.00503)	(0.00784)	(0.0810)	(0.0757)
		2,365	1,569	1,808	1,160

Robust standard errors in parentheses

Number of observations below standard errors

*** p<0.01, ** p<0.05, * p<0.1

Table C2-24: Impact of conflict on savings.

Variables	Current savings	Saved two weeks ago	Total savings
Internal binary	-0.0837*	-0.0236	-0.237
	(0.0450)	(0.0528)	(0.171)
	4,165	2,557	1,453
Internal continuous	-0.0410*	-0.0212	-0.186*
	(0.0227)	(0.0342)	(0.107)
	4,165	2,557	1,453
External binary	-0.163**	-0.0128	-0.236
	(0.0619)	(0.0854)	(0.255)
	1,847	1,356	896
External continuous	-0.0137	-0.00155	-0.0110
	(0.00883)	(0.0100)	(0.0257)
	1,847	1,356	896

Robust standard errors in parentheses

Number of observations below standard errors

*** p<0.01, ** p<0.05, * p<0.1

Table C2-25: Impact of conflict on household conditions.

Variables	People per room	Food scarcity index	Household asset index	Toilet	Good walls	Good roof	Monthly household income
Internal binary	0.729**	0.584*	-3.593***	-0.159**	-0.0807**	-0.00499	0.257
	(0.304)	(0.347)	(0.664)	(0.0734)	(0.0376)	(0.0363)	(0.301)
	4,908	5,235	5,303	4,687	4,716	4,713	4,719
Internal continuous	0.502***	0.284	-2.372***	-0.154***	-0.0455**	0.00842	0.201
	(0.178)	(0.222)	(0.332)	(0.0414)	(0.0222)	(0.0249)	(0.168)
	4,908	5,235	5,303	4,687	4,716	4,713	4,719
External binary	-0.133	-0.341	-1.330	-0.302***	-0.0958***	0.0428	-0.322
	(0.430)	(0.456)	(0.929)	(0.0571)	(0.0349)	(0.0379)	(0.186)
	2,336	2,428	2,454	4,687	4,716	4,713	2,272
External continuous	0.0194	-0.0568	-0.198**	-0.0261***	-0.00947**	0.00764*	-0.0659***
	(0.0435)	(0.0458)	(0.0893)	(0.00566)	(0.00457)	(0.00438)	(0.0180)
	2,336	2,428	2,454	4,687	4,716	4,713	2,272

Robust standard errors in parentheses
Number of observations below standard errors
*** p<0.01, ** p<0.05, * p<0.1

Table C2-26: Impact of conflict on Income Generating Activities (IGAs).

Variables	Number of IGAs	Individual monthly income
Internal binary	0.134	-0.123
	(0.0933)	(0.181)
	2,277	2,277
Internal continuous	0.0327	-0.0715
	(0.0518)	(0.115)
	2,277	2,277
External binary	-0.352***	0.237
	(0.0835)	(0.227)
	1,192	1,065
External continuous	-0.0217	0.0240
	(0.0136)	(0.0290)
	1,192	1,065

Robust standard errors in parentheses
Number of observations below standard errors
*** p<0.01, ** p<0.05, * p<0.1

Table C2-27: Impact of conflict on aspirations.

Variables	General anxiety	Ladder position
Internal binary	0.521*	-1.378***
	(0.285)	(0.344)
	2,420	2,416
Internal continuous	0.476***	-0.882***
	(0.124)	(0.222)
	2,420	2,416
External binary	0.973***	-1.134***
	(0.251)	(0.378)
	2,420	2,416
External continuous	0.0914***	-0.106**
	(0.0319)	(0.0408)
	2,420	2,416

Robust standard errors in parentheses
Number of observations below standard errors
*** p<0.01, ** p<0.05, * p<0.1

Table C2-28: Impact of conflict on empowerment.

Variables	Control index	Entrepreneurial potential	Satisfaction
Internal binary	0.153	1.014***	0.0563
	(0.303)	(0.272)	(0.211)
	4,092	4,100	4,065
Internal continuous	0.150	0.643***	-0.000584
	(0.194)	(0.195)	(0.135)
	4,092	4,100	4,065
External binary	-0.0340	1.011***	0.0183
	(0.360)	(0.279)	(0.302)
	1,806	1,805	1,791
External continuous	-0.0156	0.0657	-0.0301
	(0.0423)	(0.0453)	(0.0330)
	1,806	1,805	1,791

Robust standard errors in parentheses
Number of observations below standard errors
*** p<0.01, ** p<0.05, * p<0.1

Table C2-29: Impact of conflict on marriage related outcomes.

Variables	Empowerment	Married	Pregnant	Daughter optimist	Lost pregnancy	Children
Internal binary	0.141	0.0726*	-0.0864***	-0.0342	-0.0455**	-0.00444
	(0.242)	(0.0383)	(0.0299)	(0.0359)	(0.0223)	(0.0424)
	4,209	4,201	4,010	4,250	4,210	4,216
Internal continuous	0.235*	0.0678***	-0.0621***	-0.0309*	-0.0475***	0.00830
	(0.128)	(0.0257)	(0.0180)	(0.0184)	(0.0124)	(0.0259)
	4,209	4,201	4,010	4,250	4,210	4,216
External binary	0.603***	0.197***	-0.123**	-0.0730	-0.117***	0.0618
	(0.212)	(0.0576)	(0.0513)	(0.0506)	(0.0294)	(0.0601)
	4,209	1,843	1,752	1,854	1,836	1,840
External continuous	0.0476**	0.0201***	-0.0149**	0.00357	-0.00949***	0.00230
	(0.0230)	(0.00594)	(0.00661)	(0.00847)	(0.00306)	(0.00525)
	4,209	1,843	1,752	1,854	1,836	1,840

Robust standard errors in parentheses
Number of observations below standard errors
*** p<0.01, ** p<0.05, * p<0.1

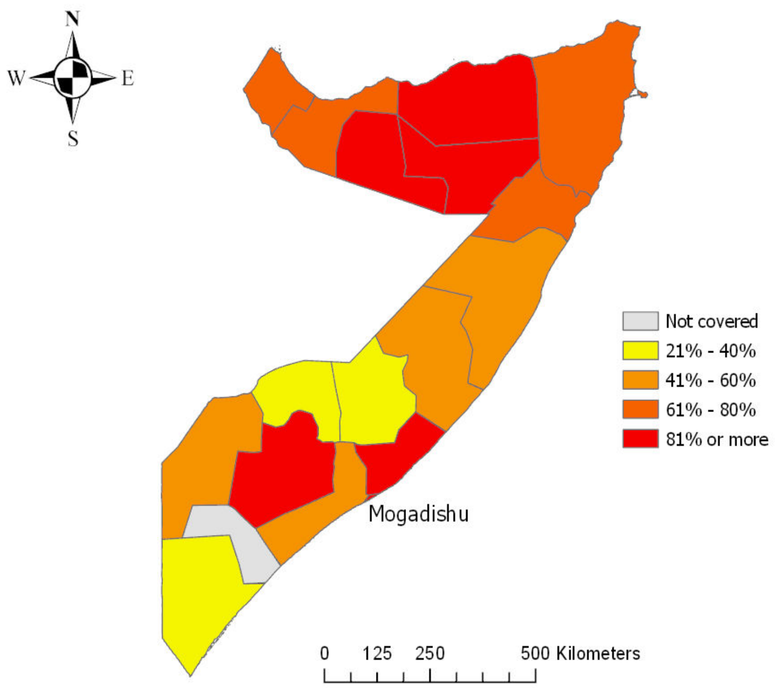
3. Poverty and violence: The immediate impact of terrorist attacks against civilians in Somalia[192]

Gonzalo Nunez-Chaim and Utz Johann Pape[193]

Introduction

Somalia is one of the poorest countries in Sub-Saharan Africa, with 69% of the population living under the standard international poverty line of US$1.90 in 2017-18 (Figure C3-1), with Pape and Karamba (2019) providing a more detailed profile of poverty. In 2004, an interim central state was established with the aim of bringing political stability across Somali regions. The political transition culminated with the establishment of the Federal Government of Somalia in 2012 and a first electoral process in 2017. The elected government has aimed to improve national security conditions, yet the opportunity to ensure a development trajectory still faces many challenges, among them terrorist attacks (World Bank 2018).

Figure C3-1: Poverty incidence in 2017-18 across Somali regions.



Note: The boundaries on the map show approximate borders of Somali pre-war regions and do not necessarily reflect official borders, nor imply the expression of any opinion concerning the status of any territory or the delimitation of its boundaries.

At a first glance, the consequences of a terrorist attack might seem small and contained given that they usually affect a small fraction of the population and the economy. Yet, several studies suggest sizable effects on economic outcomes (Abadie and Gardeazabal 2008). Further, nearly two-thirds of the poor around the world are projected to live in conflict-affected countries by 2030, including Somalia. Therefore, it is important to shed light and improve our understanding on the links between conflict and poverty.

This paper estimates the immediate (within a week) impact of terrorist attacks from Al-Shabaab against civilians in Somalia using micro-data from two waves of the Somali High Frequency Survey

---

[192] GN and UP contributed equally to the manuscript.

[193] Corresponding author: Gonzalo Nunez-Chaim (gnunez1@worldbank.org). World Bank, Poverty and Equity Global Practice, East Africa.

(SHFS), combined with geo-tagged information on attacks.[194] We exploit the spatial and time variation of interviews through a difference-in-difference identification strategy that compares outcomes of control and exposed households, before and after terrorist incidents. We also derive a shift-share instrument using changes in the number of US air/drone attacks against Al-Shabaab and employ an instrumental variables approach. We provide evidence to support the validity of our identification strategies and that our estimates are robust to different specifications, samples considered and several sensitivity checks.

Our results suggest that consumption of households exposed to terrorist incidents decreases by 33%, mainly driven by a decline in food consumption. The reduction in consumption increases poverty and the depth of poverty among the poor. The impact on consumption seems to be associated to a smaller share of household members (aged 15 to 50) working and earning income after an attack. In addition, we document that the negative impact on consumption is clustered within a 4 kilometer radius from the incident and has a heterogeneous impact, not affecting households in the top 20% of the consumption distribution. The perception of police competence also worsens as a result of a terrorist incident.

The literature models terrorists as rational actors, with terrorism having large consequences on economic outcomes, besides the loss of life, damage to persons and negative psychological effects.[195] Conflict can also lead to sharp increases in poverty and vulnerability and other adverse outcomes (Pape, Parisotto et al. 2018). Our findings are in line with the disruption that could be expected from a terrorist attack. We contribute to the literature on the intersection between poverty and adverse shocks in developing countries, as well as to the policy debate by quantifying the impact of terrorist attacks on consumption and poverty, describing which households are affected by such incidents and the mechanisms through which this is likely to occur. Most of the empirical literature on the effects of terrorism on economic outcomes has relied on data aggregated at some geographical level (district, region or country), while the growing body of research exploiting micro-data to understand the effect of various shocks on poverty has not analyzed the effect of terrorism. To our knowledge, this is the first study to measure the causal impact of terrorism on consumption and poverty using household-level data in a fragile and conflict-affected country.

The paper is structured as follows: The next section discusses the related literature on the effects of terrorism and multiple shocks on welfare conditions for households. Section 3 describes the data sources, sample considered and the definition of households exposed to terrorist incidents, besides specifying the identification strategies. Section 4 presents the results and extensions. Section 5 discusses multiple robustness checks and supplementary OLS estimates, while Section 6 contains our concluding remarks.

Literature

Terrorist incidents are different from other types of events since terrorist organizations use violence –or the threat of violence– against civilians as a tool for achieving political change (Crenshaw 1981, Kydd and Walter 2006). Under this characterization, terrorists are rational actors making tactical and strategic decisions while inflicting terror among citizens (Cornish and Clarke 2014). The United States Department of State defined terrorism in 1983 as "means premeditated, politically motivated violence perpetrated against noncombatant targets by subnational groups or clandestine agents, usually intended to influence an audience". Terrorist attacks are then part of a broader strategy with the

[194] Somalia has a strong presence of Al-Shabaab, the largest militant organization seeking to control the territory with the goal of establishing an Islamic State based on its interpretation of Shariah Law. The United States Department of State declared Al-Shabaab a foreign terrorist organization in February 2008. The group has engaged in bombings, suicide attacks and armed assaults.

[195] In Europe alone, the impact of terrorism has been estimated at around €180 billion between 2004 and 2016.

ultimate goal of undermining the government's authority, publicizing an agenda and/or creating a sense of instability (Crenshaw 1981). Therefore, terrorist incidents do not occur at random and are usually clustered in time and space (LaFree, Dugan et al. 2012).

Poverty conditions are thought to be a catalyst to develop and foster terrorist organizations. In fragile and conflict-affected situations, terrorist groups can establish themselves as alternatives to democratically elected governments, especially if governments cannot provide basic services and social safety nets.[196] However, Krueger and Malečková (2003) refute this notion as they find no evidence of a causal connection from poverty to terrorism. The authors claim terrorist activities are more likely to be associated with political conditions and social frustration, than with suboptimal welfare conditions. In line with this finding, Abadie (2004) shows that the risk of terrorism is relatively similar between developed and developing countries, after considering country-specific characteristics, and concludes that the level of political freedom is better at explaining terrorist incidents compared to economic and poverty conditions of the population.

Moreover, the consequences of a terrorist attack could be underestimated as they appear to affect only some parts of the economy. Becker and Murphy (2001) claimed that the attacks on the World Trade Center on September 11th in New York would barely affect economic outcomes since they only represented a loss of 0.06% of the stock of capital in the US. Yet, several studies indicate large effects on the economy. Abadie and Gardeazabal (2008) show that even if the threat of an attack only accounts for a small share of the overall risk, terrorism can have a substantial impact on the allocation of productive capital across nations. The risk of an attack increases uncertainty, reduces expected return to investments and induces a decline in net foreign direct investment (FDI).

Empirical research has documented the effect of terrorism on consumption and gross domestic product (GDP) due to increased uncertainty and disruption to the markets. Eckstein and Tsiddon (2004) use a long time series data and find a decrease in annual consumption per capita of around 5% after a year from a terrorist incident. Fielding (2003) describes how violence in Israel explains reductions on aggregate consumption and savings between 1987 and 1999. Using a synthetic control method, Abadie and Gardeazabal (2003) estimate a decline of 10% in GDP per capita over two decades as a result of terrorism in the Basque Country. Some authors have analyzed the consequences of terrorism on other outcomes. Nitsch and Schumacher (2004) associate terrorist actions with a reduced volume of trade across various countries, while Enders and Sandler (1991) report an annual reduction of FDI inflows by 13.5% and a loss of 140,000 tourists between 1970 and 1988 in Spain. All these negative effects can ultimately limit economic growth of a country. Gaibulloev and Sandler (2009) use panel data for Asian countries and find that terrorist incidents reduce private sector investment and increase government spending, which leads to a decline in GDP per capita growth by 1.5%. Similarly, Ruiz Estrada, Arturo et al. (2018) conclude that terrorist attacks have slowed economic growth in Turkey between 1990 and 2016.

Another stream of the literature has exploited micro-data to investigate the effect of various shocks, such as conflict and weather conditions, on multiple socio-economic characteristics. Several studies document how adverse weather conditions reduce agricultural incomes and can push households into poverty. Hill and Mejia-Mantilla (2017) describe the negative effects from droughts in Uganda. Porter (2012) finds that weather shocks reduce consumption in the long run among rural households in Ethiopia. Similarly, Hill and Porter (2016) conclude that in Ethiopia consumption declined by 9% in

[196] In Somalia, Al-Shabaab has filled this vacuum of political power and service delivery by bringing order and –to some extent– services in regions underserved by the government (Bronwyn 2015).

rural areas due to a moderate drought. For the case of Somalia, Pape and Wollburg (2019) estimate an increase in poverty of 9 percentage points among rural households attributed to a drought shock.

In terms of the drivers of conflict, Dube and Vargas (2006) use municipality-level data in a difference-in-difference framework to understand the dynamics of Colombia's civil war. They find a negative relationship between coffee prices and the incidence and intensity of conflict, while a positive relationship between oil prices and violence. Besides, large evidence supports the adverse impact of conflict-related violence on welfare conditions. Mercier, Ngenzebuke et al. (2016) use household-level panel data for Burundi and conclude that exposure to violence condemns vulnerable households into chronic poverty. Hill and Mejia-Mantilla (2017) find a negative effect from conflict and prices on poverty in Uganda. In South Sudan, Pape (2019a) investigate the effects of conflict-induced cancellation of programs on their designated beneficiaries and describe the welfare status of households displaced by violence. In the same country, Pape and Phipps (2018) analyze the impact of conflict on the socio-economic and psychosocial well-being of teenage girls, including on income opportunities, aspirations and marriage. Other authors provide evidence on how conflict increases the likelihood of chronic poverty due to disruption of income-generating activities and depletion of infrastructure and basic services (Bozzoli and Brück 2009, Bratti, Mariapia et al. 2009, Bozzoli, Brueck et al. 2016).

Empirical analysis

The World Bank implemented Wave 1 (2016) and Wave 2 (2017-18) of the Somali High Frequency Survey to better understand livelihoods, vulnerabilities and poverty across Somali regions.[197] Several terrorist attacks occurred during fieldwork of the SHFS.[198] The analysis exploits detailed household data with dates of interviews and GPS positions to evaluate the immediate impact of attacks against civilians in Somalia. The types of incidents considered correspond to attacks from Al-Shabaab against civilians.[199] We concentrate on measuring the effect within a week due to data availability given that i) the questionnaires used a recall period of 7 days for food items, which is the main component of the consumption aggregate; and ii) only one household was interviewed 8 days or more after an incident.

Data sources

The main sources of data correspond to detailed household information on socio-demographic characteristics, perceptions and poverty conditions from Waves 1 and 2 of the Somali High Frequency Survey, as well as location and time of attacks from The Armed Conflict Location & Event Data Project (ACLED). In this way, household data from the SHFS is combined with geo-tagged information on attacks to identify households exposed to terrorist incidents.

Wave 1 includes 4,117 households interviewed between February and March 2016, which are representative of 9 of the 18 Somali pre-war regions, as the remaining areas were inaccessible for security reasons at the time of fieldwork. Wave 2 expanded the coverage to 17 pre-war regions, interviewing 6,092 households between December 2017 and January 2018.[200] ACLED data records locations and intensity of armed conflict coded by researchers who collect information from secondary sources, NGO reports, local and international news reports and research publications (Raleigh and Dowd 2015). The database is unique due to its geographical level of precision when reporting the latitude and longitude of the attacks, indicating the location of an incident (Figure C3-2). All the violent

[197] Interviews were conducted between the February 10 and March 17, 2016 for Wave 1, while between the December 4, 2017 and January 16, 2018 for Wave 2.

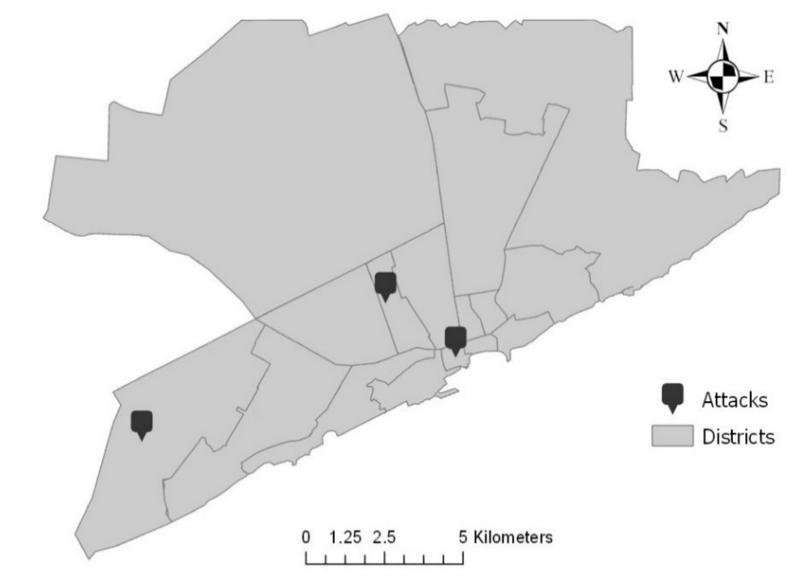
[198] The incidents did not affect data collection plans of Wave 2. The share of completed interviews increases across weeks during fieldwork, as it would be expected.

[199] Al-Shabaab perpetrated attacks outside Somalia for the first time in 2010 in Kampala, and a few years later on the Westgate mall in Nairobi, suggesting that the organization is part of the global network with strong connections to Al-Qaeda.

[200] The SHFS was not designed as a panel survey. Waves 1 and 2 correspond to repeated cross-sections, representative of Somali regions.

incidents reported in ACLED that occurred a week before the start and end of data collection of Wave 2 of the SHFS were considered as consumption data is recorded for the week preceding the interview. Attacks from Al-Shabaab against civilians were manually identified.

Figure C3-2: Terrorist attacks in Mogadishu during data collection of Wave 2.



Note: The boundaries on the map show approximate borders of districts within Mogadishu and do not necessarily reflect official borders, nor imply the expression of any opinion concerning the status of any territory or the delimitation of its boundaries.

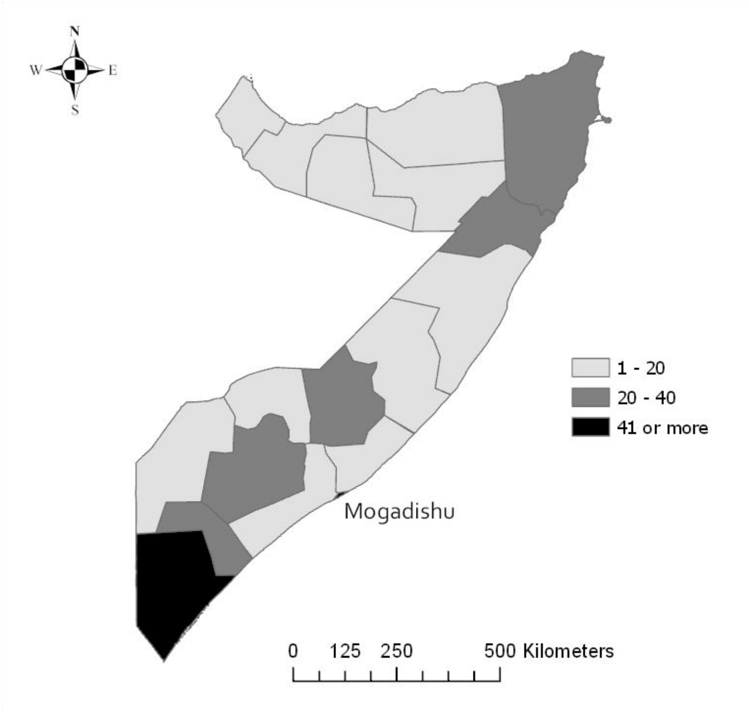
Definition of treatment and sample considered

For the econometric analysis, treated or exposed households are defined as those which meet the following criteria in time and space: i) households whose interview was conducted between 1 and 7 days after an incident occurred during data collection of Wave 2; and ii) those within a radius of 1 kilometer from the terrorist attack.[201]

From this definition we identify four groups of households. Exposed households before (2016) and after a terrorist attack (2017-18). The latter group corresponds to exposed households in Wave 2 meeting the space and time criteria, while the former group to households that are also located within 1 kilometer from the incident, but that were interviewed in Wave 1.[202] Similarly, we identify control households before and after an attack. Control households in Wave 2 are those interviewed in 2017-18 that do not meet the time and space criteria of treatment or exposed status, while control households in Wave 1 are those located more than 1 kilometer away from the incidents and that were interviewed in 2016.[203] Moreover, some households interviewed in Wave 1 were also close in time and space but to incidents that occurred in Wave 1. Those Wave 1 households interviewed up to a week after the incidents and within a 10 kilometers radius are excluded from the analysis as they are likely to be affected by a terrorist incident in a previous period.[204]

[201] We consider a period of 7 days because the consumption module of the questionnaire from the SHFS used a recall period of 1 week for food items, which is the main component of the consumption aggregate.
[202] The average distance from incidents is 1.09 kilometers for exposed households interviewed in Wave 1, while 1.12 kilometers for exposed households interviewed in Wave 2.
[203] The definition of control group does not consider an upper bound limit in terms of how far households are located from incidents.
[204] We adopt a cautionary approach and consider a larger radius of 10 kilometers to avoid including 'contaminated' households by previous incidents in our Wave 1 control group.

Figure C3-3: Number of violent incidents during data collection of Wave 2.



Note: The boundaries on the map show approximate borders of Somali pre-war regions and do not necessarily reflect official borders, nor imply the expression of any opinion concerning the status of any territory or the delimitation of its boundaries.

During fieldwork of Wave 2, a large proportion of violent incidents took place in Mogadishu, since it is home to potential targets such as government actors and international organizations (Figure C3-3). In particular, terrorist attacks from Al-Shabaab during data collection were concentrated in urban areas, mainly in Mogadishu. Furthermore, a security assessment was carried out before data collection of the SHFS and incorporated into the sampling frame such that Primary Sampling Units (PSUs) were drawn only from accessible areas, to ultimately ensure PSUs visited by enumerators were safe on the day of fieldwork (Pape, Beltramo et al. 2019b). As a result of both, the geographical clustering of incidents and the sampling strategy of the survey, only in two Somali regions interviews were conducted during Wave 2 after an incident and close to it (Table C3-7 in the Appendix).

Table C3-1: Number of households by group and Wave for each sample alternative.

Alternative		Group	Wave 1	Wave 2
(1)	Mogadishu	Exposed	21	113
		Control	664	775
(2)	Mogadishu with overlapping exposed households in Wave 1 and 2	Exposed	21	78
		Control	664	775
(3)	Mogadishu with overlapping districts in Wave 1 and 2	Exposed	21	113
		Control	519	775
(4)	All urban areas	Exposed	21	135
		Control	2,712	3,876
(5)	Urban areas with exposed and control households in Wave 1 and 2	Exposed	21	135
		Control	664	1,468

The main sample considered in the econometric analysis corresponds to Mogadishu. The capital of Somalia is one of the most fragile cities in the world (Pape and Karamba 2019). It concentrates 16 percent of Somali households and poverty is higher in Mogadishu than in other urban areas of Somalia. A few additional samples are used to provide further robustness to the results from the econometric analysis (Table C3-1). We consider a few variations within Mogadishu; one restricting the group of exposed households to overlapping Wave 1 and 2 areas (Figure C3-6 in the Appendix), and another option restricting the sample to overlapping Wave 1 and 2 districts. Then, we consider all urban households across Somali regions. This alternative includes exposed households from Mogadishu and South West urban, as well as control households from all urban areas of Somalia. Finally, the other sample considered in the econometric analysis refers to only urban areas with exposed households in Wave 2; that is, exposed and control households only from Mogadishu and South West urban.

All these different alternatives have a relatively small sample size for the group of exposed households in Waves 1 and 2. This is determined by the location and timing of interviews in relation to attacks. Further, the sampling strategy clustered households into PSUs, with a target of 12 interviews per PSU. Households within each of these geographical areas are likely to have a similar set of characteristics. Our sample of households from Mogadishu belongs to four PSUs in Wave 1 and ten in Wave 2. The small size of our exposed group of households introduces an important caveat to our estimates and findings. Nonetheless, we try to ease such concerns by considering different identification strategies and robustness checks.

Identification strategy

Using the four groups of households identified (exposed and control households before and after the incidents), we first employ a difference-in-difference (DiD) approach with repeated cross-sections to compare outcomes of households exposed to the terrorist attack against households who were not exposed to the incidents (Imbens and Wooldridge 2007). The identification strategy exploits spatial and time variation of the data. For this, we estimate the following equation:

$$Y_{it} = \alpha + \lambda Wave_t + \mu Exposed_i + \beta(Wave_t * Exposed_i) + \eta L_a + \varphi X_{it} + u_{it} \quad (5)$$

where Y_{it} refers to the outcome of interest for household *i* in period t, $Wave$ to the period t of data collection (Wave 1 in 2016 or Wave 2 in 2017-18). $Exposed$ corresponds to the status of household *i* according to our definition of treatment, L_a to location fixed effects for geography *a* –which refers regions or districts– and X_{it} to a vector of covariates capturing characteristics of the household and household head, dwelling characteristics, exposure to drought, as well as humanitarian aid received.[205]

Moreover, the probability of being exposed to an incident is likely to be associated to the location of households and the composition of regions or districts. Hence, we focus on regional comparisons (i.e., within-location variation) through the inclusion of location fixed effects.[206] For the DiD analysis we use the sample of households from Mogadishu as we can include fixed effects at the district level –which is the lowest geographical level available– providing a more precise comparison of households, as opposed to including region fixed effects.[207] Yet, we expand the analysis to the other samples as a robustness check.

Our empirical strategy relies on the parallel trend assumption of the difference-in-difference approach. It assumes that the difference in the outcome among exposed and control households would be similar had the attacks not occurred. Any difference between treatment and control is unlikely to have changed over 22 months –between Waves 1 and 2– since the identification strategy compares households within districts of Mogadishu and any other shock is likely to have affected in a similar way both exposed and control households. We also provide evidence to support the conditional independence assumption. Table C3-8 in the Appendix presents an OLS regression for the exposure of households to attacks in Mogadishu. The consumption level of households, their location and socio-economic characteristics are not associated with the propensity from being exposed to an incident. This suggests that terrorist incidents are likely to be exogenous, conditioned on location fixed effects and household characteristics.

In addition, we employ an instrumental variables (IV) identification strategy to further validate the results from the difference-in-difference approach. For this, we obtain a shift-share type of instrument based on (Bartik 1991), exploiting the spatial variation of incidents against civilians and changes in the number of US air/drone attacks against Al-Shabaab in Somalia. US air/drone attacks against Al-Shabaab were manually identified from ACLED data among all events recorded between Waves 1 and 2 of the SHFS. The US air/drone activity against Al-Shabaab is mainly concentrated on South West Somalia, with the number of attacks increasing from 4 in the first semester of 2015 to 21 in the last semester of 2017 (Figure C3-7 in the Appendix).

To support the validity of the instrument, we examine the location of both, US air/drone strikes against Al-Shabaab, and attacks from Al-Shabaab against civilians in Somalia, with an emphasis on Southern Somalia (Figure C3-8 in the Appendix). The locations of these two types of incidents are not spatially correlated. US air/drone attacks occurred in locations which are different to those of incidents against

[205] Drought exposure corresponds to drought affected status from the Standardized Precipitation Index. Humanitarian aid is the percentage of beneficiaries reached through food aid and livelihood inputs by region.

[206] It is unlikely the composition of neighborhoods changed substantially after a terrorist attack in the period of analysis. Only 1.6% of the Wave 2 sample of households in Mogadishu reported they were forced to leave their previous place of residence and moved to another region.

[207] The delimitation of PSUs was different in Wave 1 and Wave 2, in part due to the lack of census data in Somalia. As a result, PSUs from Wave 1 and Wave 2 are not comparable nor mutually exclusive. Therefore, the lowest level of aggregation comparable across surveys corresponds to districts in Mogadishu.

civilians. This could be because US air/drone strikes usually target high-profile members of Al-Shabaab and operation centers that are based in areas they already control, which is less often the case for the location of attacks against civilians.[208] US air/drone attacks are thus likely to increase the number of attacks from Al-Shabaab against civilians –as terrorist organizations rely on violence to achieve political gains– partially explaining the exposure of households to terrorist attacks, while being independent from locations where Al-Shabaab commits an attack.

For each region in Somalia, the instrument was derived as the product of i) the exposure to attacks against civilians, and ii) the rate of growth of US air/drone attacks against Al-Shabaab. We first obtain the 'initial' share of attacks on civilians as the proportion of incidents in each region from the total number of events in the period covering from Wave 1 to the mid-point between Waves 1 and 2. We then obtain the rate of growth of US air/drone attacks against Al-Shabaab between this period and the end of Wave 2 for each region. Finally, we obtain the instrument from multiplying the exposure of attacks on civilians (share) and the rate of growth of US air/drone attacks against Al-Shabab (the shift).

The IV strategy estimates the causal effect of incidents through the variation in the probability of households being exposed to an incident in Wave 2 explained by the shift-share instrument. In this way, the first stage of the IV approach is the following:

$$P_i = \omega + \theta V_i + \eta L_i + \gamma X_i + u_i \tag{6}$$

where P_i measures the likelihood of being exposed to an incident in Wave 2 for household *i*, while L_i and X_i corresponds to the same set of covariates included in the DiD approach (i.e., fixed effects, characteristics of the household and household head, dwelling characteristics, exposure to drought and humanitarian aid). V_i refers to the shift-share instrument, capturing the exposure to attacks against civilians and the growth of US air/drone attacks against Al-Shabaab, which takes the same value for all households within the same region. The second stage regression for any outcome Y corresponds to:

$$Y_i = \alpha + \delta \hat{P}_i + \eta L_i + \rho X_i + \varepsilon_i \tag{7}$$

where $\hat{P}_i$ is the predicted likelihood of household *i* being exposed to an incident in Wave 2. In this context, the coefficient of interest is δ which provides an estimate of the average effect of terrorist attacks against civilians. To obtain IV estimates we cannot use the sample of households from Mogadishu since the instrument is calculated at the region level and does not vary across households from the capital. Thus, we use the group of households from urban areas with exposed households in Wave 2 as our main IV sample, which includes Mogadishu and South West urban. However, we also expand the analysis to all urban areas as a robustness check.

Our shift-share instrument and the likelihood of households being exposed to an incident in Wave 2 show a quadratic pattern (Figure C3-9 in the Appendix). Therefore, we consider a quadratic term of the instrument in all our IV specifications. Table C3-2 presents the first stage of the IV regression. Column 1 is our basic specification, which only includes fixed effects and uses sampling weights to derive standard errors. Our most complete specification, including the full set of controls, corresponds to column 4. The estimated coefficient for the quadratic term of the shift-share instrument is always positive and statistically significant at the 1% level. Column 5 presents the coefficient with standard errors clustered at the PSU level. In all cases the F-statistic is greater than 10, which is the cutoff value

[208] The lack of spatial correlation between the location of US air/drone strikes against Al-Shabaab and attacks from the latter against civilians, as well as the different nature of these incidents suggest it is unlikely that causality runs in the opposite direction, such that US air/drone strikes target regions with high activity from Al-Shabaab against civilians.

for considering the instrument as weak. Overall, the instrument is strong at explaining the likelihood of being exposed to an incident in Wave 2 across these specifications among households in the main IV sample.

Table C3-2: First stage of the instrumental variables approach.

	Urban areas with exposed and control households in Wave 1 and 2				
	(1)	(2)	(3)	(4)	(5)
Instrument (θ)	0.0001*** (<0.001)	0.0001*** (<0.001)	0.0001*** (<0.001)	0.0001*** (<0.001)	0.0001*** (<0.001)
Fixed effects	Region	Region	Region	Region	Region
Characteristics of household & head	No	Yes	Yes	Yes	Yes
Dwelling characteristics	No	No	Yes	Yes	Yes
Drought affected status	No	No	No	Yes	Yes
F-statistic	34.9	34.6	37.7	49.8	10.7
Standard errors	Sampling weights	Sampling weights	Sampling weights	Sampling weights	Clustered by PSU
Observations	2,272	2,249	2,241	2,241	2,241

*Note: Estimated coefficients from the 1st stage of the IV regression. The dependent variable corresponds to exposure to incidents in Wave 2. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, sex and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Humanitarian assistance is not included due to collinearity with the region fixed effects. Standard errors in parenthesis; *** p<0.01, ** p<0.05, * p<0.1.*

In terms of statistical inference, the estimates from both the difference-in-difference and instrumental variables approaches are obtained with standard errors that consider the sampling weights of the surveys in all specifications, mainly due to the different sampling frame and design between Wave 1 and 2. One of the key differences is that the sampling strategy of Wave 2 included an oversampling of households in Mogadishu (Pape, Beltramo et al. 2019b). Hence, using sampling weights to derive the standard errors is needed to correct for the endogenous sampling and avoid obtaining inconsistent estimates (Solon, Haider et al. 2013). Yet, we expand the analysis and consider clustered standard errors at the PSU level, since household-level error terms within these small geographical units are likely to be correlated, given households could have a similar set of characteristics. Further, we also derive heteroskedasticity and autocorrelation-consistent (HAC) standard errors to account for spatial correlation in the data, in line with Conley (1999) and Conley (2010).

Results and extensions

For the impact on household consumption in Mogadishu, Table C3-9 in the Appendix presents various specifications with different covariates considered. Our preferred specification corresponds to column 6, which includes district fixed effects, characteristics of the household, household head and dwelling. The results indicate a decline of 33% in core consumption –an aggregate that includes both food and non-food items– after a week caused by a terrorist attack from Al-Shabaab.[209]

[209] The SHFS used a rapid consumption methodology where only a group of core food and non-food items, identified based on their consumption share, were asked to every household, while the rest of the items were algorithmically partitioned into optional modules and randomly distributed across households. After data collection, consumption of optional modules was imputed for all households. Thus, we use core consumption per capita deflated and not total imputed consumption in the econometric analysis. The core consumption aggregate represents around 75% of total consumption of Somali households.

Table C3-3 presents the estimates for various outcomes using our preferred DiD specification from equation (1) for the sample of households in Mogadishu. The negative effect of 33% on consumption (per capita deflated) from terrorist attacks seems to be concentrated on food items, as the results suggest an immediate negative effect on food consumption of around 42%.[210] Furthermore, for some households this decline in consumption brings their expenditure level below the poverty line, ultimately increasing the proportion of population living in poverty, as indicated by a positive and significant coefficient from the respective probit regression (column 3 in Table C3-3). This estimate implies an average increase of 0.3 point in the predicted probability of exposed households being poor. Among poor households, the negative effect on consumption results in consumption levels further from the poverty line due to the disruption caused by the incident. The poverty gap increases by 12% (column 4); that is, the average difference between consumption levels and the poverty line – measured as a proportion of the poverty line– increases by 12% among the poor in Mogadishu.

Table C3-3: DiD estimates for the effect of terrorist attacks against civilians in Mogadishu.

	Log of core consumption (1)	Log of food consumption (2)	Poverty status (3)	Poverty gap (4)	Experienced hunger (5)	Police competence (6)
Diff-in-diff coefficient (β)	-0.326*** (0.118)	-0.415** (0.159)	1.617*** (0.597)	0.115*** (0.039)	0.813 (0.659)	-0.881** (0.365)
Wave	Yes	Yes	Yes	Yes	Yes	Yes
Exposed/control	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics of household & head	Yes	Yes	Yes	Yes	Yes	Yes
Dwelling characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,532	1,532	1,532	1,532	1,516	1,498

*Note: Estimated coefficients from an OLS or probit regression. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, sex and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Drought affected status and humanitarian assistance are not included due to the lack of variation in the data within Mogadishu. Standard errors considering sampling weights in parenthesis; *** p<0.01, ** p<0.05, * p<0.1.*

In terms of the effect of incidents on self-reported outcomes, the coefficient for experiencing hunger is positive but not statistically significant (column 5 in Table C3-3). A reduction of food consumption is likely to be associated with a larger share of household reporting to have experienced hunger. However, a non-significant result could be explained by the recall period considered in the survey instrument. The question asked to households referred to whether they had experienced hunger over the last month. As such, hunger reported by households covers between 1 and 7 days after an incident and at least 3 weeks before. We also find a deterioration of perception on police competence (column 6 in Table C3-3). The predicted probability of households perceiving police as being competent decreases on average by 0.34 point among exposed households in Mogadishu.

For the IV estimates, we use the same preferred specification as in the DiD approach, which includes fixed effects, characteristics of the household, household head and dwelling, as well as the drought affected status of households. The results also show a negative immediate effect on core consumption

[210] The food consumption aggregate represents almost 70% of total consumption of Somali households.

(per capita deflated) attributable to terrorist attacks from Al-Shabaab against civilians (column 1 in Table C3-4).

Overall, point estimates are larger with the IV approach compared to DiD estimates. Yet, the former estimates are less precise and have larger standard errors in the second stage since it only considers part of the variation in the treatment status of households, which is induced by the instrument (the exposure to attacks combined with the rate of growth of US air/drone attacks against Al-Shabaab). Nevertheless, the IV estimates (Table C3-4) reinforce the DiD results: a negative immediate effect on consumption in urban areas with exposed and control households in Waves 1 and 2, mainly driven by a reduction of food consumption. The decline in consumption also increases the share of population with a consumption level below the poverty line, which has a similar magnitude between the IV and DiD estimates.

Table C3-4: IV estimates with exposed and control households in Wave 1 and 2

	Log of core consumption (1)	Log of food consumption (2)	Poverty status (3)	Poverty gap (4)	Experienced hunger (5)	Police competence (6)
IV coefficient (δ)	-1. 715*** (0.540)	-1.924*** (0.692)	1.564*** (0.371)	0.248 (0.208)	2.688*** (0.667)	-0.710 (0.478)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics of household & head	Yes	Yes	Yes	Yes	Yes	Yes
Dwelling characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Drought affected status	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,241	2,241	2,241	2,241	2,234	2,167

*Note: Estimated coefficients from an IV regression. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, sex and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Humanitarian assistance is not included due to collinearity with the region fixed effects. Standard errors considering sampling weights in parenthesis; *** p<0.01, ** p<0.05, * p<0.1.*

The estimated coefficient for the effect on the poverty gap is positive and larger with the IV approach (Table C3-4), compared to the DiD result (Table C3-3). Despite this, the IV estimate is not statistically significant. Contrary to this result, the IV coefficient of hunger is positive and statistically significant, unlike the DiD estimate. For the perception of police competence, the IV result also suggests a negative effect, as our DiD estimate, but the coefficient is not significant. Even though the IV and DiD coefficients are estimated from different samples of households, the differences in results seem to be related to larger standard errors from the IV approach. For the effect of incidents on the poverty gap and police competence, the 95% confidence interval of the DiD coefficient lies within the confidence interval of the respective IV estimate. The imprecision of IV estimates, combined with the fact that the point estimates are consistent on the direction on the effect, could suggest the result might not be different between DiD and IV, ultimately pointing to a positive effect of incidents on the poverty gap and a deterioration of perception of police competence.

Table C3-5: DiD and IV estimates for the effect on employment and earnings.

	Proportion of household members (aged 15-50) employed in the previous week		Proportion of household members (aged 15-50) earning income in the previous week	
	DiD (1)	IV (2)	DiD (3)	IV (4)
DiD or IV coefficient	-0.147** (0.066)	-1.952*** (0.326)	-0.150* (0.086)	-1.682*** (0.285)
Fixed effects	District	Region	District	Region
Full set of controls	Yes	Yes	Yes	Yes
Observations	1,525	2,221	1,525	2,221

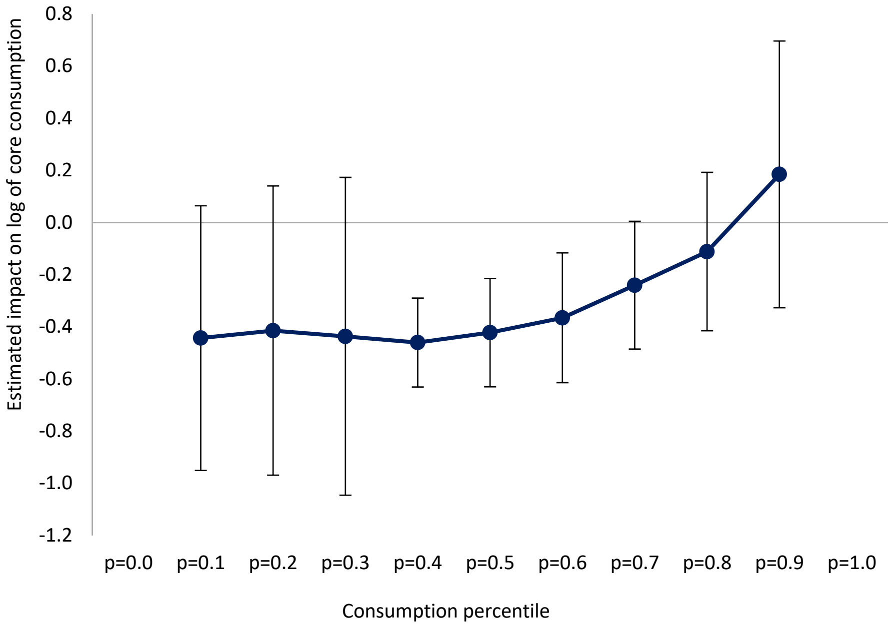
*Note: Estimated coefficients from an OLS regression for Mogadishu and IV regression for urban areas with exposed and control households in Wave 1 and 2. Estimates obtained from our most complete specification including household size, receiving remittances, age, sex and literacy of the household head, tenure, roof and floor material of the dwelling, water source and sanitation type. Drought affected status and humanitarian assistance are not included in the OLS regression due to the lack of variation in the data within Mogadishu. Humanitarian assistance is not included in the IV regression due to collinearity with the region fixed effects. Standard errors considering sampling weights in parenthesis; *** p<0.01, ** p<0.05, * p<0.1.*

We further explore the mechanisms through which a terrorist attack from Al-Shabaab against civilians can lead to a decrease in consumption. Table C3-5 presents the results for the impact on the proportion of household members aged 15 to 50 that were employed and earned income after an incident. The DiD coefficients are obtained from the sample of households in Mogadishu, while the IV coefficients from urban areas with exposed and control households in Waves 1 and 2; both from our most complete specification. All estimates are negative and statistically significant. Similar to other outcomes, IV point estimates are larger and have bigger standard errors. However, the IV estimates also reinforce the DiD findings, which indicate a decrease of around 15% in the share of both, household members working and earning income between 1 and 7 days after an incident. A reduction in employment and income could lead to a decline in consumption and exacerbate poverty and vulnerability among households exposed to an incident.

There are other supply-side mechanisms, such as limited availability of food items and higher prices, which could help explain how a terrorist attack disrupts the economy and affects welfare conditions of the population. To assess this, we compare the cost of the consumption basket –made out of 38 core food items– of exposed households from Mogadishu in Wave 2 against that from a group of households also in Wave 2 located within 1 kilometer from incidents in Mogadishu but that were interviewed before the attacks. The consumption basket of exposed households was 3% more expensive for these items, providing some evidence that higher food prices could also be another relevant mechanism through which a terrorist incident affects consumption levels.[211]

[211] It is unlikely that the difference in cost of consumption baskets is explained by a seasonal pattern as all interviews of households in these groups were conducted within a period of 13 days.

Figure C3-4: Distributional effect across consumption percentiles in Mogadishu.



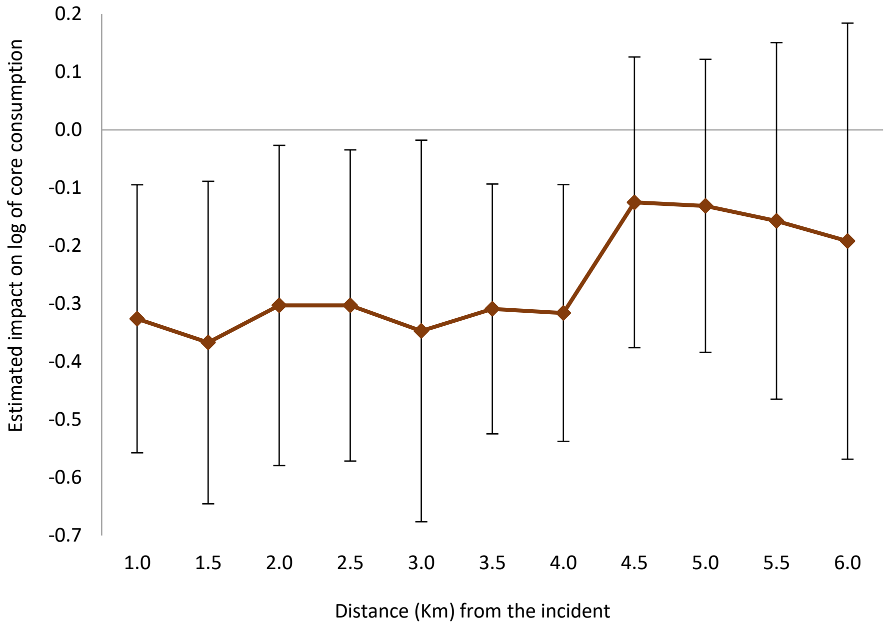
Note: The vertical lines depict 95% confidence intervals.

We extend the analysis to investigate how the impact on consumption differs across consumption percentiles and the spatial variation of this negative effect.[212] For these extensions, we use our most complete specification from the DiD approach, the sample of households from Mogadishu and the sampling weights of surveys. Figure C3-4 plots the point estimates for the effect of incidents on core consumption from a quantile regression. For the negative and immediate impact of terrorist attacks on consumption, we find a heterogeneous effect across different parts of the consumption distribution. The point estimates increase with the consumption decile. Incidents affect exposed households from most of the consumption distribution, except those in the top 20%. The estimates for deciles 1, 2 and 3 are either significant at the 10% level or not significant because these groups are underrepresented in the survey sample considered.[213] Most of the households affected experienced a decrease in consumption of similar magnitude and mainly correspond to poor households since the incidence of poverty in Mogadishu was 74% in 2017-18. Households in the top 20% of the consumption distribution are likely to have savings or other sources of income, allowing them to smooth the shock from a terrorist attack and preventing them from reducing their consumption levels.

[212] We cannot extend the analysis to measure the effect beyond 7 days after an incident occurred as only one exposed household was interviewed in this period during data collection of Wave 2.

[213] Households in the bottom 30% of the consumption distribution accumulate 40% of the sum of sampling weights from the total sample, ultimately indicating they are underrepresented in the survey data.

Figure C3-5: Spatial variation of the impact on consumption in Mogadishu.



Note: The vertical lines depict 95% confidence intervals.

Finally, we also relax the spatial criterion for the definition of treated households and classify as exposed or treated those households located from 1 to 6 kilometers away from the incidents.[214] Figure C3-5 presents the estimated coefficients for the effect on core consumption, considering as exposed those households located within the radius of each cutoff point.[215] The impact on consumption is similar for households located between 1 and 4 kilometers from the incident. After this threshold, the estimates become insignificant. The results suggest the immediate negative effect is clustered within a 4 kilometer radius from the attack. Households located within this radius suffer a decrease in consumption of similar magnitude. Conversely, those households located more than 4 kilometers away from an incident seem to be far enough, such that their consumption levels are not directly affected within a week. The impact encompasses around 10% of the area of Mogadishu and 25% of its population. Only part of the city is affected within a week, which could be associated to a localized disruption of roads and markets. Also, households located further away from the attack could still be affected after a week.

Robustness checks and additional OLS estimates

The estimated effect on consumption from our preferred DiD specification (-33%) is robust to i) the use of clustered standard errors at the PSU level and HAC standard errors (columns 3 and 4 of Table C3-9 in the Appendix); ii) seasonal patterns, after including year-month fixed effects (column 5 of Table C3-9 in the Appendix); and iii) the different samples and control groups considered (Table C3-10 in the Appendix).[216] The IV coefficients are also robust to the inclusion of clustered standard errors at the PSU level and HAC standard errors, as well as to the different samples and control groups considered; our main IV sample, composed of urban areas with exposed and control households in Waves 1 and 2, and all urban areas (Table C3-11 in the Appendix).

[214] We only consider households between 1 and 6 kilometers since there are no Wave 2 households interviewed beyond a radius of 6 kilometers from where the attacks took place.

[215] The cutoff points of the different spatial criteria used to identify exposed households correspond to geodetic distances from incidents.

[216] All estimates with HAC standard errors consider a spatial correlation within 0.5 km to allow for variation in the group of exposed households, which is defined as those within a radius of 1 km from terrorist incidents.

In addition, there is a low risk of obtaining biased DiD results due to compositional differences from using a repeated-cross section (Waves 1 and 2) representative of Mogadishu. Table C3-12 in the Appendix shows that the composition of the sample is relatively similar with respect to time-invariant characteristics when comparing exposed and control households between Wave 1 and Wave 2. Besides, we conduct a falsification test, measuring the impact before the events occurred. For this, we use the same definition for each group, and estimate equation (1) but substituting exposed households with those that were interviewed in Wave 2 before the incident took place (Table C3-13 in the Appendix). The results indicate no impact on this group of households, validating the DiD empirical strategy and the results.

To further support our findings, we employ another alternative empirical strategy. We restrict the analysis to households from Mogadishu in Wave 2 and compare exposed households –using the same definition in time and space– against a control group made of only those located within a 1 kilometer radius from an incident but that were interviewed before the attack. This alternative includes a sample of 113 exposed and 67 control households from Mogadishu in Wave 2. All households are located within the same distance from the attacks. The difference between exposed and control households is the timing of their interview in relation to when an attack occurred. However, fieldwork of Wave 2 is likely to have followed a geographical pattern or strategy when conducting interviews, due to logistical considerations. Thus, exposed and control groups and unlikely to be random but determined by the data collection schedule, which implies that households interviewed before and after the incidents are likely to be different. Yet, we find that they do not differ much in terms of observable characteristics. Exposed and control groups are relatively comparable or balanced on key observable dimensions (Table C3-14 in the Appendix).

Table C3-6: OLS estimates of Wave 2 interviewed before and after the incidents.

	Log of core consumption (per capita deflated)			
	(1)	(2)	(3)	(4)
Exposed/control	-0.215* (0.107)	-0.458*** (0.117)	-0.458*** (0.142)	-0.458*** (0.124)
District fixed effects	Yes	Yes	Yes	Yes
Characteristics of household & head	Yes	Yes	Yes	Yes
Dwelling characteristics	Yes	Yes	Yes	Yes
Standard errors	Sampling weights	Clustered by district	Clustered by PSU	HAC
Observations	180	180	180	180

*Note: Estimated coefficients from an OLS regression. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, gender and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Drought affected status and humanitarian assistance are not included due to the lack of variation in the data within Mogadishu. Standard errors in parenthesis; *** p<0.01, ** p<0.05, * p<0.1.*

We estimate a linear regression model using ordinary least squares (OLS). The coefficient of interest corresponds to the exposed or control dummy variable capturing the effect of terrorist incidents. Our specification includes district fixed effects to account for time-invariant unobservable factors. Also, there is a small risk that time-varying unobservables could affect the estimates as all interviews were conducted within a 6-week period. This alternative strategy produces similar results for Mogadishu, which are also robust to the inclusion of clustered standard errors and HAC standard errors (Table C3-6). The OLS estimate suggests a decline in core consumption of 22% attributable to attacks that occurred during data collection of Wave 2. Our main DiD estimate from the preferred specification (-33%) lies within the 95% confidence interval of this OLS point estimate.

Conclusions

After more than two decades of civil war and conflict, Somalia remains a fragile state subject to conflict and violence. The Federal Government of Somalia aims to provide the political and security conditions for improving the development trajectory of the country and increasing the welfare conditions of its population. The challenge of improving security conditions will be larger in the coming years as countries participating in The African Union Mission in Somalia (AMISOM) are considering whether to withdraw from Somalia.[217] Terrorist groups and their attacks are one of the threats to the government and its stability, representing a risk for the well-being of the population and limiting the capacity of the government and international partners to design and implement effective development policies.

This paper documents the immediate (within a week) impact of terrorist attacks from Al-Shabaab against civilians in Somalia. We combine micro-data from two waves of the Somali High Frequency Survey and employ a difference-in-difference approach comparing outcomes of households exposed to terrorist attacks against households who were not exposed to the incidents, before and after the events. Our estimates are robust to the use of clustered and HAC standard errors, different samples and control groups considered, besides that a similar composition of repeated-cross sections and a falsification test –measuring the impact before the events occurred– support the validity of our empirical strategy. We further confirm the results through an instrumental variables approach, for which we obtain a valid shift-share type of instrument that exploits the spatial variation of incidents and changes in the number of US air/drone attacks against Al-Shabaab.

Our estimates indicate a sizable immediate effect on consumption for households exposed to attacks with a decrease of 33%, mainly driven by a decline in food consumption. For some households, the reduced consumption brings their expenditure level below the poverty line, increasing the share of poor population. Among the poor, the negative effect results in consumption levels further away from the poverty line. The impact on consumption seems to be explained by a smaller share of household members (aged 15 to 50) working and earning income after an attack. In addition, we document that the negative impact on consumption is clustered within a 4 kilometer radius from the incident and has a heterogeneous effect, not affecting households in the top 20% of the consumption distribution. Besides, OLS estimates –comparing Wave 2 households in Mogadishu before and after incidents but all within 1 kilometer from them– further support our findings. We also find that perceptions of police competence worsen, which could erode trust in formal institutions and ultimately hinder the government's legitimacy and capability for implementing policies.

The results are in line with the disruption that could be expected from a terrorist attack against civilians. However, these findings cannot necessarily be extrapolated to other contexts or periods due to differences in the size, structure and operation of other criminal organizations. The stage and duration of the conflict could also lead to different results. Moreover, we only capture immediate impacts due to data limitations. Further research is needed to assess if the effect on consumption and poverty is transitory or permanent and if it varies depending on whether attacks are more common in certain regions, as well as to understand displacement and other longer term effects on welfare.

Nearly two-thirds of the world's poor will be concentrated in conflict-affected countries by 2030. Therefore, it is important to shed light and improve our understanding on the links between conflict and poverty. We contribute to the literature and policy debate by quantifying the impact on

[217] The African Union Mission in Somalia is an active, regional peacekeeping mission operated by the African Union with the approval of the United Nations Security Council.

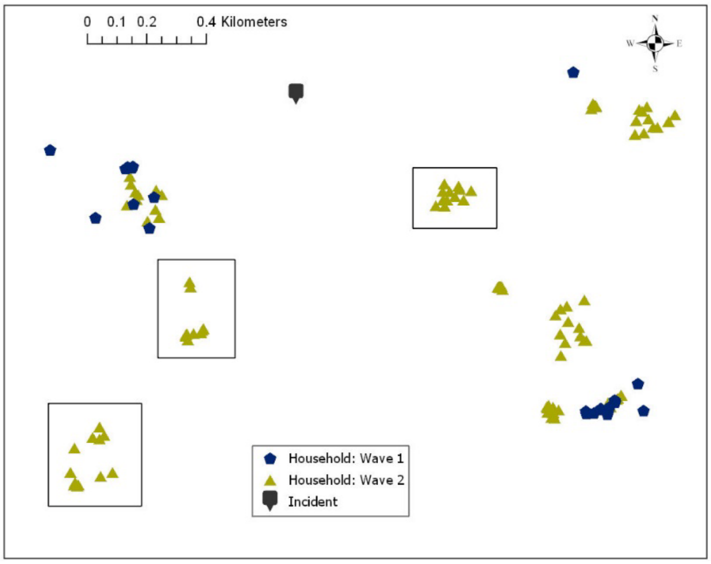
consumption and poverty, describing which households are affected by such incidents and the mechanisms through which this is likely to occur. A terrorist attack against civilians can lead to increases in poverty and vulnerability, among other adverse outcomes. In this context, policies could provide support to affected households through a combination of cash and in-kind food assistance to ameliorate the sharp decrease in consumption, mainly of food items. Beneficiaries can be identified using geographical targeting, covering those within 4 kilometers from the incident. Effective labor market interventions that support continuous employment could help by providing certainty and stability to households' incomes. Nevertheless, accelerating poverty reduction will be challenging until security conditions improve. Al-Shabaab has filled a vacuum of political power and gained control over several towns and villages across Somalia. National and international efforts should prioritize achieving peace, which is a fundamental first step for increasing welfare conditions that will also bring other wider long-term benefits in Somalia.

Appendix

Table C3-7: Wave 2 exposed households by urban region.

Somali region	Number of exposed households
Mogadishu	113
North-east Urban	0
North-east Urban	0
Central regions	0
Jubbaland Urban	0
South West Urban	22

Figure C3-6: Interviewed households closed to an incident in Wave 1 and 2.



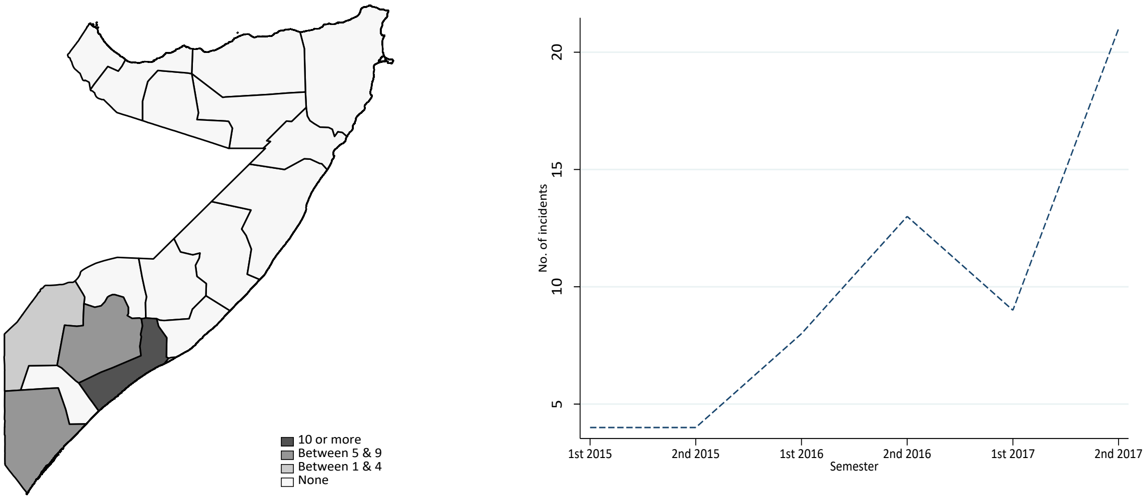
Note: Solid rectangles correspond to Wave 2 exposed households without an overlapping cluster of Wave 1 households in Mogadishu.

Table C3-8: Correlates of terrorist attacks.

	Exposure of households to Wave 2 incidents		
	(1)	(2)	(3)
Log of core consumption (per capita deflated)	0.346 (0.326)	0.029 (0.036)	0.029 (0.032)
No. of members in the household	0.002 (0.007)	0.007 (0.007)	0.007 (0.006)
Household head: sex	-0.048 (0.068)	-0.042 (0.041)	-0.042 (0.036)
Household head: literate	0.013 (0.051)	0.042 (0.037)	0.042 (0.036)
Received remittances	-0.030 (0.051)	-0.033 (0.029)	-0.033 (0.025)
Tenure of the dwelling	-0.066 (0.072)	-0.042 (0.093)	-0.042 (0.087)
Floor of cement	0.066* (0.039)	0.038 (0.039)	0.038 (0.036)
Roof of metal	0.040 (0.034)	0.001 (0.045)	0.001 (0.038)
Access to piped water	-0.029 (0.074)	0.059 (0.046)	0.059 (0.044)
Improved sanitation	0.013 (0.039)	0.022 (0.043)	0.022 (0.042)
District in Mogadishu	Yes	Yes	Yes
Standard errors	Sampling weights	Clustered by PSU	HAC
Adjusted R-squared	0.062	0.017	0.017
Observations	885	885	885

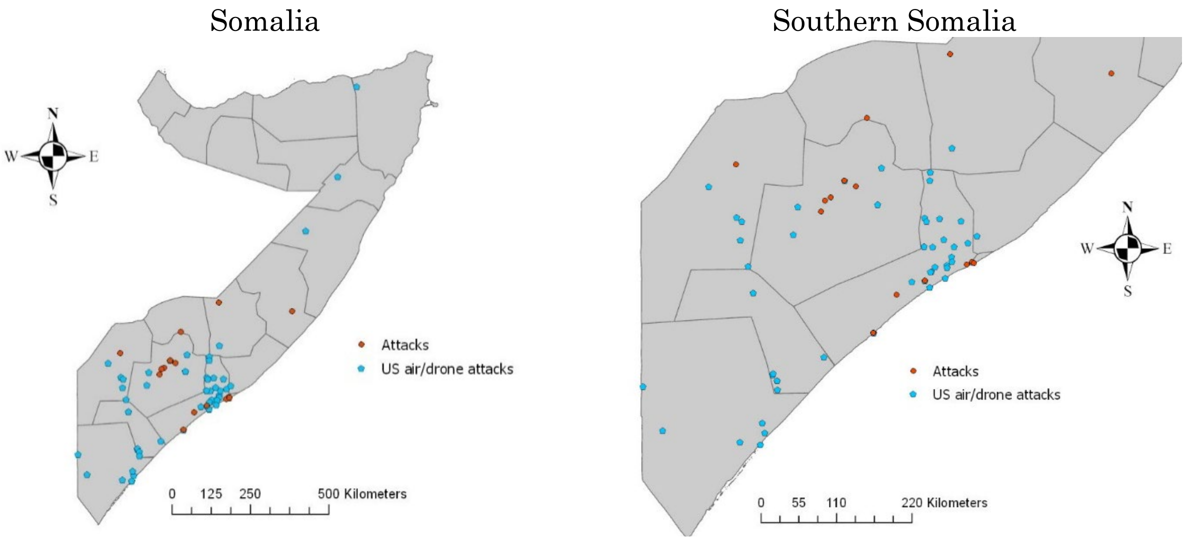
*Note: Estimated coefficients from an OLS regression. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.*

Figure C3-7: Number of US air attacks against Al-Shabaab between February 2015 and November 2017.



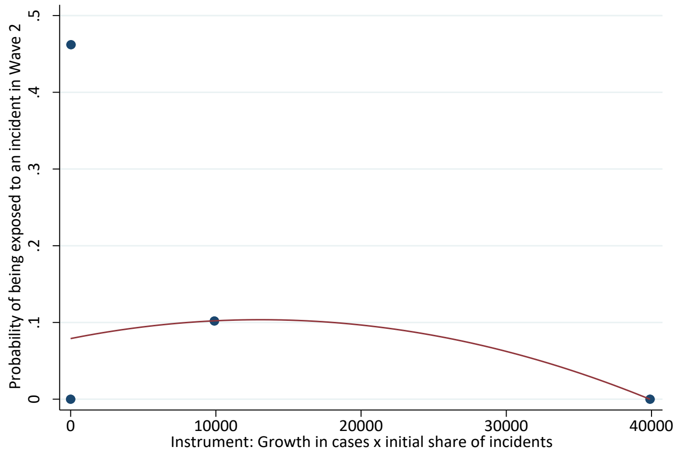
Note: The boundaries on the map show approximate borders of Somali pre-war regions and do not necessarily reflect official borders, nor imply the expression of any opinion concerning the status of any territory or the delimitation of its boundaries.

Figure C3-8: Location of incidents and US air attacks against Al-Shabaab during data collection of Wave 2.



Note: The boundaries on the map show approximate borders of Somali pre-war regions and do not necessarily reflect official borders, nor imply the expression of any opinion concerning the status of any territory or the delimitation of its boundaries.

Figure C3-9: Instrument and exposure to incidents in Wave 2 for urban areas.



***Note:** The figure presents a binned scatterplot for the relationship between the instrument and the probability of households from being exposed to an incident in Wave 2.*

Table C3-9: Different DiD specifications.

	Log of core consumption (per capita deflated)					
	(1)	(2)	(3)	(4)	(5)	(6)
Diff-in-diff coefficient (β)	-0.502*** (0.143)	-0.542*** (0.166)	-0.279** (0.130)	-0.279** (0.118)	-0.321*** (0.118)	-0.326*** (0.118)
Wave	Yes	Yes	Yes	Yes	Yes	Yes
Exposed/control	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	No	Yes	Yes	Yes	Yes	Yes
Characteristics of household & head	No	No	No	No	Yes	Yes
Dwelling characteristics	No	No	No	No	Yes	Yes
Year-month fixed effects	No	No	No	No	Yes	No
Standard errors	Sampling weights	Sampling weights	Clustered by PSU	HAC	Sampling weights	Sampling weights
Adjusted R-squared	0.018	0.045	0.100	0.100	0.342	0.342
Observations	1,557	1,557	1,557	1,557	1,532	1,532

*Note: Estimated coefficients from an OLS regression. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, sex and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Drought affected status and humanitarian assistance are not included due to the lack of variation in the data within Mogadishu. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.*

Table C3-10: DiD estimates from different samples.

	Log of core consumption (per capita deflated)				
	Mogadishu (1)	Mogadishu with overlapping exposed households in Wave 1 and 2 (2)	Mogadishu with overlapping districts in Wave 1 and 2 (3)	All urban areas (4)	Urban areas with exposed and control households in Wave 1 and 2 (5)
Diff-in-diff coefficient (β)	-0.326*** (0.118)	-0.299** (0.144)	-0.324*** (0.118)	-0.159* (0.091)	-0.315*** (0.083)
Wave	Yes	Yes	Yes	Yes	Yes
Exposed/control	Yes	Yes	Yes	Yes	Yes
Fixed effects	District	District	District	Region	Region
Characteristics of household & head	Yes	Yes	Yes	Yes	Yes
Dwelling characteristics	Yes	Yes	Yes	Yes	Yes
Drought affected status	No	No	No	Yes	Yes
Humanitarian assistance	No	No	No	Yes	Yes
Adjusted R-squared	0.342	0.344	0.329	0.333	0.368
Observations	1,532	1,497	1,396	6,560	2,241

*Note: Estimated coefficients from an OLS regression. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, sex and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Drought affected status and humanitarian assistance are not included in the first three columns due to the lack of variation in the data within Mogadishu. Standard errors considering sampling weights in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.*

Table C3-11: IV estimates from different samples.

	Log of core consumption (per capita deflated)					
	Urban areas with exposed and control households in Wave 1 and 2			All urban areas		
	(1)	(2)	(3)	(1)	(2)	(3)
IV coefficient (δ)	-1.715*** (0.540)	-2.391*** (0.857)	-2.391** (1.178)	-1.508*** (0.523)	-2.217*** (0.801)	-2.217** (1.091)
Fixed effects	Region	Region	Region	Region	Region	Region
Characteristics of household & head	Yes	Yes	Yes	Yes	Yes	Yes
Dwelling characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Drought affected status	Yes	Yes	Yes	Yes	Yes	Yes
Standard errors	Sampling weights	Clustered by PSU	HAC	Sampling weights	Clustered by PSU	HAC
Observations	2,241	2,241	2,241	6,560	6,560	6,560

*Note: Estimated coefficients from an IV regression. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, sex and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Humanitarian assistance not included due to collinearity with fixed effects by region. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.*

Table C3-12: Composition of Wave 1 and 2 samples.

Household characteristic	Exposed households			Control households		
	Difference (W2 - W1)	Significant	Obs.	Difference (W2 - W1)	Significant	Obs.
Household head without education (%)	11.9	No	134	-2.1	No	1,423
Access to piped water (%)	15.4	No	134	4.5	*	1,423
Improved sanitation (%)	-4.0	*	134	-3.7	No	1,423
Floor of cement (%)	3.9	No	134	3.4	No	1,421
Floor of tiles or mud (%)	0.6	No	134	5.3	No	1,421
Floor of wood or other material (%)	-4.0	No	134	-8.7	**	1,421
Roof of metal (%)	17.7	No	134	-5.7	No	1,423

*Note: Each row corresponds to an OLS regression of the Wave dummy over a household characteristic. Standard errors derived considering the sampling weights of the surveys. *** p<0.01, ** p<0.05, * p<0.1.*

Table C3-13: DiD falsification test measuring the impact before the terrorist attacks occurred.

	Log of core consumption (per capita deflated)			
	Exposed: Households interviewed up to a week before the incident		Exposed: All households interviewed before the incident	
	Mogadishu (1)	Urban areas with exposed and control households in Wave 1 & 2 (2)	Mogadishu (3)	Urban areas with exposed and control households in Wave 1 & 2 (4)
Diff-in-diff coefficient (β)	0.262 (0.167)	0.199 (0.131)	0.367 (0.316)	0.366 (0.343)
Wave	Yes	Yes	Yes	Yes
Exposed/control	Yes	Yes	Yes	Yes
Fixed effects	District	Region	District	Region
Characteristics of household & head	Yes	Yes	Yes	Yes
Dwelling characteristics	Yes	Yes	Yes	Yes
Drought affected status	No	Yes	No	Yes
Humanitarian assistance	No	Yes	No	Yes
Adjusted R-squared	0.346	0.362	0.341	0.361
Observations	1,419	2,106	1,419	2,106

*Note: Estimated coefficients from an OLS regression. Characteristics of household refer to size and receiving remittances, while those from household head refer to age, sex and literacy. Dwelling characteristics include tenure, roof and floor material, water source and sanitation type. Drought affected status and humanitarian assistance are not included in column 1 and 3 due to the lack of variation in the data within Mogadishu. Standard errors considering sampling weights in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.*

Table C3-14: Composition of treatment and control groups considering Wave 2 households.

Household characteristic	Difference (Exposed-Control)	Significant	Obs.
No. of dependents in the household	0.3	No	180
Share of working-age members in the household (%)	1.0	No	180
Age of household head (years)	1.3	No	180
Household head without education (%)	5.6	No	180
Access to piped water (%)	0.4	No	180
Improved sanitation (%)	13.8	**	180
Floor of cement (%)	9.7	No	180
Floor of tiles or mud (%)	-11.3	No	180
Floor of wood or other material (%)	1.6	No	180
Roof of metal (%)	-2.5	No	180

*Note: Each row corresponds to an OLS regression of the exposed/control dummy over a household characteristic. Standard errors derived considering the sampling weights of the surveys. *** p<0.01, ** p<0.05, * p<0.1.*

4. Impact of High Inflation on Household Livelihoods in Urban South Sudan[218]

Alvin Etang, Thierry Hounsa and Utz Pape[219]

June 2022

Introduction

The Republic of South Sudan gained its independence on the 9th of July 2011 following a peace agreement with the Republic of Sudan in 2005, which put an end to Africa's longest running civil war. South Sudan is a small country with a vast oil wealth, but its abysmal developmental outcomes reflect a history of conflict, characterized by a poorly functioning state and a lack of institutional services provision. Only two years after independence, civil war broke out in South Sudan and an unfavorable external macroeconomic environment triggered an economic crisis. In the years from 2015 to 2017, the South Sudanese economy displayed all characteristics of a war economy, including severe output contraction, rapid currency devaluation, and soaring inflation. Oil dependency has tied the fate of the nation to the volatility of global commodity prices.

Widespread fighting and large-scale displacement over several consecutive planting seasons have disrupted many households' normal agricultural activities, resulting in increasingly large production deficits each year and widespread food insecurity. Compounding on this, falling international oil prices triggered the rapid devaluation of the local currency driven by pressures from a low domestic supply of foreign currency, exacerbated by concurrent high domestic demand for foreign currency due to the need to supplement domestic production shortages with imported food. Falling oil prices also meant a collapse of Government revenues, which resorted to financing its deficit by printing money and incurring a growing stock of debt. Combined, these shocks have led to rapidly rising food prices, with the year-on-year CPI inflation reaching its peak at 549 percent in September 2016 (Figure C4-3). While the level of inflation almost reaches hyper-inflation, it remained – on an annual basis – still below the threshold of hyper-inflation.

Inflation has been high, but variable, across all categories of goods and services (Figure C4-4). Non-food items experienced price increase between June 2015 and June 2017. However, food prices also increased substantially during this period. This is a concern since food inflation typically hurts the poor disproportionately, due to the higher share of food in the poor's consumption basket. Given the already widespread poverty, such high food price inflation can be critical in the case of South Sudan. Although some poor rural households may be net producers of food (producing more than they consume), and thus less impacted by the high food price inflation, the very limited agricultural sector in South Sudan and the unusually high reliance on imported food suggest that the poor are also dependent on food imports, whose prices and availability have been severely affected by inflation.

An important and inevitable question is how inflation is affecting household livelihoods in South Sudan, particularly the poor.[220] High inflation can have negative impacts on household livelihoods due to increased prices for consumed goods and services with lagging wage and social assistance increases.

---

[218] AE, TH and UP contributed equally to the manuscript.

[219] Alvin Etang and Utz Pape are Senior Economists in the Poverty and Equity Global Practice of the World Bank. Thierry Hounsa is a consultant at the World Bank. Many thanks to Luca Parisotto and Ando Rahasimbelonirina for assistance with data work. Advice and comments from Nobuo Yoshida and Emmanuel Skoufias are also gratefully acknowledged. We also thank Pierella Paci for her guidance. Peer reviewer comments from Arden Finn and Nora Dihel are also gratefully acknowledged. The analysis in this paper is based on the publicly available data from three waves of the High Frequency South Sudan Survey, available on www.thepulseofsouthsudan.com. The findings, interpretations, and conclusions of this paper are those of the author and should not be attributed to the World Bank or its Executive Directors. The corresponding author Alvin Etang may be contacted at aetangndip@worldbank.org

[220] In 2015 nearly 66 percent of the population in South Sudan was poor, based on the $1.90 2011 PPP poverty line (excluding Jonglei, Unity, Upper Nile, and Warrap due to insecurity), which is a considerable increase in poverty from an already high level of 52 percent in 2009.

However, households that produce goods like food are usually less affected by high inflation as they are shielded from market prices. In fact, they can benefit from inflation if they sell products in the market. Also other characteristics like product types and market access can influence how much a household loses or benefits. For non-agricultural households, type of employment, level of education and other factors can render households more resilient against shocks.

The theoretical causes and impacts of hyperinflation are well known, and provided in the seminal work of Cagan (1956) and Nordhaus (1973). A more recent review and update was conducted by Fischer, Sahay et al. (2002). A historic overview can be found in He (2017). Recent studies focus on the causes and policy options (for example, Acemoglu, Johnson et al. 2003, Reinhart and Savastano 2003), and historic dimensions often in the context of Zimbabwe (for example, Coomer and Gstraunthaler 2011). Given the dearth of micro-data in countries with high- or hyperinflation, only very few studies look at the direct welfare impacts of high- or hyperinflation. Fajardo and Dantas (2018) study the impact of hyperinflation on investment behavior in Brazil. However, they do not touch on welfare or livelihood impacts. Larochelle, Alwang et al. (2014) uses a small-area-based approach with an asset index and finds that also rural poverty increased in Zimbabwe's hyperinflation period. In contrast, Kurasha (2021) uses micro-data from several years before and after hyperinflation, but finds that rural poverty fell while urban poverty increased, while asset inequality dropped during the hyperinflationary period. Health indicators worsened for both urban and rural as well as access to electricity, safe drinking water, improved toilets and healthcare. In contrast,

In this paper, we assess the shorter-term impacts of high inflation on household livelihoods in urban South Sudan. Longitudinal micro-data for a representative sample of households is used to understand the changes in livelihoods between 2015 and 2017, accompanied by continuous price data collected across South Sudan. The novel datasets based on a set of innovative high frequency surveys allow the use of a difference-in-difference approach providing a stronger identification than can currently be found in the literature. Furthermore, the paper identifies resilient households to draw conclusions for social protection programs and policies.

The next section of this paper presents the data and methodology in more detail. Section 3 presents descriptive results followed by results from the identification on inflation impacts on urban livelihoods in South Sudan. Section 4 concludes with a discussion and policy recommendations.

Data and Methodology

Data

This paper makes use of three waves of panel survey data from the High Frequency South Sudan Survey (HFSSS; Table C4-1).221 Wave 1 of the HFSSS was conducted largely before prices exploded, while waves 2 and 3 were implemented in the period of high inflation, and wave 4 was conducted when prices had escalated. We use location- and time-specific price differences to quantify the impact of high inflation on poverty and other livelihood indicators. The panel analysis in this paper is restricted to urban households, as it aims to identify factors that make households resilient. While the restriction to urban areas limits the scope of this paper, the panel analysis allows to gain better understanding of the impact of inflation. For urban areas, waves 1, 2 and 4 provide household panel data. The panel data will be used to analyze within household dynamics in times of high inflation. The models will be applied to changes in livelihood and determinants of the impact mainly at the household level. Since different causes affected livelihoods in this period of instability in South Sudan, the difference-in-difference approach will identify the effect of inflation on livelihoods by correlating changes in prices with changes in livelihood indicators. The datasets contain information on security, economic

[221] The High Frequency South Sudan Survey, funded by DfID, was conducted by the World Bank in collaboration with South Sudan's National Bureau of Statistics, to monitor welfare and perceptions of citizens in all accessible areas of South Sudan.

conditions, education, employment, access to services, and perceptions. They also include comprehensive information on assets and consumption, to allow estimation of poverty based on the Rapid Consumption Survey methodology as detailed in Pape and Mistiaen (2018).

Table C4-1: High Frequency South Sudan Survey, survey dates and coverage.

	Data collection dates	Geographic coverage	Rural/Urban coverage
Wave 1	February 2015 - September 2015	6 out of 10 states: Western Equatoria, Central Equatoria, Eastern Equatoria, Northern Bahr El Ghazal, Western Bahr El Ghazal, and Lakes state.	Covered urban and rural households
Wave 2	February 2016 – June 2016	7 out of 10 states: wave 1 + Warrap state. The other three former states (Jonglei, Unity, and Upper Nile) could not be surveyed due to security concerns.	Revisited urban households interviewed in Wave 1
Wave 3	September 2016 - March 2017	7 out of 10 states: Same as Wave 2	Covered a new cross-section of urban and rural households
Wave 4	May 2017 - August 2017	7 out of 10 states: Same as Waves 2 and 3.	Revisited urban households from Waves 1 and 2

Prices and Inflation

The consumption section of the household survey (HFSSS) collects information on items' unit prices and quantities. As with all data collected from sample surveys, the household-reported prices are subject to sampling errors. Item non-response and measurement error will also lead to biased estimates (Dahlhamer, Dixon et al. 2003, Garner T., McClelland R. et al. 2009). However, household-reported prices have a key strength: knowing precisely the prices paid by households who make expenditures themselves has an advantage in that it captures the parallel exchange rates, showing households' real purchasing power. This is particularly important in the context of South Sudan with a strong parallel exchange market.

Based on the strengths and weaknesses of the three price data sources, we decided to use household-reported prices because it covers the entire sample and has prices information for all items consumed by the household. Thus, for our analysis, inflation is calculated based on unit price household survey data (using Laspeyres price index). In addition to using the total inflation variable, we also break it down into food price inflation and non-food price inflation to explore which of the two might be driving the results.

Outcomes

To analyze the impact of inflation on household livelihoods, our dependent variables are household level (and individual level) outcome indicators. The variables cover a range of household social and economic indicators, which can be calculated based on the panel data (waves 1, 2 and 4). Table C4-2 shows the sample for the analysis compared to the initial sample. The outcome variables are selected from the following five categories: poverty, education, labor, hunger, and perceptions of welfare (Table C4-3).

Table C4-2: Sample size.

	Initial Sample Size	Sample Size for the Analysis
Wave 1	3550	423
Wave 2	1189	423
Wave 4	944	423

Table C4-3: Outcomes variables

Variable	**Description**
Poverty	
Poor or non-poor	Whether the household is poor or not based on the $1.90 2011 PPP poverty line
Consumption	Household consumption expenditure in real terms
Education	
School attendance	Whether children aged between 6-13 years and between 14-18 years are currently attending school
Labor[222]	
Labor force participation rate	The ratio of the active in the labor force to the total working age population (15-64 years)
Employment rate	A person is employed if he/she is of working age and has engaged in one form of employment activity.[223] The employment rate is the number of persons in employment as a percentage of the total labor force.
Unemployment rate	A person is unemployed if he/she is of working age, is not in employment during the reference period, and has been seeking employment over the past 4 weeks. The unemployment rate is the number of persons in unemployment as a percentage of the total labor force.
Outside the labor force/or inactivity	A person is outside the labor force (or "inactive") if he/she is of working-age and neither employed nor unemployed, according to the preceding definitions. An inactive person is not necessarily idle, especially in the context of a developing economy. The data breaks this group down into those who are inactive because they do household work, those who are enrolled in education, those who are discouraged, etc.
Hunger	
Hunger	How often households lacked food or lacked resources to buy food at least once in the past month
Perceptions of welfare	
Satisfaction with life	The extent to which households are satisfied with life
Living conditions	Households views about their present and future living conditions
Economic conditions	Households views about the present, past and future economic situation of South Sudan.
Control over life	The extent to which households feel that they have control over their life
Future of South Sudan	Households biggest fear about the future of South Sudan

Note: The labor force refers to the sum of persons in employment and in unemployment. It is the counterpart of the group of inactive persons, i.e. the labor force plus the inactive sum up to the entire working-age population (ILO, 2013).

Model specification

To estimate the impact of inflation on household livelihoods in urban areas of South Sudan, we use a difference-in-difference (double difference) approach to exploit both the time dimension and differences in the exposure to inflation. This identification will eliminate pre-inflation differences in the outcome variable and controls for anything that also changes over time and affects both groups.

[222] The labor market statistics presented in this paper follow closely the international standard set as per the International Labour Organisation's (ILO) Key Indicators of the Labour Market (KILM). There are two key reference periods: (a) the short observation period defined as 7 days, and (b) the long observation period defined as 12 months. Following ILO guidelines, statistics are reported for the short observation period unless explicitly stated. All persons aged 15-64 are defined as being of working age.

[223] The five employment activities are: (i) working as an apprentice, (ii) working on the household's farm, raising livestock, hunting or fishing, (iii) conducting paid or commissioned work, (iv) running a business of any size for oneself or for the household, (v) helping in a household business of any size. The definition further includes persons who are temporarily absent from their work due to training or working time arrangements such as overtime leave, and paid interns. Note that the definition excludes household work.

Hence, the assumption will be made that changes in outcomes from households in areas with high and in areas with low inflation would have been the same in the absence of the inflation shock:

$$\hat{\beta}_1^{DD} = (\bar{y}_1^{H} - \bar{y}_0^{H}) - (\bar{y}_1^{L} - \bar{y}_0^{L})$$

More specifically, the difference-in-difference estimator β_1 is computed by comparing the first-differenced values of the outcome for the high- (H) and low-inflation (L) groups. Hence, the outcome differences for the low-inflation group are differenced from the high-inflation group after taking the simple difference, which gives us the difference-in-difference estimate. The purpose of a difference-in-difference approach is to analyze whether the estimate β_1 is statistically and significantly different from zero.

To estimate the difference-in-difference effect, we use an ordinary least squares (OLS) regression model including the control vector:

$$Y_{its} = \beta_0 + \beta_1 (post_t * inflation_{st}) + \beta_X X_{its} + \gamma_s + \delta_t + \varepsilon_{its}$$

Where Y_{its} is an outcome measured for the individual or household i living in Boma s at time t ; $post_t$ is a binary variable indicating time period t (pre- or post-inflation); $Post_t = 1$ for each of waves 2 and 4 and zero otherwise (i.e. we treat waves 2 and 4 as having occurred at different times, with wave 1 being the reference period). $inflation_s$ is a continuous variable measuring the inflation rate of the Boma s; Inflation is computed as the first difference of the log price index at the Boma level. To avoid an omitted variable bias (as there are other confounding factors affecting the given outcome variables besides time-period and exposure to inflation), a control vector X_{its} for household i living in Boma s at time t is introduced; γ_s and δ_t are respectively the Boma fixed effects and the time fixed effects. Standard errors will be clustered at the Boma level to allow for within cluster correlation[224]. β_1 is the difference-in difference estimator.

Household Resilience: Triple Difference

To identify factors that make households resilient to the inflation shock, we estimate the following triple difference equation where h_i is a potential resilience factor.

$$Y_{its} = \beta_0 + \beta_1 (post_t * inflation_{st}) + \beta_2 (post_t * h_i) + \beta_3 (post_t * inflation_{st} * h_i) + \beta_X X_{its} + \gamma_s + \delta_t + \varepsilon_{its}$$

β_3 is the triple difference estimator. In this triple difference setting, β_1 is the diff-in-diff estimate for the reference group (h=0). It captures average differential change in *y* from the pre- to post-treatment period for the reference group in the treatment group *relative* to the change in *y* for the reference group (h=0) in the untreated group. The total treatment effect for both groups is $\beta_1 + \beta_3$.

Conflict indicator

Given the ongoing conflict in South Sudan, conflict will likely be one of the confounding factors affecting household livelihoods. We control for this by including a conflict variable in the regressions. We construct an exogenous conflict variable based on conflict event data from the Armed Conflict Location & Event Data (ACLED)[225] for the period of our study. The dataset codes the exact location of all political violence incidents that was reported during this time. We use proximity to a deadly conflict event to generate a continuous conflict exposure variable (i.e. the number of fatalities).

[224] Default standard errors can greatly overstate estimator precision. Instead, if the number of clusters is large, statistical inference after OLS should be based on cluster-robust standard errors. Failure to control for within-cluster error correlation can lead to very misleadingly small standard errors, and consequent misleadingly narrow confidence intervals, large t-statistics and low p-values.

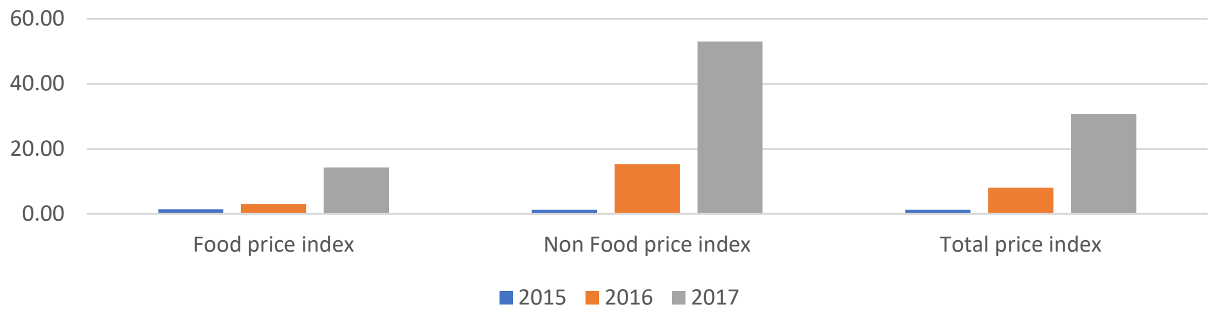
[225] Information about ACLED methodology can be found at https://www.acleddata.com/

Results

Price Changes

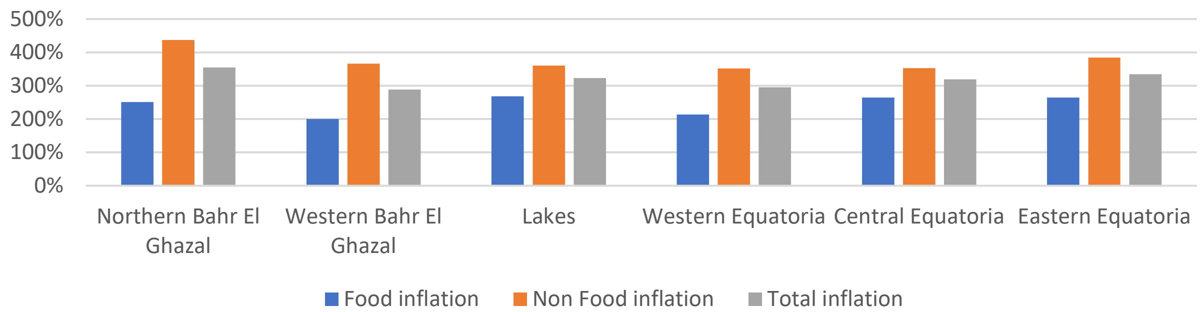
Consistent with the CPI, the HPI shows that inflation exploded between 2015 and 2017. The mean price index increased substantially from 1.31 in 2015 to 8.07 in 2016 and exploded to 30.75 in 2017 (Figure 1). The results show that non-food price inflation is drastically higher than food price inflation during this period. We use an adjusted Wald Test to test the significance of the differences in the means across the three years. The test confirms that the differences are statistically significant (p<0.01), meaning that inflation has increased significantly over time.

Figure C4-1: Recent trends in price index.



The increase in prices varies across states and have been much higher in the Eastern Equatorial, Central Equatorial, and Northern Bahr El Ghazal states than in other states (Figure C4-2). This is mainly driven by the fact that food price inflation increased much more in these states. Since the price developments have been so different across the country, it is important to consider this significant geographic variation when analyzing the impact of inflation on livelihoods.

Figure C4-2: Inflation by state - 2017 (Base year 2015=100).



Poverty and Consumption

We find strong evidence of the negative effect of inflation on household consumption and poverty.[226] The inflation impact is entirely driven by non-food price inflation. Poverty increases with the severity of exposure to inflation by the households. If inflation increases by 1 percent, the share of poor urban population (living below USD 1.90 per day PPP) increases by 0.353 percent. The results also show that the conflict variable is associated with slight decreases in consumption, with a marginal impact on poverty. We also observe that being employed on itself does not matter for poverty reduction, probably due to the low urban unemployment rates. Households whose heads are employed in the

[226] The complete difference-in-difference estimates are presented in the appendix (Table C4-5 - Table C4-11). A summary of the regression results of the impact of inflation (β1 coefficients, difference-in difference) is provided in Table 4.

services sector are less likely to be poor compared to those in the agriculture sector. This probably reflects higher wages for those in the services sector.

Some common findings similar to other poor countries are also observed here: household head being a female, and larger household size are associated with low consumption and the likelihood of being poor. On the other hand, education and land ownership help to reduce poverty. We also run a regression that includes an interaction term between inflation and land ownership. The coefficient on the interaction term is not statistically significant for both food price inflation and non-food price inflation. This suggests that the impact of inflation on poverty and real consumption is the same for both households that do own land and those that do not. Finally, university education increases consumption and reduces poverty. The impact of inflation on real consumption is significantly less for households whose heads do have university education than those who do not (the coefficient on the interaction term is 0.985, $p<0.01$). No significant differences in education level exist when it comes to the inflation impact on poverty.

School Attendance

About 3 in 4 South Sudanese children were attending primary school in 2015.[227] The primary school attendance rate remained stable in 2016 and increased to 80 percent in 2017 (Figure C4-5). Secondary school attendance remained stable from 2015 and 2016 at 78 percent but increased to 81 percent in 2017 (Figure C4-6). Primary school attendance for boys and girls increased at about the same rate between 2015 and 2017 (Figure C4-7). For the older children, attendance rate for boys declined between 2015 and 2017 by 9 percentage points (Figure C4-8). The opposite is true for girls' attendance, which has slightly increased during this period. Perhaps older boys are dropping out of school to join the labor force. This is plausible as children of working age can be expected to join the workforce to help the household support its livelihood during times of economic hardship (World Bank 2018). However, there is no evidence that this is already happening at a large scale in the states covered by the survey. This is because the difference in school attendance rate of boys aged 14 to 18 between 2015 and 2017 is not significant in a statistical sense.

Food price inflation had a negative and statistically significant impact on girls' school attendance (but no effect on boys). For girls, the likelihood of attendance diminishes with a rise in food prices. The distance to the nearest school is also important for school attendance. The chances of girls attending school diminish with increases in the distance they would have to walk to the nearest school. We run a regression for girls' attendance that includes an interaction term between inflation and distance to school. The coefficient on the interaction term is statistically significant and negative (-0.712; $p<0.01$). This means that the impact of food price inflation on school attendance is greater for girls who take more than 5 hours to walk (one way) to the nearest school from their homes compared to girls who take less than 30 minutes to do so. One explanation for this result is that when faced with an economic shock such as inflation, households become poorer (as noted above), and tend to sacrifice the education of their female children whose schools are far away from their homes as they may not be able to afford the costs related to living far away from school. In this regard, bringing schools closer to households will help to mitigate the adverse impact of inflation on girls' school attendance. The results also suggest that school attendance increases if the household head is a woman and has secondary or university education. Designing programs to promote female education will help to improve education outcomes in general, and for girls in particular.

[227] Attendance rates for children of primary school age (6-13) and secondary school age (14-18) are reported.

Labor

South Sudan's economic instability led many of the working age to drop out of the labor force between 2015 and 2016. During this period, the urban labor force participation rate dropped significantly from about one half to about one third (Figure C4-9). In 2015, the labor force participation rate remained relatively similar between poor and non-poor households and across expenditure quintiles. There are no significant differences between men and women in labor force participation in both years.[228] The urban unemployment rate was 8 percent in 2015 and 7 percent in 2016 (Figure C4-10). This number reduced substantially to 3 percent in 2017. It may not be very surprising to observe relatively high employment rates because like many other poor countries, South Sudan lacks social safety nets, which forces unemployed individuals to seek employment.

Similar to school attendance, food price inflation has a strong impact on the labor market. Increasing food prices leads to a decrease in labor force participation and increasing unemployment. In urban areas, education level is a strong determinant of unemployment for both men and women.

Hunger

Between 2015 and 2016, hunger incidence deteriorated severely for households in the poorest quintile, with the likelihood of experiencing hunger 'often' (more than 10 times per month) increasing from 4 percent to 10 percent (Figure C4-12, $p<0.05$). Due to rising prices without compensatory income increases, especially the wage-dependent urban population lost real purchasing power. Food insecurity and hunger remain a serious issue for South Sudan. For the poorest households, the likelihood of experiencing hunger 'sometimes' (3-10 times per month) has been reducing from 38 percent in 2015 to 29 percent in 2016 and rising to 40 percent in 2017. This confirms that poorest households are more vulnerable to hunger than richer households in the face of rising food prices. Richer households are much more likely to adjust their diets to cope with a lack of food, while the poorest households cope with a lack of food by going entire days without eating. This may pose serious health issues, and affect children education outcomes, with both short-term and long-term adverse effects on poverty. Resorting to more moderate strategies, households in the top 4 poverty quintiles are more likely to deal with a lack of food by reducing the number of meals or portion size, or consuming less preferred food than the poorest households.

Inflation increases hunger, and the combined effect of both food price inflation and non-food price inflation is very strong, with rising food prices having greater impact. While the pinch of inflation was felt by every household, the poorest ones were the worst affected. Rising food prices have led to growing food insecurity for the poorest households, for whom the incidence of hunger has increased sharply. The poorest households are in a vicious circle as they may become poorer due to consequences of hunger including poor health, child malnutrition and education outcomes. The finding that rapidly rising food prices is a causal source of hunger and food insecurity is consistent with findings from other poor countries (Ferreira, Chen et al. 2016). Households whose heads have university education experience less hunger than others. The coefficients on the poverty quintiles are significant and negative and get larger in magnitude as one moves up the consumption distribution from Q2 to Q3, Q4 and largest for the richest quintile. This suggests that poverty also has a significantly impact on hunger, with hunger incidence declining as consumption increases.

Perception

The deterioration of economic conditions in South Sudan, as indicated by continued high inflation, is well echoed by households' perceptions. In 2017, almost all (97 percent) South Sudanese residing in

[228] The 2017 labor force participation numbers are not entirely comparable with the previous years because of the changes in the questionnaire.

urban areas felt that economic conditions in their country were bad or very bad (Figure C4-13). This is a drastic increase compared to the previous years. This figure stood at almost two thirds of urban South Sudanese in 2015, increasing to almost 9 in 10 (63 vs. 86 percent respectively, p<0.001). However, there seems to be a growing sense of optimism among the urban residents about the future. In 2015, nearly half of households (46 percent) believed that economic conditions will be better or much better in 3 months' time. While people became less optimistic in 2016 (29 percent), the figure increased considerably to two thirds in 2017 (66 percent). It should also be noted that a sizable share of residents remains pessimistic about the future, with 21 percent of households of the view that economic conditions will get worse or much worse in 3 months' time.

The deterioration of economic conditions is also well reflected in households' perceptions of their own living conditions. In 2015, almost half (45 percent) of urban households felt that their living conditions were fairly bad or very bad (Figure C4-14). This figure increased significantly to 78 percent in 2016 and 80 percent in 2017. There does not seem to be much hope for many households who believed that their personal living conditions will deteriorate in the next 3 months. Between 2015 and 2016, the share of households that believe living conditions will get worse or much worse increased significantly from 25 percent to 46 percent (p<0.001). The figure decreased to 31 percent in 2017, though it is still high. Nevertheless, as with economic conditions, there seems to be growing optimism about the future for one half of urban households. The share of households believing that living conditions will get better or much better increased from 39 percent in 2016 (10 percentage points less, compared 2015) to 53 percent in 2017.

There seems to be a correlation between having control over people's lives and the extent to which they are satisfied with life. The share of urban residents who felt that they have no control over their lives increased from 26 percent in 2015 to 37 percent in 2016 but decreased substantially to 16 percent in 2017 (Figure C4-15). Feeling much more in control of their lives, 32 percent of households strongly agreed that they are satisfied with life (Figure C4-16). Note that in 2016 there was a general decline in life satisfaction relative to 2015, which reflected a growing feeling among urban folks that they were powerless in the face of deteriorating political and economic conditions.

Increases in inflation are associated with less satisfaction with life. Regarding satisfaction with present living conditions (ranging from 1: very good to 5: very bad), the positive coefficient means that people are less satisfied in the face of inflation. However, people's views about future living conditions are positive, consistent with optimism noted from descriptive statistics. There is a strong feeling among urban residents who are exposed to inflation that they are powerless and have no control over their lives.

Table C4-4: Summary of OLS results for each outcome indicator and inflation variable

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.252	0.00854	0.238*
Log(real consumption)	-0.829***	-0.116	-0.680***
Education			
Currently attending school (Girls)	-0.0767	-0.136***	-0.0202
Labor			
Labor force participation: Active	-0.138	-0.207***	-0.049
Unemployed	0.0841	0.126**	0.042

Hunger			
Hunger incidence	0.430***	0.329**	0.193*
Perceptions of welfare			
Life satisfaction	-1.205*	-0.178	-0.807*
Present living conditions	0.480**	0.22	0.22
Future living conditions	1.789*	-0.039	1.343**
Control over own life	-0.611**	-0.055	-0.514**
Present economic conditions	0.264	0.394**	0.053
Future economic conditions	1.370	-0.588	1.22*

*** p<0.01, ** p<0.05, * p<0.1.

Robustness

We perform robust checks of the results by running various regressions. This include regressions for wave 1 to wave 2, wave 1 to wave 4, using household fixed effects, and adding the Boma CPI in the regression to pick up the effect of high prices and high poverty. We also run regressions where households are classified as high and low inflation households using the mean inflation for urban South Sudan. All households living in an area with an inflation rate higher than the mean inflation rate are classified as "high inflation" households and are therefore in the treatment group. Households with an inflation rate lower than the mean inflation rate are in the control group.

The results, presented in the Annex Table A8, confirm the findings discussed above related to the impact of inflation on welfare in urban South Sudan. The results are generally consistent with initial specifications regardless of whether we use Boma CPI or household fixed effects or classify households as experiencing high or low inflation. Perhaps the most interesting revelation of these analyses is that the welfare impact comes from the "hyperinflation" experienced in South Sudan during wave 4 in 2017 rather than inflation between waves 1 and 2 in 2016. The coefficients for the regression model for waves 1 and 2 are generally not statistically significant, while those for the regression for waves 1 and 4 are significant with the initial specifications above. This is consistent with the inflation dynamics between waves 1 and 2 and waves 1 and 4 and changes in welfare indicators (see Figure C4-18).

Conclusion and policy recommendation

The high inflation in South Sudan continues to put many households under extreme financial stress, as they face increased prices without compensatory increases in income. This paper contributes to the available empirical evidence on micro-level impacts of inflation by analyzing the impact of high inflation on household livelihoods in urban South Sudan. We use panel data collected during the High Frequency South Sudan Survey (HFSSS) waves in 2015, 2016 and 2017. Increasing resilience to high inflation would allow to shift the inflation crisis management paradigm from a humanitarian to a development approach.

Breaking down inflation by food price inflation and non-food price inflation reveals that the latter increased drastically more than the former during this period. Although there has been significant increases in both food and non-food prices, the observed high inflation in South Sudan is mostly driven by the escalation of non-food prices. Inflation negatively impacted various household livelihood indicators related to poverty, education, labor, hunger and perceptions of welfare.

Inflation had a strong negative impact on urban poverty between 2015 and 2017, mainly driven by the increase of non-food prices. The loss of purchasing power of wages and salaries has driven many of the South Sudanese residing in urban areas into poverty. Continuous increases in inflation will only worsen the already high poverty situation. Addressing the issue of high inflation must be at the center of efforts to stability the economy and reduce poverty in South Sudan. In addition, higher education has a key role for poverty reduction because the impact of inflation on consumption is significantly lower for households whose heads do have university education.

Food price inflation had a negative and statistically significant impact on girls' primary and secondary school attendance. The probability of a girl attending school diminishes as food prices increase. Proximity to school is very important for school attendance. School attendance is less likely for girls who take more than 5 hours to walk from their home to the nearest school from compared to girls who take less than 30 minutes to do so. This corroborates earlier reports that long distance to school was one of the most cited reasons by for dropping out of primary and secondary school in South Sudan (Ministry of General Education and Instruction (MoGEI) 2016). While the cost of schooling is a major constraint for school attendance of both boys and girls, it disproportionately affects girls. In the face of limited resources, parents apparently prioritize boys for schooling over girls (World Bank 2018). Investing in female education is very important for poverty reduction and development, especially as we also find that school attendance increases if the household head is a woman and has secondary or university education. One important policy implication from this study is that bringing schools closer to households will help to mitigate the adverse impact of inflation especially on girls' school attendance. Investing in education, particularly in fragile contexts like South Sudan also helps to create resilience against such economic shocks.

Another consequence of the observed increases in food prices is a significant decline in labor force participation and a surge in unemployment among urban people. Employment programs with a focus on poverty reduction should, therefore, consider ways to mitigate the impact of rising food prices.

Inflation is exacerbating food insecurity and hunger, particularly for the poorest households who are more vulnerable to hunger. Households adopt various strategies to cope with hunger, including eating less preferred food, skipping entire days without eating and selling assets. However, these coping strategies may put them at increased risk for future spells of food insecurity. The coping strategies employed by the poor, especially selling productive assets such as livestock, typically put them at an even greater disadvantage in the future (Barrett, Beaghen et al. 2002).

Inflation has negatively affected households' perceptions of welfare. Urban residents who are exposed to inflation strongly feel that they are powerless and have no control over their lives. This has led to less satisfaction with life and present living conditions.

A key economic priority for the Government of South Sudan should be to implement urgent macroeconomic measures to reduce high inflation. Addressing the problem of high inflation will help to curb increasing poverty, crucial for progress towards achieving the first Sustainable Development Goal (SDG 1) to end poverty by 2030. In addition, for South Sudan to achieve SDG 2 (to end hunger and ensure access to food by all people, including the poor by 2030), the issue of rising inflation has to be contained very quickly as it is exacerbating hunger and food insecurity.

This paper shows that inflation has had adverse effects on the livelihoods of urban households. Because our analysis focus on urban areas, some of the results may not directly be generalized to the entire country. Even if the inflation crisis improves livelihoods of the – predominantly rural – households producing food, the rapidly increasing prices of non-food items is likely to have increased rural poverty and hunger as well.

Appendix

Figure C4-3: Trends in CPI inflation, year-on-year.

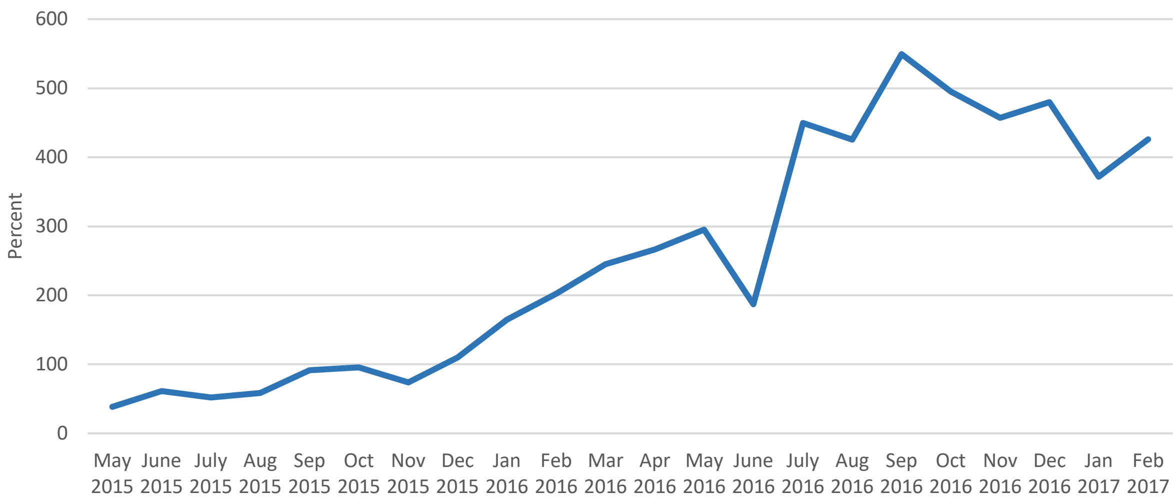


Figure C4-4: High inflation in all categories of goods between June 2015 and June 2017.

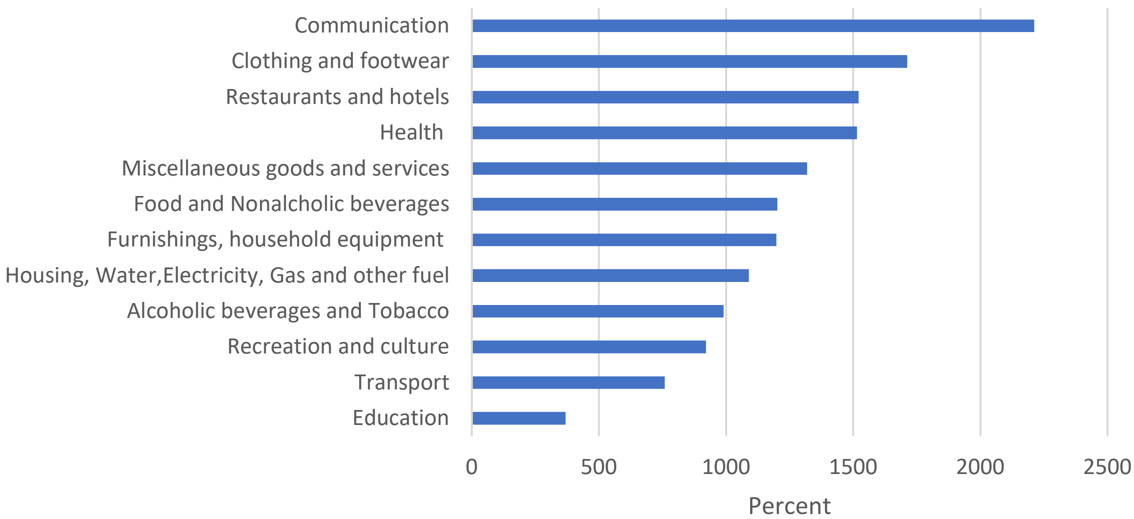


Figure C4-5: School attendance, children aged 6-13, by poverty.

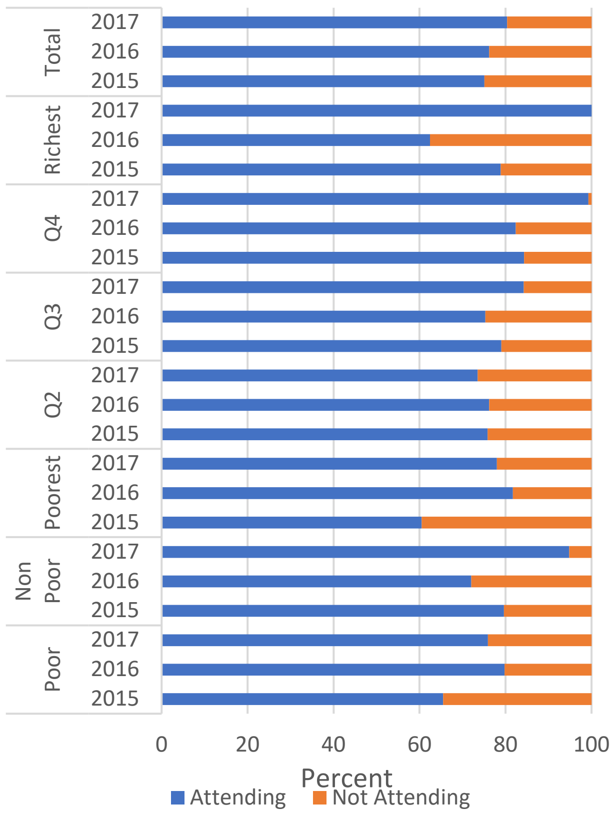


Figure C4-6: School attendance, children aged 14-18, by poverty.

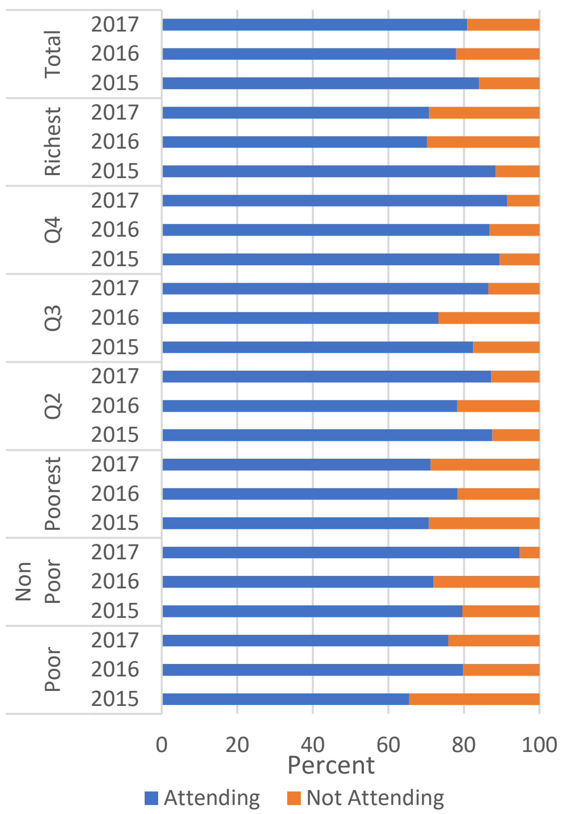


Figure C4-7: School attendance, children aged 6-13, by gender.

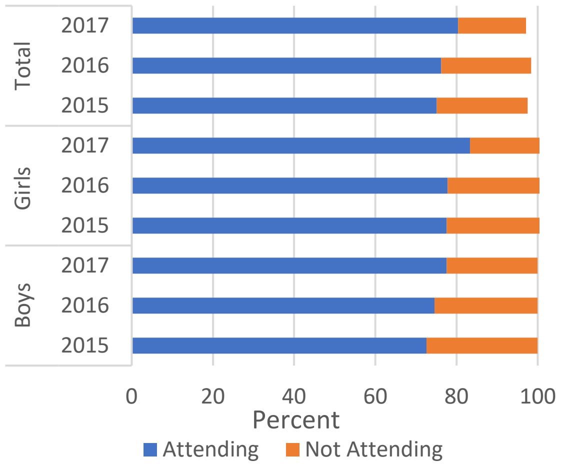


Figure C4-8: School attendance, children aged 14-18, by gender.

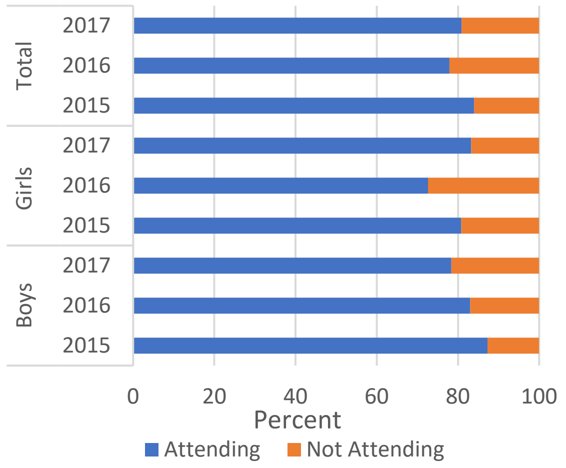


Figure C4-9: Labor force participation rate.

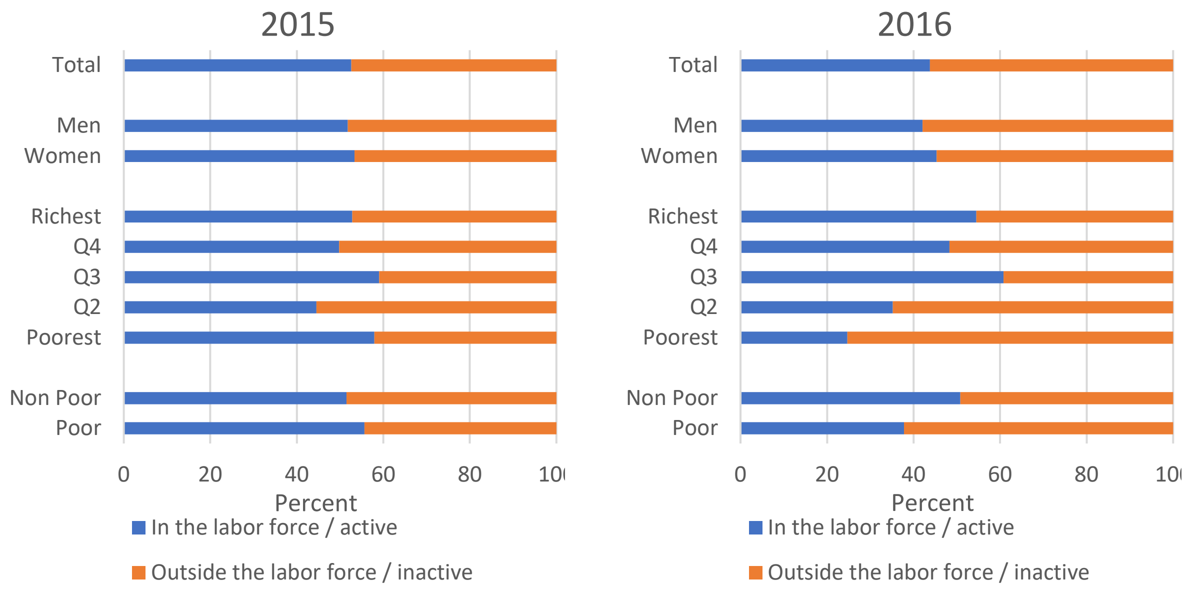
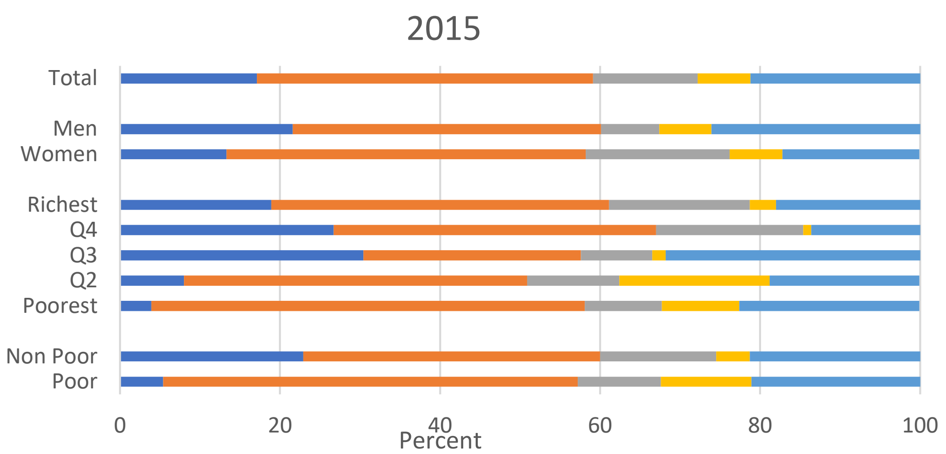


Figure C4-10: Employment and enrollment status.

Figure C4-11: Employment by type.



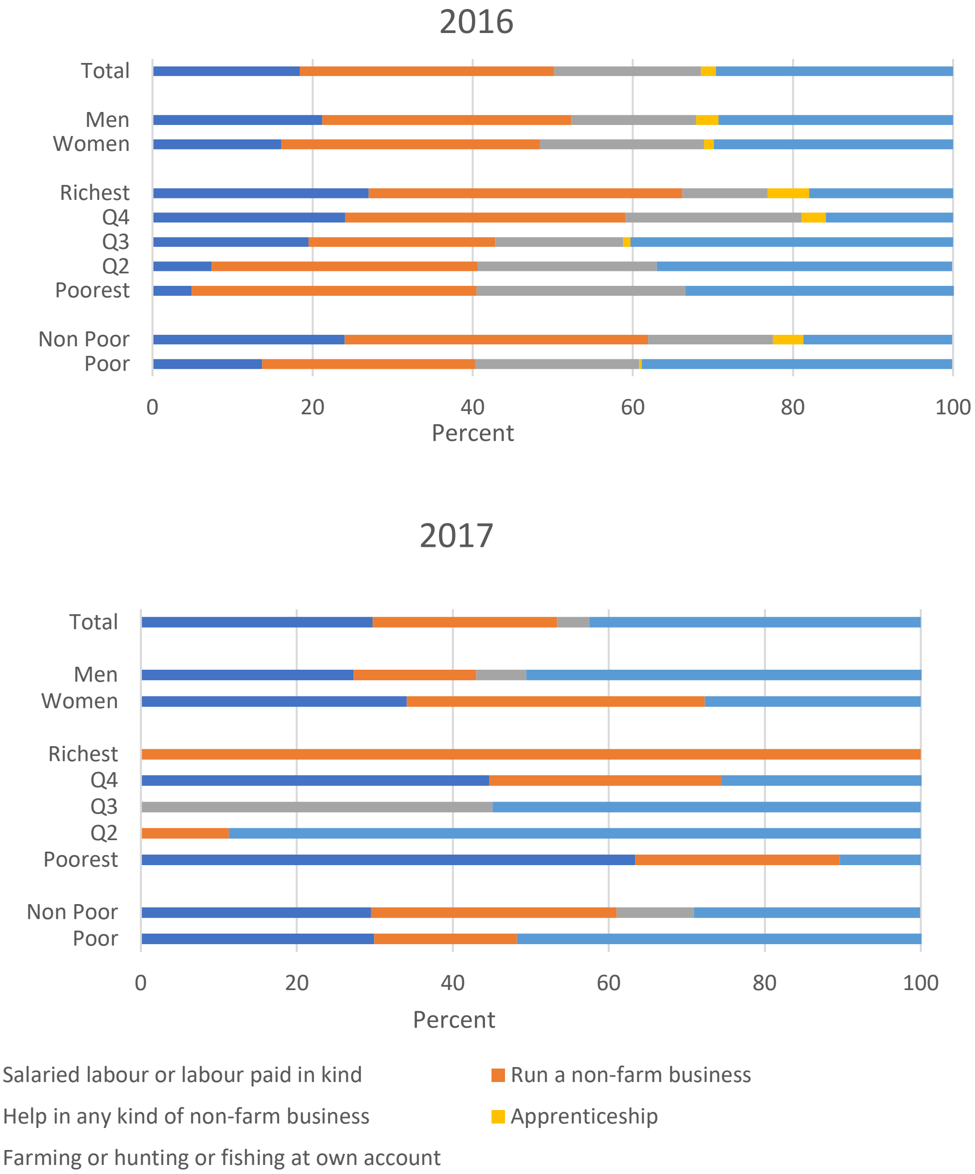


Figure C4-12: Hunger incidence over the past 4 weeks.

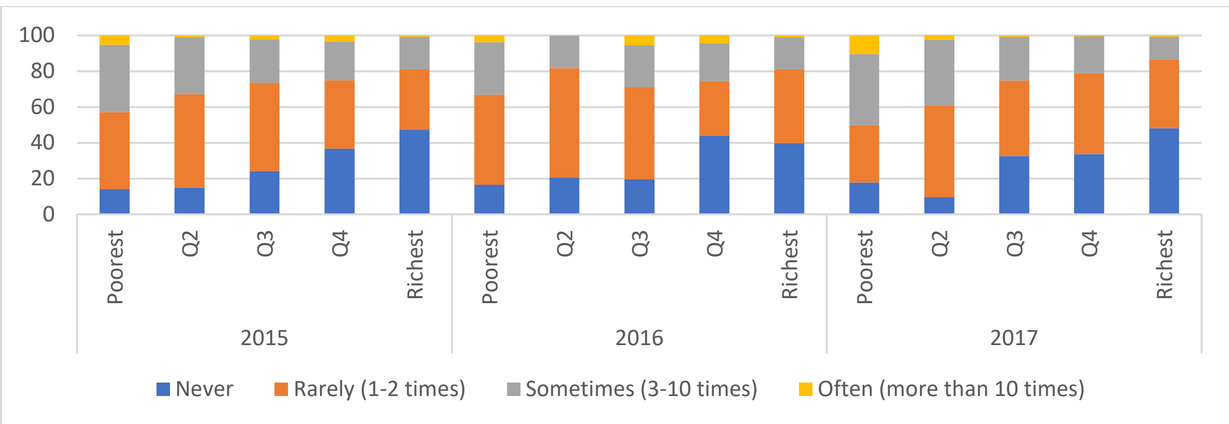


Figure C4-13: Perception of economic conditions.

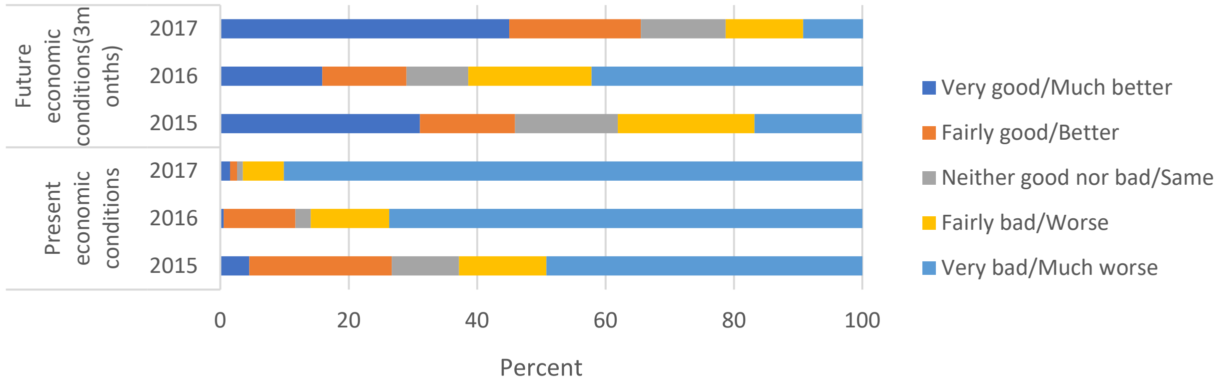


Figure C4-14: Perception of living conditions.

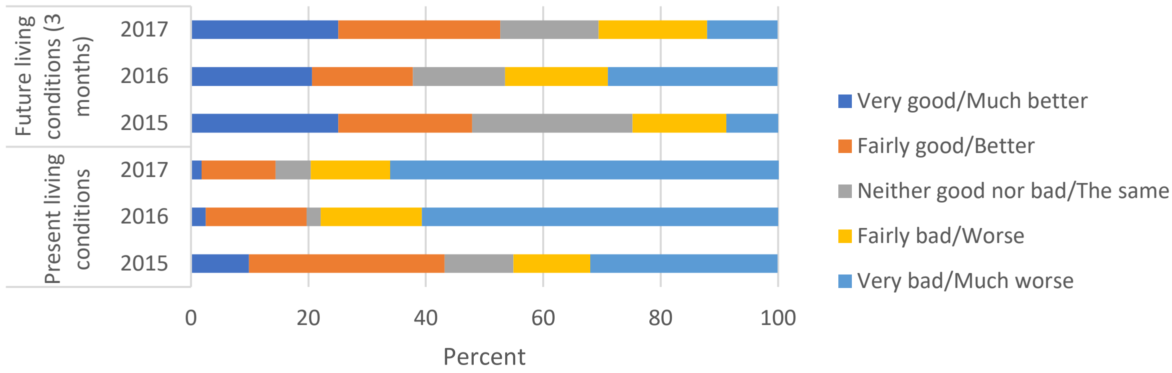


Figure C4-15: Feeling in control over own life.

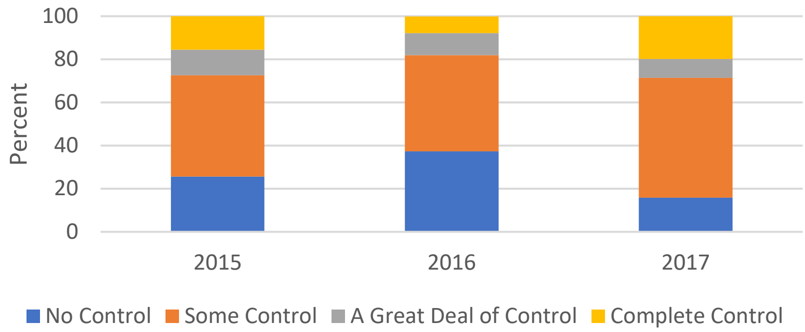


Figure C4-16: Satisfaction with life.

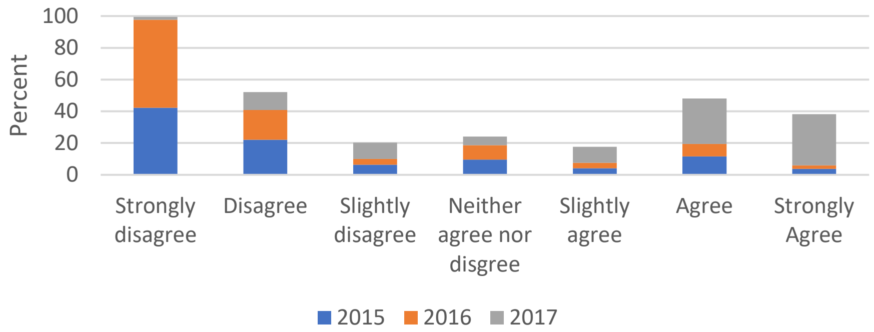


Figure C4-17: Fear for the future of South Sudan.

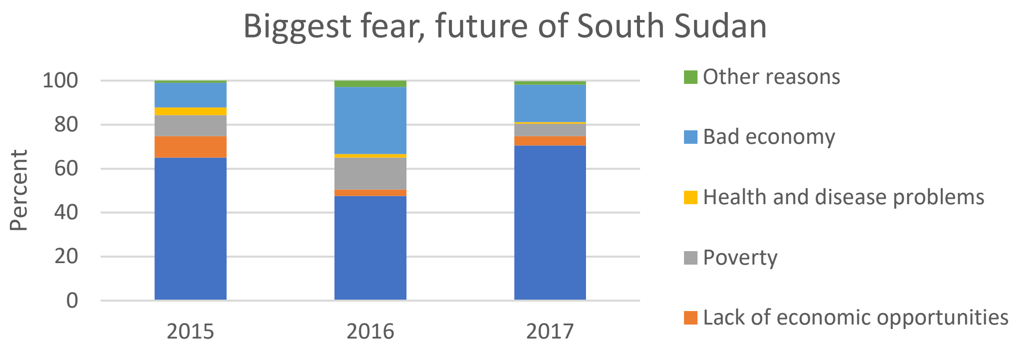


Table C4-5: OLS for poverty and consumption

VARIABLES	(1) Poor (below USD 1.90 PPP)	(2) Poor (below USD 1.90 PPP)	(3) Log(real consumption)	(4) Log(real consumption)
Survey year: 2016	-0.286	0.031	-0.524	-0.934
	(0.268)	(0.259)	(0.449)	(0.611)
Survey year: 2017	-0.051	0.260	-2.541***	-2.916***
	(0.228)	(0.221)	(0.344)	(0.455)
Inflation*Post	0.252	0.060	-0.829***	-0.575*
	(0.165)	(0.161)	(0.257)	(0.338)
Conflict	0.001	0.002	-0.005**	-0.006**
	(0.001)	(0.001)	(0.002)	(0.003)
Female household head	0.149***	0.203***	-0.030	-0.135*
	(0.051)	(0.065)	(0.050)	(0.080)
Education level of household head_Primary	-0.077	-0.081	0.066	0.120
	(0.069)	(0.068)	(0.088)	(0.103)
Education level of household head_Secondary	-0.061	-0.067	0.137	0.186
	(0.090)	(0.090)	(0.123)	(0.123)
Education level of household head_University	-0.408***	-0.416***	0.619***	0.596***
	(0.120)	(0.129)	(0.148)	(0.182)

Education level of household_Other	-0.194***	-0.273***	0.387***	0.569***
	(0.060)	(0.053)	(0.097)	(0.084)
Land ownership	-0.152**	-0.132**	0.263**	0.256***
	(0.068)	(0.058)	(0.103)	(0.092)
Household head age	-0.003	-0.002	0.036	0.025
	(0.020)	(0.008)	(0.029)	(0.019)
Household head age-squared	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Household size	0.077***	0.091***	-0.175***	-0.182***
	(0.013)	(0.013)	(0.022)	(0.024)
Household size-squared	-0.001***	-0.002***	0.004***	0.004***
	(0.000)	(0.001)	(0.001)	(0.001)
Household head Unemployed	-0.003		-0.055	
	(0.145)		(0.134)	
Household head employment: Manufacturing		0.201		-0.125
		(0.123)		(0.137)
Household head employment: Services		-0.137*		0.183**
		(0.077)		(0.087)
Household head employment: Education		-0.012		-0.112
		(0.169)		(0.157)
Household head employment: Defense/Security		-0.125		0.213**
		(0.098)		(0.091)
Household head employment: Public Administration		-0.323*		0.249
		(0.165)		(0.159)
Constant	-0.131	-0.247	1.897**	2.211***
	(0.466)	(0.274)	(0.706)	(0.498)
Observations	703	673	703	673
R-squared	0.296	0.353	0.860	0.868

*** p<0.01, ** p<0.05, * p<0.1.

Notes: Robust standard errors in parentheses. Reference is 2015 for survey year; male for gender of household head; no education for household head educational level; employed for household head employment status; and agriculture for household head's sector of employment.

Table C4-6: OLS for poverty and consumption, interacting inflation with household head education

VARIABLES	(1) Poor (below USD 1.90 PPP)	(2) Poor (below USD 1.90 PPP)	(3) Log(real consumption)	(4) Log(real consumption)
Survey year: 2016	-0.246	-0.117	0.086	0.028
	(0.260)	(0.216)	(0.365)	(0.295)
Survey year: 2017	0.000	0.147	-2.043***	-2.133***
	(0.220)	(0.192)	(0.325)	(0.279)
Inflation*Post	0.232	0.160	-1.185***	-1.162***
	(0.153)	(0.136)	(0.216)	(0.184)
inflation*Household head_University education *Post	0.036	-0.216	0.815**	1.210***
	(0.247)	(0.230)	(0.305)	(0.268)

Household head_University education*Post	-0.095	0.184	-1.193***	-1.523***
	(0.247)	(0.218)	(0.305)	(0.253)
Conflict	0.001	0.002	-0.004***	-0.005**
	(0.001)	(0.001)	(0.001)	(0.002)
Female household head	0.172***	0.233***	-0.085*	-0.192**
	(0.038)	(0.070)	(0.048)	(0.092)
Household head University education	-0.332	-0.286	0.572*	0.374
	(0.245)	(0.184)	(0.333)	(0.241)
Land Ownership	-0.153**	-0.127**	0.229**	0.214**
	(0.069)	(0.057)	(0.092)	(0.083)
Household head age	-0.003	4.90e-05	0.033	0.021
	(0.019)	(0.008)	(0.028)	(0.018)
Household head age-squared	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Household size	0.077***	0.090***	-0.168***	-0.177***
	(0.013)	(0.012)	(0.021)	(0.021)
Household size-squared	-0.001***	-0.002***	0.004***	0.004***
	(0.000)	(0.000)	(0.001)	(0.001)
Household head Unemployed	-0.0165		-0.0460	
	(0.150)		(0.140)	
Household head employment: Manufacturing		0.173		-0.028
		(0.114)		(0.124)
Household head employment: Services		-0.148**		0.207**
		(0.073)		(0.078)
Household head employment: Education		-0.028		-0.045
		(0.162)		(0.139)
Household head employment: Defense/Security		-0.124		0.203**
		(0.096)		(0.091)
Household head employment: Public Administration		-0.352**		0.355***
		(0.158)		(0.123)
Constant	-0.193	-0.344	2.007***	2.367***
	(0.454)	(0.266)	(0.684)	(0.476)
Observations	703	673	703	673
R-squared	0.293	0.353	0.863	0.876

*** p<0.01, ** p<0.05, * p<0.1. Notes: Robust standard errors in parentheses.

Table C4-7: OLS for currently attending school, boys and girls

VARIABLES	(1) Currently Attending School	(2) Currently Attending School	(3) Currently Attending School
Survey year: 2016	0.067	0.0739**	0.003
	(0.125)	(0.032)	(0.127)
Survey year: 2017	0.005	0.099	-0.041
	(0.106)	(0.071)	(0.078)
Inflation*Post	-0.035		
	(0.071)		
Inflation*distance to school more than 5 hours*Post	0.0664*		
	(0.035)		
Food Inflation*Post		-0.0865**	
		(0.038)	

Food inflation* distance to school more than 5 hours*Post		0.186**	
		(0.085)	
Non-food Inflation*Post			0.001
			(0.053)
Non-food inflation* distance to school more than 5 hours*Post			0.0447*
			(0.024)
Distance to school: More than 5 hours (Ref: Less than 30 minutes)	-0.276***	-0.279***	-0.272***
	(0.020)	(0.018)	(0.020)
Gender of school child: Girl	-0.116***	-0.115***	-0.116***
	(0.019)	(0.019)	(0.019)
Female household head	0.031	0.029	0.031
	(0.037)	(0.037)	(0.037)
Education level of household head_Primary	-0.036	-0.035	-0.036
	(0.103)	(0.103)	(0.104)
Education level of household head_Secondary	0.111**	0.108**	0.110**
	(0.045)	(0.044)	(0.045)
Education level of household head_University	0.117*	0.120**	0.116*
	(0.059)	(0.060)	(0.058)
Education level of household_Other	-0.504***	-0.478***	-0.507***
	(0.046)	(0.043)	(0.049)
Constant	0.480***	0.482***	0.481***
	(0.055)	(0.055)	(0.054)
Observations	6435	6435	6435
R-squared	0.060	0.062	0.060

*** p<0.01, ** p<0.05, * p<0.1.

Notes: Robust standard errors in parentheses.

Table C4-8: OLS for currently attending school, girls only

	(1)	(2)	(3)
VARIABLES	Currently Attending School	Currently Attending School	Currently Attending School
Survey year: 2016	0.150	0.123***	0.063
	(0.150)	(0.037)	(0.151)
Survey year: 2017	0.043	0.162**	-0.032
	(0.124)	(0.075)	(0.090)
Inflation*Post	-0.077		
	(0.086)		
Inflation*distance to school more than 5 hours*Post	-0.202***		
	(0.030)		
Food Inflation*Post		-0.136***	
		(0.041)	
Food inflation* distance to school more than 5 hours*Post		-0.745***	
		(0.063)	
Non-food Inflation*Post			-0.020

			(0.063)
Non-food inflation* distance to school more than 5 hours*Post			-0.159***
			(0.020)
Conflict	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Distance to school: More than 5 hours (Ref: < 30 minutes)	0.103***	0.0992***	0.110***
	(0.034)	(0.032)	(0.036)
Female household head	-0.0511**	-0.0538**	-0.0513**
	(0.023)	(0.023)	(0.023)
Education level of household head_Primary	-0.075	-0.074	-0.076
	(0.086)	(0.085)	(0.086)
Education level of household head_Secondary	0.080	0.076	0.078
	(0.062)	(0.060)	(0.062)
Education level of household head_University	0.086	0.088	0.082
	(0.058)	(0.059)	(0.058)
Education level of household_Other	-0.431***	-0.392***	-0.437***
	(0.052)	(0.047)	(0.056)
Constant	0.423***	0.426***	0.424***
	(0.042)	(0.041)	(0.041)
Observations	3,344	3,344	3,344
R-squared	0.064	0.066	0.063

*** p<0.01, ** p<0.05, * p<0.1.
Notes: Robust standard errors in parentheses.

Table C4-9: OLS for labor indicators

VARIABLES	(1) Labor Force participation (Ref: Inactive)	(2) Labor Force participation (Ref: Inactive)	(3) Unemployed (Ref: employed)
Survey year: 2016	0.250	0.174***	-0.101**
	(0.163)	(0.052)	(0.042)
Survey year: 2017	0.440***	0.590***	-0.250**
	(0.122)	(0.110)	(0.097)
Inflation*Post	-0.138		
	(0.094)		
Food Inflation*Post		-0.207***	0.126**
		(0.070)	(0.051)
Conflict	0.000	-0.001	0.000744**
	(0.001)	(0.001)	(0.000)
Respondent is a woman	0.000	-0.016	0.0474**
	(0.035)	(0.033)	(0.020)
Education level of household head_Primary	0.022	0.039	0.029
	(0.036)	(0.031)	(0.020)
Education level of household head_Secondary	-0.060	-0.054	0.0783*
	(0.068)	(0.064)	(0.043)
Education level of household head_University	-0.036	-0.057	0.0803**
	(0.052)	(0.056)	(0.035)
Education level of household_Other	-0.165**	-0.0793**	0.015

	(0.075)	(0.033)	(0.015)
Age	0.0103***	0.0685***	-0.005
	(0.001)	(0.007)	(0.003)
Age-squared		-0.000842***	0.000
		(0.000)	(0.000)
Constant	0.152**	-0.695***	0.138**
	(0.067)	(0.114)	(0.061)
Observations	3,838	3,838	2,011
R-squared	0.204	0.277	0.154

*** p<0.01, ** p<0.05, * p<0.1.
Notes: Robust standard errors in parentheses.

Table C4-10: OLS for hunger

VARIABLES	(1) Hunger	(2) Hunger	(3) Hunger
Survey year: 2016			
	-0.822***	-0.322*	-0.525**
Survey year: 2017	(0.255)	(0.187)	(0.198)
	-0.726***	-0.718***	-0.403***
Inflation*Post	(0.199)	(0.255)	(0.148)
	0.430***		
Food price inflation*Post	(0.145)		
		0.329**	
Non-food price inflation*Post		(0.143)	
			0.193*
Conflict			(0.108)
	0.00362**	0.00563***	0.00424**
Female household head	(0.002)	(0.002)	(0.002)
	(0.036)	(0.020)	(0.042)
Education level of household head_Primary	(0.072)	(0.074)	(0.073)
	-0.151*	-0.146*	-0.152**
Education level of household head_Secondary	(0.077)	(0.075)	(0.075)
	-0.209*	(0.195)	-0.213*
Education level of household head_University	(0.121)	(0.119)	(0.123)
	-0.437***	-0.449**	-0.450***
Education level of household_Other	(0.154)	(0.169)	(0.162)
	0.011	(0.025)	0.034
Land ownership	(0.122)	(0.127)	(0.123)
	-0.264**	-0.235**	-0.253**
Household head Unemployed	(0.104)	(0.109)	(0.107)
	0.246	0.222	0.251
Household size	(0.182)	(0.182)	(0.191)
	(0.008)	(0.012)	(0.010)
Household size-squared	(0.021)	(0.022)	(0.022)
	0.000	0.000	0.000
2nd welfare quintile (Ref=1st quintile)	(0.000)	(0.000)	(0.000)
	0.036	0.040	0.030
3rd welfare quintile	(0.151)	(0.147)	(0.146)
	(0.182)	(0.164)	(0.186)
4th welfare quintile	(0.153)	(0.155)	(0.152)

	-0.388***	-0.407***	-0.408***
5th welfare quintile (Richest)	(0.082)	(0.079)	(0.082)
	-0.507***	-0.532***	-0.515***
Constant	(0.091)	(0.083)	(0.090)
	2.499***	2.488***	2.513***
Observations	(0.118)	(0.115)	(0.115)
R-squared	702	702	702

*** p<0.01, ** p<0.05, * p<0.1.
Notes: Robust standard errors in parentheses.

Table C4-11: OLS for perceptions of welfare

VARIABLES	(1) Life satisfaction	(2) Present living conditions	(3) Future living conditions	(4) Present economic conditions	(5) Future economic conditions	(6) Control over own life
Survey year: 2016	1.176	0.038	-3.906**	0.244	-3.203*	0.886
	(1.058)	(0.399)	(1.755)	(0.405)	(1.870)	(0.530)
Survey year: 2017	1.045	0.373	-2.569*	0.711***	-1.143	1.078**
	(0.791)	(0.338)	(1.427)	(0.257)	(1.646)	(0.412)
Inflation*Post	-1.205*	0.480**	1.789*	0.264	1.370	-0.611**
	(0.612)	(0.239)	(0.941)	(0.208)	(0.976)	(0.282)
Conflict	-0.0177**	0.00874**	0.004	0.001	0.000	-0.002
	(0.008)	(0.004)	(0.004)	(0.002)	(0.003)	(0.002)
2nd welfare quintile (Ref=1st quintile)	0.526**	-0.212	-0.102	0.022	0.115	-0.108
	(0.237)	(0.159)	(0.215)	(0.127)	(0.300)	(0.144)
3rd welfare quintile	0.504*	-0.269	-0.257	-0.228	-0.198	-0.00252
	(0.291)	(0.191)	(0.231)	(0.142)	(0.283)	(0.113)
4th welfare quintile	0.714**	-0.328	0.0364	-0.0757	0.142	-0.074
	(0.324)	(0.209)	(0.334)	(0.133)	(0.310)	(0.100)
5th welfare quintile (Richest)	0.656**	-0.496***	-0.601**	-0.0836	-0.336	0.00515
	(0.322)	(0.177)	(0.248)	-0.137	-0.337	-0.136
Constant	3.283***	3.328***	3.679***	3.814***	3.209***	2.157***
	(0.210)	(0.124)	(0.273)	(0.0792)	(0.381)	(0.127)
Observations	1,221	1,210	851	1,146	849	1,162
R-squared	0.276	0.292	0.255	0.234	0.285	0.226

*** p<0.01, ** p<0.05, * p<0.1.
Notes: Robust standard errors in parentheses.

Table C4-12: Robustness checks

Robustness check 1: Adding the initial level of CPI at the Boma Level to the initial specification

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.353**	0.0332	0.322***
Log(real consumption)	-0.833***	-0.173	-0.685***
Education			
Currently attending school (Girls)	-0.0243	-0.130***	0.0149
Labor			
Labor force participation: Active	-0.126	-0.206***	-0.0309
Unemployed	0.0200	0.0877*	0.0107
Hunger			
Hunger incidence	0.509***	0.325**	0.243**
Perceptions of welfare			
Life satisfaction	-1.218*	-0.180	-0.807*
Present living conditions	0.479*	0.218	0.225
Future living conditions	1.779*	-0.0515	1.349**
Control over own life	-0.600**	-0.0495	-0.516**
Present economic conditions	0.272	0.399**	0.0531
Future economic conditions	1.367	-0.593	1.227*

*** p<0.01, ** p<0.05, * p<0.1.

Robustness check 2: Household Fixed effects with robust standard errors

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.130*	0.0592	0.117*
Log(real consumption)	-0.521***	-0.243**	-0.397***
Education			
Currently attending school (Girls)	-0.114**	-0.139***	-0.0557
Labor			
Labor force participation: Active	-0.178***	-0.208***	-0.0774***
Unemployed	-0.0296	0.0620***	-0.0168
Hunger			
Hunger incidence	0.495***	0.493***	0.209*
Perceptions of welfare			
Life satisfaction	-0.944*	-0.287	-0.564
Present living conditions	0.308	0.315	0.0691
Future living conditions	1.472**	-0.276	1.157**
Control over own life	-0.647***	0.0601	-0.579***
Present economic conditions	0.319**	0.469***	0.0801
Future economic conditions	0.813	-0.792*	0.802

*** p<0.01, ** p<0.05, * p<0.1.

Robustness check 3: Initial Specification for only waves 1 and 2

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.115	-0.0840	0.0623
Log(real consumption)	-0.860*	-0.141	-0.193
Education			
Currently attending school (Girls)	0.0971	-0.130*	0.0688
Labor			
Labor force participation: Active	0.0374	-0.112	0.137*
Unemployed	0.0390	-0.0883	0.115
Hunger			
Hunger incidence	-0.0397	0.570	-0.339
Perceptions of welfare			
Life satisfaction	-0.538	2.496***	-0.437
Present living conditions	0.507	-0.224	0.0303
Future living conditions	1.367*	0.263	0.561
Control over own life	-0.0384	0.128	-0.304
Present economic conditions	-0.276	0.279	-0.227
Future economic conditions	1.339*	-0.267	0.928**

*** p<0.01, ** p<0.05, * p<0.1.

Robustness check 4: Initial Specification for only waves 1 and 4

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.596***	0.354**	0.521***
Log(real consumption)	-1.117***	-0.724***	-0.971***
Education			
Currently attending school (Girls)	0.0503	-0.0373	0.0646
Labor			
Labor force participation: Active	-0.185	-0.368***	-0.0806
Unemployed	-0.0123	0.117**	-0.0384
Hunger			
Hunger incidence	0.525**	0.239	0.451***
Perceptions of welfare			
Life satisfaction	-1.593**	-1.957**	-1.041**
Present living conditions	0.575**	0.539**	0.416*
Future living conditions	1.366	-0.0110	1.198
Control over own life	-0.904*	-0.310	-0.795*
Present economic conditions	0.334	0.364	0.240
Future economic conditions	1.325	-0.497	1.256

*** p<0.01, ** p<0.05, * p<0.1.

Robustness check 5: Initial Specification for only waves 1 and 4 with inflation computed with wave 1 prices

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.645***	0.649***	0.186
Log(real consumption)	-1.684***	-1.509***	-0.618**
Education			
Currently attending school (Girls)	0.116	0.0678	0.0516
Labor			
Labor force participation: Active	-0.000360	-0.318***	0.186**
Unemployed	-0.0266	-0.0890	0.0175
Hunger			
Hunger incidence	0.0485	0.332	-0.0918
Perceptions of welfare			
Life satisfaction	-0.466	-1.007	0.229
Present living conditions	0.349	0.493*	0.0594
Future living conditions	1.665	2.951**	0.638
Control over own life	-0.949*	-0.783	-0.531*
Present economic conditions	0.587**	0.344	0.364**
Future economic conditions	1.121	2.153	0.500

*** p<0.01, ** p<0.05, * p<0.1.

Robustness check 6: Adding the initial level of CPI at the Boma Level for only waves 1 and 2

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.114	-0.0846	0.0618
Log(real consumption)	-0.861*	-0.143	-0.190
Education			
Currently attending school (Girls)	0.105	-0.124*	0.0658
Labor			
Labor force participation: Active	0.0447	-0.110	0.141*
Unemployed	0.0400	-0.0860	0.118
Hunger			
Hunger incidence	-0.0352	0.580	-0.342
Perceptions of welfare			
Life satisfaction	-0.542	2.557***	-0.454
Present living conditions	0.489	-0.254	0.0395
Future living conditions	1.296	0.172	0.595
Control over own life	-0.0220	0.146	-0.307
Present economic conditions	-0.257	0.302	-0.223
Future economic conditions	1.345*	-0.251	0.893*

*** p<0.01, ** p<0.05, * p<0.1.

Robustness check 7: Adding the initial level of CPI at the Boma Level for only waves 1 and 4

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.599***	0.358**	0.521***
Log(real consumption)	-1.118***	-0.725***	-0.972***
Education			
Currently attending school (Girls)	0.0539	-0.0290	0.0660
Labor			
Labor force participation: Active	-0.184	-0.367***	-0.0807
Unemployed	-0.0122	0.117**	-0.0379
Hunger			
Hunger incidence	0.524**	0.235	0.456***
Perceptions of welfare			
Life satisfaction	-1.592**	-1.946**	-1.030**
Present living conditions	0.575**	0.541**	0.408*
Future living conditions	1.364	-0.000547	1.194
Control over own life	-0.905*	-0.324	-0.791*
Present economic conditions	0.333	0.356	0.241
Future economic conditions	1.322	-0.484	1.240

*** p<0.01, ** p<0.05, * p<0.1.

Robustness check 8: High inflation households for waves 1 and 2 (dummy variable)

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.0505	-0.0507	-0.0439
Log(real consumption)	-0.0947	-0.229	0.111
Education			
Currently attending school (Girls)	0.0801	-0.136**	0.131**
Labor			
Labor force participation: Active	0.0520	-0.101**	0.120**
Unemployed	-0.0219	0.0222	-0.0258
Hunger			
Hunger incidence	-0.136	0.654***	-0.421
Perceptions of welfare			
Life satisfaction	-0.666	1.358***	-0.605
Present living conditions	0.303	-0.413**	0.274
Future living conditions	0.457	0.212	0.369
Control over own life	-0.0858	-0.0888	-0.303
Present economic conditions	-0.152	0.144	-0.0478
Future economic conditions	0.417	-0.130	0.594

*** p<0.01, ** p<0.05, * p<0.1.

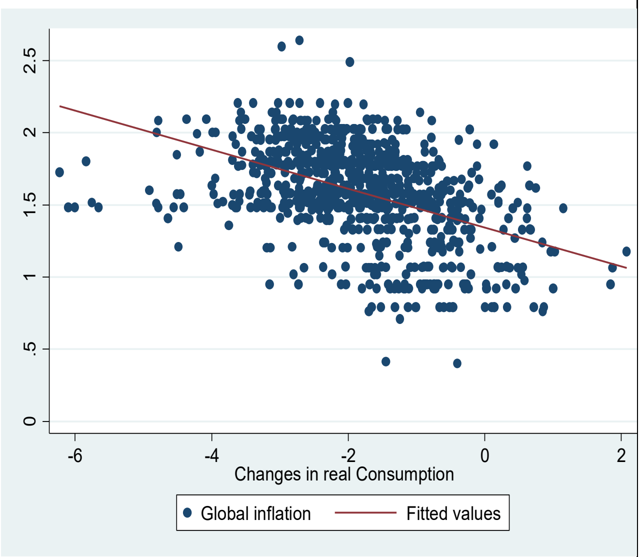
Robustness check 9: High inflation households for waves 1 and 4 (dummy variable)

Outcomes	Total Inflation	Food price inflation	Non-food price inflation
Poverty			
Poor (below USD 1.90 PPP)	0.275**	0.454***	0.0824
Log(real consumption)	-0.787***	-0.957***	-0.399*
Education			
Currently attending school (Girls)	0.0668*	0.0500	0.0250
Labor			
Labor force participation: Active	0.0187	-0.173**	0.104
Unemployed	-0.00484	-0.0517	0.0238
Hunger			
Hunger incidence	-0.0107	0.265	-0.0883
Perceptions of welfare			
Life satisfaction	-0.0178	-0.632	0.0778
Present living conditions	0.102	0.303*	0.0946
Future living conditions	0.542	1.929***	0.0101
Control over own life	-0.356	-0.669**	-0.286
Present economic conditions	0.296**	0.227	0.333**
Future economic conditions	0.0931	1.576	-0.302

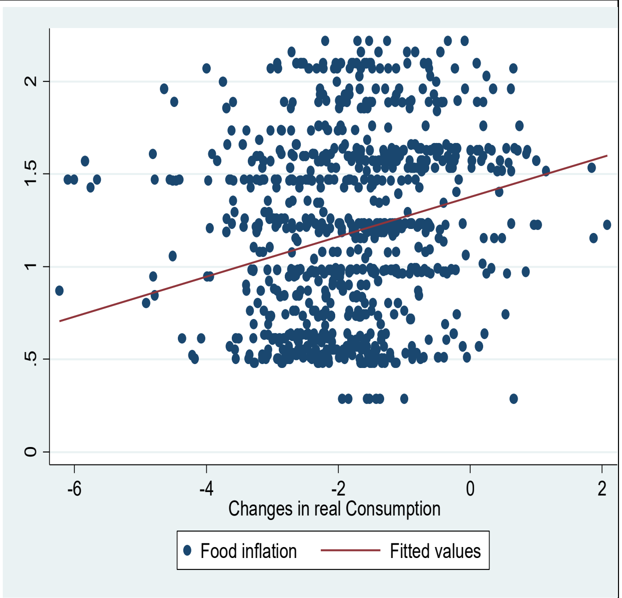
*** p<0.01, ** p<0.05, * p<0.1.

Figure C4-18: Robustness checks regression results

Total (global) inflation and changes in real consumption (Waves 1, 2, 4)

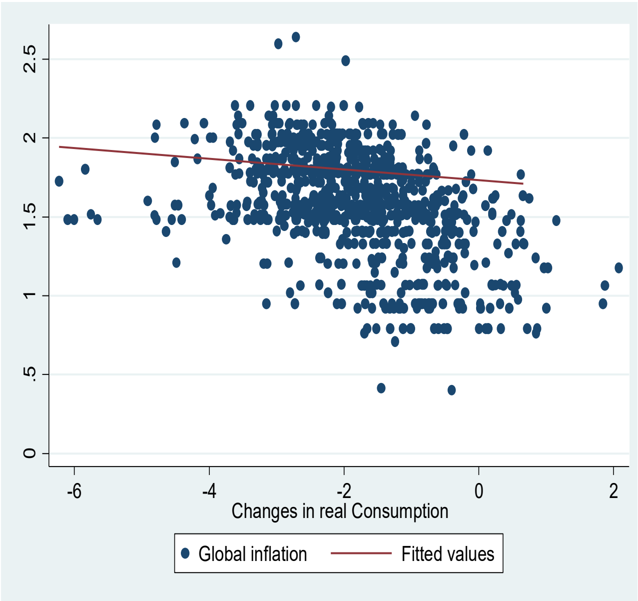


Food Inflation and Changes in real consumption (Waves 1, 2, 4)

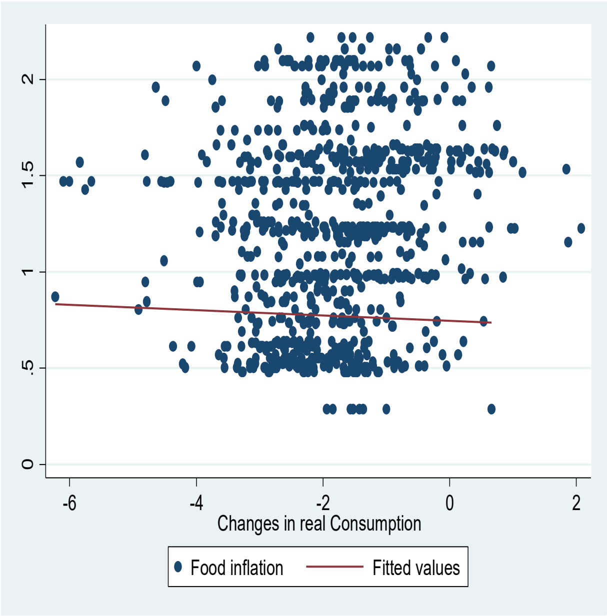


Non-food Inflation and changes in real consumption (Waves 1, 2, 4)

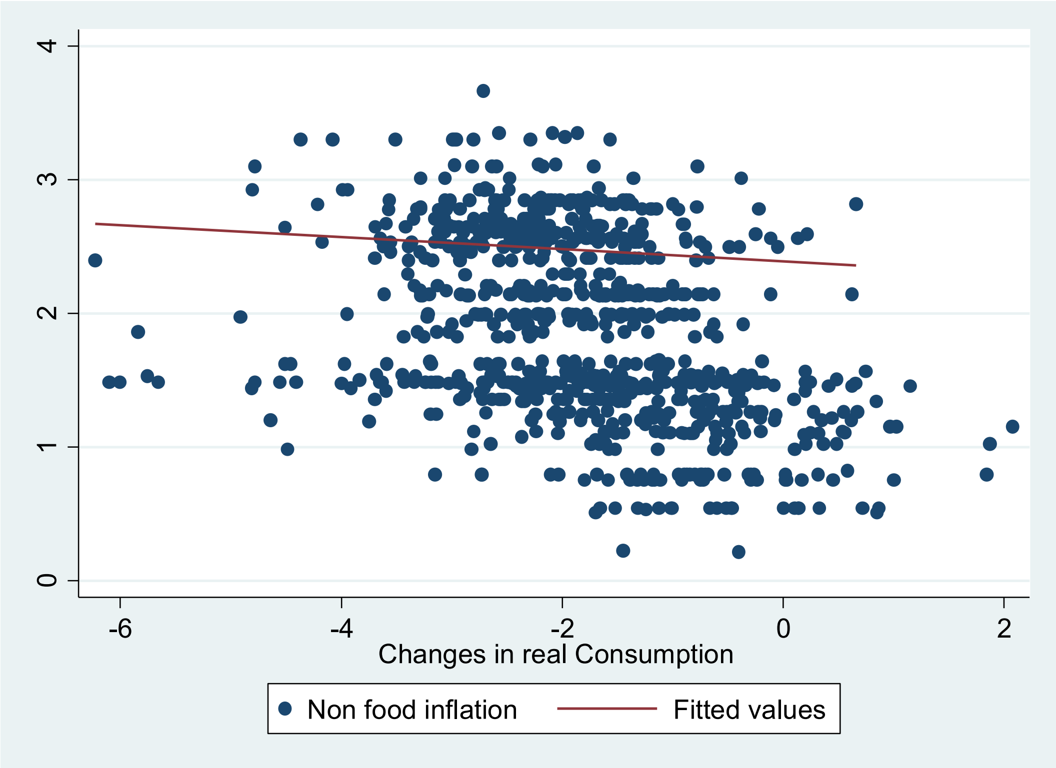
Total (global) and changes in real consumption (Waves 1, 2)



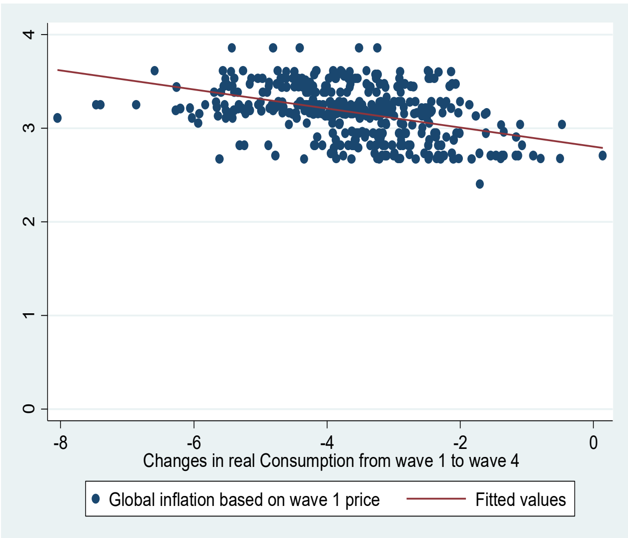
Food inflation and changes in real consumption (Waves 1, 2)



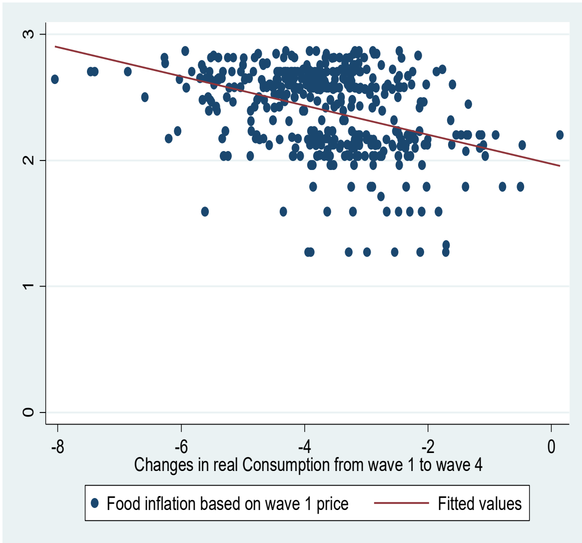
Non-Food inflation and changes in real consumption (Waves 1, 2)



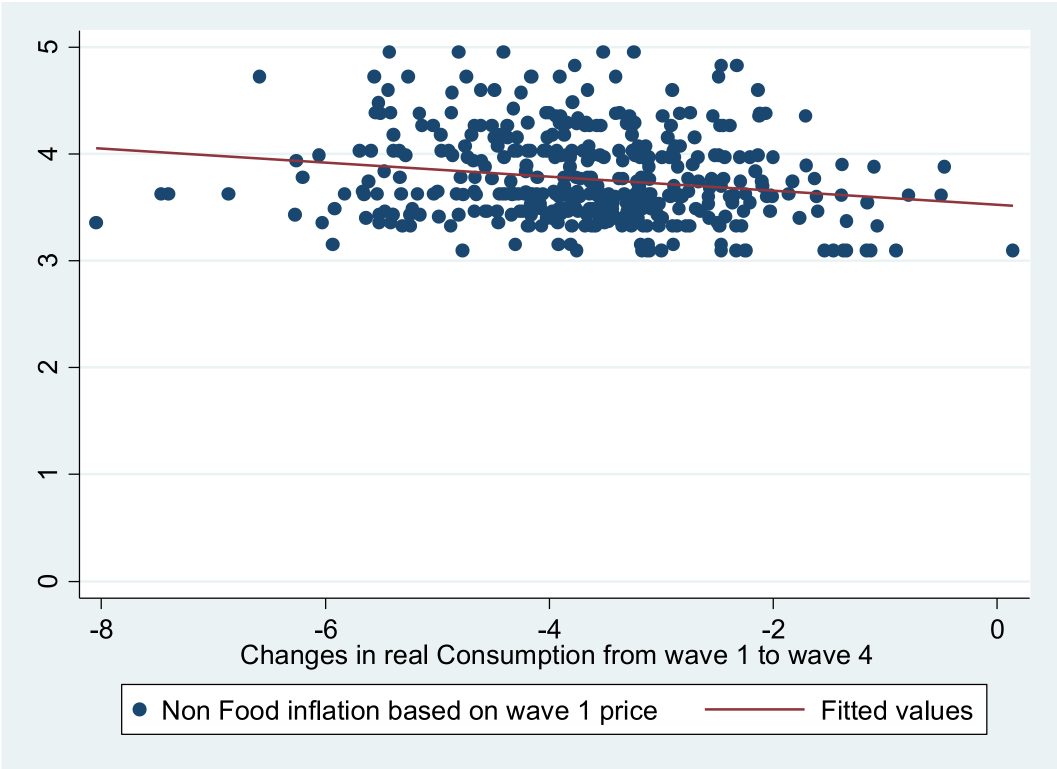
Total (global) and changes in real consumption (Waves 1, 4)



Food inflation and changes in real consumption (Waves 1, 4)



Non-Food inflation and changes in real consumption (Waves 1, 4)



5. Impact of Drought on Poverty in Somalia[229]

Utz Pape and Philip Wollburg[230]

Introduction

The impact of adverse climatic, and other, income shocks on household and individual welfare in developing countries is an issue of considerable policy interest. Understanding the magnitude and importance of income shocks in causing and perpetuating poverty is critical to designing measures aimed at building resilience, contributing towards the goal of ending poverty. A growing body of literature provides empirical evidence of the micro-level impacts of adverse shocks in developing countries. Dercon and Krishnan (2000), Dercon (2004), and Porter (2012) find that weather shocks have a negative and long-lasting effect on consumption outcomes in rural Ethiopia. Hill and Porter (2016) and Makoka (2008) show that drought and price shocks reduce consumption and especially farm income, while increasing vulnerability to poverty in rural Ethiopia and Malawi, respectively. Similarly, Alem and Soderbom (2012) conclude that high food prices adversely affect households in urban Ethiopia, especially those relying on casual work and with low asset levels. Hill and Mejia-Mantilla (2017) find negative effects of drought, conflict, and prices on poverty levels in Uganda, and Parisotto and Pape (forthcoming) find a large and significant impact of conflict on poverty in South Sudan. Hoddinott and Kinsey (2001) and Alderman, Konde-Lule et al. (2006) show the causal relation between rainfall shocks and reduced human capital formation.

This paper contributes to the existing literature, by focusing on the impact of drought on poverty in Somalia. Four consecutive seasons of poor rains between April 2016 and December 2017 resulted in a severe drought across Somalia (FEWSNET 2018). The drought exacerbated preexisting food insecurity, as half of the population faced acute food insecurity in mid-2017 (FEWSNET 2016, FSNAU 2017). The drought threatened the livelihoods of many Somalis. Lack of water and pasture led to high livestock deaths and low birth rates, and induced distress selling caused the 26 percent of Somalis relying on livestock for their livelihoods to lose between 25 and 75 percent of their herds in the first half of 2017 (FEWSNET 2018). Households depleted productive assets and food stocks to cope with the rising food and water prices, while weak demand for labor in the agricultural sector led to lower wage levels (FEWSNET 2017). As a result, the drought displaced close to one million people between 2016 and 2017. Large-scale humanitarian interventions provided critical relief to up to 3 million people to reduce the risk of famine (FEWSNET 2017).

Using data from two waves of the Somali High Frequency Survey (SHFS), this analysis employs a regression framework to measure the micro-level impact of the 2016/17 drought on poverty. It exploits spatial variation in the intensity of drought that different households experienced and compares consumption before and after the drought. Households' level of drought exposure is measured by using the Normalized Difference Vegetation Index (NDVI). The temporal difference is provided by the timing of the first two waves of the SHFS. The first wave took place before the onset of the drought in early 2016, while the second wave surveyed households in late 2017, when the drought had surpassed its peak.

The remainder of this paper is structured as follows. Section 0 describes the data used to measure the impact of drought on poverty. Section 0 outlines the identification and estimation strategy. Section 0

[229] UP developed the research question and designed as well as supervised the field work. PW and UP jointly conducted the analysis, interpreted results, and drafted as well as finalized the manuscript.

[230] Authors in alphabetically order. Corresponding author: Utz Pape (upape@worldbank.org). The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. The authors would like to thank Ruth Vargas Hill, Wendy Karamba, and Gonzalo Nunez for discussions.

presents the results, and section 0 tests the robustness of these results. Section 0 concludes with a discussion of the results and discusses policy recommendations.

Data

This analysis uses cross-sectional household-level data from two waves of the SHFS. Wave 1 interviewed 4,117 urban, rural, and IDP households in February and March of 2016, representative of 9 of 18 Somali pre-war regions, excluding inaccessible areas in the south. Wave 2 expanded coverage to all but one, inaccessible pre-war region, and included the nomadic population, interviewing a total of 6,092 households in December 2017 (Table C5-1).

Table C5-1: Number of interviews by population type.

Population type	Wave 1 households	Wave 2 households
Urban	2,864	4,011
Rural	822	1,106
IDP	431	468
Nomadic	0	507
Total	4,117	6,092

The analysis excludes nomadic households and IDPs within and outside IDP settlements to avoid invalid comparisons between wave 1 and wave 2. Large-scale drought-related displacement implies that IDP populations before the drought in wave 1 were different from IDP populations surveyed during the drought in wave 2. Nomadic households do not have a permanent place of residence, so that a geographical exposure measure cannot be assigned in a meaningful way. The final data set contains 5,852 urban and 1,594 rural households. The data include information on consumption and key household and individual characteristics and perceptions, as well as information on shocks and vulnerabilities. Poverty is measured against the international poverty line of US$ 1.90 per capita per day, derived from the spatially and intertemporally deflated consumption aggregate (Pape and Wollburg 2019).

The Normalized Deviation Vegetation Index (NDVI) is used to determine the exposure of households to the drought. The NDVI is derived from satellite images measuring the health of vegetation. Below-average NDVI values imply drier-than-usual conditions, indicating the vegetation health is also below-average. NASA's MODIS Terra and Aqua platform provides the daily global NDVI data at 500m resolution, which serve as the source of data for this analysis (Schaaf 2015). While four consecutive rainy seasons delivered poor rains in 2016 and 2017, the severe rainfall deficits in the second rainy season of 2016 and first rainy season of 2017 were the key drivers of the 2016/17 drought in Somalia (FEWSNET 2018).[231] Hence, each household's level of drought exposure is defined in this analysis as the percentage deviation of the NDVI during these two seasons from the pre-drought 2012 to 2015 average,[232] within a 25km radius around each household. The levels of drought exposure range from NDVI values of 6 percent above average to 20 percent below average in wave 1, and from 4 percent above average to 36 percent below average in wave 2, reflecting the overall spectrum of drought severity (Figure C5-1).

[231] Somalia has two main rainy seasons: the main Gu rains from April to June and the short Deyr rains from October to December. Significantly below-average rainfall started with the 2016 Gu rains and extended to the 2017 Deyr rains.

[232] In 2010-11, there was a severe drought in Somalia. The reference period was chosen to start after the 2010-11 drought to preclude this unusual event from interfering with the series average.

Figure C5-1: Distribution of NDVI distribution, all Somalia, wave 1 and wave 2 households

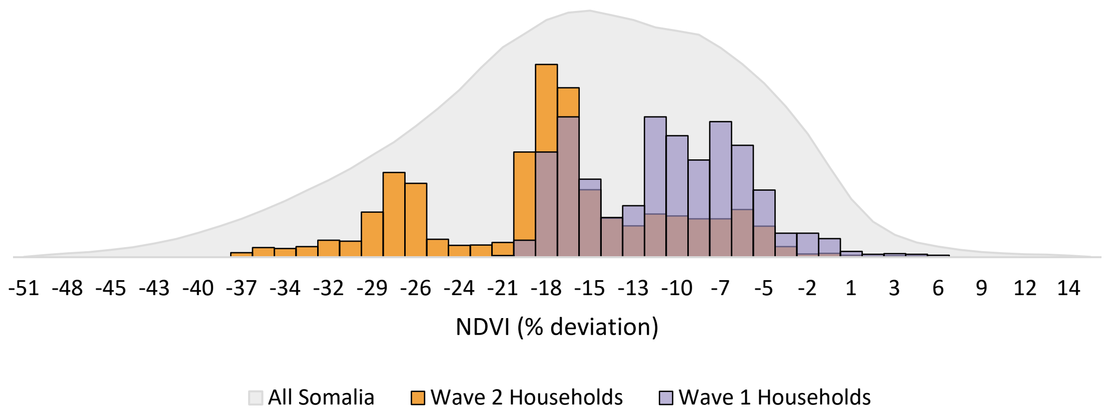


Figure C5-2: NDVI deviation, 2016 Deyr season

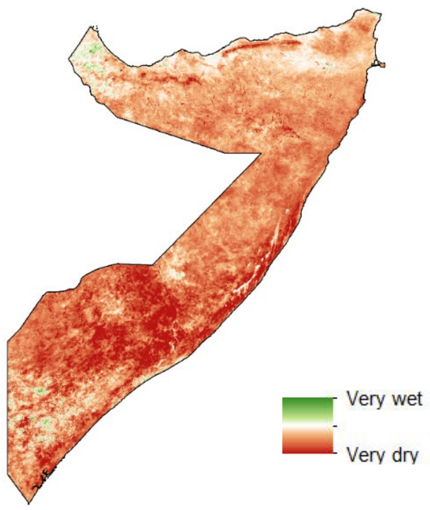
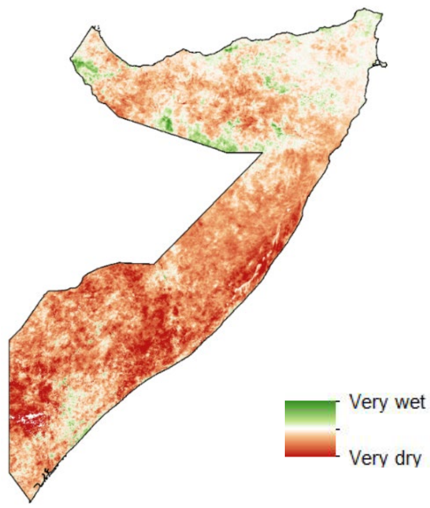


Figure C5-3: NDVI deviation, 2017 Gu season



In controlling for potential confounding factors, we rely on geo-coded conflict fatality data provided by the Armed Conflict Location Event Dataset (ACLED) and on data on the percentage of target beneficiaries reached with aid by pre-war region coming from the Food Security Cluster Somalia.

Methodological approach

This analysis uses a regression framework similar to Hill and Porter (2016) to estimate the effect of the drought on poverty and consumption. To isolate the drought effect, the analysis exploits two characteristics of the SHFS data set. First, fieldwork timing was such that data were collected before the drought shock (wave 1) and during the drought (wave 2), allowing for a before-and-after comparison. Second, there was spatial variation in households' exposure to drought, with some in highly and others in less drought-affected areas (Figure C5-1; Figure C5-2; Figure C5-3). The analysis compares how much poverty and consumption changed between wave 1 and wave 2 for households in highly drought-exposed areas relative to households in less drought-exposed areas, which can be written as

$$Y_{it} = \beta_0 + \beta_1 post_t + \beta_2 DroughtIntensity_i + \beta_3 post_t * DroughtIntensity_i + \varepsilon_{it} \quad (8)$$

Here, Y_{it} denotes outcomes of interest for household *i* at time *t*, primarily the poverty headcount rate. $post_t$ is a binary variable indicating time period *t* (wave 1, wave 2) and $DroughtIntensity_i$ is the continuous treatment variable, indicating the level of drought exposure of household *i* in standard deviations of NDVI anomaly from the 2012 to 2015 average. ε_{it} denotes the error term. β_0 is the intercept, β_1 is the expected mean change in outcome from wave 1 to wave 2. The coefficient of the drought exposure variable, β_2, is the estimated mean difference in outcomes prior to the drought: it represents whatever baseline differences existed between households before exposure to the drought. The coefficient of interest is $\beta_{3,}$ which estimates the drought effect.

In (1), if households in highly drought-exposed areas experienced a larger increase in poverty than households in less drought-exposed areas, the interpretation is that drought increased poverty. The validity of this conclusion rests on the assumption that households in wave 1 and wave 2 and in highly and less drought-affected areas make for good comparison groups, so that exposure to drought can be thought of as exogenous.

This assumption may be violated for several reasons. First, there may be factors that affect the outcome variables at the same time as the drought, such as conflict or humanitarian assistance. Second, the use of repeated cross-sectional data does not allow for household-level fixed effects to control for all baseline differences. Critically, some regions may be inherently more likely than others to experience drought. Therefore, a vector of control variables $\boldsymbol{X}_{it}$ is introduced, such that equation (8) becomes

$$Y_{it} = \beta_0 + \beta_1 post_t + \beta_2 DroughtIntensity_i + \beta_3 post_t * DroughtIntensity_i + \beta_4 \boldsymbol{X}_{it} + \varepsilon_{it} \tag{9}$$

As a proxy for the region's propensity to experience drought, $\boldsymbol{X}_{it}$ includes a measure of the medium-term (2002 to 2013) deviation from the NDVI average for each region surveyed (Hill and Mejia-Mantilla, 2017). Price levels are a further potential confounding factor and are therefore included in $\boldsymbol{X}_{it}$. Further control variables fall into five categories: regional and population-type controls, household characteristics, dwelling characteristics, exposure to conflict, and humanitarian assistance. Equation (9) is implemented with OLS, Probit, or quantile regressions, depending on the objective at hand: when the dependent variable is binary, as is the case with poverty and hunger, Probit is used. When the depending variable is continuous, as with consumption, OLS is more appropriate. Quantile regressions are used to understand the drought impact along the entire consumption distribution, to gauge whether the drought affected households at different welfare levels differentially.

The drought impact is estimated from the full set of urban and rural households surveyed in wave 1 and wave 2 of the SHFS. Geographical coverage across waves was different, as additional regions were surveyed in wave 2 (Figure C5-8; Figure C5-9). The lack of complete geographical overlap impedes controlling for regional idiosyncrasies of regions covered in wave 2 only at baseline. As a robustness check, we present a specification of only overlapping wave 1 and wave 2 areas, allowing for a genuine region fixed effect. The additional specification restricts the analysis to urban households in Mogadishu and the north-west and to rural households only in the north-west. This limits the appeal of the additional specification because it reduces the analysis to estimating a localized rather than global drought-effect.

Results

We find that, in rural areas, more drought-exposed households experienced a significant reduction in consumption and increase in poverty. An increase of one standard deviation in drought exposure during the 2016/17 drought led to a decline in household consumption of 26 percent, based on the preferred regression specification with the full set of controls. One standard deviation increase in drought exposure corresponds to a seven percentage-point negative anomaly in NDVI, relative to the pre-drought average. This reduction in consumption corresponds to an increase of 15 percent in the probability of being poor (Table C5-2, column III). The drought had no significant effect on poverty in consumption in urban areas, nor in the combined sample of urban and rural households (Table C5-2, column I and II; Table C5-6).

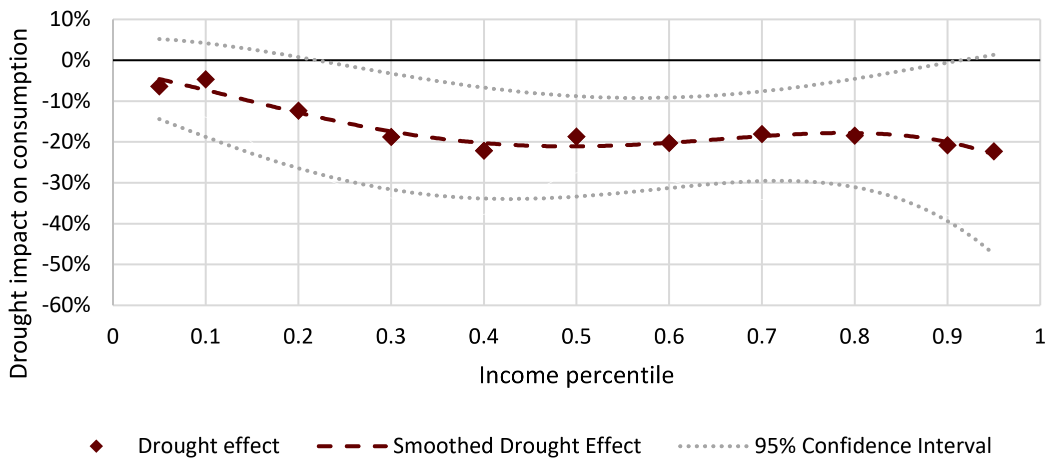
Table C5-2: Drought impact on poverty and consumption.

	(I)	(II)	(III)
Sample	Full urban + rural sample	Full urban sample	Full rural sample
Outcome variable (Yit)		Poverty Status	
Drought Impact	-0.00535	-0.0201	0.264***
S.E.	(0.0508)	(0.0568)	(0.0812)
Outcome variable		ln(Core Consumption)	
Drought Impact	1.59e-05	0.0286	-0.146**
S.E.	(0.0386)	(0.0345)	(0.0665)
Controls (Xit)	Yes	Yes	Yes
Observations	7,214	5,678	1,536
R-squared	0.352	0.345	0.522

Note: ***p<0.01, **p<0.05, *p<0.1. Poverty status results estimated using Probit, Consumption results estimated using OLS. Drought effect expressed in standard deviations of NDVI loss.

Implementing equation (2) with controls through quantile regressions allows assessing the drought's impact on consumption at different points along its distribution. In rural areas, the drought's impact on consumption was smaller for the poorest households. Higher drought exposure had no significant impact on consumption for the poorest 10 percent of rural households, reduced consumption by 17 percent for rural households at the twentieth percentile, and between 20 and 25 percent for the top 80 percent of rural households (Figure C5-4). In urban areas, the impact is around zero across the income distribution (Figure C5-10).

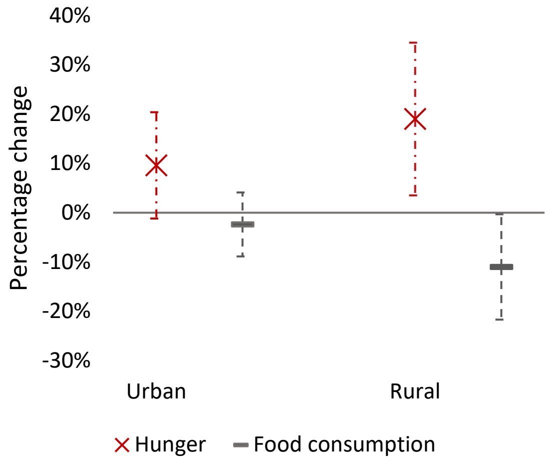
Figure C5-4: Drought effect along the consumption expenditure distribution, rural areas.



Varying levels of drought exposure along the consumption distribution do not explain these differences, as the median drought intensity among the poorest 10 percent of households is similar to the overall average drought exposure. With an average poverty gap of 72 percent, this group is very poor. These households may have relied on humanitarian assistance already before the onset of the drought, thus being isolated from the impact of the drought on consumption.

More drought exposed households were also more likely to experience hunger. As levels of hunger rose across all Somali regions, rural households in highly drought-exposed areas were most severely affected. Higher drought exposure led to an 11 percent decrease in food consumption, accompanied by a 19 percent increase in the probability of experiencing hunger in December 2017. The effect is much less pronounced among urban households (Table C5-7; Table C5-8).

Figure C5-5: Drought effect on hunger and food consumption.



Robustness

The robustness of this analysis' main findings on the drought impact on rural poverty is verified in several ways. First, the results' sensitivities to the inclusion and exclusion of various groups of control variables are tested. The results are robust across all tested specifications, and do not depend on the inclusion of certain groups of control variables. With the inclusion of subsequent groups of control variables, the drought impact point estimates vary between 19 and 30 percent increase in the probability of being poor.

Table C5-3: Robustness of results across various specifications.

	(I)	(II)	(III)	(IV)	(V)	(VI)
Sample	Full rural sample					
Outcome variable	PoorPPP					
Drought Impact	0.192***	0.256***	0.222***	0.230***	0.301***	0.264***
S.E.	(0.0629)	(0.0733)	(0.0679)	(0.0739)	(0.0769)	(0.0812)
Outcome variable	ln(Core Consumption)					
Drought Impact	-0.107**	-0.189***	-0.146***	-0.152***	-0.169**	-0.146**
S.E.	(0.0428)	(0.0598)	(0.0555)	(0.0551)	(0.0668)	(0.0665)
Controls						
Regional	No	Yes	Yes	Yes	Yes	Yes
Household	No	No	Yes	Yes	Yes	Yes
Dwelling	No	No	No	Yes	Yes	Yes
Conflict	No	No	No	No	Yes	Yes
Assistance	No	No	No	No	No	Yes
Observations	1,591	1,591	1,563	1,536	1,536	1,536
R-squared	0.032	0.226	0.359	0.487	0.501	0.522

Note: ***p<0.01, **p<0.05, *p<0.1. Poverty status results estimated using Probit, Consumption results estimated using OLS. Drought effect expressed in standard deviations of NDVI loss.

Second, the analysis is replicated on households only in regions overlapping between wave 1 and wave 2 of the SHFS. This reduces the geographical scope of the analysis to rural areas in the north-west of Somalia and to urban areas in the north-west and Mogadishu. The results from this subsample are in line with the main findings from the full sample of households (Table C5-9). In rural areas, more drought-exposed households are 36 percent more likely to be poor, experiencing a 15 percent reduction in consumption, slightly higher than the 26 and 15 percent, respectively, in the full sample. No significant effect was found in urban areas.

Third, the sample is restricted by iteratively removing from the analysis households from north-eastern regions (Table C5-4, columns I and II), central regions (Table C5-4, columns III and IV), and south-western regions (Table C5-4, columns V and VI). The main results are largely unchanged when removing north-eastern regions, while a larger drought effect is found when removing south-western households. In contrast, the results are weaker when excluding central regions, though point estimates are still in a similar range. This indicates that the drought effect was weaker in south-west and particularly strong in central regions.

Table C5-4: Regression results with restricted samples.

Sample	Rural, NE excluded		Rural, Central excluded		Rural, SW excluded	
Outcome variable			Poverty			
Drought Impact	0.197***	0.251***	0.137**	0.201***	0.224***	0.424***
S.E.	(0.066)	(0.082)	(0.038)	(0.051)	(0.059)	(0.074)
Outcome variable			ln(Core Consumption			
Drought Impact	-0.129***	-0.149**	-0.051	-0.071	-0.128***	-0.195***
S.E.	(0.048)	(0.075)	(0.038)	(0.051)	(0.043)	(0.057)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,511	1,456	1,087	1,035	1,319	1,277
R-squared	0.054	0.508	0.029	0.507	0.065	0.561

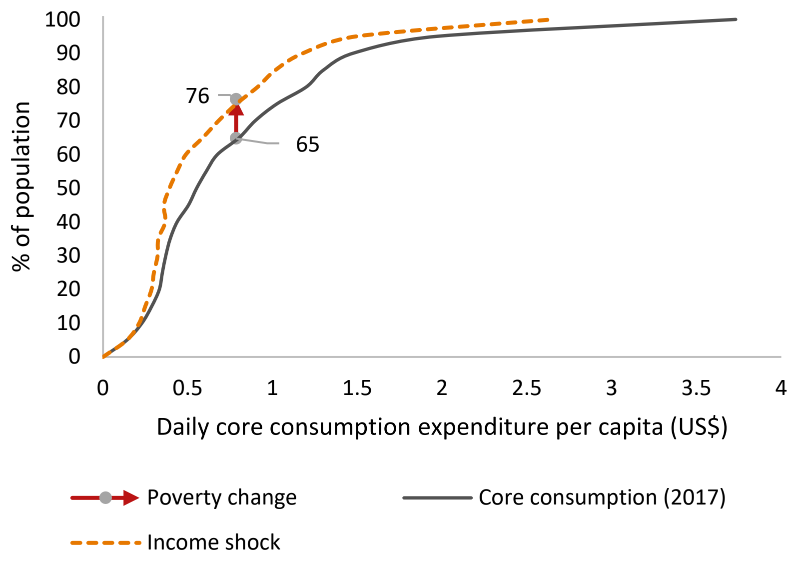
*Note: ***p<0.01, **p<0.05, *p<0.1. Poverty status results estimated using Probit, Consumption results estimated using OLS. Drought effect expressed in standard deviations of NDVI loss.*

Discussion and conclusions

The results show that the 2016/17 drought severely affected consumption in rural households. The magnitude of the drought impact is generally in line with findings in the literature, but on the upper end of the reported effects. For example, Hill and Porter (2016) find that a moderate drought shock leads to a 9 percent reduction in consumption in Uganda, while this analysis finds an effect almost double that size. However, given the severity of the 2016/17 drought in Somalia, the results appear consistent.

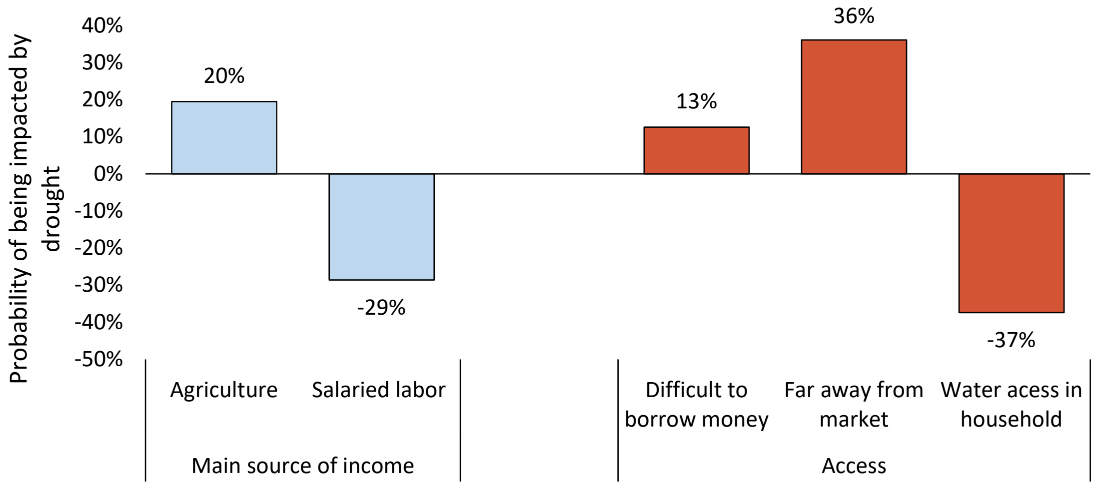
Droughts are cyclical events in Somalia and the Horn of Africa region in general. Recently, severe droughts affected Somalia in 2011 and 1991. A renewed drought shock is therefore likely very to occur at some point in the future. The detailed results from the regression analysis allow to simulate how a renewed income shock of the same magnitude as the 2016/17 drought would affect rural households. To model another income shock, the quantile regression estimates of the drought's effect on household consumption at different points along its distribution are applied to the SHFS data. Based on this simulation, a renewed income shock could increase rural poverty by nine percentage points, from 65 to 76 percent (Figure C5-6).

Figure C5-6: Simulation of income shock among rural households.



The simulation results emphasize that a sustainable poverty reduction strategy should involve making rural households more resilient to climatic income shocks. To guide such efforts, it is instructive to analyze the characteristics of the most affected rural households. To do this, we focus on which households self-reported having been affected by the 2016/17 drought, regressing it on household characteristics while controlling for location, income, and households' exogenous level of drought exposure as measured with NDVI. First, households relying on agricultural income are 20 percent more likely to report being affected by the drought. In contrast, households relying on salaried labor are significantly less likely to report being affected by the drought (Figure C5-7). The fact that households relying on agricultural income are mainly in rural areas is likely part of the reason why no drought effect was found in urban areas. The particular vulnerability to drought shocks of agricultural households is also well-documented in the literature (e.g. Hill and Mejia-Mantilla, 2017). This set of findings suggests that agricultural households may benefit from insurance products, such as agricultural index insurance (see Berhane et al., 2012; Dercon et al., 2014). Further, measures facilitating the diversification of income sources, especially the shifting of household members towards wage jobs, could help cushion the effect of climatic shocks (Alem and Soderbom, 2012).

Figure C5-7: Correlates of drought-impacted rural households.



Note: Coefficients from Probit regression with self-reporting to be impacted by the drought as dependent variable. Regression with controls for drought-exposure measured by NDVI, household income, and region. All reported results significant at the 5%-level.

Second, rural households without access to water in the dwelling, agricultural households more than an hour away from the nearest food market, and households who struggle to borrow money in an emergency were also more likely to be impacted by the drought (Figure C5-7). While the latter is another indication of the usefulness of agricultural insurance, these results show that households lacking access to infrastructure and services are also particularly vulnerable. Investments in infrastructure and basic services could thus improve rural households' resilience.

Appendix

Figure C5-8: Coverage wave 1

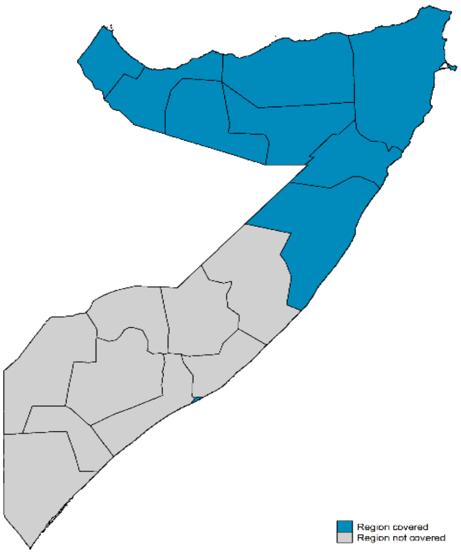


Figure C5-9: Coverage wave 2

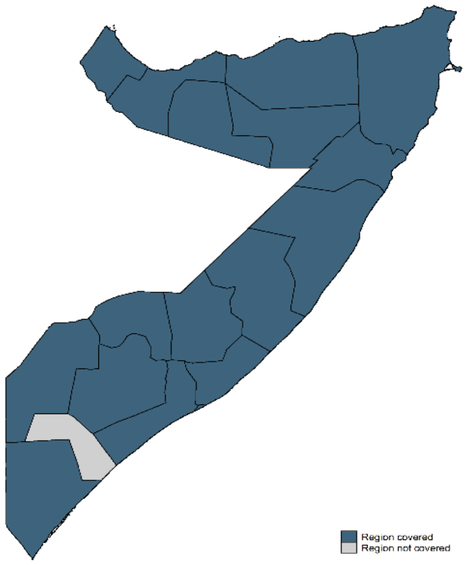


Table C5-5: List of control variables for the regression analysis.

Variable	Description	Source
Average NDVI	Average value of NDVI at the district-level, 2002-2013.	MODIS NDVI data from WFP VAM
Price level	Price level at the disaggregation of analytical strata.	SHFS data
Regional and population type controls		
Region x type	Interaction takes the following values: Mogadishu-urban, NE-urban, NE-rural, NW-urban, NW-rural, Central regions - urban, Central regions - rural, Jubbaland-urban, SW-urban, SW-rural.	
Type	Urban, rural indicator.	
Household characteristics		
Household size	Number of members in the household.	
Remittances	Household remittances receipt status (Yes/No).	
Household head age	Age of the household head (years).	
Household head literacy	Literacy of the household head (Yes/No)	
Gender composition	Gender composition of the household (Share of males).	
Dwelling characteristics		
Tenure	Tenure status of household (own, rent, other).	
Dwelling type	Type of the dwelling (Shared, separate, other).	
Roof material	Roof material of the dwelling (Metal sheets, Tiles, Harar, Wood, Plastic, Other).	
Floor material	Floor material of the dwelling (Concrete, Tiles or Mud, Other).	
Improved sanitation	Access to improved sanitation.	
Conflict controls		
Conflict fatalities	Conflict fatalities in district in past 12 month according to ACLED.	ACLED
Conflict x drought	Interaction of drought intensity and conflict fatalities.	ACLED; MODIS NDVI
Assistance controls		
Assistance in region	Percentage of beneficiaries reached through food aid and livelihood inputs in 2017 in region.	Food Security Cluster

Table C5-6: Regression results, consumption and poverty, full sample.

Outcome variable		Consumption			Poverty	
Population	urban + rural	urban	Rural	urban + rural	urban	rural
Post	-0.155***	-0.174***	-0.202***	0.272***	0.357***	0.425***
	(0.036)	(0.036)	(0.051)	(0.051)	(0.063)	(0.076)
Drought Intensity	-0.050**	-0.066***	0.100***	0.068*	0.084*	-0.132***
	(0.024)	(0.024)	(0.029)	(0.040)	(0.045)	(0.050)
Drought Effect	0.000	0.029	-0.146**	-0.005	-0.020	0.264***
	(0.039)	(0.035)	(0.066)	(0.051)	(0.057)	(0.081)
Average NDVI	0.413***	0.170**	0.814**	-0.450***	-0.255*	-0.049
	(0.136)	(0.082)	(0.381)	(0.157)	(0.134)	(0.407)
Price level	-0.192	-0.410**	0.375	0.572***	0.475*	0.411

	(0.165)	(0.164)	(0.399)	(0.167)	(0.245)	(0.386)
Regional controls						
NE-urban	0.051	0.200***			-0.469***	
	(0.077)	(0.072)			(0.101)	
NW-urban	-0.117*	0.001			-0.115	
	(0.069)	(0.055)			(0.082)	
NE-rural	-0.244***					
	(0.067)					
NW-rural	-0.210***		0.124***			0.367***
	(0.076)		(0.042)			(0.076)
Central-urban	0.153**	0.201***			-0.466***	
	(0.066)	(0.072)			(0.112)	
Central-rural	0.205		0.793***			-0.198
	(0.184)		(0.224)			(0.190)
Jubbaland-urban	0.589***	0.484***			-1.127***	
	(0.119)	(0.079)			(0.152)	
SW-urban	0.372***	0.199***			-0.386***	
	(0.101)	(0.070)			(0.127)	
SW-rural	0.282**		1.283***			-0.459
	(0.124)		(0.346)			(0.298)
Household controls						
HH head literacy	0.047***	0.066***	0.011	-0.051*	-0.062*	-0.040
	(0.016)	(0.015)	(0.028)	(0.030)	(0.033)	(0.056)
HH head age	0.001**	0.001	0.002**	-0.001	-0.000	-0.003**
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
Received remittances	0.065***	0.073***	0.039	-0.136***	-0.143***	-0.118
	(0.014)	(0.015)	(0.031)	(0.022)	(0.021)	(0.079)
Household size	-0.058***	-0.056***	-0.057***	0.081***	0.084***	0.068***
	(0.003)	(0.003)	(0.008)	(0.005)	(0.005)	(0.017)
Gender composition	0.030	-0.005	0.112*	-0.072	-0.028	-0.195*
	(0.031)	(0.032)	(0.057)	(0.057)	(0.062)	(0.101)
Dwelling controls						
Dwelling tenure: Rent	0.010	0.008	0.023	-0.027	-0.042	0.038
	(0.014)	(0.016)	(0.027)	(0.026)	(0.028)	(0.051)
Dwelling tenure: Other	-0.051*	-0.075**	0.029	0.107**	0.162***	-0.011
	(0.027)	(0.029)	(0.053)	(0.045)	(0.053)	(0.078)
Dwelling floor: Tiles or mud	-0.005	0.025*	-0.192***	-0.018	-0.055*	0.205***
	(0.016)	(0.015)	(0.049)	(0.027)	(0.029)	(0.062)
Dwelling floor: Other	-0.064***	-0.061**	-0.193***	0.044	0.064	0.202***
	(0.023)	(0.024)	(0.043)	(0.037)	(0.040)	(0.075)
Dwelling type: Separate	0.020	0.029	-0.058	-0.034	-0.043	-0.009
	(0.025)	(0.021)	(0.044)	(0.039)	(0.039)	(0.083)
Dwelling type: Other	0.022	0.002	0.069	-0.040	-0.030	-0.083
	(0.021)	(0.018)	(0.042)	(0.031)	(0.030)	(0.087)
Dwelling roof: Tiles	0.015	-0.057	0.529***	0.093	0.155*	-0.227**
	(0.061)	(0.048)	(0.128)	(0.067)	(0.080)	(0.087)
Dwelling roof: Harar	-0.051*	-0.127***	0.035	0.070	0.228***	-0.048
	(0.031)	(0.029)	(0.057)	(0.052)	(0.058)	(0.077)
Dwelling roof: Raar	-0.206***	-0.291***	-0.169	0.160	0.421**	0.083
	(0.079)	(0.082)	(0.122)	(0.142)	(0.190)	(0.184)
Dwelling roof: Wood	-0.038	-0.068**	-0.015	0.100*	0.096	0.201*
	(0.031)	(0.030)	(0.059)	(0.052)	(0.062)	(0.108)
Dwelling roof: Plastic	-0.083**	-0.166***	-0.046	0.038	0.304***	-0.075
	(0.035)	(0.042)	(0.068)	(0.074)	(0.076)	(0.073)
Dwelling roof: Concrete	0.020	0.051	-0.021	0.068	0.072	0.130

	(0.055)	(0.067)	(0.058)	(0.081)	(0.102)	(0.091)
Dwelling roof: Other	-0.133*	-0.106	-0.248*	0.127*	0.072	0.328**
	(0.078)	(0.093)	(0.130)	(0.071)	(0.093)	(0.158)
Improved sanitation	0.019	0.026	0.054	-0.054	-0.085***	-0.033
	(0.025)	(0.030)	(0.036)	(0.037)	(0.033)	(0.074)
Conflict Controls						
Conflict fatalities in district	0.000	-0.000	-0.000	-0.000	-0.000*	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Conflict x drought	0.000	0.000*	-0.001**	-0.000	-0.000	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Assistance						
Asisstance (% of beneficiaries reached)	-0.347***	-0.337***	-0.307***	0.570***	0.581***	0.472***
	(0.050)	(0.040)	(0.062)	(0.078)	(0.075)	(0.107)
Observations	7,214	5,678	1,536	7,214	5,678	1,536
R-squared	0.348	0.347	0.520			

*Note: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses. Poverty status results estimated using Probit, Consumption results estimated using OLS. Drought effect expressed in standard deviations of NDVI loss.*

Figure C5-10: Drought effect along the consumption distribution, urban areas.

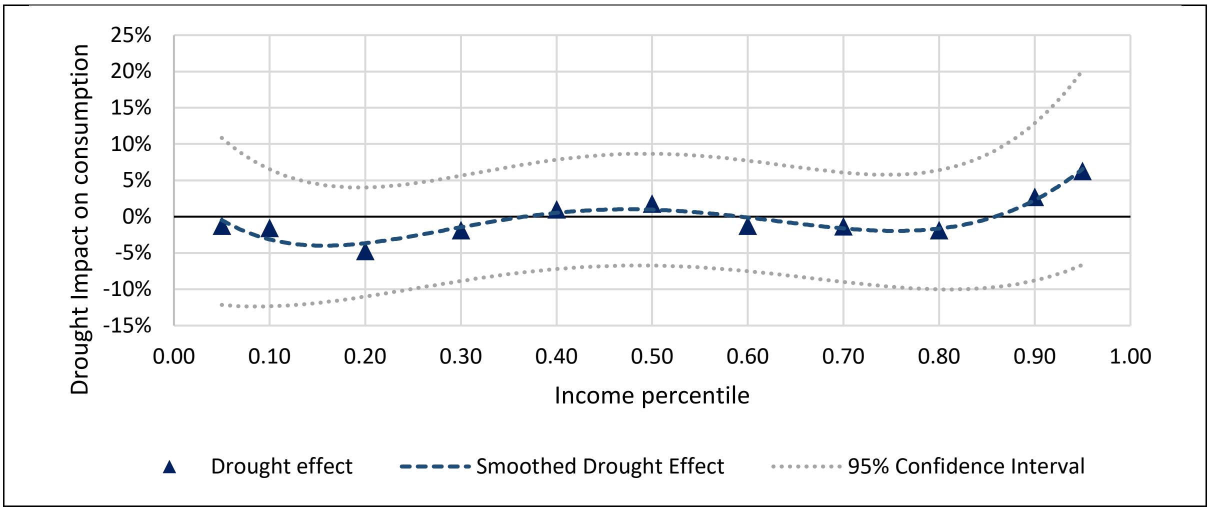


Table C5-7: Regression results, hunger.

	All regions			Overlapping regions		
Outcome variable			Hunger			
Sample	urban + rural	urban	rural	urban + rural	urban	rural
Post	0.087	0.130**	0.114	0.117***	0.123***	-0.005
	(0.058)	(0.060)	(0.123)	(0.033)	(0.034)	(0.059)
Drought Intensity	-0.050	-0.084*	-0.034	-0.084***	-0.118***	-0.037
	(0.039)	(0.048)	(0.060)	(0.032)	(0.044)	(0.031)
Drought Effect	0.092**	0.096*	0.190**	0.161***	0.116***	0.588***
	(0.045)	(0.055)	(0.079)	(0.038)	(0.038)	(0.134)
Average NDVI	0.034	-0.092	0.964**	-0.573*	-0.680**	-0.286
	(0.144)	(0.153)	(0.446)	(0.318)	(0.302)	(0.309)
Regional controls						
NE-urban		-0.030				
		(0.083)				
NW-urban		-0.225***			-0.098	
		(0.067)			(0.084)	
NE-rural						

NW-rural			-0.336**			
			(0.135)			
Central-urban		-0.015				
		(0.095)				
Central-rural			0.329			
			(0.233)			
Jubbaland-urban		-0.127				
		(0.176)				
SW-urban		-0.149				
		(0.128)				
SW-rural			0.213			
			(0.352)			
Household controls						
HH head literacy	-0.051**	-0.032	-0.119**	-0.027	-0.017	-0.137***
	(0.024)	(0.026)	(0.058)	(0.026)	(0.027)	(0.038)
HH head age	-0.001	-0.001*	0.002	-0.001	-0.001	0.000
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Received remittances	-0.002	-0.033	0.170***	-0.020	-0.033	-0.015
	(0.024)	(0.025)	(0.044)	(0.024)	(0.025)	(0.021)
Household size	-0.006	-0.001	-0.021	-0.012**	-0.009	-0.007
	(0.006)	(0.005)	(0.017)	(0.005)	(0.005)	(0.009)
Gender composition	-0.003	0.024	-0.019	-0.007	0.012	0.010
	(0.051)	(0.051)	(0.130)	(0.048)	(0.054)	(0.033)
Dwelling controls						
Dwelling tenure: Rent	0.029	0.018	0.074	0.004	0.013	-0.101***
	(0.022)	(0.020)	(0.065)	(0.020)	(0.019)	(0.028)
Dwelling tenure: Other	0.212***	0.116*	0.246**	0.153*	0.109*	0.024
	(0.070)	(0.060)	(0.112)	(0.079)	(0.063)	(0.079)
Dwelling floor: Tiles or mud	-0.010	-0.016	0.036	-0.010	-0.006	0.106**
	(0.031)	(0.030)	(0.082)	(0.028)	(0.030)	(0.042)
Dwelling floor: Other	0.003	0.051	-0.027	0.058	0.053	0.124**
	(0.041)	(0.038)	(0.087)	(0.042)	(0.043)	(0.051)
Dwelling type: Separate	-0.068	-0.087*	-0.063	-0.056	-0.085**	0.156**
	(0.054)	(0.050)	(0.100)	(0.034)	(0.036)	(0.066)
Dwelling type: Other	-0.036	-0.026	-0.130*	-0.036	-0.030	0.065
	(0.044)	(0.037)	(0.071)	(0.031)	(0.028)	(0.054)
Dwelling roof: Tiles	0.036	-0.001	0.296**	-0.143***	-0.216**	
	(0.126)	(0.125)	(0.114)	(0.034)	(0.089)	
Dwelling roof: Harar	0.130**	0.174***	0.140*	0.065	0.093	0.047
	(0.058)	(0.060)	(0.076)	(0.057)	(0.061)	(0.049)
Dwelling roof: Raar	0.070	0.099	0.120*	-0.052	0.006	0.056
	(0.068)	(0.077)	(0.067)	(0.046)	(0.072)	(0.064)
Dwelling roof: Wood	-0.059	-0.042	-0.109	-0.077*	-0.097	-0.007
	(0.064)	(0.079)	(0.146)	(0.046)	(0.072)	(0.055)
Dwelling roof: Plastic	0.091	0.076	0.124	-0.004	0.048	-0.052
	(0.063)	(0.092)	(0.075)	(0.065)	(0.086)	(0.058)
Dwelling roof: Concrete	-0.053	-0.033	-0.228			
	(0.104)	(0.119)	(0.148)			
Dwelling roof: Other	0.075	-0.000	0.161	-0.010	-0.019	-0.020
	(0.077)	(0.083)	(0.109)	(0.060)	(0.074)	(0.078)
Improved sanitation	-0.002	0.015	-0.043	-0.031	-0.045	0.013
	(0.039)	(0.043)	(0.052)	(0.030)	(0.031)	(0.035)
Conflict fatalities in district	0.000	0.000	0.001***			
	(0.000)	(0.000)	(0.000)			
Conflict x drought	-0.000	-0.000	-0.000			
	(0.000)	(0.000)	(0.000)			

Assistance (% of beneficiaries reached)	-0.052	-0.191**	0.198	0.078	-0.010	0.039
	(0.094)	(0.094)	(0.122)	(0.068)	(0.099)	(0.055)
Observations	7,153	5,637	1,516	3,962	3,292	663

Note: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses. Results estimated with Probit. Drought effect expressed in standard deviations of NDVI loss.

Table C5-8: Regression results, food consumption.

	All regions			Overlapping regions		
Outcome variable	Food Consumption					
Sample	urban + rural	urban	rural	urban + rural	urban	rural
Post	-0.069**	-0.080**	-0.182***	-0.049	-0.020	-0.230***
	(0.033)	(0.036)	(0.051)	(0.033)	(0.044)	(0.053)
Drought Intensity	-0.013	-0.017	0.091***	0.000	-0.016	0.068
	(0.018)	(0.020)	(0.030)	(0.017)	(0.020)	(0.042)
Drought Effect	-0.026	-0.024	-0.110**	0.006	0.033	-0.105**
	(0.031)	(0.033)	(0.054)	(0.028)	(0.039)	(0.043)
Average NDVI	0.362***	0.179**	0.737**	0.145	0.153	-0.338
	(0.120)	(0.087)	(0.342)	(0.163)	(0.168)	(0.313)
Price level	0.034	-0.293	0.501	0.164	0.394**	0.113
	(0.160)	(0.265)	(0.326)	(0.160)	(0.181)	(0.224)
Regional controls						
NE-urban	-0.003	0.155*				
	(0.080)	(0.089)				
NW-urban	-0.036	0.125*		0.071	-0.032	
	(0.062)	(0.075)		(0.055)	(0.060)	
NE-rural	-0.440***					
	(0.055)					
NW-rural	-0.074		0.345***	0.066		
	(0.065)		(0.058)	(0.065)		
Central-urban	0.201***	0.311***				
	(0.075)	(0.104)				
Central-rural	0.210		0.918***			
	(0.147)		(0.177)			
Jubbaland-urban	0.393***	0.375***				
	(0.099)	(0.101)				
SW-urban	0.293***	0.260**				
	(0.092)	(0.108)				
SW-rural	0.240**		1.204***			
	(0.102)		(0.331)			
Household controls						
HH head literacy	0.032**	0.054***	-0.001	0.054***	0.058***	0.031
	(0.013)	(0.011)	(0.024)	(0.013)	(0.014)	(0.042)
HH head age	0.001	0.000	0.002**	0.001	0.000	0.003**
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Received remittances	0.038***	0.046***	0.002	0.049***	0.056***	-0.047*
	(0.014)	(0.016)	(0.026)	(0.016)	(0.017)	(0.028)
Household size	-0.048***	-0.046***	-0.049***	-0.047***	-0.044***	-0.071***
	(0.002)	(0.003)	(0.007)	(0.003)	(0.002)	(0.009)
Gender composition	0.006	-0.035	0.110*	-0.014	-0.031	0.051
	(0.032)	(0.028)	(0.058)	(0.029)	(0.032)	(0.062)
Dwelling controls						
Dwelling tenure: Rent	0.010	0.004	0.015	0.010	0.009	0.009
	(0.012)	(0.013)	(0.023)	(0.015)	(0.016)	(0.028)
Dwelling tenure: Other	-0.043*	-0.060**	0.023	-0.022	-0.038	0.187***

	(0.024)	(0.028)	(0.041)	(0.031)	(0.030)	(0.068)
Dwelling floor: Tiles or mud	-0.003	0.019	-0.114***	-0.001	0.018	-0.223***
	(0.016)	(0.016)	(0.042)	(0.018)	(0.018)	(0.057)
Dwelling floor: Other	-0.033*	-0.017	-0.128***	-0.048**	-0.054**	-0.218***
	(0.020)	(0.023)	(0.038)	(0.023)	(0.026)	(0.060)
Dwelling type: Separate	0.015	0.015	-0.091**	-0.011	-0.002	-0.055
	(0.019)	(0.017)	(0.037)	(0.018)	(0.018)	(0.060)
Dwelling type: Other	0.019	0.002	0.019	-0.012	-0.012	-0.010
	(0.018)	(0.016)	(0.030)	(0.017)	(0.017)	(0.060)
Dwelling roof: Tiles	0.045	-0.006	0.295***	-0.018	-0.035	0.364***
	(0.046)	(0.037)	(0.073)	(0.043)	(0.043)	(0.094)
Dwelling roof: Harar	-0.033	-0.077***	0.015	-0.026	-0.081***	0.144*
	(0.026)	(0.028)	(0.055)	(0.033)	(0.029)	(0.080)
Dwelling roof: Raar	-0.168***	-0.204***	-0.126	-0.158	-0.268***	0.038
	(0.064)	(0.054)	(0.093)	(0.106)	(0.044)	(0.130)
Dwelling roof: Wood	0.026	0.011	0.024	-0.008	0.004	-0.003
	(0.028)	(0.031)	(0.050)	(0.033)	(0.033)	(0.090)
Dwelling roof: Plastic	-0.027	-0.097**	-0.016	-0.082*	-0.116***	0.090
	(0.030)	(0.038)	(0.052)	(0.043)	(0.038)	(0.077)
Dwelling roof: Concrete	0.064**	0.097***	-0.062	0.046	0.090***	-0.065
	(0.027)	(0.035)	(0.075)	(0.041)	(0.028)	(0.095)
Dwelling roof: Other	-0.068	-0.063	-0.161	-0.053	-0.047	0.028
	(0.073)	(0.090)	(0.111)	(0.072)	(0.087)	(0.107)
Improved sanitation	0.003	-0.000	0.039	-0.022	-0.025	0.013
	(0.027)	(0.036)	(0.032)	(0.033)	(0.040)	(0.046)
Conflict fatalities in district	-0.000**	-0.000**	-0.000**	-0.000**	-0.000***	-0.013
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)
Conflict x drought	0.000	0.000	-0.000*	0.000***	0.000***	-0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)
Asisstance (% of beneficiaries reached)	-0.197***	-0.232***	-0.240***	-0.186***	-0.174***	-0.078
	(0.035)	(0.041)	(0.081)	(0.035)	(0.034)	(0.081)
Observations	7,214	5,678	1,536	4,044	3,348	696
R-squared	0.347	0.304	0.591	0.297	0.312	0.461

*Note: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses. Results estimated with OLS. Drought effect expressed in standard deviations of NDVI loss.*

Table C5-9: Regression results, consumption and poverty, overlapping sample.

Outcome variable	Consumption			Poverty		
Sample	urban + rural	urban	Rural	urban + rural	urban	rural
Post	-0.223***	-0.241***	-0.237***	0.454***	0.386***	0.570***
	(0.040)	(0.047)	(0.064)	(0.093)	(0.116)	(0.080)
Drought Intensity	-0.045**	-0.079***	0.047	0.046	0.111***	-0.046
	(0.020)	(0.022)	(0.041)	(0.033)	(0.040)	(0.076)
Drought Effect	-0.045	-0.033	-0.137**	0.141*	0.021	0.356***
	(0.039)	(0.041)	(0.055)	(0.081)	(0.101)	(0.090)
Average NDVI	-0.055	-0.064	-0.490	-0.247	-0.383	1.040

	(0.163)	(0.157)	(0.426)	(0.373)	(0.369)	(0.778)
Regional controls						
NW-urban	0.155*	0.124			-0.049	
	(0.079)	(0.083)			(0.187)	
NE-rural						
NW-rural	0.090					
	(0.097)					
Household controls						
HH head literacy	0.078***	0.081***	0.044	-0.092***	-0.078**	-0.145***
	(0.014)	(0.014)	(0.050)	(0.030)	(0.032)	(0.044)
HH head age	0.001	0.000	0.003*	-0.001	0.000	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Received remittances	0.073***	0.076***	-0.029	-0.138***	-0.145***	-0.091
	(0.015)	(0.016)	(0.042)	(0.024)	(0.026)	(0.056)
Household size	-0.058***	-0.056***	-0.078***	0.084***	0.080***	0.136***
	(0.003)	(0.003)	(0.010)	(0.005)	(0.005)	(0.012)
Gender composition	0.012	0.000	0.040	-0.057	-0.023	-0.198***
	(0.033)	(0.036)	(0.075)	(0.067)	(0.069)	(0.071)
Dwelling controls						
Dwelling tenure: Rent	-0.001	0.000	0.035	-0.005	-0.022	0.063
	(0.016)	(0.015)	(0.028)	(0.024)	(0.023)	(0.048)
Dwelling tenure: Other	-0.051	-0.065*	0.198***	0.076	0.114**	-0.246**
	(0.034)	(0.036)	(0.069)	(0.050)	(0.053)	(0.119)
Dwelling floor: Tiles or mud	0.012	0.044**	-0.283***	-0.024	-0.057*	0.259***
	(0.020)	(0.018)	(0.070)	(0.032)	(0.033)	(0.061)
Dwelling floor: Other	-0.064***	-0.082***	-0.223***	0.081**	0.145***	0.141
	(0.024)	(0.026)	(0.073)	(0.039)	(0.043)	(0.085)
Dwelling type: Separate	0.009	0.016	-0.025	-0.017	-0.028	-0.000
	(0.020)	(0.020)	(0.070)	(0.041)	(0.042)	(0.069)
Dwelling type: Other	-0.007	-0.010	0.058	-0.030	-0.033	0.005
	(0.019)	(0.019)	(0.079)	(0.029)	(0.029)	(0.072)
Dwelling roof: Tiles	-0.061	-0.079	0.377***	0.159**	0.196**	-0.206
	(0.049)	(0.051)	(0.104)	(0.068)	(0.080)	(0.157)
Dwelling roof: Harar	-0.044	-0.117***	0.160	0.107**	0.194***	-0.063
	(0.036)	(0.033)	(0.098)	(0.052)	(0.061)	(0.062)
Dwelling roof: Raar	-0.213*	-0.364***	0.035	0.245	0.730***	-0.007
	(0.117)	(0.066)	(0.176)	(0.170)	(0.091)	(0.161)
Dwelling roof: Wood	-0.084**	-0.074**	-0.041	0.152***	0.124*	0.297**
	(0.034)	(0.031)	(0.104)	(0.052)	(0.064)	(0.116)
Dwelling roof: Plastic	-0.132***	-0.179***	0.103	0.203***	0.292***	-0.089
	(0.049)	(0.044)	(0.096)	(0.061)	(0.079)	(0.089)
Dwelling roof: Concrete	0.036	0.091	-0.033	0.082	0.044	0.075
	(0.064)	(0.061)	(0.121)	(0.072)	(0.081)	(0.157)
Dwelling roof: Other	-0.115	-0.091	-0.041	0.127*	0.062	0.100
	(0.078)	(0.092)	(0.118)	(0.075)	(0.094)	(0.122)
Improved sanitation	0.025	0.017	0.059	-0.058	-0.055	-0.064
	(0.027)	(0.034)	(0.050)	(0.039)	(0.035)	(0.070)

Conflict fatalities in district	0.000	0.000	-0.012	-0.000***	-0.000*	0.003
	(0.000)	(0.000)	(0.009)	(0.000)	(0.000)	(0.009)
Conflict x drought	0.000*	0.000*	-0.001	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.007)	(0.000)	(0.000)	(0.011)
Asisstance (% of beneficiaries reached)	-0.207***	-0.219***	-0.125	0.371***	0.407***	0.076
	(0.046)	(0.036)	(0.091)	(0.093)	(0.086)	(0.122)
Observations	4,044	3,348	696	4,044	3,348	696
R-squared	0.332	0.349	0.474			

*Note: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses. Poverty status results estimated using Probit, Consumption results estimated using OLS. Drought effect expressed in standard deviations of NDVI loss.*

6. The Labor Market Implications of Restricted Mobility during the COVID-19 Pandemic in Kenya[233]

Markus Heemann, Utz Johann Pape, Sebastian Vollmer

Introduction

The pandemic of Coronavirus Disease-19 (Covid-19) has been an unprecedented situation for the world. To this date, estimates are that more than 230 million people have been infected and around 4.7 million people have died from the COVID-19 pandemic across the globe (WHO 2021, Johns Hopkins University 2021). At the same time, the pandemic has had significant labor market implications, with an estimated 225 million full-time jobs lost worldwide between the fourth quarter of 2019 and the first quarter of 2021 (ILO 2021). These COVID-19 related labor market costs are driven by many factors, such as peoples' behavior in uncertain times as well as the policies and guidelines governments impose to curb the spread of the virus.

As a response to the pandemic many governments have imposed two types of measures. Firstly, measures aimed at restricting mobility and social interaction to reduce the speed of further infection as well as, secondly, measures to mitigate the economic consequences on businesses and households. The consequences from the pandemic and restrictions on personal mobility have severely disrupted economic activities, as between one and four in five workers reside in countries with required workplace closures (ILO 2021).

Particularly for households in developing countries, the labor market implications of the pandemic can be dire. The lack of economic safety nets especially in the informal sector but also increased risk of infection and related expenses, especially for poor people living in high density areas with daily hands-on income, can exacerbate the consequences of losing parts of the income or the job entirely (Bargain and Ulugbek 2021, Gupta, Bavinck et al. 2021). Given the additional challenges households in developing countries face in coping with the crisis, it is elementary for policy makers to understand which socio-economic consequences any countermeasures aimed at curbing the spread of the virus may have. As governments react and impose restrictions to save lives, people subsequently change their behavior (e.g. reduce mobility) and this in turn affects labor markets. Therefore, a better understanding of the causal relationships between human behavior and labor market outcomes is vital to crafting better, more effective and targeted policies in future situations in which there is the joint goal of slowing down everyday life to save lives while minimizing the negative economic and societal effects.

Kenya's first case of COVID-19 was recorded in March 2020. Since then, reported infections have considerably increased, peaking on October 31, 2020 with 1,395 new infections per day (Ritchie et al. 2020). Following Kenya's first case of confirmed COVID-19 in March 2020, the Government of Kenya quickly put in place multiple policies and measures to contain the spread of the virus. In March 2020 for instance, the Government of Kenya introduced a series of restrictions ranging from the closure of educational institutions to directing public and private sector workers to home-based work, except for essential workers (Bowmans 2020, Nechifor, Ferrari et al. 2020). Entry into Kenya was limited to citizens and residents but required quarantine for 14 days while local air travel was suspended and resumed on July 15. These measures were followed by fast reductions in average mobility outside of residential areas but with an increase in residential movement (Figure C6-1).

[233] MH and UP jointly conceptualized the research, obtained and processed the data as well as drafted the manuscript. MH conducted the statistical analysis, while UP and SV provided supervision.

Many studies in different contexts have shown that COVID-19-related containment measures aiming to reduce mobility and social contacts are a key tool in slowing the spread of the virus and as such, saving lives and buying vital time to develop vaccines and flatten the curve such that a country's health infrastructure is not overwhelmed (Jarvis, Van Zandvoort et al. 2020, Yilmazkuday 2021). Additionally, studies have used Google Mobility Data to demonstrate these policies' successes in reducing mobility compared to pre-COVID-19 levels (Drake, Docherty et al. 2020, Saha, Barman et al. 2020, Vinceti, Filippini et al. 2020). However, as the disease is better understood, socioeconomic effects of the COVID-19 pandemic have started receiving increased attention. Multiple studies have looked into COVID-19 effects on different dimensions of household livelihoods both in the developed world (Auriemma and Iannaccone 2020, Bonaccorsi, Pierri et al. 2020) as well as developing countries (Josephson, Kilic et al. 2020, Khamis, Prinz et al. 2021). Using data from high-frequency phone surveys, Khamis, Prinz et al. (2021) for example estimated the early impact of COVID-19 on the labor markets of 39 countries. Their findings show that the pandemic has negatively affected labor market outcomes in these countries (job and income losses, lack of payment, job changes), with more pronounced impacts among workers in manufacturing (40%) and services (38%) than in agriculture (22%) as well as among self-employed (46%) compared to employees (39%).

While there is extensive literature on the aggregated socio-economic effects of the COVID-19 pandemic both in developed countries (Auriemma and Iannaccone 2020, Bonaccorsi, Pierri et al. 2020) as well as developing countries (Khamis, Prinz et al. 2021) including Kenya (Janssens, Pradhan et al. 2021, Kansiime, Tambo et al. 2021, Pape, Delius et al. 2021) little research has been conducted looking into the specific mechanisms through which the pandemic affected labor market outcomes in developing countries. In particular, the channel of changing mobility has not been investigated extensively yet most likely due to both measurement difficulties and identification issues.

Mobility is an outcome of labor market activity as well as something that drives labor market activity, for example by providing jobs in the transportations sector. Likewise, the ability to move determines whether people have access to markets to sell their goods, as well as whether customers can attain the goods that they would like to have. Finally, supply chains as well as trade rely on frictionless mobility, which in turn may impact production and thus labor markets further downstream (Espitia, Mattoo et al. 2022). Given that mobility was severely impacted by policy to curb the spread of the virus in Kenya, it is an interesting shock-like mechanism driving labor market outcomes to look at. We intend to quantify the changes in labor market outcomes that were driven by changing mobility levels over the course of the pandemic in Kenya by applying IV analyses.

Figure C6-1: Development of Kenyan Policy Stringency and Mobility Types since February 2020

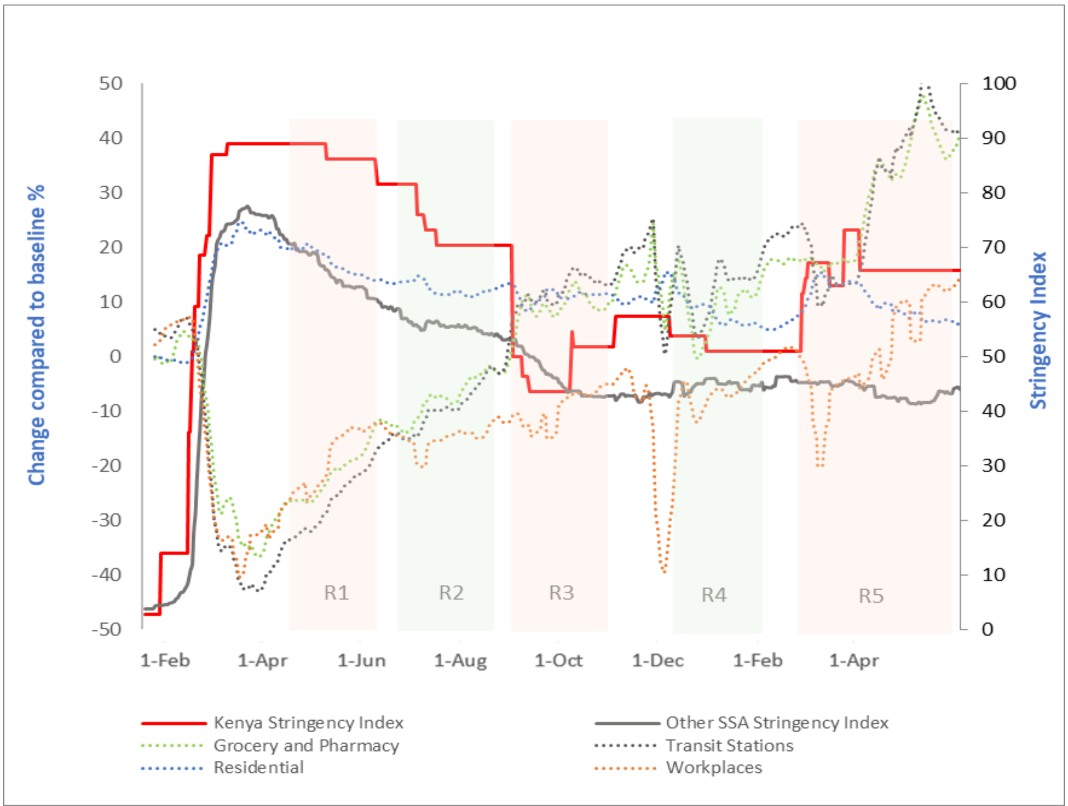


Figure C6-1 shows how the policy stringency and different types of mobility changed over time. The graph highlights another important factor determining the actual observed mobility levels, i.e. the peoples' adherence to the implemented policies and the government's ability to enforce them. In the beginning, mobility changes followed the changes of policy stringency with opposite direction. However, by the time mobility levels recovered to pre-pandemic levels at the end of 2020, this relationship became much less clear. Therefore, to better understand mobility levels as mechanism that drives labor market outcomes, it is important to also better understand what drives policy adherence of citizens in the respective setting. Many studies have looked at determinants of mobility restriction and COVID-19 guidelines. However, most of them were either placed in developed countries (Al-Hasan, Yim et al. 2020, Carlucci, D'ambrosio et al. 2020, Coroiu, Moran et al. 2020) or lacked a representative sample size (Ahmed, Siewe Fodjo et al. 2020, Usman, Ssempijja et al. 2020). Given the importance of policy adherence to understand mobility levels, we complement our analysis by determining which factors were associated with respondents self-reported mobility reduction in Kenya over the course of the pandemic.

We aim to add to the literature by examining labor market effects driven by changing mobility levels that can be attributed both to the measures imposed by the Kenyan government as well as people's adherence to these policies, combining data on policy restrictions with insights from Google Mobility Reports and large-scale household surveys. As far as we are aware, this is the first paper to investigate the causal effects of changing mobility levels on labor market outcomes over the course of the pandemic in a developing country. This study would be the first to do this in a nationally representative setting in a developing country with panel data reaching into early 2021. By estimating these causal effects, our findings will inform both researchers aiming to establish direct links from mobility to labor market outcomes as well as policy makers looking to balance the trade-off between curbing the spread of the virus and containing the magnitude of socioeconomic costs. In line with this, our analysis of factors associated with adherence to mobility restrictions add important information on how to design, target and communicate mobility restrictions in Kenya more effectively in order to increase the restrictions' ability to slow the spread of the virus.

Data Sources and Variables Used

Rapid Response Household Surveys

To conduct our analyses of the mobility-related labor market effects of the COVID-19 containment measures, we leverage multiple sources of data. Central to our analyses, we use the Kenya COVID-19 Rapid Response Phone Household Surveys (RRPS) to measure labor market effects of the pandemic on households on a county-level for multiple survey waves between 2020 and 2021. The Kenya COVID-19 RRPS was structured as a five-waves bi-monthly panel survey that targeted nationals, refugees and stateless persons and has representative weights for national as well as county (admin-1) levels. Five rounds of the survey were completed between May 2020 and February 2021 (Table C6-7) The sampling frame of telephone numbers was composed of two groups of households. The first was based on a randomly drawn subset of the 2015/16 Kenya Integrated Household Budget Survey (KIHBS) with 9,009 households which covered urban and rural areas and was designed to be representative of the population of Kenya using cell phones. The household head or a knowledgeable person within the household was interviewed via Computer Assisted Personal Interviews (CAPI) and were asked to provide telephone numbers. Given that this sampling frame was five years old at the time of the first RRPS wave, an additional group was added by applying Random Digit Dialing (RDD). This method contacted households from a list of mobile phone numbers that was created using a random number generator from the 2020 Numbering Frame produced by the Kenya Communications Authority. The initial sampling frame consisted of 92,999,970 randomly ordered phone numbers assigned to three networks: Safaricom, Airtel, and Telkom. There was no stratification, and individuals, regardless of their household head status, that were reached through the selected phone numbers were asked about the households they live in. Household reached via RDD make up between 18.7% and 20.4% of our sample in the five survey waves (Table C6-7).

The questionnaire covered multiple topics, such as behavior in response to the COVID-19 pandemic and mobility, changes in employment, income, food security, subjective well-being, access to education and health services, knowledge of COVID-19 and mitigation measures as well as perceptions of the government's response and coping strategies. The questionnaire was translated into Swahili, Luo, Arabic, French, Kirundi, Luganda, Oromo, Somali, Kinyarwanda, Tigrinya, Nuer and Dinka to ensure all respondents can be interviewed in a language they are comfortable with. Our analysis focuses on working adults between 14 and 65 years old. We attain nationally representative RRPS data from 24,340 respondents. Out of these, 22,708 respondents gave complete information on employment status, 11,045/ 11,860 respondents on agricultural hours/income, 4,486/3,197 respondents on wage hours/income and 1,681 respondents on self-employment hours as well as the other covariates we consider. Sample characteristics are consistent across survey waves (Supplement Table 1). For the analyses of determinants of self-reported mobility reduction, we attain complete data from a total of 12,563 respondents.

Mobility Development

To determine mobility trends during the time of the pandemic, we use Google Community mobility reports (Google 2021). These mobility reports provide insights into how mobility changes during the pandemic and into policies' effectiveness aimed at reducing mobility. Google mobility reports tracks aggregated, anonymized sets of GPS data for changes in mobility from users who opted-in/ did not opt out of location history for their Google Account. The data shows how visits to (or time spent in) categorized places change compared to a baseline. The baseline is the median value for the specific weekday from the 5-week period Jan 3 – Feb 6, 2020. Data is recorded for a total of six different location types, residential, grocery and pharma, transit, workplaces, retail and recreation and parks and leisure and collected on a county level (admin 1) as is our RRPS data. We consider five of them, excluding parks and leisure as we want to focus on dimensions of social and economic life (Chen, Igan et al. 2020) to construct the average mobility change. The average mobility change is computed by taking weekly overall average mobility change of the four location types (multiplying residential mobility change with minus one to attain a negative value for overall mobility reduction outside of home).

Policy Stringency

To determine the degree of mobility restrictions in Kenya, we use the COVID-19 Government Response Tracker from the Blavatnik School of Government which tracks and collects systematic information on policy responses from governments during the pandemic for multiple countries (Hale, Angrist et al. 2021). The tracker traces health policies, economic policies and containment and closure policies of governments and assigns them an ordinal value ranging from 0 to 100 depending on severity and penetration across the country. We consider the latter type, i.e. containment and closure policies enacted by the Government of Kenya. Among the containment that are part of the index and that are assigned ordinal values are school closures, workplace closures, cancellation of public events, restrictions on gatherings, closure of public transport, stay at home requirements, as well as restrictions on national and international travel. The index is calculated using these ordinal containment and closure policy indicators, plus an indicator recording public information campaigns (Hale, Angrist et al. 2021). Data for Kenya is aggregated on a national level for each day starting January 1, 2020, ranging from 0 to 88.89. For our analyses, we calculate weekly average policy stringency levels to match the granularity of data of mobility and labor market outcomes.

Confirmed COVID-19 Cases in Kenya

As part of our analyses, we also consider confirmed COVID-19 cases in Kenya, both national aggregates and county cases. National confirmed COVID-19 cases were obtained from both published government briefs as well as the data set on Policy Stringency, that also included national reported confirmed COVID-19 cases. For state specific confirmed cases, we used regular updates by the Kenyan Ministry of Health from the respective homepage and Twitter.

Labor Market Outcomes of Interest

Labor market outcomes from the RRPS can be allocated into three categories: A) employment status; B) hours worked in past 7 days; C) income earned in past 14 days per adult and thus combine both extensive margins of employment (category A) and intensive margins of employment (categories B and C). Within these categories, we look at a total of 8 different labor market outcomes: 1) % employed, 2) % unemployed, 3) % not in the labor force, 4) hours worked in agriculture, 5) hours worked in wage employment, 6) hours worked in self-employment, 7) agricultural earnings and 8) wage earnings (Supplement Table 2). The wage indicators combined both formal and informal employment. We take weekly averages for all adults for which we have data available and aggregate them on a per county per-week level, which reflects the sampling and data collection strategy of the RRPS. County specific weekly datapoints range from 1 to 51, with 75% of week averages comprising depending on the labor market outcomes between more than 2- 6 observations per county. For three of the eight variables, i.e 4) hours in agriculture, 5) hours in wage employment and 8) wage earnings, the RRPS survey also asks recall questions for levels prior to COVID-19 in February 2020, which we include into our analysis as additional week averages in the last week of February, giving us additional pre-pandemic datapoints.

Statistical Analyses and Estimation Strategy

Causal Impact of Mobility on Labor Market Outcomes

Regression Results

We start our analysis by running OLS and county fixed effects regression for the average weekly mobility change and average weekly labor market outcomes in a simple model and a model including additional covariates averages of economic uncertainty, fear of illness, knowing someone who had an infection, the change in national confirmed COVID-19 cases compared to the previous week in %. All models yield significant correlations between mobility levels for the extensive margins of employment as well as the number of hours worked both in formal and informal wage employment. Coefficients are similar for the extensive margins of employment with a correlation coefficient of ~0.004 for outcome employed, implying that a 1 percent increase of mobility is associated with an increase in employment of 0.4 percentage points (Table C6-1). Including the set of additional covariates yields significant results for both outcomes related to agriculture.

Table C6-1: OLS and FE estimates for labor market outcomes on changing mobility levels

	OLS (1)	OLS incl. covariates (2)	FE (3)	FE incl. covariates (4)
Employment (% of Hh members)				
Employed	0.004*** (0.00)	0.003*** (0.00)	0.004*** (0.00)	0.003*** (0.00)
n	1649	1555	1649	1555
Unemployed	-0.001** (0.00)	-0.002*** (0.00)	-0.001** (0.00)	-0.001* (0.00)
n	1649	1555	1649	1555
Not in labor force	-0.002*** (0.00)	-0.002*** (0.00)	-0.003*** (0.00)	-0.003*** (0.00)
n	1649	1555	1649	1555
Hours Worked in past 7 days				
Agriculture	-0.003 (0.02)	0.004 (0.02)	0.014 (0.02)	0.002 (0.03)
n	1441	1440	1441	1440
Wage Job (formal and informal)	0.064** (0.03)	0.040 (0.03)	0.166*** (0.03)	0.142*** (0.03)
n	1161	1161	1161	1161
Self-Employment	0.038 (0.04)	0.043 (0.04)	0.050 (0.05)	-0.001 (0.08)
n	780	779	780	779
Income in past 14 days in KSH				
Agriculture	10.635 (5.90)	0.659 (7.84)	11.895 (6.3)	10.866 (10.67)
n	1495	1493	1495	1493
Wage Job (formal and informal)	14.742 (11.70)	4.492 (2.32)	13.149 (12.38)	22.108 (19.76)
n	1018	1018	1018	1018

*Note: Aggregated on weekly levels, *** is significant at the 1% level, ** is significant at the 5% level and *is significant at the 10% level*

However, plain OLS regression results (including fixed effects regression) can hardly be interpreted as causal. At first, it is easy to find third variables that have explanatory power for both, such as overall levels of fear of economic and health consequences. Our surveys ask specifically for these sentiments of uncertainty and fears of health and economic consequences. However, even if we control for these sentiments, the main problem of reverse causality remains, i.e. the fact that mobility does not only explain changes in labor market outcomes but that labor market outcomes and overall economic activity themselves have impacts on observed mobility. Therefore, the regression results in Table 1 cannot be considered causal in any direction.

Identification Strategy

To address these issues and given that mobility levels are highly interlinked with economic activity, we leverage policy stringency as exogenous shock in an IV estimation framework to overcome the issue of reverse causality and determine the causal impact of varying mobility levels on labor market outcomes in Kenya. As such, we use the overall policy stringency levels as instrument for observed mobility levels. We apply the following first stage regression controlling for the percentual change of confirmed national cases:

$$M_{tc} = PSI_t + C_t + \omega_{tc},$$

Mtc refers to the average mobility change on a county-level, PSIt to the Policy Stringency Index on national level, and Ctc to the % change in confirmed cases in week t compared to week t-1 on the national level. We also considered county-level case changes, however these did not prove useful, given the low figures and large uncertainty between reported vs. actual numbers. We incorporate the % change in confirmed cases compared to the prior already in the first stage, to filter out "fear" effects that were not driven by public policy changes.

The second stage of our analysis is a county fixed effects regression at the county-week level. We include responses on concerns about the disease in terms of concerns about the illness itself, as well as fear of economic consequences. Households were asked if the pandemic was cause for concern, and if so, they were asked to provide the specific source of concern. Furthermore, we control for age and education (ranging from no formal education to postgraduate university degree). For respondents that provided us with recall-baselines, we assumed the education as well as the age to be the same at the time of the baseline, given that recall values were from February and survey data was available as of June of the same year. To control for the overall development of the pandemic, we include changes in Kenya's weekly reported COVID-19 cases as well as answers to the questions, whether a household knew of someone who had been infected with COVID-19. This latter control was added, because reported cases can be expected to be much lower than actual cases and therefore nationally representative surveys asking about known cases may serve as important addition to representing the overall course of a pandemic. A full overview of the covariates can be found in Supplement Table 2. This yields our second stage regression:

$$Y_{tc} = \widehat{M}_{tc} + X_{tc} + C_t + \delta_c + \varepsilon_{tc}$$

With Yct being our labor market outcomes of interest, δ_c denoting the county fixed effect and Xtc capturing the county/week specific averages of economic uncertainty, fear of illness, age and education levels of respondents and the overall progress of the pandemic.

Threats to Identification Strategy

Our identification strategy relies on two assumptions. The first, our exclusion restriction is that the reduction of mobility is the only channel through which the government's policies aimed at curbing the spread of the virus effected labor market outcomes. Clearly this is only possible when we can control for any signaling effect and concerns that the imposed policies may have had on households. As part of the RRPS survey data, we have representative data on fear of the illness as well as self-perceived economic uncertainty, which allows us to control for these sentiments. Additionally, our estimation strategy relies on the assumption that the IV is exogenous, i.e. that there is no causal impact running from labor market outcomes to our instrument, the policy stringency index itself. There are a couple of observations that we believe justify this assumption. At first, the Kenyan government immediately implemented very strong measures including a national curfew at a time, where only a handful COVID-19 cases had been confirmed in the country. Secondly, the government quickly enacted several economic relief policies which can be taken as anecdotal evidence that the mobility policy's primary concern was to curb the spread of the virus (see Presidential Announcement from April 16th, 2020) and economic considerations were tried to be addressed otherwise. We investigate this idea by looking at survey responses for questions on whether households had received transfers from government or politicians including the amounts. The share of people self-reporting receiving transfers from government programs ranged from 1.3% in wave 4 to 4.1% in wave 5 yet with no clear patterns across the waves. However, looking at the magnitude of transfers compared to pre-pandemic levels, there is anecdotal evidence that of increases in all survey waves (n=381) compared to pre-pandemic levels with increases ranging from an additional 913 KSH on average in wave 2 to 2,120 KSH in wave 4. Additionally, we look at the development of people's trust in the government's ability to deal with the pandemic as proxy for public sentiment about the government's performance that could reflect increasing pressure on politicians to take economic consequences more into consideration. Indeed, average scores changed from 1.51 during wave 1 of the RRPS to 1.40 during wave 5. However, given that trust levels were on average high (distrust was coded as 0, neutral as 1 and trust as 2), we do not believe this change to have made much of a difference. Overall, it does not seem that more severe labor market conditions were associated with increased political pressure, enabling the Government of Kenya to form mobility policies that were solely aimed at saving lives and containing the spread of the virus.

Factors Associated with Self-Reported Mobility Restrictions

Our second set of analyses looks at whether households self-reported any behavioral change that could be attributed to self-restricting mobility and interaction. The outcome variable is a binary variable "Any self-reported mobility restriction" that was given a value of 1, if respondents stated that due to COVID-19, they had either avoided groups more often, stay at home more, traveled outside less, gone to work less, or returned home earlier at night (Table C6-9).

Looking at factors that are associated with any self-reported mobility restriction, we – as above – consider the number of confirmed COVID-19 cases and the overall policy stringency. In addition, we incorporate a set of 10 covariates recorded in the RRPS. The co-variates include respondents' answers on questions about their trust in the government in handling the pandemic, trust in their fellow citizens, characteristics such as sex, education level, age, employment status, location (urban vs rural) and household heads status and whether they know someone who was infected or whether they were worried about having enough food (Table C6-9). To determine factors that influence any self-reported mobility reducing behavior, we run a multilevel logit model at the household level, where week and county form our two levels of analysis:

$$m_{it} = x_{it} + PSI_t + C_t + \vartheta_{it},$$

With mit being self-reported mobility for household i in week t, xit household characteristics, Ct the % change in confirmed cases for week t compared to t-1 and ϑ_{it} the error term.

Results

Policy stringency on a national level and average mobility changes in the individual counties are significantly and negatively associated with one another. Table C6-2 shows the results of our first stage regression, which is significant not just for policy stringency but also negatively and statistically significantly related to the weekly change of national confirmed COVID-19 cases. We see that in terms of magnitude however, a one-point Policy Stringency Index increase is associated with a more than 8 times decrease of mobility compared to a percentage point increase in national weekly confirmed COVID-19 cases.

Table C6-2: First Stage Regression Results

Weekly Mobility Change Levels from Feb 2020-June 2021, n=2,617	Coefficient (S.E.)	95% Confidence Interval
Policy Stringency Index	-0.252*** (0.014)	[-0.279;-0.223]
Weekly Change Confirmed COVID-19 cases (national)	-0.029*** (0.003)	[-0.035;-0.023]

Note: Aggregated on weekly levels, *** is significant at the 1% level

There is a significant impact of changing mobility on the overall employment and labor force participation of household members, with positive effects of increasing mobility on employment and unemployment and negative effects on not being in the labor force. Roughly three quarter of people entering the labor force entered employment following increases in overall mobility, while a bit more than a third entered unemployment. A 10% increase in mobility caused a 12 percentage points of people to return to the workforce. Hence, we see that the mobility restrictions mainly affected peoples' participation in the labor force and thus affected extensive margins of employment. Given that our RRPS data commences in May at a time where mobility recovery was already underway, this can be interpreted as increased mobility signaling people that things are returning to being back to normal which causes them to look for jobs again. Surprisingly, these changes are consistent across urban and rural areas with minor yet statistically significant differences in employment, unemployment and not in the labor force.

Table C6-3: IV estimation for labor market outcomes using changing mobility levels as explaining variable

	OLS- full sample	IV- full sample	IV- rural	IV-urban
	(1)	(2)	(3)	(4)
Employment (% of Hh members)				
Employed	0.004*** (0.00)	0.009*** (0.00)	0.009*** (0.00)	0.010*** (0.00)
n	1649	1555	1470	1467
Unemployed	-0.001** (0.00)	0.002 (0.00)	0.004*** (0.00)	0.003** (0.00)
n	1649	1555	1470	1467
Not in labor force	-0.002*** (0.00)	-0.012*** (0.00)	-0.014*** (0.00)	-0.013*** (0.00)

n	1649	1555	1470	1467
Hours Worked in past 7 days				
Agriculture	-0.003 (0.02)	0.127* (0.07)	-0.130 (0.08)	0.399*** (0.08)
n	1441	1440	1287	1233
Wage Job (formal and informal)	0.064** (0.03)	1.143*** (0.38)	1.772*** (0.50)	0.958*** (0.20)
n	1161	1161	721	910
Self-Employment	0.038 (0.04)	0.413** (0.16)	0.269 (0.24)	0.239 (0.15)
n	780	779	400	567
Income in past 14 days in KSH				
Agriculture	10.635 (5.90)	66.154 (52.78)	16.141 (85.31)	113.101** (50.78)
n	1495	1493	1342	1299
Wage Job (formal and informal)	14.742 (11.70)	134.636 (86.72)	62.042 (205.26)	131.075 (112.40)
n	1018	1018	591	761

*Note: Aggregated on weekly levels, *** is significant at the 1% level, ** is significant at the 5% level and *is significant at the 10% level*

Looking at the intensive margins of employment, i.e. the indicators that provide context about existing employment, we find that the most significant effects were for the hours worked by household members. Here, a 10% increase in mobility was associated with an increase of 11 wage hours per week (formal and informal). Overall, there seem to be more significant effects for wage professions (both formal and informal). The coefficient of hours worked in agriculture is only statistically significant at the 10% level and has a much lower coefficient than hours in wage jobs. Self-employment hours seem to have been positively affected by the recovery of overall mobility as well. For income generated from wage work and agriculture, we find no statistically significant effects of recovering mobility. Comparing urban vs. rural, we find that employment effects (from entering the labor force) and wage hours worked were larger in the rural setting, while agricultural employment in terms of hours worked and income generated was significantly affected in the urban setting. Finally, the estimated coefficients using our IV approach yield much higher results and higher statistical significance for amount if hours worked compared to our previous OLS estimates that were subject to reverse causality.

Looking at the other correlates that we included into our analyses (Table C6-4), we find that economic uncertainty is inversely related to people working in self-employment. Additionally, age and education seem are positively associated with (re-)entering employment, indicating that the overall labor market recovery was more pronounced for older, more experienced and educated workers.

Table C6-4: IV estimation results for whole set of covariates used in regression model

	Employed	Unemployed	Not in Labor Force	Agri Hours (7days)	Wage Hours (7days)	Self-employment Hours (7days)	Agri Income (14days)	Wage Income (14days)
IV Mobility	0.009***	0.002	-0.012**	0.127*	1.143***	0.413**	66.154	134.636
Economic Uncertainty	0.037	-0.040	0.003	0.993	-1.118	-6.826**	956.343	189.959
Fear of Illness	-0.001	0.020	-0.035	2.272	8.024**	9.628***	-263.304	2304.881
Know s/o Infected	0.041	-0.162**	0.152	-0.385	-22.184**	-3.000	-1659.48	-4327.56
Age	0.010***	-0.001	-0.000	-0.037	-0.091	-0.043	0.182	66.676**
Education	0.033***	0.011**	0.017**	-0.612**	-1.230**	0.112	136.823	2255.364***

*Note: Aggregated on weekly levels, * is significant on 10% level, ** significant on 5% level, ***significant on 1% level*

It is possible that our results are driven by different mechanisms that played a role at varying stages of the pandemic. For this reason, we also compare results for different stages of the pandemic. Specifically, we split our sample into a "recovery" and "post-recovery sample", the first reflecting waves 1 and 2, in which mobility returned to pre-pandemic levels and a post-recovery phase, in which mobility exceeded pre-pandemic levels (Table C6-5). Our results show first that the effects on extrinsic margins of employment differed quite substantially between the two phases. Most of the re-entering the labor force in the beginning led to re-employment, while between wave 3-5 the most pronounced employment effects came from people leaving unemployment. Likewise, for the intrinsic margins of employment, most significant outcomes of hours worked, and income generated are significant in the past-recovery phase. Looking at the split for urban and rural, we find that the initial employment recovery is mainly due to recovery in rural areas. At the same time, recovery of hours worked both in agriculture as well as wage professions seems as well as agricultural income generated seems to be driven by urban areas.

Table C6-5: IV estimation results for our outcomes for different stages of the pandemic

Wave 1-2 (initial recovery) Wave 3-5 (post-recovery)	National Wave 1-2 (1)	National Wave 3-5 (2)	Rural Wave 1-2 (3)	Rural Wave 3-5 (4)	Urban Wave 1-2 (5)	Urban Wave 3-5 (6)
Employment (% of Hh members)						
Employed	0.009 (0.00)	0.017*** (0.01)	0.016*** (0.01)	0.012*** (0.00)	0.003 (0.01)	0.016*** (0.00)
n	492	1061	460	1002	456	1005
Unemployed	0.001 (0.000)	-0.011*** (0.00)	-0.001 (0.00)	-0.009*** (0.00)	0.006 (0.00)	-0.012*** (0.00)

n	492	1061	460	1002	456	1209
Not in labor force	-0.010** (0.00)	-0.007*** (0.00)	-0.016*** (0.01)	-0.006** (0.00)	-0.008 (0.01)	-0.003 (0.00)
n	492	1061	460	1002	456	1005
Hours Worked in past 7 days						
Agriculture	0.473** (0.19)	0.243** (0.09)	0.222 (0.23)	0.248** (0.11)	0.644*** (0.21)	0.280** (0.11)
n	452	986	398	884	380	849
Wage Job (formal and informal)	0.653 (0.43)	-0.348 (0.22)	1.000 (0.71)	-0.373 (0.25)	0.864* (0.44)	-0.374 (0.26)
n	302	858	147	574	215	693
Self-Employment	0.808 (0.45)	-0.228 (0.22)	0.394 (0.96)	-0.208 (0.32)	0.369 (0.51)	-0.095 (0.35)
n	240	539	120	280	178	389
Income in past 14 days in KSH						
Agriculture	97.558 (152.36)	-153.623** (71.66)	263.063 (384.59)	-68.036 (42.38)	-14.600 (144.97)	213.656** (98.31)
n	458	1033	404	932	393	901
Wage Job (formal and informal)	116.376 (171.69)	-69.228 (114.33)	-301.697 (228.41)	-94.535 (131.30)	170.208 (260.21)	-64.162 (121.68)
n	244	773	111	480	157	602

*Note: Aggregated on weekly levels, * is significant on 10% level, ** significant on 5% level, ***significant on 1% level*

Among the broad set of potential determinants of self-reporting any form of mobility reduction, we find that the trust in the government handling the pandemic well (driven by urban areas), knowing someone who had been infected (driven by rural areas) and the overall policy stringency level are statistically significant (Table C6-6). Interestingly, both the trust in the government's ability to handle the pandemic as well as knowing someone who had been infected has a negative sign, implying either that a good trust in the government's ability to deal with the pandemic reduces the individual households need to comply with recommended mobility restrictions or that having someone infected in immediate reach implied increased need of support which translates into mobility. However, overall seems that one of the main drivers of self-reported mobility reduction is the overall severity of mobility restriction policy in Kenya. Given that the policy stringency is a continuous variable running from 0-100, a seven-point increase of the stringency index has a similar effect compared to being employed.

Table C6-6: Determinants of self-reported mobility restricting behavior

Self-reported mobility restriction	National n=11,351	Rural n=5,318	Urban n=6,033
Trust in Government	-0.31**	-0.24	-0.46**
Trust in fellow citizens	0.49	0.73**	0.39
Sex (Female)	-0.26	-0.21	-0.27
Education Level	-0.06	0.12	-0.37**
Household Head	-0.11	-0.12	0.10
Age	-0.00	-0.00	-0.01
Urban/Rural	0.04	N/A	N/A
Know someone who is/was infected	-1.40**	-2.30***	-0.42
Employed	0.31	0.78**	-0.22
Worried about food	0.21	0.50**	-0.23
Policy Stringency Index	0.06***	0.06***	0.06***
Weekly Change COVID-19 cases (%)	0.00	0.00	-0.00

*Note: *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level*

Comparing urban vs rural outcomes, we find that there seem to be different drivers of reducing mobility. While in rural areas, self-reported mobility reduction was associated with trust in fellow citizens, being employed and worrying about food, in urban areas education levels and the trust in the government's were associated with less self reported mobility restricting behavior.

Discussion

Our study has a few salient findings. First, recovering mobility levels in Kenya following the initial declines in early 2020 have caused people to enter the labor force again, three-quarters of them re-entering into employment. Second, while increased mobility caused an increase in hours worked for the different sectors, no effects can be found for generated incomes. Potential reasons for this observation may be that employers continued to support workers for a while up until their re-entry, or otherwise lowered payments at the beginning of the pandemic and did not increase payment as the number of hours worked went up again either due to financial distress or with the promise of later repayment. We allow ourselves a cautious interpretation by leveraging asset information for a total of seven assets (radio, mattress, charcoal jiko, refrigerator, television, landline telephone and computer/tablet/laptop) that became available during wave 4 and 5 of the RRPS surveys for a total of 10,785 households, which also incorporated baseline values from February 2020. Comparing wave averages to pre-pandemic levels show that overall asset ownership reduced over the course of the pandemic until wave 4 and wave 5 with a slight recovery between wave 4 and wave 5. These results are consistent when incorporating the full set and sub-sets of the seven assets. We interpret this as evidence that household had to sell assets to cope with income and job loss as well as health-related expenses, which makes the idea, that employers continued payments or that social safety nets were at play rather implausible. However, given that we lack precise income baseline data, understanding the exact dynamics over the full course of the pandemic will be a subject for future research.

Comparing urban vs rural, we do find additional statistically significant effects of mobility on agriculture in an urban setting, which may be due to the fact that the agricultural workplace in the rural setting is often directly linked to the place of living i.e., farms or plantations connected with villages. At the same time in rural settings, the number of wage hours worked increased more which may be explained by an increased reliance on commuting to the workplace or an increased elasticity of job availability in downturn times compared to urban areas.

Looking at different stages of the pandemic, we find that particularly in rural settings people quickly re-entered employment already during the pre-recovery phase. During the post-recovery phase both in rural and urban areas people left unemployment more than re-entering the labor force. This implies that people that lost their jobs left the labor force and quicky re-entered into employment while people that could not afford to leave the labor force stayed unemployed for a longer period. Overall recovery in agriculture seems to be driven mainly by urban dynamics.

Thinking about safety nets and mitigation measures, awareness of differential impacts across sectors in urban and rural areas carries important insights into target groups and economic costs of restriction measures in these specific areas. To determine causal effects of mobility not just during a recovery phase but for overall economic and labor market activity, future research will rely on researchers' ability to attain high-frequency data covering not only the course of a pandemic but also the time prior to the outbreak. Furthermore, given that this is a country case, it will be interesting to see how estimates of the causal impact of mobility on economic recovery compare to findings from other countries or regions.

Finally, we find that peoples' trust in the Kenyan government's ability to deal with the pandemic, employment status and overall level of stringency significantly influence people's self-reported reductions of mobility. There are differences between urban and rural households. While for rural households the level of stringency and worry about food, knowing someone who was infected, employment and trust in fellow citizens were of significance, in the urban setting additional factors are statistically relevant such as education, and the trust in the government's ability to handle the pandemic. This may suggest that urban educated citizens generally perceive less risk for themselves and therefore are more receptive to the perception of the government doing the job. At the same time, the coefficient for employment in the rural setting is much larger than for urban households, pointing towards increased opportunity costs of illness. The significant negative impact of knowing someone who has been infected could point towards the need to support the person that falls ill which translates into additional mobility. This increased relevance of social ties is also backed by the relevance of peopl's trust in their fellow citizens in the rural setting. While we are aware that self-reported behavior data needs to be treated with caution (Jakubowski, Egger et al. 2021), we nevertheless believe that our large sample allows for important insights into determinants of self-restricting behavior during the time of a pandemic. Comparing coefficients, a 10-point increase of policy stringency outweighs most of the other coefficients, highlighting potential signaling or enforcement mechanisms that come with more severe government measures. These insights underscore the importance of strong government measures to save lives. However, they also show that different messages and different channels need to be applied to convince citizens to self-reduce mobility and social interaction.

Our study has a few limitations that are mostly due to data availability. At first, given that the RRPS started in May, we lack baseline data for pre-pandemic levels. While for three of the five labor market outcomes, we do have retrospective recall values, this data is subject to the innate bias that recall data carries. While the lack of baseline data does not directly affect our message, the interpretation needs to be cautious as the causal effect of mobility recovery may differ from the causal effect of mobility on labor market outcomes in non-pandemic times. Another limitation is the fact that we do not have county-level stringency index data but had to rely on national aggregates to instrument for county-specific mobility changes. However, given that a) only very few policies were implemented on county-levels and b) the national index score is an average of stringency across the country, we believe that this is justifiable. In case that county-specific stringency index data for Kenya is released, it will be necessary to compare these results. Due to the nature of the RRPS survey waves and the fact, that due to COVID-19, interviews had to be conducted via phone, there is a potential bias due to the selection at baseline and the attrition of the selected population in the follow-up waves. Phone surveys can only reach respondents using a phone in an area with network coverage, therefore statistics are only representative for this part of the population, potentially excluding to some extent the poorest households who do not own phones or live-in areas with no network coverage. RRPS weights were adjusted by the World Bank in a two-step approach (Himelein, Eckman et al. 2014) to make sure the RRPS is as representative as possible for the entire population and adjusting for attrition. We therefore do not believe this bias to be significant. Finally, the instrumental variable approach hinges on the assumption that the policy stringency index has no direct causal relationship to the outcome measures, which are not mitigated through mobility changes or other measures that we control for as well as that the economic environment itself did not affect the policies put in place to reduce mobility. While we present anecdotal evidence that is in favor of this idea, we realize that it is indeed possible that decision makers worried about containing the spread of the virus did also factor in economic concerns, particularly at later stages of the pandemic as the virus was better understood.

As final sanity check we used weekly lags of explaining variables, given that low mobility levels may take a bit of time to translate into labor market outcomes. However, we do not find this to impact our results.

Conclusion

We examined the impact of increasing mobility on household labor market outcomes over the course of the pandemic following the initial steep declines in March and April 2020 and determined which factors influenced people's self-reported adherence to recommended mobility restricting behavior.

Over the course of the pandemic from May 2020 until June 2021, a 10 % of recovering mobility leads to a 12 percentage points recovery of labor force participation and an increase of 9 percentage points of household members being employed. At the same time, a 10% of recovering mobility causes an increase of 11 wage hours per week (formal and informal). Particularly the results for extrinsic margins of employment are consistent for urban and rural over the course of the past year, with differences regarding the timing of the recoveries. Looking at the intrinsic margins of employment, wage work was more affected in rural areas, while agricultural work was more affected in urban areas.

Among the factors influencing self-reported mobility and thus, nationwide mobility levels, the trust in the government's ability to deal with the pandemic leads to less self-restriction, while country wide policy stringency level leads to higher self-restriction, the overall policy stringency being of specific importance.

Knowing about the sectors affected most by mobility and at which stage of the pandemic this affect takes place is important knowledge for policy makers. Policy makers in future pandemics will need to carefully evaluate policies aimed at reducing mobility with the economic costs that are associated with them. We find that labor market recovery in terms of employment levels and hours worked comes quickly with increasing mobility, with strong effects on wage work across the country and agricultural work in urban areas. Income however does not seem to be causally influenced by recovering mobility. Finally, providing safety nets and working to save employment status in formal and informal wage employment will continue to be important measures to shield people from the most severe consequences of the pandemic but based on self-reported behavior can also be beneficial especially to people's adherence in rural areas to officially recommended mobility reductions.

Supplements

Table C6-7: Sociodemographic comparison of different RRPS waves

	Wave 1 (14/5/2020-8/7/2020)	Wave 2 (16/7/2020-18/9/2020)	Wave 3 (28/9/2020-30/11/2020)	Wave 4 (15/1/2021-25/3/2021)	Wave 5 (29/3/2021-25/6/2021)
Average Age of Respondent	35.03	35.19	34.71	36.1	36.22
Share of Female Respondents	50%	53%	51%	50%	49%
Average Education of Respondent*	3.29	3.31	3.39	3.25	3.31
Household size	4.13	4.15	3.4	3.65	3.26
Average Age of Household Head	39.53	40.08	37.42	37.7	37.67
Share of Female Household Heads	33%	36%	37%	41%	39%
Share Urban	35.9%	36.0%	37.0%	36.4%	40.0%
Sample Size	4,062	4,504	4,993	4,906	5,874
Share RDD	18.9%	18.7%	20.2%	17.2%	19.8%
Response Rate	36%	41%	45%	43%	51%

*An education level of 3 equals to completed post-primary, vocational, a score of 4 equals completed secondary education

Table C6-8: Variables for causal effect of mobility on labor market outcomes analysis

Role in Analyses	Category	Variables	Coding	Pre-COVID-19 Recall Data?
Outcome Variables	Employment Status	Respondent Employed (%)	Binary (Yes/No)	
		Respondent Unemployed (%)	Binary (Yes/No)	
		Respondent Not in Labor Force (%)	Binary (Yes/No)	
	Hours worked	Working Hours in Agriculture per Working Household Member in past 7 days	Ordinal	Yes
		Working Hours in Wage Employment per Working Household Member in past 7 days	Ordinal	Yes
		Working Hours in Self Employment per Working Household Member in past 7 days	Ordinal	Yes
	Income earned	Agricultural Earnings (KSH past 14 days)	Ordinal	Yes
		Wage Earnings (KSH past 14 days)	Ordinal	Yes
Explaining Variables	Fear of Illness	Yes to the question “Are you feeling nervous or anxious due to the coronavirus outbreak?” and statement of one of the following reasons: Fear of myself or family getting infected by coronavirus Fear of myself or family dying due to coronavirus Fear of me infecting others in the community Fear of losing access to health facilities	Binary (Yes/No)	N/A
	Economic Uncertainty	Yes to the question “Are you feeling nervous or anxious due to the coronavirus outbreak?” and statement of one of the following reasons: Loss of employment / business Fear of being unable to feed or provide for family Effect on education system and school closures Economic Crisis/Paralyzed Movement Uncertainty of when lockdown will end / things will return to normal	Binary (Yes/No)	N/A
	Know s/o Infected	Do you know anyone that has, or has had, COVID-19/coronavirus?	Binary (Yes/No)	N/A

Table C6-9: Variables for analysis of determinants of self-reported mobility reduction behavior

Role in Analyses	Category	Explanation	Coding
Outcome Variables	Self-reported behavior change	Any self-restricted mobility behavior (at least one answer with yes to the following questions): - Avoid groups more often? - Stay at home more? - Travel outside less? - Go to work less? - Return home earlier at night?	Binary (Yes/No)
Explaining Variables	Trust in Government	The Government is trustworthy in the way it manages the Coronavirus crisis?	Binary (Yes/No)
	Trust in fellow citizens	Generally speaking, would you say that most people can be trusted?	Binary (Yes/No)
	Sex (Female)	Gender Dummy	Binary (Male)
	Education Level	No education=0, University postgraduate=8	Ordinary
	Household Head	Household Head Status Dummy	Binary (Yes, No)
	Age		Ordinary
	Urban/Rural	Urban Dummy	Binary
	Know s/o infected	Do you know anyone that has, or has had, COVID-19/coronavirus?	Binary
	Employed	Employment Dummy	Binary
	Worried about food	Household missing/cutting meals in past 7 days (%) (at least one yes answer to the following 2 questions): - In the past 7 DAYS, how many days have ADULTS in your household skipped meals or cut the number of meals?	Binary

7. Broken Promises: Evaluating an incomplete Cash Transfer Program[234]

Angelika Müller, Utz Pape and Laura Ralston[235]

Introduction

An increasing share of the world's poor live in fragile states, which poses new challenges to programs that seek to raise their incomes. One major risk associated with an insecure and fragile context is the unintended and unplanned interruption or cancellation of the program. Despite the prevalence of these cases, little is known about the effect of a program cancellation on intended beneficiaries. However, knowing about these risks would help policy makers make informed decisions about the costs and benefits of an intervention a priori. In addition, information on the consequences of failed implementation can help reduce detrimental impacts at the program design stage.

To our best knowledge, this study is the first to analyze what happens if an intended intervention is canceled. The Youth Startup Business Grant Program in South Sudan that was canceled due to erupting violence in 2016 provides us with the opportunity to study the impacts on socio-economic, behavioral and psychological outcomes on intended beneficiaries. In particular, we are interested in understanding effects on participants who were promised to receive a cash grant but did not ultimately receive it. Economic theory lends multiple reasons why outcomes for these participants could differ from outcomes in the absence of the program. Overall, our results suggest that the impact of failed interventions is mixed and depends on the gender of participants and their ex post treatment status. In this instance, on average across all participants, the invention was largely ineffective, but some sub-groups were negatively affected. Given that applicants for the intervention were on average more educated than the average youth in South Sudan, the average population might have displayed reduced skills to cope with the program cancelation. In that sense, findings present a lower bound.

The Youth Startup Business Grant Program consisted of an unconditional cash grant combined with a business and life skills training exercise and was particularly targeted at young women. South Sudan has suffered from political instability and latent conflict since its inception in 2011. In this context, the youth struggled with declining livelihoods and a lack of economic opportunities. This put them at risk of participating or becoming victims of criminal or violent activities. Young women were at particular risk. In response, the program was designed by the World Bank in collaboration with the Ministry of Commerce to offer a cash grant worth US$ 1,000. Existing evidence suggests that injections of capital are the most effective means of raising income in poor and fragile states (Blattman and Ralston 2015). Beneficiaries could access the grants denominated in local currency through a commercial bank account. Although the cash grant was aimed towards promoting (self-) employment and business development, beneficiaries were free to decide on its use. The program also entailed a one-week business and life-skill training, which participants needed to attend in order to access the grant.

In late 2014, the program randomly selected 1,200 beneficiaries out of a pool of more than 6,000 applications to receive the grant. More than 60 percent of the grants were awarded to young women. A similarly sized control group was selected to enable the assessment of the program in a rigorous impact evaluation. Baseline data from both treatment groups were collected before grant beneficiaries received their business and life skill training in April and May 2015. Almost all selected

---

[234] UP and LR contributed equally, while AM conducted data cleaning, implemented the analysis and provided the first draft of the write-up.
[235] Authors in alphabetically order. Corresponding authors: Utz Pape (upape@worldbank.org) and Laura Ralston (lralston@worldbank.org). The authors are grateful for contributions from Mollie Foust, Luca Parisotto, Nadia Selim, Jeremy Shapiro and James Walsh as well as Nicola Pontara. We also thank Bledi Celiku, Axel Dreher, Arevik Gnutzmann-Mkrtchyan, Markus Goldstein and seminar participants at Heidelberg University and UC Davis for useful comments. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

beneficiaries attended the 1-week training. After the training, participants were asked to open a commercial bank account in which the grant would be deposited.

Escalating violence at the end of 2015 forced the program to terminate the disbursement of the grants before all participants had accessed them. Completion of the program was first postponed and finally canceled to mitigate the perceived risk for beneficiaries to become the target of crime. In addition, there was the risk that the conflict might be exacerbated if grant money got into the wrong hands and was used to purchase arms. Delays in communication and in the processing of the grants meant that the timing at which disbursement was stopped varied across regions and bank branches.

Interventions in highly fragile and insecure states are often at risk of failing to be rolled out as originally planned. Obvious ethical objections make it impossible to study this effect in the form of a randomized-controlled trial. This study takes advantage of the circumstances under which the Youth Startup Business Grant Program was canceled to identify the socio-economic and behavioral consequences of projects that fail to be implemented as intended. Those originally assigned to the treatment group but who did not end up receiving grants show few systematic differences, except their location, from those who accessed the grants. We exploit this natural variation in location in interaction with the original assignment to the treatment group as an instrument for those who obtained the grants versus those who did not.

Hence, this study distinguishes between two *de facto* treatments. “Training but no grant” consists of having participated in the business skills training and been informed of receiving a US$ 1,000 grant, but later having to experience that the grant disbursement was stopped. To assess the treatment effect, this group will be compared to the control group of the original intervention who was informed of not having been selected to receive the grant. In addition, this study analyzes the effect of the originally planned intervention. “Training and grant” consists of having participated in the life-skills training and successfully having accessed the cash grant.

On average, across all participants most socio-economic, and behavioral and psychological indicators were neither negatively nor positively affected by the intervention. However, when considering ex post treatment groups and gender, some groups were detrimentally affected by the intervention. For example, participants who only received the training, but expected the grant also, seem to have experienced small consumption declines relative to the control group. Female participants among this group also showed a strong reduction in their trust level. We also found some evidence that these women were less likely to migrate. Given that large shares of the population in South Sudan migrated in the period of our analysis to escape conflict affected areas, it is possible that women who expected the grant stayed back who would have migrated in the absence of the intervention. While we do not have direct information on this unintended consequence, one could be concerned of the potential detrimental outcomes to these participants.

Positive impacts were only detected on some outcomes and only to those who received the grants. For example, consumption, savings and reductions in debt, as well as reported levels of psychological well-being increased among the participants receiving both the training and the grant. These positive effects seemed to be independent of gender. Given these results, we argue that greater concern should be taken when planning programs in these volatile environments, as there is at least some evidence from our results on unintended negative consequences on program participants who did not receive the full set of benefits anticipated at the program outset.

The remainder of the paper proceeds as follows. Section 2 reviews theoretical considerations and the related literature. Section 3 discusses our study design. Section 4 describes our empirical specifications and discusses the reasoning behind our instrumental variable estimations. In Section 5, we describe the main results on the socio-economic, and the psychological and behavioral outcomes. Section 6 concludes.

Theoretical considerations and existing literature

The benefits of conditional cash transfers in non-fragile environments are well documented. For instance, multiple studies analyze the benefit of cash grants for education and health (see, Manley, Gitter et al. 2013 for szstematic overviews, Baird, Ferreira et al. 2014). A large body of literature evaluates the benefits of cash grants for the profits and growth of microenterprises (De Mel, McKenzie et al. 2008, De Mel, McKenzie et al. 2012, Fafchamps, McKenzie et al. 2014). More recent studies also document the benefits of cash transfers on self-employment and income. Banerjee, Rema et al. (2017) re-analyze results from six randomized trials on cash transfer program to show that cash grants do not discourage work as standard economic theory would suggest. In addition, Bianchi and Bobba (2013) find that a cash transfer program in rural Mexico significantly increased the probability to start an enterprise. Most relevant to our context, a study in the conflict-affected north of Uganda finds that a cash grant program targeted at generating self-employment among youth significantly increased their earnings (Blattman, Fiala et al. 2014). A recent analysis of an invention in Kenya also suggests that cash grant programs can have positive effects not only on the economic well-being of beneficiaries, but also on their psychological well-being (Haushofer and Shapiro 2016).

Despite the strong evidence on benefits of cash transfer programs, Baird, McKenzie et al. (2018) argue in their systematic literature review that *"cash transfers that are made without an explicit employment focus […] tend to result in little to no change in adult labor*". To address this concern, the Youth Startup Business Grant Program in South Sudan combined the unconditional cash grant with a business skills training. Research on the impact of business trainings is generally mixed and the evidence is in the African context is scarce. An early study on microfinance clients in Peru found no economically significant effect of complementing a loan with a business training (Karlan and Valdivia 2011). In contrast, a randomized trial in India showed that business trainings could be effective in overcoming restrictions based on gender-norms that held female entrepreneurs back (Field and al. 2013). In addition, a study conducted in Ghana finds a rudimentary management training for micro and small enterprises can significantly improve their performance, because many business practices that are standard in developed countries are unknown to the participants (Mano, Iddrisu et al. 2012). Furthermore, two studies on the same business training program for entrepreneurs in Tanzania both found that business knowledge significantly increased (Bjorvatn and Tungodden 2010, Berge, Oppedal et al. 2014). These diverging findings highlight that there is still a lack of evidence on the type of content that shows the best results (McKenzie and Woodruff 2014). In their meta-analysis on the effectiveness of entrepreneurship programs, Cho and Honorati (2014) find that business trainings are most effective among beneficiaries that already own a business and when combined with financial support.

Despite the extensive seminal work on cash grant programs and business skills trainings, the existing literature does not offer clear predictions on how beneficiaries are affected if a program has to be canceled. There are multiple reasons why the false expectation of receiving a cash grant could have detrimental effects on socio-economic outcomes. First, the existing literature on cash grants suggests that these are an effective way to overcome credit constraints. If beneficiaries commit to an investment in the belief to receive a grant in the future, it is likely to be welfare reducing if the grant disbursement never happens. Second, most seminal work on cash grants finds increases in consumption (See Baird, McKenzie et al. (2018) for an overview of the literature). Again, if program participants already increase consumption before having accessed the grant, they might suffer from reductions in savings and consumption if their expectation does not come true. Finally, beneficiaries might decline employment opportunities that they would have accepted in the absence of the program.

In addition, existing research lends multiple explanations of how the program cancellation might affect psychological and behavioral indicators. Psychological theory suggests that mental health depends strongly on external stressors as well as a person's resources to cope with these. Research on transfer program has shown that receiving grant payments can improve indicators of psychological well-being (Ozer, Fernald et al. 2011, Baird, de Hoop et al. 2013, Haushofer and Shapiro 2013). Interestingly, Baird, de Hoop et al. (2013) find also increased psychological distress among untreated study participants in treatment areas. These findings are consistent with the theory that psychological wellbeing depends on not only absolute economic status, but also relative economic status compared to one's peer group (Luttmer 2005). In consequence, it is likely that participants who knew about others who received the grant experienced a reduction in their personal utility. Recent evidence in experimental economics shows that the experience of being lied to significantly reduces participants' trust level as well as their trustworthiness (Gawn and Innes 2018). The "broken promise" that the cancellation of the program created could erode social capital in a similar way. Subsequently, other outcomes such as employment or engagement in crime and violence could suffer negative impacts. This mechanism would be particularly concerning, given new evidence that international organizations such as the World Bank sometimes enjoy more trust than governments – particularly if governments are seen as corrupt (Milner, Nielson et al. 2016, Findley, Harris et al. 2017). In addition, theory suggests particular risk for female participants. One study found that female transfer beneficiaries of the Oportunidades program in Mexico were more likely to receive violent threats from their partners, indicating that threats were used to extract rents (Bobonis, Gonzâlez-Brenes et al. 2013). Hence, there is a particular risk associated with the possibility that women who failed to access the grant could not convince their partners of the program cancellation. This might put them at increased risk to experience domestic violence. What is more, it is possible that the failed implementation influenced the migration decision of participants. Due to the conflict about a quarter of the South Sudanese population are currently internally displaced or have left the country. It is possible that the expectation of receiving the grant incentivized participants to stay back in their region of origin.

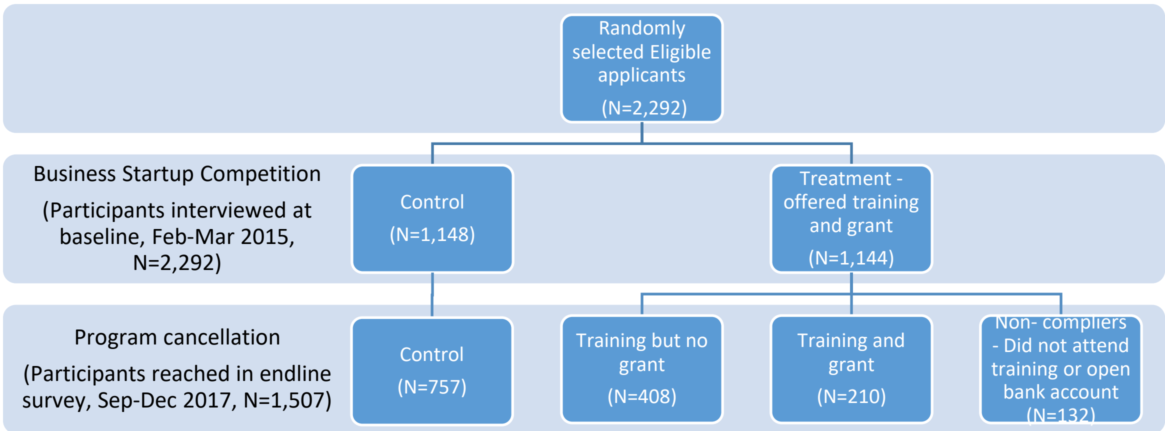
Study Design

The eligible population of the grant program was the youth in six states in South Sudan with focus on young women. The program was implemented in the least conflict-affected states in South Sudan at the time of its launch: Central Equatoria, Eastern Equatoria, Western Equatoria, Northern Bahrel Ghazel, Western Bahrel Ghazel, Lakes State. Eligible individuals had to be aged between 18 and 34 and be of South Sudanese nationality. Originally, 200 individuals were selected from each of the six states. A share of 60 percent of the grants was targeted at women.

The program received approximately 6,000 applicants. Interested applicants had to submit a one-page written proposal for a new business idea. The document had to be written in English, although communication materials were also provided in Juba Arabic. In addition, the applications required proof of their South Sudanese nationality and documents needed to open a bank account. This application process was designed to incentivize positive self-selection into the sample. In this sense, the program participants may be better equipped to use the cash grant successfully to improve their business or employment situation than the average population. From the received 8,240 applications, 4,699 were found to be eligible.[236] From these eligible applicants, 1,200 were randomly selected to receive the grant and 1,200 were randomly selected for study in the control group, with equal proportions per state and by gender. The baseline survey was conducted between January to March 2015 and data were collected from 1,144 treatment participants and 1,148 control participants. Approximately 4.5 percent of initially identified study participants could not be tracked and did not participate in either the baseline survey or the program. The baseline survey was concluded prior to the commencement of the one-week training that was held across the 6 states between April and May 2015.

The intensification of violence between 2015 to 2017 forced many study participants to migrate reducing the number of participants that could be located for the endline survey. About a quarter of the population of South Sudan was displaced during the study period, which made it difficult to locate all participants of the original control and treatment group. Before the endline survey, the World Bank

Figure C7-1: Treatment streams of original and new intervention.



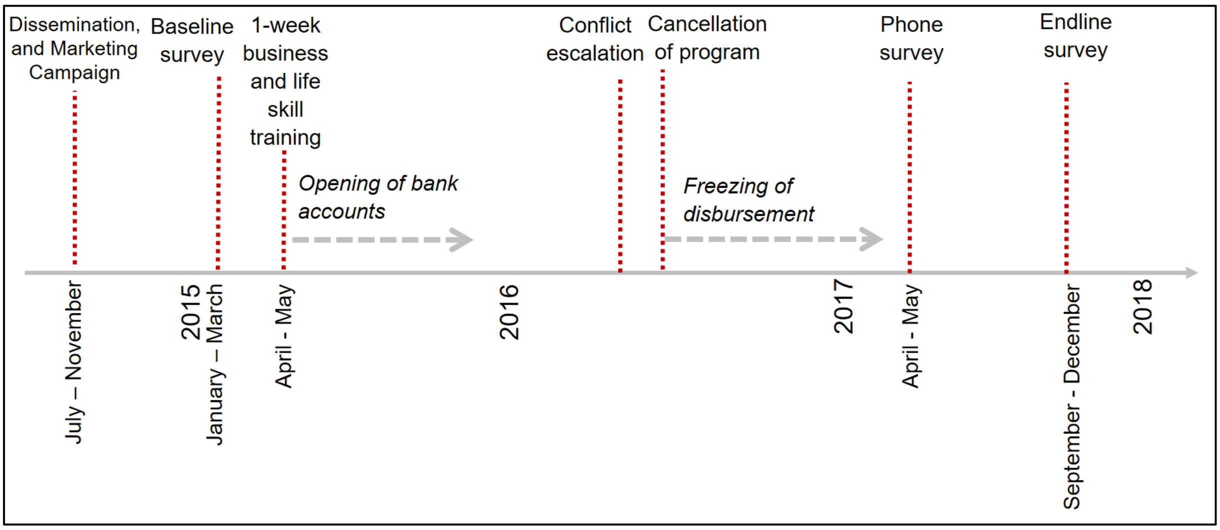
conducted a phone survey in May 2017 that informed the grant beneficiaries of the halt of the program and assessed the feasibility of collecting endline data. The phone survey managed to reach around 55 percent of the study participants (1,264: 642 from the control group and 622 from the original treatment group), from which 99 percent agreed to participate in the endline.

Due to budget and logistical considerations, the endline survey targeted a sample size of 1,800 individuals randomly chosen from the list of participants after prioritizing the phone survey respondents who had agreed to be interviewed again. Endline data collection activities commenced

[236] Of those deemed ineligible, the desired purchase of land was the main reason. Other explanations included blank or unrealistic business ideas, age listed outside target range (18-35 years), no identification attached, or not being South Sudanese.

in September 2017. After intensive tracking efforts over a period extending to four months,[237] 1,524 participants were located, and 1,507 participants completed the interviews. The respondents interviewed in the endline survey were given the opportunity to voice their concerns and opinions about the cash grant program, through short video testimonials that are publicly available online.[238] Out of these 1,507 respondents, 1,045 had been reached in the phone survey and 462 had been located through intensive tracking efforts based on information provided in the baseline.[239] Figure 2 illustrates the time line of the data collection and intervention steps.

Figure C7-2: Timeline of program implementation, cancellation and data collection.



At the end of the endline survey, there was approximately equal representation between the treatment (750) and control (757) groups, with 394 and 391 fewer observations from each group respectively. This was despite ongoing conflict keeping enumerators from going to a few counties due to insecurity.[240] As a robustness check on results, the difficult-to-reach study participants (those reached only through intensive tracking) are reweighted to regenerate the baseline sample. This recognizes that the hard-to-reach may be more representative of the outstanding missing observations in the endline data than those reachable through the phone survey.

The main approach for measuring outcome variables was through face-to-face interviews that were conducted as part of baseline and endline surveys described above. In addition, risk preferences, trust attitudes and engagement in crime and violence were assessed using experimental data collected during these face-to-face interviews from decisions reported over lotteries, trust games and elicited through list experiments (see appendix 1 for full methodological details).

The hypotheses of this study are grouped into 2 main families of outcomes – socio-economic outcomes, and psychological and behavioral outcomes. It is possible that these families of outcomes were differently affected by the intervention. For instance, it is possible that participants who failed

[237] The majority of data was collected between September to November 2017, but field teams remained on the ground until end of December 2017 trying to locate and interview participants.
[238] The video testimonials from the BSCIE as well as other surveys conducted in South Sudan during this period are available at: www.thepulseofsouthsudan.com
[239] Intensive tracking efforts included returning to the GPS coordinates for the baseline survey and looking for participants, contacting other contacts listed by the participant in their program application and through the baseline survey, asking other respondents, local officials, at the Chambers of Commerce and trade unions about the location of difficult to find participants, and making at least five attempts to find persons over a period of several weeks.
[240] In WEQ: Mvolo, Mundri East, and Mundri West ; in CEQ: Kajo Keji, Morobo, and Lainya ; in Lakes: Rumbek North (flooding during time of data collection).

to access the grant but participated in the training did not experience any negative effects on their socio-economic situation while suffering negative effects on their psychological and behavioral well-being.

Methodology

Selection into treatment arms

Selection into treatment arms was a two-stage process. In the first stage participants from the control group and the original treatment group were randomly selected according to the originally planned experiment. A balance test on these baseline study participants shows no systematic differences between the original control and the treatment groups (Table C7-2). We find weak evidence that the control group was slightly less affected by the conflict. Our measure of conflict exposure is based on geo-referenced data by UCDP (Sundberg and Melander 2013). It consists in an average of deaths by event within a 300 km radius weighted by geographic proximity to participants' baseline location.[241] Figure C7-4 in the appendix displays maps of conflict events before program initiation and between baseline and endline survey. The difference in conflict exposure between control and treatment group is only significant at the 90 % level. To control for potential selection bias, we include conflict exposure as control variable in our regressions. In addition, we find no evidence that attrition depended on the selection of the original treatment group (Table C7-10), nor differential attrition across covariates between these two groups (Table C7-12). Importantly, there is also no evidence that participants in the control group accessed either the training or the grants. What is more, the low geographic concentration of program participants makes spill-over effects unlikely. Hence, control group outcomes can plausibly be regarded as counterfactual outcomes for beneficiaries in the absence of the program.

The second stage of the selection process decided which *de facto* treatment participants of the original treatment group received. Since the cancellation of the program was not planned, this process was not systematically controlled. Among the original treatment group participants reached through the endline survey, we have three *de facto* groups. The "Training but no grant" group consists of the 408 individuals that had not accessed their grants when erupting violence forced the program to terminate in late 2015, the "Training and grant" group consists of the 210 individuals who successfully accessed the grant, and the "Non-compliers" group consists of 132 individuals who did not attend the training and therefore could not access the grant. [242] The assignment process to these de facto treatment arms poses some challenges for identification. Participants had no reason to anticipate that the grant disbursement might be frozen in the future, the assignment to "Training and grant" and "Training, but no grant" was partly random. However, participants who tried to access the grant right after the training had a higher chance of receiving it than those who waited, so individual characteristics of the participants might have created some self-selection into these two groups.

To assess the degree of endogenous selection into the "training but no grant" and "training and grant" groups, we examine the balance on covariates between these two groups using the baseline data. We find that older, married participants with larger families were more likely to access the grants (Table C7-3). In addition, participants who received the grant were more likely to already own a business and hold a bank account, reported higher consumption levels and reached higher ranks in our literacy and numeracy evaluations. Hence, there is some evidence that those who accessed the grants were endogenously equipped to have better access to grants (via prior formal banking experience) or

[241] 300 km equals the maximum distance of any program participants to a KCB bank branch. We also tested 50 km, 100 km, 200 km radii, but the results remain extremely similar.

[242] Attending the training program was a pre-requisite to accessing the grant. Part of the training included financial literacy around opening and using the bank accounts, so only those participating had bank accounts opened for them.

predisposed to benefit more greatly from them in terms of the socioeconomic outcomes we measured. Therefore, when analyzing the treatment effects the estimations will address endogenous selection into training and grants group not only by the inclusion of covariates, but also with an instrumental variable approach that focuses on an exogenous channel through which some participants were better enabled to access the grants. Importantly, we do not find strong evidence that exposure to conflict events determined whether participants could access the grant or not.

Despite some correlations with individual covariates as discussed above, the main determinant whether participants had access to the grant was their location. In Lakes and Western Bahr El Ghazal the majority of the eligible participants received the cash grants, while in Eastern Equatoria and Western Equatoria the majority did not receive the grants (Table C7-3). The difference between states can be explained by failures in the coordination between different bank branches across the six states. While some branches moved quickly with halting disbursements of grants, other branches did not. In addition, selection into de facto treatment arms depended on the distance to the closest KCB bank branch. Figure C7-3 in the appendix displays participants' baseline location, major cities in the project states and locations of KCB bank branches. Since not every large city in South Sudan had a branch of this particular bank, we regard this variable as plausibly exogenous and exploit it for an instrumental variable approach.

Estimation strategy

We begin our analysis by estimating an intent-to-treat effect which gives us the average effect of the intervention on all participants that were selected for the original treatment group. Since assignment to the original treatment group was randomized the coefficient of the estimate has a causal interpretation. It tells us whether there was a negative net average effect of the intervention on any of the main outcomes. This gives us a first indication of whether the intervention created more "harm" than "good". The specification for the intent-to-treat effect is as follows:

$$(1)\ y_{ij} = \beta Z_i + X_i'\gamma + s_j + \varepsilon_{ij}$$

where y_{it} is a vector of outcomes for individual *i* in strata *j*, Z_i is a dummy variable that takes a value of 1 if individual *i* was originally selected for the cash grant program, s_j are strata fixed effect and ε_{ij} is the error-term clustered at baseline boma level. X_i' are individual-level covariates that were collected at baseline and might affect outcome variables. In addition, we run treatment-on-the-treated (TOT) estimations to understand the effects of receiving "Training and grant" or "Training, but no grant". The specification for the treatment-on-the-treated effect is as follows:

$$(2)\ y_{ij} = \alpha Treatment1_i + \beta Treatment2_i + X_i'\gamma + s_j + \varepsilon_{ij}$$

where y_{it} is a vector of outcomes for individual *i* in strata *j*, s_j are strata fixed effect, X_i' are individual-level covariates that were collected at baseline, ε_{ij} is the error-term clustered at boma level, $Treatment1_i$ is a dummy variable that takes a value of 1 if individual *i* participated in the business skills training, but did not receive the grant due to the cancellation of the program and $Treatment2_i$ is a dummy variable that takes a value of 1 if individual *i* participated in the training and also received their grant. Thus, participants that received no treatment because they were either part of the control group or were invited but did not attend the training build the baseline of this estimation. TOT effects of treatment 1 and treatment 2 are estimated by parameters α and β respectively. As discussed above, the treatment-on-the-treated effects has no causal interpretation, because participants assignment to "training, but no grant" and "training and grant" partly depended on the time at which participants tried to access the grant as well as willingness to attend the training and could therefore be endogenous.

To address endogenous selection into "training and grant" versus "training, but no grant", we run instrumental variable regressions. As described previously, the instrumental variable relies on the fact that receiving the grant was conditional on holding a formal bank account at KCB bank. KCB bank operated only 16 bank branches in that time in South Sudan, out of which 15 were in the states of our program. A KCB bank branch was not in every large city. This leads to some variation in how convenient it was to access the grant which is uncorrelated with participants' personal characteristics. Of course, the distance to the closest KCB bank branch may correlate directly with many other geographic variables that affect our outcomes. Fortunately, we also have observations from the control group, that was randomly selected, and can therefore control for outcome differences conditional on the distance to the closest bank branch. To do so, we interact the logarithmic distance to the closest KCB bank branch with a dummy variable that marks the original assignment to the treatment group. The local average treatment effect (LATE) then estimates the effect of having received training and grant, because one was selected for the original treatment group and lived close to a KCB bank branch while controlling for the average outcome levels at the location.

Our first stage regressions in Table 7 demonstrate that even after controlling for distance to the closest KCB branch the interaction term remains a strong predictor of whether participants received "training and grant" or "training, but no grant". In addition, we argue that the instrument is excludable. We exploit the variation generated through the interaction between distance to the closest KCB branch and exogenous assignment to the treatment group, while controlling for the main effect of the potentially endogenous distance to closest KCB branch. This interaction term can be considered an exogenous regressor under some mild assumptions (Bun and Harrison 2018).[243] The exclusion restriction requires that conditional on the distance to the closest KCB branch the instrument affects outcomes not directly, but only by making it more or less likely that a participant will receive the grant. It would only be violated if the interaction term had a direct effect on outcomes, i.e. if being selected for the treatment group had different effects for participants closer to a KCB branch that did not result from a higher probability to receive the grant. We do not think this is plausible.

The specification for the local average treatment effects are as follows.

Second stage equation:

$$(3)\ y_{ij} = \alpha Tre\widehat{atmen}t1_i + \beta Tre\widehat{atmen}t2_i + \delta KCBDistance_i + X_i'\gamma + s_j + \varepsilon_{ij}$$

First stage equations:

$$(4a)\ Treatment1_i = \alpha Z_i + Z_i \times KCB\ Distance_i'\sigma + \delta KCBDistance_i + X_i'\gamma + s_j + \varepsilon_{ij}$$

$$(4b)\ Treatment2_i = \alpha Z_i + Z_i \times KCB\ Distance_i'\sigma + \delta KCBDistance_i + X_i'\gamma + s_j + \varepsilon_{ij}$$

where y_{it} is a vector of outcomes for individual *i* in strata *j*, s_j are strata fixed effect, X_i' are individual-level covariates that were collected at baseline, ε_{ij} is the error-term clustered at boma level, and $Treatment1_i$ and $Treatment2_i$ are dummy variables indicating treatment streams as described above. Equations (4a) and (4b) display the first-stage equations, which instruments $Treatment1_i$ and $Treatment2_i$ with the original assignment to treatment Z_i as well as the

[243] In particular, the identifying assumption is that the outcome variables and the endogenous variable distance are jointly independent of the exogenous variable of original treatment assignment.

interaction between Z_i and the logarithmic distance to the closest KCB branch $KCB\ Distance_i$. Like the TOT estimation, the LATE of treatment 1 and treatment 2 will be estimated by parameters α and β respectively.

In addition, we establish a separate method to estimate experimental data on sensitive outcomes that tend to be under-reported when asked about directly through survey data. For example, we used list experiments to collect information on conflict and crime. The list experiment relies on a separate treatment group to assess the true propensity of positive answers in the study population. As we are not only interested in the propensity of positive answers in average across all study participants but would like to know whether participation in either treatment group changed the propensity of positive responses, we deploy a difference-in-difference estimation. This estimator calculates the difference in positive responses between positive responses in the control group and treatment group. It consists of an interaction between treatment status of our invention and treatment status in the list experiment. While list experiments often use simple difference-in-means estimators, it can be useful to use regression in order to control for covariates (Imai 2011). In favor of simplicity, this study opts for a linear regressor with an interaction between original intervention and list experiment treatment (Blair and Imai 2017). This allows us to estimate not only the intent-to-treat effect, but also the local average treatment effect, similar to the estimation strategy used for other indicators.

Outliers and indicators with limited variation were excluded from the final sample. In order to exclude outliers, indicators were winsorized all continuous non-negative indicators at 99 percent at the top-end. In addition, indicators were tested for limited variation as determined by the pre-analysis plan. This implied that questions for which 95 percent of observations have the same value within the relevant sample were omitted from the analysis. This resulted in the exclusion of only 6 indicators.[244]

The exploratory nature of this study makes it necessary to test a large number of outcomes. However, testing a multitude of hypotheses makes it more likely to identify an effect in any one of the outcomes. To control for this type of bias the study uses two approaches. First, the number of tested hypotheses is reduced by summarizing outcome variables into grouped indices. To create indices, we combine indicators related to each primary group of outcomes, by creating standardized indexes following a method championed by Haushofer and Shapiro (2016). The indexes consist of a weighted average of a number of standardized outcome variables within a outcome group, e.g. employment or consumption. Weights are calculated by the inverse of the covariance matrix of outcomes within one group. This approach maximizes the variance of the final index (See Appendix 5). It allows us to keep the number of outcome variables low and allow for greater statistical power. Since combining individual outcome variables in indexes as described above still leaves multiple groups or families of key outcomes of interests, regressions also report p-values adjusted by false discovery rate following the two-step procedure introduced by Benjamini and Hochberg (1995). It controls for the expected proportion of rejections that are type I errors within a family of outcomes. The group of socio-economic and behavioral/psychological outcomes are employed as the two main families of outcomes.

In addition, we include an analysis on gender heterogeneity. We split the sample across gender and report all estimates for both subsamples separately. We also test whether male and female point estimates are significantly different by means of a Wald-test. The gender analysis can be found in Appendix 3.

[244] Indicators excluded due to limited variation are: Engagement in cattle raids and frequency of cattle raids, number of times having been beaten during the past month, in-kind payment for wage employment, remaining amount from a formal loan and remaining amount from an informal loan.

Results

Socio-economic outcomes

The intervention had no effect on employment. Study participants showed no positive improvement in the employment index on average (Table C7-6). What is more, none of the estimators that differentiate between the two ex post treatments reach conventional levels of statistical significance. When looking at gender heterogeneity, the treatment-on-the-treated estimates show a statistically significant improvement of the employment indicator of about 0.3 standard deviations for female participants over male participants in either treatment group (Table C7-21). These findings are consistent with the idea that the 1-week business skills training – in which both treatment groups participated – was particularly effective for women. However, the 2SLS results are weaker. Although the coefficients are large and positive for both female ex post treatment groups, large standard errors keep the estimates from reaching the 10 percent significance level.

For the consumption indicator we find no impact, on aggregate (Table C7-5). Once we control for selection into the two ex post treatments, we find a large (up to 1 standard deviation) and statistically significant increase in consumption for participants that received both training and grant. This effect matches findings of existing literature on cash grants that typically find large consumption effects (e.g. Manley, Gitter et al. 2013, Haushofer and Shapiro 2016). The pattern is consistent across genders and point estimates are not significantly different between male and female participants (Table C7-20). Furthermore, participants that only received the training, but expected the grant also, seem to have experienced small consumption declines relative to the control group. These findings suggest that grant money was partly used to boost consumption, while unexpectedly not receiving the grant had a negative impact on this measure of welfare.

The intention-to-treat estimates indicate a positive aggregate impact of the program on the savings indicator (Table C7-5). However, when we analyze the different ex post treatments, we can see a positive impact only for those who received training and grants. The effect is large at up to about 1 standard deviation and significant at the 1 percent level. The gender analysis shows the same trend for males and females (Table C7-20). Although males display a slightly larger point estimate, the difference is not significant. This finding is again in line with existing evidence on the effects of cash transfers (e.g. Banerjee, Duflo et al. 2015).

Business skills did not improve on aggregate (Table C7-5). The results provide some evidence that skills improved for those that were able to access the grants when calculated as a treatment-on-the-treated effect. When we control for self-selection through the LATE, the effect becomes close to zero and statistically insignificant. This suggests that business skills improved only for participants who had already a greater propensity to benefit from the intervention and were more likely to receive the grant due to their personal characteristics (e.g., higher business savviness).

Psychological and behavioral outcomes

Turning to psychological well-being, the results display no impact of the intervention on average (Table C7-5). Yet, when looking at the differences between the two ex post treatments, the LATE estimate shows a statistically significant increase of about 0.8 standard deviations in this indicator for participants that received both training and grant (Table C7-9). This result echoes new findings on the psychological benefits of cash transfers (Ozer, Fernald et al. 2011, Haushofer and Shapiro 2016). Importantly, we find no significant increase for participants that went to the training but were not able to access the grant (Table C7-9). While statistically insignificant both the TOT and the LATE estimates for this group show positive point estimates. Therefore, it appears very unlikely that the program negatively affected psychological well-being as suggested by literature on relative economic status and well-being (Luttmer 2005). Possibly participants of the "training and grant" group still perceived themselves relatively well-off compared to their peer-group because they had the chance to participate in the business-and-life skills training. We also find no significant difference between genders (Table C7-22).

For risk preferences, the results draw a similar picture. While there is no effect on risk preferences on average (Table C7-6), according to the LATE estimates risk preferences increased by about 0.7 standard deviations for participants that got training and grant (Table C7-9). Still, this result should be interpreted cautiously, since it reaches only the 10 percent significance level. Again, there seems to be no large difference in the effect on male or female participants (Table C7-22). The trust indicator is another indicator where we find a negative impact of the program cancellation. Although there is no effect on trust on the average treatment group participant (Table C7-6), participants who received the training, but were not able to access the grant, show a reduction in trust by about 0.5 standard deviation (Table C7-9). The effect seems to be driven by female participants. While male participants of both ex post treatments display positive insignificant effects, female participants who failed to access the grant show a highly significant reduction in their trust indicator by about 0.9 standard deviation (Table C7-22). Surprisingly, women who also received the grant as well as the training display a negative LATE estimate, although this result is only significant at the 10 percent level. This empirical finding is consistent with two alternative theoretical explanations. First, this could echo the findings of Gawn and Innes (2018) from the lab that the experience of being lied to erodes trust. An alternative explanation suggested by the literature would be that cash transfers put women at increased risk of violent threats which in turn reduces their general trust level. The latter explanation could also account for the fact that reductions in trust levels were experienced by both women who received the grant and those who failed to receive it.

The effect on crime and violence is complex. On average, the results show a weakly significant negative effect on the crime and violence indicator (Table C7-6), indicating lower levels of vulnerability to crime and violence and lower participation in security groups. Turning to the two ex post treatments, some of the LATE estimates suggest a weakly significant negative effect on participants that received the training, but not the grant (Table C7-9). This pattern can also be found in the female sub-sample, not, however, in the male (Table C7-22). Yet, the most rigorous specification fails to reach the 10 percent significance level. On net, it does not seem that the intervention had impacts on these measures of crime and violence.

We find no significant effect on migration propensity either on average (Table C7-6) or among both ex post treatment groups (Table C7-9). For female participants, we find some evidence that those who failed to access the grant are slightly less likely to migrate (Table C7-22). Again, the most rigorous specification, however, falls short of reaching the 10 percent significance level.

Finally, turning to results of the list experiment, there seems to be on average a slight increase in the propensity of cattle raiding, but not in aggressive arguments (Table C7-6). In particular, the increase in cattle raiding is prominent among the group that did not receive the grant (Table C7-9). However, after controlling for self-selection into this group through the LATE estimates, the effect becomes smaller and loses statistical significance. It is likely the effect is not causally driven by the disappointment of not receiving the grant or whether this group was initially more likely to engage in cattle raiding. When observing differences across genders we even find weak evidence that cattle raiding increased among men that received both the training and the grant (Table C7-22). In contrast, aggressive arguments seem to have fallen among participants that received only the training, but not the grant (Table C7-9). Altogether, it is difficult to determine a strong impact of the intervention on the propensity to engage in cattle raiding or aggressive arguments.

Robustness

Since our study showed some degree of attrition, we test the robustness of our ITT estimates by calculating upper and lower bounds. These correct for attrition by making the extreme assumption about missing information. We report the results in Appendix 2. Even after extreme assumptions about attritors the ITT effect on the savings, investment and debt index remains statistically significant – overall, savings increased among those assigned to receive training and grants by about 0.3 standard deviations. In contrast, the crime and violence index loses its significance when assuming that attritors reached higher values of the outcome distribution. Nevertheless, the effect size does not change much which makes us confident that estimated reduction is not an artifact of the data selection.

In addition, we address potential attrition bias by re-weighting observations based on their likelihood to be included in the final sample. Control group participants that were reached during the phone survey had an 82 percent likelihood to be reached for the final survey, while control group participants that were not reached during the phone survey had only a 46 percent likelihood to be reached for the final survey. Likelihoods to be reached in the final survey differ similarly for the treatment group. We thus attach sampling weights to all observations based on the inverse likelihood to be successfully interviewed for the final survey. Results are reported in Table C7-15 and following.

All intention-to-treat estimates proof robust to our re-weighing exercise. The weighted regressions confirm a positive average effect on the savings, investment and debt index, a negative effect on the likelihood to be vulnerable to crime and violence, and a positive effect on the list experiment on cattle raiding.

Turning to results that distinguish between “training and grant” and “training, but no grant” largely confirm our main results. Our main results show positive effects on consumption and savings for participants that received training and grant, and this finding is confirmed in the weighted regressions. What is more, the weighted regressions confirm our positive finding on the effect of grant and training on psychological well-being. For participants that only received the training, but not the grant, we can also confirm that estimates on trust, on crime and violence and on the list experiment regarding aggressive arguments show a weakly significant negative effect.

Conclusion

Our study used the example of the unplanned cancellation of the South Sudan Youth Business Start-Up Grant Program to evaluate the impacts of interventions that fail to be implemented as planned. Overall, our results suggest that the impact of failed interventions is mixed and depends on the gender of participants and their ex post treatment status. In this instance, on average across all participants, the intervention was largely ineffective. Most socio-economic or psychological and behavioral indicators neither worsened nor improved.

However, when considering ex post treatment groups and gender, some groups were detrimentally affected by the intervention. In particular, female participants that had expected to receive the cash grant but did not due to the cancellation of the program showed a strong reduction in their trust level. We also found some evidence that these women were less likely to migrate. Given that large shares of the population in South Sudan migrated in the period of our analysis to escape conflict affected areas, it is possible that women that expected a grant stayed back in expectation of the grant that would have migrated in the absence of the intervention. While we do not have direct information on this unintended consequence, one could be concerned regarding the potential detrimental outcomes to these program participants.

Where positive impacts were detected, for example on savings and on consumption, these tended to accrue to those that received the grants. Although the group that received the grant was smaller than the group that only received training, the positive impacts on the savings indicator was large enough to lift the average effect above a statistically significant level, but not for the consumption indicator. In addition, psychological well-being improved for the group receiving both training and grants. These positive effects seemed to be independent of gender.

The most unexpected result is the reported reduction of crime and violence experienced by women who did not receive the grant. Equally puzzling is the reduction in aggressive arguments among both men and women of the group who did not receive the grant money. Potentially this finding is due to a reporting bias on this indicator. For example, those who did not receive the grant but had expected to, became more wary about reporting on sensitive events, given that they may have perceived the program to be less responsive to their needs and vulnerabilities.

This paper is the first study that shows how failed intervention can have a negative impact on intended beneficiaries. While we did not identify clear socio-economic disadvantages for participants that vainly expected to receive the grant money, the negative impact on female trust levels and migration behavior should warn policy makers to pay more attention to unintended damage from failed interventions. Since the main negative effect appears only for the female subgroup, the external validity of the result should be confirmed by further research on failed inventions and heterogenous effects across gender. Although most indicators showed no significant net improvements, participants who did receive the treatment as intended seemed to benefit economically and psychologically. While it remains to be argued whether these positive impacts outweigh the negative impacts on participants who did not receive the complete treatment, our study makes it clear that interventions should consider the consequences of potential failure in the planning stages. For example, future interventions in risky environments might want to explicitly flag the potential of a program cancellation to pro-actively mitigate against trust loss.

Supplemental Tables

Table C7-1: Main outcomes of interest

	Outcomes Name	Details
		Socio-economic outcomes – survey based
1	Employment index	Standardized weighted average of the number of hours spend on wage employed activities in the past 7 days, (log) cash wage received in the past 7 days, (log) outstanding wage from the past 7 days, (log) total wage in past 7 days, number of activities on wage employment in the past 7 days, number of hours spend on self-employed activities in past 7 days, (log) self-employed cash earnings in the past 7 days, (log) self-employed in-kind earnings in the past 7 days, (log) outstanding earnings from the past 7 days, (log) total self-employed earnings in the past 7 days, number of self-employed activities in the past 7 days, total number of employees, (log) business revenue during the past 4 weeks, (log) business sales yesterday, (log) aggregated business costs in the past 4 weeks
2	Consumption index	Standardized weighted average of the number of different food items consumed in the past 7 days, (log) total food expenditure in the past 7 days, (log) value of self-produced food in the past 7 days, (log) expenditure on non-food items in past 1month, (log) expenditure on assets in past 1 month

3	Savings, investment and debt index	Standardized weighted average of having or sharing a formal bank account, currently saving any money, (log) amount held at bank account, (negatively coded) number of formal loans received, (negatively coded) other debt, (negatively coded) number of informal loans received in the past 1 month, (negatively coded) (log) total amount of formal loans, (negatively coded) (log) total amount of informal loans, business ownership, participation in training during the past 12 months, number of trainings done in the past 12 months
4	Business skills index	Standardized weighted average of frequency of visiting competitors, frequency of asking customers about other products they would like to be sold, frequency of setting sales targets, frequency of comparing targets to performance, frequency of recording purchase and sales, knowledge of the business register, knowledge of fees to register a business at cashier's office of the Business Register, knowledge of operating license from State government, knowledge of inspections from payam authorities, knowledge of taxes, knowledge of bribes (rashua), knowledge of paying an intermediate person to take care of taxes, registration of company name at business register, registration at cashier's office of the Business Register, obtainment of operation license from the State government, experienced inspection by payam authorities, payment of formal taxes, payment of bribes (rashua), payment of intermediary person to take care of taxes
		Psychological and behavioral outcomes
5	Psychological wellbeing index	Standardized weighted average of happiness with education level, with family, with job and work, with earnings or income, with house they live in, with life as a whole, with community they live in, with security and with friends, ladder of life rating self now, ladder of life rating household now, ladder of life rating self in 5 years, ladder of life rating household in 5 years, internal locus of control score on the possibility to become a leader based on ability, on general events in life, on influencing the number of friends, on control over future events, on feeling protected, on planning ahead, on pleasing people above to get ahead, on (negatively coded) dependence on luck to become a leader, on working hard to get ahead, on the belief that own actions matter most, empowered decisions on food/clothing purchases for children, on opening a business, on taking a loan, on visiting a friend, on traveling to another town, on staying overnight at another town, on getting a child vaccinated, on purchasing small items, on paying school fees for relatives
6	Risk index	Standardized weighted average of (negatively coded) likelihood of sleeping under a mosquito net, likelihood to walk alone at night, (negatively coded) likelihood to spend an afternoon waiting for a medical exam, likelihood to take a boda boda if the driver is unknown, likelihood to engage in unprotected sex, (negatively coded) likelihood to invest in a safe business accepting low profits, likelihood to invest into a business that has high profits but equal chance of failing, likelihood to take a loan if there were no restrictions, experimental data on number of times the more risky lottery was chosen

7	Trust index	Standardized weighted average of 13 trust items: trust to people in general, trust that people are helpful, (negatively coded) belief that people seek their own advantage, willingness to lend money, willingness to lend possessions, trust in family, trust in friends, trust in neighbors, trust in police, trust in NGO, trust in elders, trust in local government, trust in state government, experimental data on amount send to the WB in trust game and amount send to another player in the trust game
8	Crime and violence index	Standardized weighted average of participation in a security group, frequency of participation in a security group, hours participated in a security group last week, experience of own cattle been stolen, number of times own cattle had been stolen in the past 1 year, knowledge of a least 1 home/market stall robbery, number of known home/market stall robberies, experience of harassment during past 1 month, number of times been harassed during past 1 month, experience of having been physically punished or beaten, feeling concerned that receiving money might foster crime or violence
10	Migration index	Standardized weighted average of having moved since baseline, living outside SSD in the past 1 year, living in a refugee camp in the past 1 year, living in an IDP camp in the past 1 year, having the wish to move
11	List experiment cattle index	Standardized average of the two list experiment questions on cattle raiding
12	List experiment argument index	Standardized average of the two list experiment questions on arguments

Table C7-2: Balancing original control and treatment group at baseline

		Control group		ITT group		Difference	p-value
		N	Mean	N	Mean	in means	
Individual and household characteristics							
Age		1,148	27.417	1,144	27.683	0.265	0.2001
Gender		1,148	0.602	1,144	0.611	0.009	0.6559
Married		1,148	0.666	1,143	0.649	-0.016	0.4103
Employment status		1,148	0.612	1,144	0.624	0.012	0.5626
Business ownership		1,148	0.642	1,144	0.659	0.017	0.3907
Consumption food		1,148	5.330	1,144	5.400	0.070	0.1740
Consumption nonfood		1,148	2.418	1,144	2.429	0.010	0.8547
Formal bank account		1,148	0.373	1,144	0.369	-0.004	0.8452
(Log) amount formal loans		1,139	-0.332	1,141	-0.367	-0.036	0.6339
(Log) amount informal loans		1,134	-1.329	1,124	-1.225	0.104	0.4432
Education level	No education	1,148	0.191	1,144	0.206	0.016	0.3517
	Some Primary	1,148	0.315	1,144	0.330	0.015	0.4401
	Some Secondary	1,148	0.404	1,144	0.373	-0.031	0.1289
	Some University or Higher	1,148	0.090	1,144	0.090	0.000	0.9791
Literacy	No English	1,148	0.247	1,144	0.263	0.016	0.3882
	Some English	1,148	0.273	1,144	0.295	0.022	0.2443
	Good English	1,148	0.480	1,144	0.442	-0.038*	0.0706
Numeracy	Low	1,148	0.238	1,144	0.247	0.010	0.5931
	Medium	1,148	0.160	1,144	0.198	0.037**	0.0199
	High	1,148	0.602	1,144	0.555	-0.047**	0.0231
Household size		1,148	7.310	1,144	7.260	-0.050	0.7257
Number of children		1,148	3.107	1,144	3.241	0.134	0.1635
Number of elderly		1,148	0.109	1,144	0.087	-0.021	0.1292
Number of rooms		1,148	3.180	1,144	3.087	-0.093	0.1935
Number of buildings		1,148	3.676	1,144	3.538	-0.138*	0.0830
(Log) distance to KCB branch		1,130	2.395	1,126	2.396	0.001	0.9871
Conflict exposure 2011-2014 (300km buffer)		1,148	0.000	1,144	0.084	0.084	0.3283
Conflict exposure 2015-2017 (300km buffer)		1,148	0.000	1,144	0.128	0.128*	0.0953
State at baseline							
Central Equatoria		1,148	0.169	1,144	0.167	-0.002	0.8966
Eastern Equatoria		1,148	0.160	1,144	0.152	-0.008	0.5898
Lakes		1,148	0.158	1,144	0.159	0.001	0.9256
Northern Bahr El Ghazal		1,148	0.170	1,144	0.176	0.006	0.7118
Western Bahr El Ghazal		1,148	0.172	1,144	0.171	-0.000	0.9861
Western Equatoria		1,148	0.172	1,144	0.175	0.003	0.8386

*Note: All indicators were measured at baseline. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Table C7-3: Balancing between "training, no grant" vs "training and grant"

		„training, no grant“		„training and grant“		N
		Mean	SD	Coeff.	SE	
Individual and household characteristics						
Age		27.570	4.691	5.594***	1.280	626
Married		0.606	0.489	0.160***	0.053	626
Employment status		0.656	0.476	0.134**	0.058	626
Business ownership		0.642	0.480	0.215***	0.049	626
Consumption food		5.390	1.150	0.908***	0.216	626
Consumption nonfood		2.398	1.322	0.676***	0.137	626
Formal bank account		0.421	0.494	0.137***	0.047	626
(Log) amount formal loans		-0.338	1.756	-0.140	0.171	625
(Log) amount informal loans		-0.972	2.892	-0.522*	0.267	614
Education level	No education	0.173	0.379	-0.056	0.044	626
	Some Primary	0.308	0.462	0.078*	0.044	626
	Some Secondary	0.399	0.490	0.164***	0.046	626
	Some University or Higher	0.120	0.326	-0.002	0.029	626
Literacy	No English	0.233	0.423	-0.080*	0.042	626
	Some English	0.269	0.444	0.131***	0.041	626
	Good English	0.498	0.501	0.133**	0.056	626
Numeracy	Low	0.192	0.395	-0.022	0.036	626
	Medium	0.216	0.412	-0.006	0.040	626
	High	0.591	0.492	0.212***	0.058	626
Household size		7.058	3.215	1.648***	0.508	626
Number of children		3.171	2.239	0.628**	0.284	626
Number of elderly		0.072	0.332	0.039	0.036	626
Number of rooms		3.240	1.698	0.533***	0.191	626
Number of buildings		3.639	2.029	0.783***	0.293	626
(Log) distance to KCB branch		2.749	2.089	0.078	0.170	617
Conflict exposure 2011-2014 (300km buffer)		0.208	4.393	-0.149	0.166	626
Conflict exposure 2015-2017 (300km buffer)		0.136	3.728	-0.195	0.157	626
State at baseline						
Central Equatoria		0.188	0.391	0.019**	0.009	626
Eastern Equatoria		0.240	0.428	0.000***	0.000	626
Lakes		0.063	0.242	0.000***	0.000	626
Northern Bahr El Ghazal		0.125	0.331	0.178***	0.043	626
Western Bahr El Ghazal		0.091	0.288	-0.004	0.004	626
Western Equatoria		0.293	0.456	-0.009*	0.006	626

Note: Differences between treatment group participants that received that grant and those who did not use baseline characteristics. Column (1) reports mean values of baseline covariates for participants that received training but no grant". Column (2) reports OLS estimates on receiving "training and grant" and strata fixed effect. Standard errors are clustered at boma level and reported below coefficients in parenthesis. (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level."*

Table C7-4: Summary statistics of outcome variables for the control group

VARIABLES	N	mean	sd	min	max
Main outcomes (survey-based)					
Employment index	763	0	1	-2.314	6.401
Consumption index	763	0	1	-1.580	5.037
Savings, investment and debt index	763	0	1	-4.013	2.984
Business skills index	763	0	1	-2.971	2.569
Psychological wellbeing index	763	0	1	-2.625	3.606
Risk index	763	0	1	-2.789	3.142
Trust index	763	0	1	-2.982	3.147
Crime and violence index	763	0	1	-1.214	5.667
Migration index	763	0	1	-0.838	3.740
List experiment cattle index	763	0	1	-3.360	3.095
List experiment argument index	763	0	1	-3.666	4.163

Note: Higher values of all indicators refer to higher scores in the respective outcome. For instance, higher values in the risk index imply a higher preference for risky behavior. Higher values in the list experiment cattle index imply a higher propensity to engage in cattle raiding, while higher values in the list experiment argument index imply a higher propensity to engage in arguments. Higher values of the migration index mark a higher propensity of having, or planning to migrate.

Table C7-5: ITT effects of the original intervention on main socio-economic outcomes

	(1) ITT (no controls)	(2) ITT (controls)
Main outcomes – Socioeconomic		
Employment index	0.063	0.067
	(0.281)	(0.242)
	[0.375]	[0.323]
Consumption index	0.094	0.086
	(0.120)	(0.153)
	[0.240]	[0.307]
Savings, investment and debt index	0.274***	0.271***
	(0.000)	(0.000)
	[0.001]	[0.001]
Business skills index	0.016	0.018
	(0.747)	(0.735)
	[0.748]	[0.735]
Observations	1,523	1,495

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg p-values are reported in square brackets. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline.*

Table C7-6: ITT effects of the original intervention on main psychological and behavioral outcomes

	(1) ITT (no controls)	(2) ITT (controls)
Main outcomes – Psychological and behavioral		
Psychological wellbeing index	-0.009	0.002
	(0.845)	(0.965)
	[0.845]	[0.965]
Risk index	-0.043	-0.052
	(0.501)	(0.383)
	[0.692]	[0.537]
Trust index	-0.035	-0.055
	(0.482)	(0.274)
	[0.692]	[0.480]
Crime and violence index	-0.080	-0.089*
	(0.119)	(0.090)
	[0.343]	[0.315]
Migration index	-0.026	-0.015
	(0.593)	(0.767)
	[0.692]	[0.896]
List experiment cattle index	0.172*	0.169**
	(0.075)	(0.050)
	[0.343]	[0.315]
List experiment argument index	-0.135	-0.132
	(0.147)	(0.149)
	[0.343]	[0.322]
Observations	1,523	1,495

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg p-values are reported in square brackets. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict between baseline and endline.*

Table C7-7: First stage results from LATE estimation of Table 8 and Table 9

		(1)	(2)	(3)	(4)	(5)	(6)
		„Training, no grant“	„Training and grant“	„Training, no grant“	„Training and grant“	„Training, no grant“	„Training and grant“
Instrument 1	Treatment	0.4226***	0.3860***	0.4196***	0.3875***	0.4414***	0.4254***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)
Instrument 2	Treatment x (log) distance to KCB branch	0.0517***	-0.0450***	0.0523***	-0.0442***	0.0716***	-0.0620***
		(0.002)	(0.001)	(0.003)	(0.001)	(0.000)	(0.000)
	(log) Distance to KCB branch	-0.0032	0.0050	-0.0050	0.0081	-0.0093	0.0143
		(0.661)	(0.454)	(0.549)	(0.261)	(0.418)	(0.107)
	Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
	Individual controls	No	No	Yes	Yes	Yes	Yes
	Geography controls	No	No	No	No	Yes	Yes
	Observations	1,500	1,500	1,474	1,474	1,474	1,474

*Note: This table displays the first stage results for LATE estimates of Table 11. Columns (1) and (2) correspond to LATE estimates of column (3) in Table 11. Column (3) and (4) correspond to LATE estimates in column (4) in Table 11 and columns (5) and (6) to column (5) respectively. We report the effect of our two instrumental variables – original assignment to the treatment group and its interaction with distance to the closest KCB bank branch – on our two main regressors of interest. All regression control for gender-state fixed effects and for the level effect of distance to the closest KCB bank branch. Control variables of column (3)-(6) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between endline and baseline. Column (5) and (6) also controls for geographic features since the estimation strategy relies on the distance to the closest KCB bank branch, which might correlate with other geographic characteristics. Geography controls include distance to the closest city, distance to the closest road, average land gradient and their interactions with selection to the original treatment group, and the interaction of conflict exposure with the original treatment group. P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level*

Table C7-8: Effects of "training and grant" vs "training, but no grant" on socio-economic outcomes

		(1)	(2)	(3)	(4)	(5)	(6)
		TOT (no controls)	TOT (controls)	TOT (controls + geography controls)	LATE (no controls)	LATE (controls)	LATE (controls + geography controls)
Main outcomes – Socio-economic							
Employment index	Training, no grant	0.087	0.086	0.081	-0.069	-0.050	0.064
		(0.149)	(0.134)	(0.424)	(0.766)	(0.818)	(0.833)
		[0.238]	[0.215]	[0.679]	[0.968]	[0.988]	[0.989]
	Training and grant	0.057	0.062	0.044	0.369	0.338	0.082
		(0.580)	(0.554)	(0.752)	(0.384)	(0.391)	(0.820)
		[0.595]	[0.676]	[0.813]	[0.655]	[0.626]	[0.989]
Consumption index	Training, no grant	0.046	0.037	0.026	-0.389**	-0.350**	-0.136
		(0.489)	(0.591)	(0.810)	(0.019)	(0.029)	(0.659)
		[0.595]	[0.676]	[0.813]	[0.071]	[0.077]	[0.989]
	Training and grant	0.178**	0.157**	0.169	1.042**	0.933**	1.049**
		(0.023)	(0.048)	(0.166)	(0.027)	(0.027)	(0.015)
		[0.046]	[0.096]	[0.332]	[0.071]	[0.077]	[0.060]
Savings, investment and debt index	Training, no grant	0.221***	0.205***	0.127	-0.166	-0.171	-0.200
		(0.000)	(0.000)	(0.147)	(0.275)	(0.278)	(0.483)
		[0.001]	[0.001]	[0.332]	[0.572]	[0.556]	[0.989]
	Training and grant	0.434***	0.420***	0.327***	1.282***	1.270***	0.992***
		(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.007)
		[0.001]	[0.001]	[0.008]	[0.006]	[0.005]	[0.054]
Business skills index	Training, no grant	-0.031	-0.022	0.024	-0.113	0.003	0.024
		(0.594)	(0.727)	(0.813)	(0.520)	(0.988)	(0.929)
		[0.595]	[0.728]	[0.813]	[0.974]	[0.988]	[0.989]
	Training and grant	0.240***	0.220***	0.299**	0.267	0.046	-0.005
		(0.004)	(0.010)	(0.017)	(0.442)	(0.903)	(0.988)
		[0.012]	[0.027]	[0.069]	[0.968]	[0.988]	[0.989]
Observations		1,523	1,495	1,474	1,500	1,474	1,474
F-stat					23.88	21.61	36.61

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg p-values are reported in square brackets. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. Column (3) also controls for geographic features since the estimation strategy relies on distance to the closest KCB bank branch which might correlate with other geographic characteristics. Geography controls include distance to the closest city, distance to the closest road, average land gradient and their respective interactions with selection to the original treatment group.*

Table C7-9: Effects of "training and grant" vs "training, but no grant" on other outcomes.

		(1) TOT (no controls)	(2) TOT (controls)	(3) TOT (controls + geography controls)	(4) LATE (no controls)	(5) LATE (controls)	(5) LATE (controls + geography controls)
				Main outcomes – Psychological and behavioral			
Psychological wellbeingindex	Training, no grant	0.029	0.035	0.127	-0.238	-0.080	0.196
		(0.585)	(0.490)	(0.139)	(0.151)	(0.614)	(0.373)
		[0.745]	[0.624]	[0.324]	[0.389]	[0.749]	[0.746]
	Training and grant	0.027	-0.014	0.064	0.397	0.131	0.701**
		(0.716)	(0.847)	(0.505)	(0.229)	(0.672)	(0.027)
		[0.795]	[0.913]	[0.708]	[0.389]	[0.749]	[0.372]
Risk index	Training, no grant	0.016	0.000	0.106	-0.441	-0.408	-0.109
		(0.839)	(0.998)	(0.322)	(0.103)	(0.104)	(0.794)
		[0.840]	[0.998]	[0.574]	[0.389]	[0.369]	[0.955]
	Training and grant	-0.068	-0.076	0.028	0.702	0.605	0.712
		(0.365)	(0.327)	(0.780)	(0.194)	(0.231)	(0.106)
		[0.640]	[0.464]	[0.840]	[0.389]	[0.462]	[0.372]
Trust index	Training, no grant	-0.077	-0.096	-0.038	-0.020	-0.020	-0.501*
		(0.182)	(0.105)	(0.740)	(0.920)	(0.923)	(0.064)
		[0.365]	[0.378]	[0.840]	[0.921]	[0.924]	[0.372]
	Training and grant	0.128	0.131	0.211*	-0.098	-0.153	-0.021
		(0.122)	(0.111)	(0.095)	(0.792)	(0.695)	(0.955)
		[0.365]	[0.378]	[0.324]	[0.853]	[0.749]	[0.955]
Crime and violence index	Training, no grant	-0.051	-0.061	0.010	-0.470*	-0.554*	-0.277
		(0.414)	(0.331)	(0.905)	(0.100)	(0.062)	(0.361)
		[0.645]	[0.464]	[0.906]	[0.389]	[0.369]	[0.746]
	Training and grant	-0.104	-0.103	-0.089	0.578	0.692	0.155
		(0.170)	(0.190)	(0.356)	(0.250)	(0.185)	(0.669)
		[0.365]	[0.378]	[0.574]	[0.389]	[0.432]	[0.937]
Migration index	Training, no grant	-0.080	-0.078	-0.148*	-0.258	-0.292	-0.307
		(0.150)	(0.167)	(0.098)	(0.119)	(0.105)	(0.251)
		[0.365]	[0.378]	[0.324]	[0.389]	[0.369]	[0.703]
	Training and grant	0.029	0.018	-0.053	0.449	0.543	-0.043
		(0.738)	(0.823)	(0.594)	(0.223)	(0.157)	(0.890)
		[0.795]	[0.913]	[0.757]	[0.389]	[0.432]	[0.955]
List experiment cattle index	Training, no grant	0.207*	0.222**	0.250**	0.108	0.112	0.166
		(0.052)	(0.040)	(0.026)	(0.666)	(0.688)	(0.558)
		[0.365]	[0.378]	[0.324]	[0.777]	[0.749]	[0.904]
	Training and grant	-0.089	-0.103	-0.080	0.410	0.392	0.336
		(0.565)	(0.216)	(0.369)	(0.457)	(0.519)	(0.581)
		[0.745]	[0.378]	[0.574]	[0.582]	[0.749]	[0.904]
List experiment argument index	Training, no grant	-0.157	-0.130	-0.176*	-0.431*	-0.352*	-0.334*
		(0.112)	(0.191)	(0.082)	(0.060)	(0.084)	(0.089)
		[0.365]	[0.378]	[0.324]	[0.389]	[0.369]	[0.372]
	Training and grant	-0.224	-0.256	-0.245	0.400	0.217	0.099
		(0.148)	(0.121)	(0.130)	(0.432)	(0.662)	(0.838)
		[0.365]	[0.378]	[0.324]	[0-582]	[0.749]	[0.955]
Observations		1,523	1,495	1,474	1,500	1,474	1,474
F-stat					23.88	21.61	36.61

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg p-values are reported in square brackets. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. Column (3) also controls for geographic features since the estimation strategy relies on the distance to the closest KCB bank branch that might correlate with other geographic characteristics. Geography controls include distance to the closest city, distance to the closest road, average land gradient and their respective interactions with selection to the original treatment group.*

Appendix 1 – Additional Balance Tables

Table C7-10: Attrition - Difference in attrition probability between original treatment and control group

	Control mean (SD)	Treatment	N
Attrition	0.335	0.002	2,292
	(0.472)	(0.018)	

*Note: Difference in attrition probability between original treatment vs. control group, estimated with an OLS regression of the attrition dummy on the treatment dummy and strata fixed effects. The standard error of the treatment dummy is clustered at boma level and reported in parentheses. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Table C7-11: Attrition - Baseline difference between attritors and non-attritors

		Non-attritors		Attritors		N
		Mean	SD	Coeff.	SE	
Individual and household characteristics						
Age		27.632	4.826	-0.281	0.236	2,292
Married		0.661	0.473	-0.025	0.026	2,291
Employment status		0.619	0.486	0.004	0.020	2,292
Business ownership		0.649	0.478	0.011	0.018	2,292
Consumption food		5.405	1.170	-0.108**	0.051	2,292
Consumption nonfood		2.432	1.325	0.004	0.063	2,292
Formal bank account		0.397	0.489	-0.068***	0.021	2,292
(Log) amount formal loans		-0.290	1.626	-0.183**	0.088	2,280
(Log) amount informal loans		-1.360	3.323	0.275**	0.132	2,258
Education level	No education	0.210	0.408	-0.035*	0.019	2,292
	Some Primary	0.307	0.462	0.059***	0.019	2,292
	Some Secondary	0.379	0.485	0.019	0.022	2,292
	Some University or Higher	0.104	0.305	-0.042***	0.011	2,292
Literacy	No English	0.261	0.440	-0.012	0.020	2,292
	Some English	0.286	0.452	0.003	0.020	2,292
	Good English	0.453	0.498	0.009	0.023	2,292
Numeracy	Low	0.252	0.434	-0.026	0.018	2,292
	Medium	0.173	0.378	0.028	0.017	2,292
	High	0.575	0.494	-0.002	0.020	2,292
Household size		7.384	3.342	-0.299**	0.144	2,292
Number of children		3.248	2.294	-0.211**	0.104	2,292
Number of elderly		0.098	0.344	-0.002	0.014	2,292
Number of rooms		3.179	1.691	-0.125	0.078	2,292
Number of buildings		3.611	1.989	-0.016	0.077	2,292
(Log) distance to KCB branch		2.338	1.938	0.227*	0.132	2,256
Conflict exposure 2011-14 (300km buffer)		0.074	2.427	-0.083*	0.049	2,292
Conflict exposure 2015-2017 (300km buffer)		0.083	2.139	-0.056	0.057	2,292
State at baseline						
Central Equatoria		0.171	0.376	0.008	0.007	2,292
Eastern Equatoria		0.154	0.361	-0.001	0.001	2,292
Lakes		0.147	0.354	0.001	0.001	2,292
Northern Bahr El Ghazal		0.175	0.380	-0.005	0.004	2,292
Western Bahr El Ghazal		0.171	0.376	-0.002	0.003	2,292
Western Equatoria		0.183	0.387	-0.001	0.003	2,292

*Note: Differences between attritors and non-attritors in baseline characteristics estimated by an OLS on the attrition dummy and strata fixed effects. Standard errors are clustered at boma level. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Table C7-12: Baseline difference between attritors from original control vs original treatment group

		Control group		ITT group		N
		Mean	SD	Coeff.	SE	
Individual and household characteristics						
Age		27.226	5.186	0.387	0.363	769
Married		0.670	0.471	-0.026	0.031	768
Employment status		0.644	0.479	-0.050	0.036	769
Business ownership		0.670	0.471	-0.022	0.031	769
Consumption food		5.223	1.334	0.136	0.086	769
Consumption nonfood		2.447	1.287	-0.043	0.099	769
Formal bank account		0.322	0.468	0.016	0.030	769
(Log) amount formal loans		-0.386	1.859	-0.176	0.148	765
(Log) amount informal loans		-1.017	2.913	-0.220	0.212	758
Education level	No education	0.190	0.393	-0.023	0.024	769
	Some Primary	0.340	0.474	0.034	0.031	769
	Some Secondary	0.410	0.493	-0.017	0.029	769
	Some University or Higher	0.060	0.237	0.006	0.015	769
Literacy	No English	0.249	0.433	-0.009	0.032	769
	Some English	0.249	0.433	0.062**	0.030	769
	Good English	0.501	0.501	-0.052	0.033	769
Numeracy	Low	0.231	0.422	-0.012	0.026	769
	Medium	0.190	0.393	0.008	0.030	769
	High	0.579	0.494	0.003	0.035	769
Household size		7.182	3.463	-0.143	0.258	769
Number of children		3.026	2.301	0.069	0.168	769
Number of elderly		0.117	0.360	-0.036	0.023	769
Number of rooms		3.091	1.784	-0.070	0.099	769
Number of buildings		3.670	1.836	-0.111	0.112	769
(Log) distance to KCB branch		0.174	0.380	-0.012	0.011	769
Conflict exposure 2011-2014 (300km buffer)		0.003	1.177	-0.051	0.061	769
Conflict exposure 2015-2017 (300km buffer)		-0.004	1.130	0.011	0.068	769
State at baseline						
Central Equatoria		0.169	0.375	-0.000	0.000	769
Eastern Equatoria		0.164	0.370	0.003	0.003	769
Lakes		0.174	0.380	0.008	0.007	769
Northern Bahr El Ghazal		0.148	0.356	0.000	0.004	769
Western Bahr El Ghazal		0.171	0.377	0.000	0.004	769
Western Equatoria		0.169	0.375	-0.000	0.000	769

*Note: Differences between the original control vs ITT group in baseline characteristics estimated by an OLS on the ITT group dummy and strata fixed effects. Standard errors are clustered at boma level. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Appendix 2 – Robustness Checks Tables

Table C7-13: Lee bounds for the ITT effects on main socio-economic outcomes

	(1) Lower bound	(2) Upper bound
Main outcomes – Socio-economic		
Employment index	0.045	0.047
	(0.610)	(0.810)
Consumption index	0.093	0.098
	(0.173)	(0.538)
Savings, investment and debt index	0.261**	0.268**
	(0.031)	(0.047)
Business skills index	0.007	0.009
	(0.942)	(0.926)
Observations	2292	

*Note: Outcome variables are listed on the left. Column (1) reports the lower bound. Column (2) reports the upper bound. P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Table C7-14: Lee bounds for the ITT effects on main psychological and behavioral outcomes

	(1) Lower bound	(2) Upper bound
Main outcomes (survey-based) – Psychological and behavioral		
Psychological wellbeing index	-0.005	-0.002
	(0.961)	(0.989)
Risk index	-0.052	-0.049
	(0.595)	(0.645)
Trust index	-0.055	-0.050
	(0.590)	(0.641)
Crime and violence index	-0.253***	-0.105
	(0.000)	(0.553)
Migration index	-0.027	-0.027
	(0.641)	(0.826)
Observations	2292	

*Note: Outcome variables are listed on the left. Column (1) reports the lower bound. Column (2) reports the upper bound. P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level*

Table C7-15: Weighted ITT effects of the original intervention on socio-economic outcomes.

	(1) ITT (no controls)	(2) ITT (controls)
Main outcomes – Socio-economic		
Employment index	0.065	0.075
	(0.285)	(0.211)
Consumption index	0.095	0.094
	(0.146)	(0.146)
Savings, investment and debt index	0.266***	0.265***
	(0.000)	(0.000)
Business skills index	0.012	0.018
	(0.814)	(0.744)
Observations	1523	1507

*Note: Observations are weighted by their inverse likelihood to be in the final sample, based on who was easy to reach during the phone survey. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Table C7-16: Weighted ITT effects of the original intervention on other outcomes.

	(1) ITT (no controls)	(2) ITT (controls)
Main outcomes – Psychological and behavioral		
Psychological wellbeing index	-0.036	-0.022
	(0.476)	(0.646)
Risk index	-0.054	-0.062
	(0.394)	(0.292)
Trust index	-0.013	-0.033
	(0.811)	(0.548)
Crime and violence index	-0.110**	-0.119**
	(0.029)	(0.023)
Migration index	-0.045	-0.036
	(0.363)	(0.482)
List experiment cattle index	0.215**	0.210**
	(0.034)	(0.037)
List experiment argument index	-0.125	-0.120
	(0.179)	(0.201)
Observations	1523	1495

*Note: Observations are weighted by their inverse likelihood to be in the final sample, based on who was easy to reach during the phone survey. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, and number of buildings at baseline. P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Table C7-17: Weighted TOT and ATE estimates on socio-economic outcomes

		(1) TOT (no controls)	(2) TOT (controls)	(3) TOT (controls + geography controls)	(4) LATE (no controls)	(5) LATE (controls)	(6) LATE (controls + geography controls)
				Main outcomes – Socio-economic			
Employment index	Training, no grant	0.090	0.095*	0.059	-0.011	0.011	-0.016
		(0.126)	(0.093)	(0.593)	(0.957)	(0.958)	(0.956)
	Training and grant	0.040	0.051	0.019	0.277	0.262	0.165
		(0.717)	(0.649)	(0.903)	(0.506)	(0.503)	(0.664)
Consumption index	Training, no grant	0.009	0.004	-0.060	-0.434**	-0.374**	-0.280
		(0.889)	(0.953)	(0.605)	(0.019)	(0.038)	(0.372)
	Training and grant	0.194**	0.174*	0.134	1.145**	1.017**	0.986**
		(0.029)	(0.052)	(0.340)	(0.026)	(0.027)	(0.021)
Savings, investment and debt index	Training, no grant	0.200***	0.186***	0.107	-0.194	-0.207	-0.283
		(0.000)	(0.001)	(0.203)	(0.219)	(0.218)	(0.327)
	Training and grant	0.460***	0.444***	0.360***	1.349***	1.356***	1.161***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
Business skills index	Training, no grant	-0.038	-0.038	-0.025	-0.122	0.002	0.098
		(0.524)	(0.524)	(0.690)	(0.528)	(0.993)	(0.727)
	Training and grant	0.294***	0.294***	0.274***	0.275	0.050	0.004
		(0.000)	(0.000)	(0.001)	(0.453)	(0.901)	(0.991)
Observations		1,500	1,495	1,474	1,500	1,474	1,474
F-stat					20.62	18.22	34.41

*Note: Observations are weighted by their inverse likelihood to be in the final sample, based on who was easy to reach during the phone survey. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. Column (3) also controls for geographic features since the estimation strategy relies on the distance to the closest KCB bank branch that might correlate with other geographic characteristics. Geography controls include distance to the closest city, distance to the closest road, average land gradient and their respective interactions with selection to the original treatment group, and the interaction between conflict exposure and the original treatment group. P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Table C7-18: Weighted TOT and ATE estimates of the on other outcomes

		(1) TOT (no controls)	(2) TOT (controls)	(3) TOT (controls + geography controls)	(4) LATE (no controls)	(5) LATE (controls)	(6) LATE (controls + geography controls)
		Main outcomes (survey-based) – Psychological and behavioral					
Psychological wellbeing index	Training, no grant	0.018	0.020	0.111	-0.285	-0.100	0.179
		(0.751)	(0.701)	(0.214)	(0.115)	(0.582)	(0.429)
	Training and grant	-0.005	-0.044	0.053	0.400	0.093	0.779**
		(0.945)	(0.554)	(0.578)	(0.268)	(0.789)	(0.017)
Risk index	Training, no grant	0.007	-0.014	0.056	-0.425	-0.381	-0.146
		(0.925)	(0.868)	(0.617)	(0.127)	(0.146)	(0.712)
	Training and grant	-0.097	-0.104	-0.025	0.625	0.508	0.639
		(0.211)	(0.187)	(0.806)	(0.264)	(0.328)	(0.129)
Trust index	Training, no grant	-0.059	-0.075	-0.014	-0.070	-0.077	-0.471*
		(0.337)	(0.240)	(0.903)	(0.752)	(0.740)	(0.082)
	Training and grant	0.174**	0.177**	0.253**	0.072	0.028	0.023
		(0.041)	(0.036)	(0.040)	(0.856)	(0.946)	(0.950)
Crime and violence index	Training, no grant	-0.094	-0.105*	-0.030	-0.514*	-0.603*	-0.278
		(0.118)	(0.091)	(0.712)	(0.097)	(0.067)	(0.408)
	Training and grant	-0.123*	-0.128*	-0.096	0.557	0.682	0.427
		(0.091)	(0.086)	(0.302)	(0.290)	(0.217)	(0.287)
Migration index	Training, no grant	-0.083	-0.085	-0.134	-0.246	-0.286	-0.397
		(0.125)	(0.128)	(0.113)	(0.183)	(0.154)	(0.160)
	Training and grant	-0.006	-0.018	-0.066	0.355	0.455	-0.029
		(0.947)	(0.821)	(0.473)	(0.369)	(0.272)	(0.935)
List experiment cattle index	Training, no grant	0.247**	0.257**	0.282**	0.108	0.143	0.182
		(0.021)	(0.031)	(0.020)	(0.666)	(0.627)	(0.533)
	Training and grant	-0.032	-0.044	-0.018	0.410	0.484	0.471
		(0.832)	(0.593)	(0.841)	(0.457)	(0.443)	(0.445)
List experiment argument index	Training, no grant	-0.134	-0.109	-0.147	-0.431*	-0.282	-0.291
		(0.190)	(0.301)	(0.168)	(0.060)	(0.188)	(0.160)
	Training and grant	-0.208	-0.232	-0.212	0.400	0.093	0.071
		(0.186)	(0.167)	(0.200)	(0.432)	(0.856)	(0.889)
Observations		1,500	1495	1474	1,500	1,474	1,474
F-stat					20.62	18.37	34.65

*Note: Observations are weighted by their inverse likelihood to be in the final sample, based on who was easy to reach during the phone survey. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. Column (3) also controls for geographic features since the estimation strategy relies on the distance to the closest KCB bank branch that might correlate with other geographic characteristics. Geography controls include distance to the closest city, distance to the closest road, average land gradient and their respective interactions with selection to the original treatment group, and the interaction of conflict exposure and the original treatment group. P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.*

Appendix 3 – Gender heterogeneity

Table C7-19: ITT effects of the original intervention on socio-economic outcomes by gender

	(1)	(2)	(3)	(4)	(5)
	ITT for males		**ITT for females**		Coefficient equality (2) vs (4)
	(no controls)	(controls)	(no controls)	(controls)	
Main outcomes – Socioeconomic					
Employment index	0.034	0.020	0.080	0.084	0.064
	(0.764)	(0.847)	(0.203)	(0.176)	(0.595)
Consumption index	0.056	0.028	0.116*	0.110	0.082
	(0.574)	(0.783)	(0.098)	(0.117)	(0.476)
Savings, investment and debt index	0.387***	0.349***	0.210***	0.209***	-0.140
	(0.000)	(0.000)	(0.001)	(0.001)	(0.167)
Business skills index	0.082	0.090	-0.022	-0.025	-0.114
	(0.263)	(0.242)	(0.736)	(0.716)	(0.241)
Observations	555	547	968	948	

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. Column (5) reports tests for coefficient equality between estimates from males and females in column (2) and (4). Displayed are differences of coefficient p-values of the test in parenthesis.*

Table C7-20: ITT effects of the original intervention on other outcomes by gender

	(1)	(2)	(3)	(4)	(5)
	ITT for males		ITT for females		Coefficient equality (2) vs (4)
	(no controls)	(controls)	(no controls)	(controls)	
Main outcomes – Psychological and behavioral					
Psychological wellbeing index	0.099	0.079	-0.071	-0.045	-0.125
	(0.204)	(0.306)	(0.216)	(0.414)	(0.190)
Risk index	0.004	-0.003	-0.069	-0.061	-0.058
	(0.960)	(0.964)	(0.391)	(0.430)	(0.565)
Trust index	0.038	-0.004	-0.076	-0.102	-0.098
	(0.653)	(0.963)	(0.235)	(0.127)	(0.407)
Crime and violence index	0.007	0.007	-0.129**	-0.152**	-0.159*
	(0.939)	(0.927)	(0.024)	(0.011)	(0.087)
Migration index	-0.050	-0.002	-0.013	-0.025	-0.022
	(0.478)	(0.975)	(0.845)	(0.716)	(0.834)
Main outcomes (experiments) – Psychological and behavioral					
List experiment cattle index	0.269*	0.258	0.108	0.117	-0.142
	(0.094)	(0.174)	(0.382)	(0.206)	(0.507)
List experiment argument index	0.140	0.135	-0.303**	-0.304**	-0.439**
	(0.378)	(0.419)	(0.017)	(0.014)	(0.038)
Observations	555	547	968	948	

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. Column (5) reports tests for coefficient equality between estimates from males and females in column (2) and (4). Displayed are differences of coefficient p-values of the test in parenthesis.*

Table C7-21: Effects of on socio-economic outcomes by gender

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
		TOT for males			**LATE for males**			**TOT for females**			**LATE for females**			Coeff equality (6) vs (12)
		(no controls)	(controls)	(controls + geo controls)	(no controls)	(controls)	(controls + geo controls)	(no controls)	(controls)	(controls + geo controls)	(no controls)	(controls)	(controls + geo controls)	
Employment index	Training, no grant	-0.025	-0.064	-0.416*	0.105	-0.001	-0.505	0.145**	0.152**	0.273**	-0.131	-0.049	0.454	0.959
		(0.850)	(0.601)	(0.066)	(0.781)	(0.998)	(0.444)	(0.017)	(0.010)	(0.010)	(0.581)	(0.807)	(0.139)	(0.173)
	Training and grant	-0.095	-0.104	-0.471*	-0.040	0.074	-0.374	0.158	0.169	0.286*	0.637	0.456	0.750*	1.124
		(0.542)	(0.495)	(0.072)	(0.952)	(0.901)	(0.506)	(0.258)	(0.228)	(0.087)	(0.191)	(0.284)	(0.096)	(0.124)
Consumption index	Training, no grant	0.069	0.058	0.187	-0.323	-0.319	0.326	0.032	0.017	-0.072	-0.374*	-0.303	-0.543	-0.870
		(0.524)	(0.605)	(0.317)	(0.307)	(0.327)	(0.483)	(0.674)	(0.837)	(0.610)	(0.067)	(0.133)	(0.236)	(0.213)
	Training and grant	0.058	0.029	0.164	0.616	0.521	0.785*	0.264***	0.248**	0.242*	1.241**	1.073**	0.750	-0.035
		(0.623)	(0.812)	(0.467)	(0.302)	(0.361)	(0.100)	(0.008)	(0.012)	(0.059)	(0.025)	(0.036)	(0.122)	(0.959)
Savings, investment and debt index	Training, no grant	0.373***	0.303***	0.114	-0.429	-0.450	-0.561	0.140**	0.135**	0.104	-0.049	-0.021	-0.286	0.275
		(0.001)	(0.006)	(0.453)	(0.203)	(0.174)	(0.200)	(0.014)	(0.018)	(0.343)	(0.768)	(0.898)	(0.421)	(0.592)
	Training and grant	0.408***	0.388***	0.205	1.715***	1.603***	1.029**	0.460***	0.439***	0.421***	0.975**	0.903**	0.913*	-0.116
		(0.002)	(0.001)	(0.211)	(0.008)	(0.008)	(0.027)	(0.000)	(0.000)	(0.000)	(0.013)	(0.019)	(0.065)	(0.862)
Business skills index	Training, no grant	0.042	0.026	0.091	-0.046	-0.035	0.540	-0.070	-0.059	-0.013	-0.167	-0.016	0.058	-0.482
		(0.660)	(0.789)	(0.590)	(0.895)	(0.929)	(0.179)	(0.313)	(0.417)	(0.922)	(0.406)	(0.941)	(0.898)	(0.451)
	Training and grant	0.285**	0.291**	0.357**	0.340	0.320	0.104	0.212*	0.182	0.269*	0.250	-0.090	0.195	0.092
		(0.019)	(0.023)	(0.041)	(0.551)	(0.603)	(0.873)	(0.061)	(0.110)	(0.078)	(0.584)	(0.846)	(0.704)	(0.924)
Observations		555	547	541	547	541	541	968	948	933	953	933	933	
F-stat					4.510	4.568	19.85				5.330	5.268	15.76	

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and exposure to conflict events between baseline and endline. Column (3) also controls for geographic features since the estimation strategy relies on the distance to the closest KCB bank branch that might correlate with other geographic characteristics. Geography controls include distance to the closest city, distance to the closest road, average land gradient and their respective interactions with selection to the original treatment group. Column (13) reports tests for coefficient equality between estimates from males and females in column (6) and (12). Displayed are differences of coefficient p-values of the test in parenthesis*

Table C7-22: Effects of other outcomes by gender

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
		TOT for males			**LATE for males**			**TOT for females**			**LATE for females**			Coeff equality (6) vs (12)
		(no controls)	(controls)	(controls + geo controls)	(no controls)	(controls)	(controls + geo controls)	(no controls)	(controls)	(controls + geo controls)	(no controls)	(controls)	(controls + geo controls)	
Main psychological and behavioral outcomes (survey measures)														
Psychological wellbeing index	Training, no grant	0.108	0.079	0.170	-0.013	0.089	-0.159	-0.012	0.012	0.126	-0.370**	-0.153	0.292	0.452
		(0.303)	(0.475)	(0.366)	(0.969)	(0.802)	(0.721)	(0.814)	(0.806)	(0.153)	(0.042)	(0.342)	(0.433)	(0.510)
	Training and grant	0.154	0.122	0.173	0.257	0.075	0.475	-0.059	-0.099	0.009	0.527	0.130	0.760*	0.285
		(0.146)	(0.224)	(0.257)	(0.618)	(0.885)	(0.268)	(0.538)	(0.296)	(0.941)	(0.290)	(0.757)	(0.081)	(0.671)
Risk index	Training, no grant	-0.117	-0.123	-0.160	-0.371	-0.438	0.265	0.088	0.077	0.208	-0.470	-0.366	-0.496	-0.761
		(0.197)	(0.221)	(0.305)	(0.306)	(0.238)	(0.569)	(0.389)	(0.493)	(0.120)	(0.113)	(0.147)	(0.370)	(0.168)
	Training and grant	0.087	0.069	0.061	0.577	0.639	0.529	-0.186**	-0.170*	-0.030	0.792	0.605	0.698*	0.169
		(0.468)	(0.552)	(0.693)	(0.333)	(0.290)	(0.255)	(0.044)	(0.077)	(0.804)	(0.235)	(0.294)	(0.093)	(0.761)
Trust index	Training, no grant	-0.001	-0.062	0.213	-0.095	-0.133	0.402	-0.117*	-0.124*	-0.098	-0.012	-0.053	-0.896***	-1.298**
		(0.992)	(0.556)	(0.352)	(0.792)	(0.729)	(0.467)	(0.096)	(0.098)	(0.434)	(0.948)	(0.767)	(0.008)	(0.049)
	Training and grant	0.257**	0.238*	0.493**	0.210	0.177	0.677	0.041	0.038	0.055	-0.277	-0.278	-0.791*	-1.468**
		(0.047)	(0.074)	(0.030)	(0.679)	(0.734)	(0.236)	(0.716)	(0.725)	(0.704)	(0.557)	(0.545)	(0.083)	(0.042)
Crime and violence index	Training, no grant	0.093	0.082	-0.017	-0.371	-0.434	-0.309	-0.127**	-0.151**	-0.043	-0.543*	-0.642**	-0.631	-0.321
		(0.499)	(0.508)	(0.932)	(0.332)	(0.306)	(0.561)	(0.034)	(0.018)	(0.667)	(0.092)	(0.044)	(0.111)	(0.605)
	Training and grant	-0.180	-0.176	-0.262	0.625	0.669	0.399	-0.042	-0.049	0.004	0.614	0.745	0.103	-0.296
		(0.126)	(0.163)	(0.218)	(0.324)	(0.285)	(0.456)	(0.663)	(0.623)	(0.973)	(0.340)	(0.253)	(0.814)	(0.662)
Migration index	Training, no grant	-0.147*	-0.098	-0.214	0.048	0.010	0.156	-0.045	-0.072	-0.127	-0.376**	-0.403**	-0.692*	-0.848
		(0.081)	(0.287)	(0.150)	(0.866)	(0.975)	(0.745)	(0.552)	(0.319)	(0.264)	(0.040)	(0.026)	(0.086)	(0.217)
	Training and grant	0.051	0.047	-0.070	-0.181	-0.008	-0.038	0.009	-0.003	-0.044	0.835	0.846	-0.002	0.036
		(0.653)	(0.687)	(0.654)	(0.693)	(0.986)	(0.930)	(0.937)	(0.973)	(0.716)	(0.126)	(0.106)	(0.997)	(0.958)
Main psychological and behavioral outcomes (experimental measures)														
List experiment cattle index	Training, no grant	0.371**	0.376	0.435	-0.577	-0.611	-0.614	0.117	0.152	0.181	-0.577	-0.611	-0.614	1.112
		(0.020)	(0.147)	(0.124)	(0.325)	(0.336)	(0.419)	(0.380)	(0.212)	(0.129)	(0.325)	(0.336)	(0.419)	(0.207)
	Training and grant	0.015	0.059	0.114	1.745*	1.739	2.088*	-0.128	-0.146	-0.096	1.745*	1.739	2.088*	-2.650*
		(0.946)	(0.643)	(0.474)	(0.078)	(0.104)	(0.097)	(0.554)	(0.202)	(0.462)	(0.078)	(0.104)	(0.097)	(0.093)
List experiment argument index	Training, no grant	0.024	0.025	0.031	-1.133	-0.787	-0.473	-0.265**	-0.245**	-0.291**	-1.133	-0.787	-0.473	0.213
		(0.900)	(0.899)	(0.874)	(0.139)	(0.292)	(0.399)	(0.016)	(0.032)	(0.011)	(0.139)	(0.292)	(0.399)	(0.746)
	Training and grant	0.191	0.131	0.192	2.195**	1.662	1.306	-0.488**	-0.501**	-0.451*	2.195**	1.662	1.306	-2.011*
		(0.390)	(0.573)	(0.411)	(0.044)	(0.108)	(0.134)	(0.029)	(0.036)	(0.052)	(0.044)	(0.108)	(0.134)	(0.085)
Observations		555	547	541	547	541	541	968	948	933	953	933	933	
F-stat					4.510	4.568	19.85				5.330	5.268	15.76	

*Note: P-values are in parenthesis displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. All regression control for gender-state fixed effects. Control variables of column (2) include all baseline controls that were significant determinants of attrition and of selection between receiving "training and grant" vs "training, but no grant". In particular, these are age, marital status, employment status, business ownership, food consumption, non-food consumption, formal bank account, formal loans, informal loans, education level, literacy level, numeracy level, household size, number of children, number of rooms, number of buildings at baseline, and conflict exposure between baseline and endline. Column (3) also controls for geographic features since the estimation strategy relies on the distance to the closest KCB bank branch that might correlate with other geographic characteristics. . Geography controls include distance to the closest city, distance to the closest road, average land gradient and their respective interactions with selection to the original treatment group and the interaction between conflict exposure and original treatment group. Column (13) reports tests for coefficient equality between estimates from males and females in column (6) and (12). Displayed are differences of coefficient p-values of the test in parenthesis.*

Appendix 4 – Methodological details on experimental games

Lotteries

This study uses choices over lotteries that vary in expected return and variance to extract risk preferences. In the endline, data collection respondents were asked to choose between two or three alternative lotteries. The design of this experiment involved eight rounds, building on research design by (Jakiela and Ozier 2015). After choosing one option, the chosen lottery was played as a flip of a fair coin (50 percent chance of each outcome). The game started with two practice rounds to make participants familiar with the rules. After that, the participants had to play six additional rounds. At the end of the game, one round was selected at random and the lottery chosen by the participants was played and paid out. Participants were informed about these rules at the beginning of the game. The lotteries are set up as described below in Table C7-23.

The number of times respondents chose the riskiest lottery can be used as a proxy for their risk preferences. Given that respondents in these types of experiments often display choices that are inconsistent with CRRA utility a non-parametric approach to measure risk aversion is more appropriate. Thus, following the approach put forward by Jakiela and Ozier (2015) the set of lottery choices can also be used to infer risk preferences in a less stringent and non-theoretic manner. One measure is created by counting how many times respondents choose the riskiest lotteries, i.e. lotteries with the largest spread, or the safest lotteries. In addition, the likelihood to choose the riskier lottery during each decision round was evaluated individually. The results are then compared to survey answers on risk preferences.

Test questions were included to detect biased answers that resulted from a lack of understanding. Due to the relatively low numeracy skills and the complexity of the lotteries, the study included 3 questions to test for monotonicity, i.e. if participants behaved like utility-maximizers (Andreoni and Sprenger 2010). If participants answered more than 1 of these test questions in a way inconsistent with utility maximization, it is likely that they simply did not understand the nature of the decision problem.

Table C7-23: Pay-outs of lotteries, expected utility

	Lottery A		Lottery B		Lottery C	
	Heads	Tails	Heads	Tails	Heads	Tails
Practice						
Decision 1	100	100	150	150		
Decision 2	100	150	200	250		
Game						
Decision 3	100	100	100	120		
Decision 4	100	100	0	400		
Decision 5	30	340	100	100	0	400
Decision 6	100	100	55	240	30	340
Decision 7	30	230	60	170	90	110
Decision 8	10	200	70	160	90	110

Trust game

Trust attitudes towards the World Bank were assessed using a trust game. The basic structure of a trust game developed by Berg, Dickhaut et al. (1995) involves Player A receiving an endowment of X and choosing how much of this endowment to send to Player B, Y ϵ [0, X]. Player B receives 3Y – i.e., three times whatever A send him – and must decide how much of this endowment to send back to A, Z ϵ [0, 3Y]. A receives a payout of X-Y+Z and B receives a payout of 3Y-Z. Y/X is used as a measure of trust. Z/3Y is used as a measure of trustworthiness. The table below summarizes payouts for the two players:

Table C7-24: Trust game payouts

Player 1			**Player 2**		
Endowment	Sends	Payout	Endowment	Sends	Payout
X	Y	$X - Y + Z$	$3Y$	Z	$3Y - Z$

In our study, participants were asked to play several rounds of a trust game. In the first game, Player B was framed as the World Bank to extract a measure of trust toward the World Bank or official institutions in general. Participants may hold the World Bank responsible for the (non-) payment of the business start-up grants. This framing of Player B as the World Bank allows for a direct measure of how willing participants are to partake in an interaction with the World Bank that could have financial consequences. Hence, it can act as a measure of how not receiving the promised grant had influenced their level of trust and their willingness to interact with the World Bank. The reciprocal behavior of Player B was modeled to mirror the probability of non-disbursement of the cash grant. In 34 percent of the cases documented by the phone survey, participants received the grant. This information was used to define the reciprocal behavior of Player B. Player B played fairly 34 percent of the time – that is, returns back exactly half of what they obtain from the study participant. Player B 66 percent of the time acted unfairly and kept all that is sent to them, regardless of what the respondent sent. In the end, the participant was paid out the budget of Player A.

To obtain a more general measure of the respondents' trust levels, and to accompany the first measure, a second game was played which pit the participants against each other. The survey respondents were equally and randomly selected as players A and B, stratified by treatment groups and treatment strands. Regarding the implementation of the games and pairing of the players, a lab-in-the-field experimental setup was impossible to organize because respondents had to be interviewed individually. This was primarily due to the complicated logistical circumstances surrounding fieldwork in South Sudan, in no small part due to rapidly deteriorating security conditions, but also due to constraints on the respondents' time. Respondents were, therefore, playing the games against a pre-loaded hypothetical distribution of responses. Enumerators explained to the respondents that the other player would be another survey respondent elsewhere in South Sudan. The set of possible responses, in terms of the fraction of the endowment sent or returned, was equally distributed between [0.1,1] in increments of 0.1. In no cases was the fraction of endowment sent or returned equal to zero.

List-experiment

Based on the results from the baseline survey, it was determined that the reporting of sensitive behaviors might have been untruthful. Methods to elicit more truthful responses were therefore employed in the endline questionnaire. For example, the rates at which respondents reported even simply knowing someone who may have participated in cattle raiding were close to zero, despite 63 percent of respondents reporting cattle raiding in their area in the baseline. Rates of reporting respondents' own sensitive behaviors were even lower. Therefore, a set of list questions – also commonly known as the "item count technique" introduced by Miller (1984) – were added to the endline questionnaire. In these questions, the sample is split into a treatment and control group, and respondents in the control group are given a set of N statements and asked to answer with how many of these statements do they agree with/or would say yes to, without explicitly stating which ones.

Respondents in the treatment group are given the same N statements + a sensitive item. The estimate of the true rate at which respondents agree with the sensitive statements is simply the difference in means, in terms of the number of statements, between the treatment and control groups. In the context of the endline survey, the sensitive behaviors pertained to violent behavior, including domestic violence, as well as cattle raiding. Direct questions were asked to the control group alongside the list question without the sensitive item, so as to compare results obtained through the list-method. Below we report the list of sensitive items included in the experiment.

Table C7-25: List of sensitive statements included in the list experiment

Cattle raiding: 1. I know someone who has participated in cattle raiding, including myself. 2. I have participated in cattle raiding. **Violent behavior:** 1. I have had a verbal disagreement in the last month where the other person threatened me with violence. 2. I have had a verbal disagreement in the last month where I threatened the other person with violence. 3. I have had a verbal disagreement with someone in the last month which ended with violence.

Appendix 5 – Index creation

Indexes s_{ji} are defined as a weighted average of all standardized outcomes *k* within outcome group *j*.

$$s_{ij} = \frac{1}{W_{ij}} \sum_{k} w_{ik} \frac{y_{ijk} - \bar{y}_{jk}}{\sigma_{jk}^{y}}$$

Weight w_{jk} of each outcome *k* is derived from the inverted covariance matrix of all standardized outcomes *k.*

$$\sum\nolimits_{j}^{-1} = \begin{bmatrix} c_{j11} & \cdots & c_{j1K} \\ \vdots & \ddots & \vdots \\ c_{jK1} & \cdots & c_{jKK} \end{bmatrix}$$

Weight w_{jk} then consists of the row sum of the inverted covariance matrix.

$$w_{jk} = \sum_{l=1}^{K_j} c_{jkl}$$

Appendix 6 – Additional figures

Figure C7-3: Map of participants' baseline locations and major cities of project states

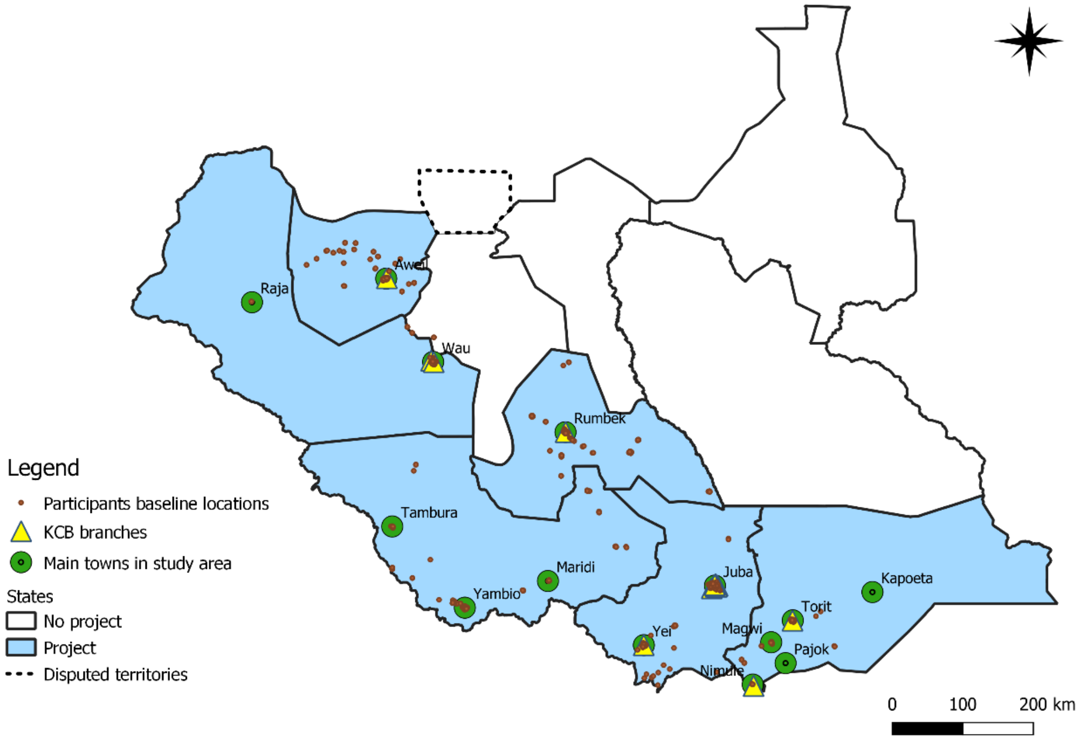
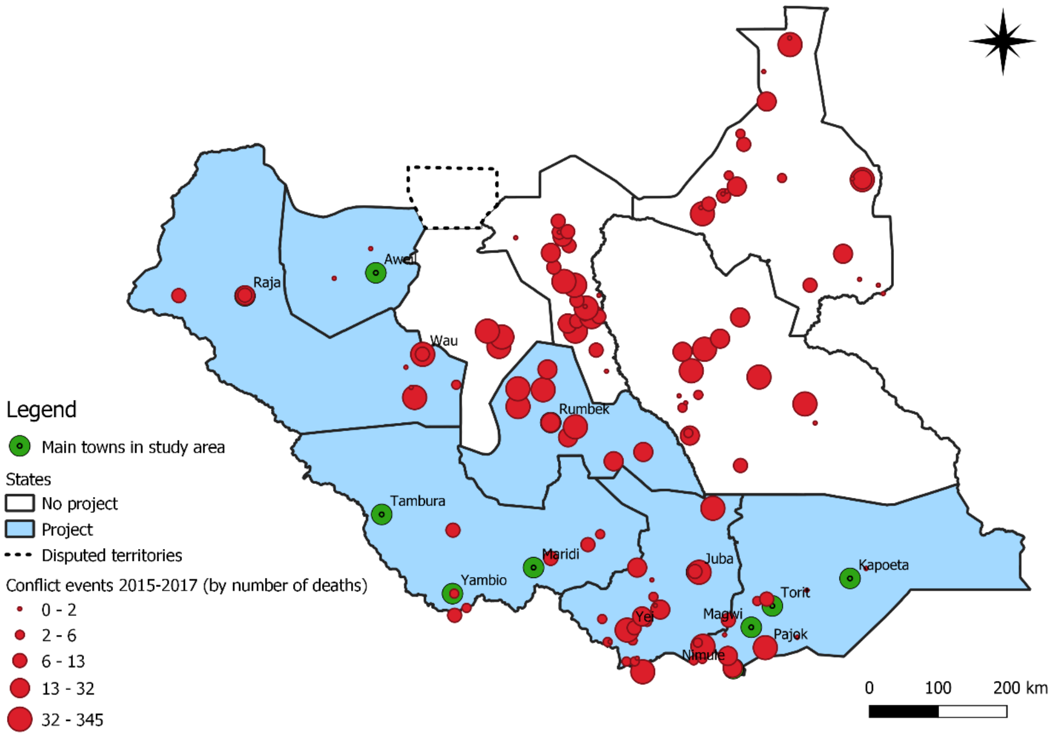
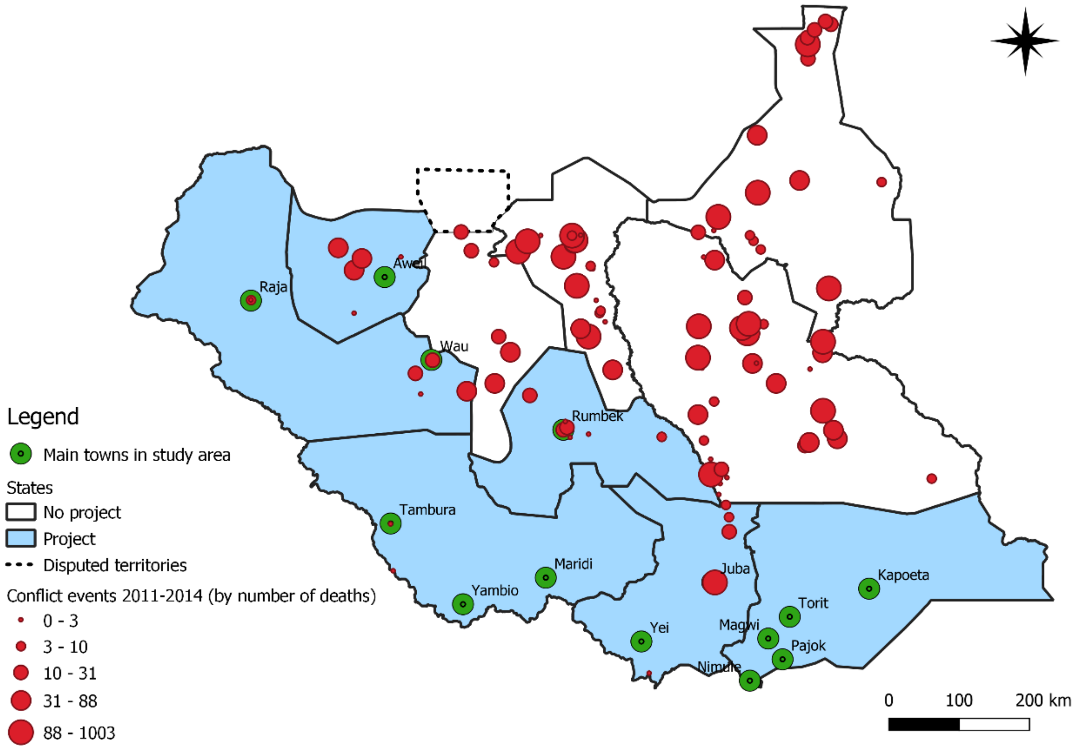


Figure C7-4: Map of conflict events before and during project period





D. References

(2017). United Nations Office for the Coordination of Humanitarian Affairs. South Sudan.

Abadie, A. (2004). Poverty, Political Freedom, and the Roots of Terrorism. Working Paper 10859, National Bureau of Economic Research.

Abadie, A. and J. Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country." American Economic Review **93**(1): 113-132.

Abadie, A. and J. Gardeazabal (2008). "Terrorism and the world economy." European Economic Review **52**(1): 1-27.

Abelson, B., R. Varshney and J. Sun (2014). Targeting direct cash transfers to the extremely poor. Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining**:** 1563 - 1572.

Acemoglu, D., S. Johnson, J. Robinson and Y. Thaicharoen (2003). "Institutional causes, macroeconomic symptoms: volatility, crises and growth." Journal of Monetary Economics **50**(1).

Acharya, S. (2008). "Poverty Alleviation and the Industrial Employment of Women (The Case of Nepal)." Journal of International Development **20**(5): 670 - 685.

African Development Bank (2013). South Sudan Infrastructure Action Plan - A Program for Sustained Strong Economic Growth.

Agostini, C. A., P. H. Brown and A. C. Roman (2010). Poverty and Inequality Among Ethnic Groups in Chili.

Aguilera, A., N. Krishnan, J. Muñoz, F. R. Riva, D. Sharma and T. Vishwanath (2020). Sampling for Representative Surveys of Displaced Populations. Data Collection in Fragile States. J. H. a. U. Pape, World Bank.

Ahmed, M. A., J. N. Siewe Fodjo, A. A. Gele, A. A. Farah, S. Osman, I. A. Guled, A. M. Ali and R. Colebunders (2020). "COVID-19 in Somalia: adherence to preventive measures and evolution of the disease burden." Pathogens **9**(9): 735.

Akresh, R., S. Bhalotra, M. Leone and U. O. Osili (2012). "War and Stature: Growing Up during the Nigerian Civil War." American Economic Review **102**(3): 273-277.

Akresh, R., P. Verwimp and T. Bundervoet (2011). "Civil War, Crop Failure, and Child Stunting in Rwanda." Economic Development and Cultural Change **59**(4): 777 - 810.

Al-Hasan, A., D. Yim and J. Khuntia (2020). "Citizens' adherence to COVID-19 mitigation recommendations by the government: a 3-country comparative evaluation using web-based cross-sectional survey data." Journal of medical Internet research **22**(8): e20634.

Alderman, H., J. Konde-Lule, I. Sebuliba, D. Bundy and A. Hall (2006). "Effect on weight gain of routinely giving albendazole to preschool children during child health days in Uganda: cluster randomised controlled trial." BMJ : British Medical Journal **333**(7559): 122.

Alem, Y. and M. Soderbom (2012). "Household-Level Consumption in Urban Ethiopia: The Effects of a Large Food Price Shock." World Development **40**(1): 146-162.

Almond, D. and J. Currie (2011). Human Capital Development before Age Five. Handbook of Labor Economics. O. Ashenfelter and D. Card. **4:** 1315 - 1486.

Alt, C., W. Bien and D. Krebs (1991). "Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe." ZUMA-Nachrichten **28**: 65 - 72.

Altindag, O., S. O'Connell, A. Şaşmaz, B. Zeynep, P. Cadoni, M. Jerneck and A. K. Foong (2021). "Targeting humanitarian aid using administrative data: Model design and validation." Journal of Development Economics **148**(C).

Ambler, G., R. Z. Omar and P. Royston (2007). "A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome." Statistical methods in medical research **16**(3): 277-298.

Amendola, N. and G. Vecchi (2014). Durable Goods and Poverty Measurement, Policy Research Working Papers, The World Bank.

Aminipouri, M., R. Sliuzas and M. Kuffer (2009). Object-oriented analysis of very high resolution orthophotos for estimating the population of slum areas, case of Dar-Es-Salaam, Tanzania. Proc. ISPRS XXXVIII Conf.**:** 1 - 6.

Andreoni, J. and C. Sprenger (2010). Certain and Uncertain Utility: The Allais Paradox and Five Decision Theory Phenomena. Levine's Working Paper Archive**:** 1 - 21.

Angrist, J. and J.-S. Pischke (2009). Mostly Harmless Econometrics - An Empiricist's Companion. Princeton, N.J., Princeton University Press.

Angrist, J. D. and A. D. Kugler (2008). "Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia." The Review of Economics and Statistics **90**(2): 191 - 215.

Annan, J., C. Blattman, D. Mazurana and K. Carlson (2009). Women and Girls at War: "Wives", Mothers, and Fighters in the Lord's Resistance Army, Brighton: Households in Conflict Network, The Institute of Development Studies - at the University of Sussex.

Araujo, M. C., F. H. G. Ferreira, P. Lanjouw and B. Özler (2008). "Local Inequality and Project Choice: Theory and Evidence From Ecuador." Journal of Public Economics **92**(5-6): 1022-1046.

Athey, S. and G. W. Imbens (2017). "The state of applied econometrics: Causality and policy evaluation." Journal of Economic Perspectives **31**(1): 3 - 32.

Auriemma, V. and C. Iannaccone (2020). "COVID-19 pandemic: socio-economic consequences of social distancing measures in Italy." Front **Sociol 5**.

Ayazi, T., L. Lien, A. Eide, L. Swartz and E. Hauff (2014). "Association between exposure to traumatic events and anxiety disorders in a post-conflict setting: a cross-sectional community study in South Sudan." BMC Psychiatry **14**(1): 6.

Bachman, J. G. and P. M. O'Malley (1984). "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles on JSTOR." Public Opinion Quarterly **48**(2): 491 - 509.

Baird, S., J. de Hoop and B. Özler (2013). "Income Shocks and Adolescent." The Journal of Human Resources **48**(2): 370 - 403.

Baird, S., F. H. G. Ferreira, B. Özler and M. Woolcock (2014). "Conditional, Unconditional and Everything in Between : A Systematic Review of the Effects of Cash Transfer Programs on Schooling Outcomes." Journal of Development Effectiveness **6**(1).

Baird, S., J. Hamory and E. Miguel (2008). Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1), UC Berkeley: Center for International and Development Economics Research.

Baird, S., D. McKenzie and B. Özler (2018). The Effects of Cash Transfers on Adult Labor Market Outcomes. 8404 Policy Research Working Paper.

Baker, J. L. (2000). Evaluating the impact of development projects on poverty: A handbook for practitioners, The World Bank.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert and C. Udry (2015). "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." Science **344**(6236).

Banerjee, A. V., H. Rema, G. E. Kreindler and B. A. Olken (2017). Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs. World Bank Research Observer. **32:** 155 - 184.

Bangladesh Bureau of Statistics, World Bank and United Nations World Food Programme (2010). Updating Poverty Maps: Bangladesh Poverty Maps for 2005.

Banks, R. and H. Laurie (2000). "From Papi to Capi: The Case of the British Household Panel Survey." Social Science Computer Review **18**(4): 397 - 406.

Baranov, V., S. Bhalotra, P. Biroli and J. Maselko (2017). Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Large Randomized Control Trial. Bonn, Institute for the Study of Labor (IZA).

Bargain, O. and A. Ulugbek (2021). "Poverty and COVID-19 in Africa and Latin America." World Development **142**.

Barrett, D. F., M. Beaghen, D. Smith and J. Burcham (2002). Census 2000 housing unit coverage study. Proceedings of the Section on Survey Research Methods, American Statistical Association

Barry, M. and H. Rüther (2005). "Data collection techniques for informal settlement upgrades in Cape Town, South Africa." Urisa Journal **17**: 43 - 52.

Bartik, T. (1991). "Who Benefits from State and Local Economic Development Policies?" Upjohn Press.

Bauer, J. J. (2014). "Selection errors of random route samples." Sociological Methods & Research **43**(3): 519 - 544.

Bauer, J. J. (2016). "Biases in random route surveys." Journal of survey statistics and methodology **4**(2): 263 - 287.

Becker, G. and K. Murphy (2001). "Prosperity Will Rise Out of the Ashes." Wall Street Journal.

Beegle, K., L. Christiaensen, A. Dabalen and I. Gaddis (2016). Poverty in a Rising Africa. Washington, DC, World Bank Publications.

Beegle, K., J. De Weerdt, 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.

Beltramo, T., H. Dang, I. Sarr and P. Verme (2021). "Estimating Poverty among Refugee Populations: A Cross-Survey Imputation Exercise for Chad." IZA Discussion Papers **14606**.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society **57**(1): 289 - 300.

Berg, J., J. Dickhaut and K. McCabe (1995). "Trust, Reciprocity, and Social History." Games and Economic Behavior **10**: 122 - 142.

Berge, I., L. Oppedal, K. Bjorvatn and B. Tungodden (2014). "Human and Financial Capital for Microenterprise Development : Evidence from a Field and Lab Experiment." Management Science.

Bertrand, M., E. Duflo and S. Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?" The Quarterly Journal of Economics **119**(1): 249 - 275.

Bianchi, M. and M. Bobba (2013). "Liquidity, Risk, and Occupational Choices." The Review of Economic Studies **80**(2): 491 - 511.

Bien, W., D. Bender and D. Krebs (1997). DJI-Familiensurvey: Der Zwang, mit unterschiedlichen Stichproben zu leben. Stichproben in der Umfragepraxis, Springer: 127 - 147.

Biscaye, P. E., D. Egger and U. J. Pape (2022). "Balancing Work and Childcare - Evidence from COVID-19 School Closures and Reopenings in Kenya." World Bank Policy Research Working Paper(9958).

Bjorvatn, K. and B. Tungodden (2010). "Teaching Business in Tanzania: Evaluating Participation and Performance." Journal of the European Economic Association **8**(2): 561 - 570.

Blackorby, C. and D. Donaldson (1987). "Welfare Ratios and Distributionally Sensitive Cost-Benefit Analysis." Journal of Public Economics **34**(3): 265 - 290.

Blair, G. and K. Imai (2017). "Statistical Analysis of List Experiments." Political Analysis **20**(1): 47 - 77.

Blattman, C., N. Fiala and S. Martinez (2014). "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda." The Quarterly Journal of Economics: 697 - 752.

Blattman, C., J. Jamison, T. Koroknay-Palicz, K. Rodrigues and M. Sheridan (2016). "Measuring the measurement error: A method to qualitatively validate." Journal of Development Economics **120**: 99 - 112.

Blattman, C. and A. Jeannie (2007). The Consequences of Child Soldiering. HiCN Working Paper 22, Brighton: Households in Conflict Network, The Institute of Development Studies - at the University of Sussex.

Blattman, C. and E. Miguel (2010). "Civil War." Journal of Economic Literature **48**(1): 3 - 57.

Blattman, C. and L. Ralston (2015). Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs.

Blohm, M. (2006). Data Quality in Nationwide Face-to-face Social Surveys.

Blumenstock, J., G. Cadamuro and R. On (2015). "Predicting poverty and wealth from mobile phone metadata." Science 350 **626**: 1073 -1076.

Bobonis, G. J., M. Gonzâlez-Brenes and R. Castro (2013). "Public Transfers and Domestic Violence : The Roles of Private Information and Spousal Control." American Economic Journal: Economic Policy **5**(1).

Bollinger, C. and P. Hagstrom (2004). Poverty rates of refugees and immigrants, UKCPR: University of Kentucky Center for Poverty Research, report.

Bonaccorsi, G., F. Pierri, M. Cinelli, A. Flori and A. Galeazzi (2020). " Economic and social consequences of human mobility restrictions under COVID-19." PNAS **117**(27).

Boom, B., A. Halsema and V. Molini (2015). Are We Confusing Poverty with Preferences? World Bank Policy Research Working Paper. **7247**.

Boothby, N., J. Crawford and J. Halperin (2006). "Mozambique child soldier life outcome study: Lessons learned in rehabilitation and reintegration efforts." Global Public Health **1**(1): 87 - 107.

Bowmans (2020). "COVID-19: Tracking Government Response in Kenya."

Bozzoli, C. and T. Brück (2009). "Agriculture, Poverty, and Postwar Reconstruction: Micro-Level Evidence from Northern Mozambique:." Journal of Peace Research.

Bozzoli, C. and T. Brück (2009). Identifying Conflict and its Effects Using Micro-Level Surveys. HiCN Research Design Note 13. University of Sussex, Brighton: Households in Conflict Network, The Institute of Development Studies.

Bozzoli, C., T. Brück and T. Muhumuza (2015). "Activity Choices of Internally Displaced Persons and Returnees: Quantitative Evidence from Post-War Northern Uganda." Bulletin of Economic Research.

Bozzoli, C., T. Brück and N. Wald (2011). "Self-Employment and Conflict in Colombia." Journal of Conflict Resolution **57**(1).

Bozzoli, C., T. Brueck and T. Muhumuza (2016). "Activity Choices of Internally Displaced Persons and Returnees: Quantitative Evidence from Post-War Northern Uganda." Bulletin of Economic Research **68**(4): 329 - 347.

Brandolini, A. and G. D'Alessio (2001). Measuring Well-Being in the Functioning Space. Debating Global Society: Reach and Limits of the Capability Approach. M. Chiappero-Martinetti and A. Sen. Milano, Feltrinelli.

Bratti, M., M. Mariapia and M. Alfonso (2009). Hard to Forget: The Long-Lasting Impact of War on Mental Health. Design Note 206. University of Sussex, Brighton: Households in Conflict Network, The Institute of Development Studies.

Bratti, M., M. Mendola and M. Alfonso (2016). Hard to forget: The long-lasting impact of war on mental health. HiCN WP 206. University of Sussex, Brighton: Households in Conflict Network, The Institute of Development Studies.

Brown, R. and A. Velásquez (2015). The Effect of Violent Crime on the Human Capital Accumulation of Young Adults. Population Association of America Annual Meeting. San Diego, CA.

Brück, T. and M. d'Errico (2019). "Food security and violent conflict: Introduction to the special issue." World Development **117**: 167-171.

Brück, T., Patricia Justino, Philip Verwimp and A. Avdeenko (2010). Identifying Conflict and Violence in Micro-Level Surveys. IZA Discussion Paper. Bonn, Institute for the Study of Labor (IZA). **5067**.

Brück, T. and K. Schindler (2009). "The Impact of Violent Conflicts on Households: What Do We Know and What Should We Know about War Widows?" Oxford Development Studies **37**(3): 289 - 309.

Brück, T. and M. Vothknecht (2011). "Impact of Violent Conflicts on Women's Economic Opportunities." Women and War: Power and Protection in the 21st Century: 85-114.

Bui, T. D. and C. V. Nguyen (2017). "Spatial Poverty Reduction in Vietnam : An Application of Small Area Submission Number." Economics Bulletin **37**(3): 1785-1796.

Bun, M. J. G. and T. D. Harrison (2018). "OLS and IV Estimation of Regression Models Including Endogenous Interaction Terms." Econometric Reviews.

Bundervoet, T., P. Verwimp and R. Akresh (2009). "Health and Civil War in Rural Burundi." The Journal of Human Resources **44**(2): 536 - 563.

Buvinic, M., M. Das Gupta, U. Casabonne and P. Verwimp (2012). Violent Conflict and Gender Inequality: An Overview. HiCN Working Papers, Households in Conflict Network.

Caeyers, B., N. Chalmers and J. De Weerdt (2012). "Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment." Journal of Development Economics **98**: 19 - 33.

Cagan, P. (1956). The Monetary Dynamics of Hyperinflation. Studies in the Quantity Theory of Money. M. Friedman. Chicago, University of Chicago Press: 25 - 117.

Cagoco-Guiam, R. (2013). Gender and Livelihoods among Internally Displaced Persons in Mindanao, Philippines, The Brookings - London School of Economics.

Camacho, A. (2008). "Stress and Birth Weight: Evidence from Terrorist Attacks." American Economic Review **98**(2): 511-515.

Cameron, E., A. Delius, A. Devercelli, U. Pape and S. Siewers (2022). "The Consequences of the COVID-19 Pandemic for Children in Kenya." World Bank Policy Research Working Paper(10003).

Card, D. (1990). "The Impact of the Mariel Boatlift on the Miami Labor Market." Industrial and Labor Relations Review **43**: 245 - 257.

CARE (2014). "'The Girl Has No Rights': Gender-Based Violence in South Sudan."

Carlucci, L., I. D'ambrosio and M. Balsamo (2020). "Demographic and attitudinal factors of adherence to quarantine guidelines during COVID-19: the Italian model." Frontiers in psychology **11**(2702).

Carpenter, J. R., M. G. Kenward and I. R. White (2007). "Sensitivity analysis after multiple imputation under missing at random: a weighting approach." Statistical methods in medical research **16**(3): 259-275.

Chen, S., D. O. Igan, N. Pierri, A. F. Presbitero, M. Soledad and M. Peria (2020). Tracking the Economic Impact of COVID-19 and Mitigation Policies in Europe and the United States. IMF Working Papers. **125**.

Chen, X., H. Hu, X. Xu, J. Gong, Y. Yan and F. Li (2018). "Probability Sampling by Connecting Space with Households Using GIS/GPS Technologies." Journal of Survey Statistics and Methodology **6**(2): 149 - 168.

Chew, R. F., S. Amer, K. Jones, J. Unangst, J. Cajka, J. Allpress and M. Bruhn (2018). "Residential Scene Classification for Gridded Population Sampling in Developing Countries Using Deep Convolutional Neural Networks on Satellite Imagery." International Journal of Health Geographics **17**(1): 1 - 17.

Cho, Y. and M. Honorati (2014). "Entrepreneurship Programs in Developing Countries : A Meta Regression Analysis." Labour Economics **28**: 110 - 130.

Christiaensen, L., P. Lanjouw, J. Luoto and D. Stifel (2010). The Reliability of Small Area Estimation Prediction Methods to Track Poverty.

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.

Cialdini, R. B. (2003). "Crafting Normative Messages to Protect the Environment." Current Directions in Psychological Science **12**(4): 105-109.

Cilliers, J., O. Dube and B. Siddiqi (2015). "The white-man effect: How foreigner presence affects behavior in experiments." Journal of Economic Behavior and Organization **18**: 397 - 414.

Cohen, M. J. and P. E. R. Pinstrup-Andersen (1999). "Food Security and Conflict." Social Research **66**(1): 375-416.

Cohn, A., E. Fehr and M. A. Maréchal (2014). "Business culture and dishonesty in the banking industry." Nature **516.7529**(86).

Cohn, A., M. A. Maréchal and T. Noll (2010). "Bad Boys: How Criminal Identity Salience Affects Rule Violation." Review of Economic Studies **82**(4): 1289 – 1308.

Collier, P. and A. Hoeffler (2007). Civil War. K. Hartley and T. Sandler, Elsevier. **2:** 711-739.

Comin, F., M. Qizilbash and S. Alkire (2008). The Capability Approach. Concepts, Measures, and Applications. Cambridge, Cambridge University Press.

Conley, T. G. (1999). "GMM Estimation with Cross Sectional Dependence." Journal of Econometrics **92**(1): 1 - 45.

Conley, T. G. (2010). Spatial Econometrics. Microeconometrics The New Palgrave Economics Collection. Steven N. Durlauf and L. E. Blume. London, Palgrave Macmillan UK**:** 303 – 313.

Cook, T. D. (2008). "Waiting for Life to Arrive: A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics." Journal of Econometrics **142**: 636 - 654.

Coomer, J. and T. Gstraunthaler (2011). "The Hyperinflation in Zimbabwe." Quarterly Journal of Austrian Economics **14**(3).

Cornish, D. B. and R. V. Clarke (2014). "The Reasoning Criminal: Rational Choice Perspectives on Offending." Transaction Publishers.

Coroiu, A., C. Moran, T. Campbell and A. C. Geller (2020). "Barriers and facilitators of adherence to social distancing recommendations during COVID-19 among a large international sample of adults." PloS one **15**(10): e0239795.

Corral, P., A. Irwin, N. Krishnan, D. G. Mahler and T. Vishwanath (2020). Fragility and Violence: On the Front Lines of the Fight against Poverty. Washington, D.C., World Bank.

Coudouel, A., J. S. Hentschel and Q. T. Wodon (2002). Poverty measurement and analysis. A Sourcebook for poverty reduction strategies **1**.

Crenshaw, M. (1981). "The Causes of Terrorism." Comparative Politics **13**(4): 379 - 399.

Cronbach, L. J. (1946). "Response Sets and Validity. Educational and psychological measurement." (6): 672-683.

Cuddeback, G., E. Wilson, J. G. Orme and T. Combs-Orme (2004). "Detecting and statistically correcting sample selection bias." Journal of Social Service Research **30**(3): 19 - 33.

Cuevas, P. F., I. O. Kaan, A. Twose and C. Celik (2019). "Vulnerability and Protection of Refugees in Turkey: Findings from the Rollout of the Largest Humanitarian Cash Assistance Program in the World."

D'Souza, A. and D. Jolliffe (2013). "Conflict, food price shocks, and food insecurity: The experience of Afghan households." Food Policy **42**: 32-47.

Dabalen, A. and S. Paul (2012). Estimating the Causal Effects of Conflict on Education in Côte d'Ivoire, Brighton: Households in Conflict Network, The Institute of Development Studies - at the University of Sussex.

Dahlhamer, J., J. Dixon, P. Doyle, J. Eargle, D. Griffin and P. McGovern (2003). Quality at the Item Level: Terms, Methods, and Guidelines for Cross-Survey Comparisons. Washington, DC., Bureau of Labor Statistics.

Dang, H.-A. H. and C. V. Nguyen (2021). "Gender inequality during the COVID-19 pandemic: Income, expenditure, savings, and job loss." World Development **140**: 105296.

Dang, H. V., P. (2021). Estimating Poverty for Refugees in Data-scarce Contexts: An Application of Cross-Survey Imputation. ECINEQ, Society for the Study of Economic Inequality. **578**.

Das, S. and R. Chambers (2017). "Robust Mean-Squared Error Estimation for Poverty Estimates Based on the Method of Elbers, Lanjouw and Lanjouw." Journal of the Royal Statistical Society. Series A: Statistics in Society **180**(4): 1137-1161.

De Leeuw, E. D., J. J. Hox and G. Snijkers (1995). "The effect of computer-assisted interviewing on data quality. A review." Market Research Society. Journal **37**: 1 - 19.

De Mel, S., D. McKenzie and C. Woodruff (2008). "Returns to Capital in Microenterprises: Evidence from a Field Experiment." The Quarterly Journal of Economics **123**(4): 1329 - 1372.

De Mel, S., D. McKenzie and C. Woodruff (2012). "One-Time Transfers of Cash or Capital Have Long-Lasting Effects on Microenterprises in Sri Lanka." Science **335**: 962 - 966.

de Vries, L. and M. Schomerus (2017). "Fettered Self-determination: South Sudan's Narrowed Path to Secession." Civil Wars **19**(1): 26-45.

de Waal, A. (2014). "When kleptocracy becomes insolvent: Brute causes of the civil war in South Sudan." African Affairs **113**(452): 347-369.

Deaton, A. (1980). <u>The Measurement of Welfare: Theory and Practical Guidelines</u>. Washington, D.C., The World Bank.

Deaton, A. (2006). Measuring poverty. <u>Understanding poverty</u>**:** 3 - 15.

Deaton, A. and M. Grosh (2000). "Consumption in designing household survey questionnaires for developing countries." <u>Designing Household Survey Questionnaires for Developing Countries: Lessons from from</u> **15**.

Deaton, A. and V. Kozel (2005). "Data and dogma: The great Indian poverty debate." <u>World Bank Research Observer</u> **20**(2): 177-199.

Deaton, A. and J. Muellbauer (1980). <u>Economics and Consumer Behavior</u>. Cambridge, Cambridge University Press.

Deaton, A. and S. Zaidi (2002). Guidelines for constructing consumption aggregates for welfare analysis. <u>Living Standards Measurement Study</u>. Washington, D.C., World Bank.

Delius, A., K. Himelein and U. Pape (2020). Conducting Rapid Response Phone Surveys (RRPS) to Fill Data Gaps, The World Bank.

Demobynes, G. and K. T. Sofia (2016). "What Has Driven the Delince of Infant Mortality in Kenya." <u>Economics & Human Biology</u>(21): 17-32.

Demombynes, G. and B. Özler (2005). "Crime and Local Inequality in South Africa." <u>Journal of Development Economics</u> **76**(2): 265-292.

Dercon, S. (2004). "Growth and shocks: evidence from rural Ethiopia." <u>Journal of Development Economics</u> **74**(2): 309–329.

Dercon, S. and J. Hoddinott (2004). "Health, shocks and poverty persistence." <u>Insurance against poverty</u>: 124–136.

Dercon, S. and P. Krishnan (2000). "Vulnerability, seasonality and poverty in Ethiopia." <u>The Journal of Development Studies</u> **36**(6): 25-53.

Devi, S. (2017). "Famine in South Sudan." <u>The Lancet</u> **389**(10083): 1967 - 1970.

Di Maio, M. N. F. (2020). "Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect." <u>The World Bank Economic Review</u> **34**(3): 654 - 669.

Diaz de Rada, V. and V. M. Martín (2014). " Random route and quota sampling: Do they offer any advantage over probably sampling methods?".

Diehr, P., L. Chen, D. Patrick, Z. Feng and Y. Yasui (2005). "Reliability, effect size, and responsiveness of health status measures in the design of randomized and cluster-randomized trials." <u>Contemporary Clinical Trials</u> **26**(1): 45-58.

Dijkman, N., B. Catrien and V. Philip (2014). Sexual Violence in Burundi: Victims, perpetrators, and the role of conflict, Brighton: Households in Conflict Network, The Institute of Development Studies - at the University of Sussex.

Dillon, B. (2012). "Using mobile phones to collect panel data in developing countries." <u>Journal of international development</u> **24**(4): 518–527.

Dong, E., H. Du and L. Gardner (2020). "An interactive web-based dashboard to track COVID-19 in real time." <u>The Lancet Infectious Diseases</u> **20**(5): 533 - 534.

Douarin, E., J. Litchfield and R. Sabates-Wheeler (2011). Poverty, Livelihoods and War Legacies: The Case of Post-War Rural Kosovo. <u>IDS Working Papers</u>. University of Sussex, Brighton: Institute of Development Studies. **380**.

Douidich, M., A. Ezzrari, R. van der Weide and P. Verme (2013). "Estimating Quarterly Poverty Rates Using Labor Force Surveys." World Bank Policy Research Working Paper(6466).

Drake, T., A. Docherty, T. Weiser, S. Yule, A. Sheikh and E. Harrison (2020). "The effects of physical distancing on population mobility during the COVID-19 pandemic in the UK." The Lancet Digital Health **2**(8): e385 - e387.

Draper, N. R. and H. Smith (2014). "Applied regression analysis."

Dube, O. and J. F. Vargas (2006). Resource Curse in Reverse: The Coffee Crisis and Armed Conflict in Colombia. Discussion Papers in Economics, Royal Holloway, University of London. **6**.

Dupas, R. (2012). "The (hidden) costs of political instability: Evidence from Kenya's 2007 election crisis." Journal of Development Economics.

Dwyer, L. and R. Cagoco-Guiam (2011). Gender and conflict in Mindanao. New York, The Asia Foundation.

Eckman, S. (2013). "Do different listers make the same housing unit frame? Variability in housing unit listing." Journal of Official Statistics **29**(2): 249 - 259.

Eckman, S. and A. Koch (2016). Are High Response Rates Good for Data Quality? Evidence from the European Social Survey. Paper under review.

Eckman, S., F. Kreuter, A. Kirchner, A. Jäckle, R. Tourangeau and S. Presser (2014). "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys." Public Opinion Quarterly **78**(3): 721-733.

Eckstein, Z. and D. Tsiddon (2004). "Macroeconomic Consequences of Terror: Theory and the Case of Israel." Journal of Monetary Economics **51**(5): 971 - 1002.

Egger, D., E. Miguel, S. S. Warren, A. Shenoy, E. Collins, D. Karlan, D. Parkerson, A. M. Mobarak, G. Fink, C. Udry, M. Walker, J. Haushofer, M. Larreboure, S. Athey, P. Lopez-Pena, S. Benhachmi, M. Humphreys, L. Lowe, N. F. Meriggi, A. Wabwire, C. A. Davis, U. J. Pape, T. Graff, M. Voors, C. Nekesa and C. Vernot (2021). "Falling living standards during the COVID-19 crisis: Quantitative evidence from nine developing countries." Science Advances **7**(6): eabe0997.

EGRIS (2018). International Recommendations on Refugee Statistics.

EGRIS (2020). International Recommendations on IDP Statistics (IRIS).

Elangovan, A., G. Elavarsu, R. Ezhil, R. Prabu and J. Yuvaraj (2016). "Geospatial Techniques in Health Survey to Overcome the Lack of Sampling Frame." International Journal of Advanced Remote Sensing and GIS: pp.pp - 1908.

Elbers, C., T. Fujii, P. Lanjouw, B. Özler and W. Yin (2007). "Poverty Alleviation Through Geographic Targeting: How Much Does Disaggregation Help?" Journal of Development Economics **83**(1): 198-213.

Elbers, C., J. O. Lanjouw and P. Lanjouw (2003). "Micro-Level Estimation of Poverty and Inequality." Econometrica **71**(1): 355-364.

Elbers, C., Lanjouw, J.O., Lanjouw, P. (2002). Micro-level estimation of welfare. Policy Research Working Paper Washington DC, World Bank. **2911**.

Elbers, C. and R. van der Weide (2014). "Estimation of Normal Mixtures in a Nested Error Model with an Application to Small Area Estimation of Poverty and Inequality." Policy Research Working Paper **6962**.

Elia, L. (2007). Fighting gender-based violence in South Sudan.

Elia, L. (2007). Sexual violence: weapon of war, impediment to peace.

Emwanu, T., J. G. Hoogeveen and P. Okiira Okwi (2006). "Updating Poverty Maps with Panel Data." World Development **34**(12): 2076-2088.

Enders, W. and T. Sandler (1991). "Causality between Transnational Terrorism and Tourism: The Case of Spain." Terrorism **14**(1): 49 - 58.

Engstrom, R., J. Hersh and D. Newhouse (2017). Poverty from space: using high-resolution satellite imagery for estimating economic well-being, The World Bank.

Engstrom, R., J. Hersh and D. L. Newhouse (2021). "Poverty from space: Using high resolution satellite imagery for estimating economic well-being." The World Bank Economic Review.

Escamilla, V., Emch, M., Dandalo, L., Miller, W. C., Martinson, F., & Hoffman, I. (2014). "Sampling at community level by using satellite imagery and geographical analysis." Bulletin of the World Health Organization **92**(9): 690 - 694.

Espitia, A., A. Mattoo, N. Rocha, M. Ruta and D. Winkler (2022). "Pandemic trade: COVID-19, remote work and global value chains." The World Economy **45**(2): 561-589.

Etang, A. and K. Himelein (2020). Monitoring the Ebola Crisis Using Mobile Phone Surveys. Data Collection in Fragile States: Innovations from Africa and Beyond. J. Hoogeveen and U. Pape, Springer International Publishing**:** 15-32.

Etang, A., T. Hounsa and U. J. Pape (2022). Impact of High Inflation on Household Livelihoods in Urban South Sudan. World Bank Policy Research Working Paper. Washington,D.C., World Bank.

Euler, M., V. Krishna, S. Schwarze, H. Siregar and M. Qaim (2017). Oil Palm Adoption, Household Welfare, and Nutrition Among Smallholder Farmers in Indonesia. World Development**:** 219 - 235.

Fafchamps, M., D. Mckenzie, S. Quinn and C. Woodruff (2010). "Using PDA consistency checks to increase the precision of profits and sales measurement in panels."

Fafchamps, M., D. McKenzie, S. Quinn and C. Woodruff (2014). "Microenterprise Growth and the Flypaper Effect: Evidence from a Randomized Experiment in Ghana." Journal of Development Economics **106**: 211 - 226.

Fajardo, J. and M. Dantas (2018). "Understanding the impact of severe hyperinflation experience on current household investment behavior." Journal of Behavioral and Experimental Finance **17**.

FAO (2017). South Sudan Situation Report.

FAO, IFAD, UNICEF, WFP and WHO (2017). The state of food security and nutrition in the world 2017. Building resilience for peace and food security. Rome, FAO.

Farhood, L. and H. Dimassi (2012). "Prevalence and predictors for posttraumatic stress disorder, depression and general health in a population from six villages in South Lebanon." Soc Psychiatry Psychiatr Epidemiol **47**(4): 639–649.

Ferreira, F. H. G., S. Chen, A. Dabalen, Y. Dikhanov, N. Hamadeh, D. Jolliffe, A. Narayan, E. B. Prydz, A. Revenga, P. Sangraula, U. Serajuddin and N. Yoshida (2016). "A global count of the extreme poor in 2012: data issues, methodology and initial results." The Journal of Economic Inequality **14**(2): 141-172.

FEWSNET (2016). More than 1.1 million people in Somalia face acute food security crisis. Food Security and Nutrition Analysis Unit Somalia and Famine Early Warning Systems Network.

FEWSNET (2017). Somalia Food Security Alert. Severe drought, rising prices, continued access limitations, and dry forecasts suggest Famine is possible in 2017. Food Security and Nutrition Analysis Unit, Somalia and Famine Early Warning Systems Network.

FEWSNET (2017). Somalia Food Security Outlook. Risk of Famine (IPC Phase 5) persists in Somalia. Food Security and Nutrition Analysis Unit, Somalia and Famine Early Warning Systems Network.

FEWSNET. (2018, 2018). "South Sudan Food Security Outlook - October 2017 to May 2018."

FEWSNET, F. (2018). Somalia Food Security Outlook. Despite improvements, 2.7 million people need emergency assistance through the lean season, Food Security and Nutrition Analysis Unit – Somalia and Famine Early Warning Systems Network.

Fiedler, J. L. and D. M. Mwangi (2016). Improving household consumption and expenditure surveys' food consumption metrics: Developing a strategic approach to the unfinished agenda. IFPRI Discussion Paper, Intl Food Policy Res Inst.

Field, E. and e. al. (2013). "Do Traditional Institutions Constrain Female Entrepreneurship ? A Field Experiment on Business Training in India." American Economic Review: Papers & Proceedings **100**(2): 125 - 129.

Fielding, D. (2003). "Counting the Cost of the Intifada: Consumption, Saving and Political Instability in Israel." Public Choice **116**(3): 297 - 312.

Filmer, D. and L. H. Pritchett (2001). "Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India." Demography: 115 - 132.

Findley, M. G., A. S. Harris, H. V. Milner and D. L. Nielson (2017). "Who Controls Foreign Aid? Elite versus Public Perceptions of Donor Influence in Aid-Dependent Uganda." International Organization: 1 - 31.

Finn, A. and V. Ranchhod (2015). "Genuine Fakes: The Prevalence and Implications of Data Fabrication in a Large South African Survey." World Bank Economic Review(129203).

Firchow, P. and R. Mac Ginty (2016). "Including Hard-to-Access Populations Using Mobile Phone Surveys and Participatory Indicators." Sociological Methods & Research: 0049124117729702.

Firpo, S., N. M. Fortin and T. Lemieux (2009). "Unconditional quantile regressions." Econometrica **77**(3): 953 - 973.

Fischer, S., R. Sahay and C. A. Vegh (2002). "Modern Hyper- and High Inflations." Journal of Economic Literature **40**: 837 - 880.

Foster, J., J. Greer and E. Thorbecke (1984). "A Class of Decomposable Poverty Measures." Econometrica **52**(3): 761-766.

FSNAU (2017). Focus on Post Gu 2017 Early Warning. Food Security and Nutrition Analysis Unit. Somalia.

Fujii, T. and R. Van der Weide (2013). Cost-effective estimation of the population mean using prediction estimators, The World Bank.

Fujii, T. and R. van der Weide (2016). "Is Predicted Data a Viable Alternative to Real Data?" World Bank Policy Research Working Paper(7841).

Gaibulloev, K. and T. Sandler (2009). "The Impact of Terrorism and Conflicts on Growth in Asia." Economics & Politics **21**(3): 359 - 383.

Galway, L., N. Bell, A. S. Sae, A. Hagopian, G. Burnham, A. Flaxman, W. M. Weiss, J. Rajaratnam and T. K. Takaro (2012). "A Two-Stage Cluster Sampling Method Using Gridded Population Data, a GIS, and Google Earth Imagery in a Population-Based Mortality Survey in Iraq." International Journal of Health Geographics **11**(12).

Garner T., McClelland R. and P. W. (2009). Strengths and Weaknesses of the Consumer Expenditure Survey from a BLS Perspective. Washington, DC., the Bureau of Labor Statistics.

Gawn, G. and R. Innes (2018). "Do Lies Erode Trust?" International Economic Review **59**(1): 137 - 161.

Gibson, J. (2018). "Forest Loss and Economic Inequality in the Solomon Islands: Using Small - Area Estimation to Link Environmental Change to Welfare Outcomes." Ecological Economics **148**(February): 66-76.

Gilens, M., P. M. Sniderman and J. H. Kuklinski (1998). "Affirmative Action and the Politics of Realignment." British Journal of Political Science **28**(1): 159-183.

Gneezy, U. (2005). "Deception: The Role of Consequences." American Economic Review **95**(1): 384-394.

Goodall, C. R. (1993). "13 Computation using the QR decomposition."

Google (2021). "Google COVID-19 Community Mobility Reports."

Graham, J. W., S. M. Hofer and D. P. MacKinnon (1996). "Maximizing the Usefulness of Data Obtained with Planned Missing Value Patterns: An Application of Maximum Likelihood Procedures." **31**(2): 197 - 218.

Grais, R. F., A. M. Rose and J.-P. Guthmann (2007). Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys. Emerging Themes in Epidemiology 4. **8**.

Grosh, M. E. and J. Munoz (1996). A Manual for Planning and Implementing the Living Standards Measurement Study Survey. LSMS Working Paper. Washington, D.C., The World Bank. **126**.

Guadarrama, M., I. Molina and J. N. K. Rao (2016). "A Comparison of Small Area Estimation Methods for Poverty Mapping." Statistics in Transition new series **1**(17): 41-66.

Gupta, J., M. Bavinck, M. Ros-Tonen, K. Asubonteng, H. Bosch, E. van Ewijk, M. Hordijk, Y. Van Leynseele, M. L. Cardozo and E. Miedema (2021). "COVID-19, poverty and inclusive development." World Development **145**: 105527.

Hale, T., N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar and H. Tatlow (2021). "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)." Nature Human Behaviour.

Hallegatte, S., M. Bangalore, L. Bonzanigo, M. Fay, T. Kane, U. Narloch, J. Rozenberg, D. Treguer and A. Vogt-Schilb (2016). Shock Waves: Managing the Impacts of Climate Change on Poverty. Climate Change and Development Series. Washington, D.C., World Bank.

Hallsworth, M., J. List, R. Metcalfe and I. Vlaev (2017). "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance." Journal of Public Economics **148**(C): 14-31.

Hamilton, D. (1968). "Personality attributes associated with extreme response style. - PubMed - NCBI." sychological Bulleting **69**(3): 192-203.

Haslett, S. J. (2016). Small Area Estimation Using Both Survey and Census Unit Record Data. Analysis of Poverty Data by Small Area Estimation. Chichester, UK, John Wiley & Sons, Ltd**:** 325-348.

Haslett, S. J., M. C. Isidro and G. Jones (2010). "Comparison of Survey Regression Techniques in the Context of Small Area Estimation of poverty." Survey Methodology **36**(2): 157-170.

Haughton, J. and S. R. Khandker (2009). Handbook on Poverty & Inequality. Washington DC, World Bank Publications.

Haushofer, J. and J. Shapiro (2013). Household Response to Income Changes: Evidence from an Unconditional Cash Transfer Program in Kenya, Massachusetts Institute of Technology **24:** 1 - 57.

Haushofer, J. and J. Shapiro (2016). "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." The Quarterly Journal of Economics **131**(4): 1973 - 2042.

He, L. (2017). "Hyperinflation: A world history." Routledge.

He, L. (2020). Hyperinflation: A world history, Routledge.

Healy, A. J., S. Jitsuchon and Y. Vajaragupta (2003). "Spatially Disaggregated Estimates of Poverty and Inequality in Thailand." Massachusetts Institute of Technology and Thailand Development Research Institute.

Heemann, M., U. Pape and S. Vollmer (2022). "The Labor Market Implications of Restricted Mobility during the COVID-19 Pandemic in Kenya." World Bank Policy Research Working Paper(9963).

Hejoj, I. (2007). "A profile of poverty for Palestinian refugees in Jordan: The case of Zarqa and Sukhneh camps." Journal of Refugees Studies **20**(1): 120 _ 145.

Henderson, J. V., A. Storeygard and D. N. Weil (2012). "Measuring economic growth from outer space." American economic review **102**(2): 994–1028.

Heß, S. (2017). "Randomization inference with Stata: A guide and software." Stata Journal: 630 - 651.

Hill, R. and C. Mejia-Mantilla (2017). "With a little help: shocks, agricultural income, and welfare in Uganda." Policy Research Working Paper, World Bank(7935).

Hill, R. and C. Porter (2016). "Vulnerability to Drought and Food Price Shocks." Policy Research Working Paper, World Bank(7920).

Himelein, K., S. Eckman and S. Murray (2014). "Sampling nomads: a new technique for remote, hard-to-reach, and mobile populations." Journal of official statistics **30**: 191 - 213.

Himelein, K., S. Eckman, S. Murray and J. Bauer (2016). "Second-stage sampling for conflict areas: methods and implications."

Himelein, K., S. Eckman, S. Murray and J. Bauer (2017). "Alternatives to full listing for second stage sampling: Methods and implications." Statistical Journal of the IAOS **33**(3): 701 - 718.

Himelein, K., U. Pape and M. Wild (forthcoming). "Implications of Choice of Second Stage Selection Method on Sampling Error and Non-Sampling Error: Evidence from an IDP camp in South Sudan." World Bank Policy Research Working Paper(0000).

Hoddinott, J. and B. Kinsey (2001). "Child growth in the time of drought." Oxford Bulletin of Economics and statistics **63**(4): 409–436.

Hoffmeyer-Zlotnik, J. H. (2003). New sampling designs and the quality of data. Developments in Applied Statistics, Ljubljana: FDV Methodoloski Zvezki**:** 205 - 2017.

Hoogeveen, J. and N. T. V. Nguyen (2017). "Statistics Reform in Africa: Aligning Incentives with Results." The Journal of Development Studies: 1–18.

Hoogeveen, J. and U. J. Pape (2020). Data Collection in Fragile Countries, Palgrave Macmillan.

Hoogeveen, J., M. Rossi and D. Sansone (2019). "Leaving, Staying, or Coming Back? Migration Decisions During the Northern Mali Conflict." The Journal of Development Studies **T 55**(10): 2089 – 2105.

Hurd, M. D. (1999). "Anchoring and Acquiescence Bias in Measuring Assets in Household Surveys." Journal of Risk and Uncertainty **19**(1-3): 111-136.

Ibáñez, A. M., V. Calderón and M. Gafaro (2011). "Forced Migration, Female Labour Force Participation, and Intra-Household Bargaining: Does Conflict Empower Women?" Institute of Development Studies.

ICRC (2001). Women Facing War. Geneva, International Committee of the Red Cross.

IFC (2012). Demographic and health survey sampling and household listing manual. Calverton, MD. USA, ICF International.

ILO (2021). ILO Monitor: COVID-19 and the world of work. Seventh edition, Updated estimates and analysis. , ILO.

Imai, K. (2011). "Multivariate Regression Analysis for the Item Count Technique." Journal of the American Statistical Association **106**(494): 407 - 416.

Imbens, G. and J. Wooldridge (2007). Difference-in-Differences Estimation. Research Working Paper, National Bureau of Economics.

IMF (2016). "Article IV Consultation Report, February 2017." **4**.

Isidro, M. C. (2010). Intercensal Updating of Small Area Estimates: a Thesis Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics at Massey University, Palmerston North, New Zealand.

Isidro, M. C., S. Haslett and G. Jones (2016). "Extended Structure Preserving Estimation (ESPREE) for Updating Small Area Estimates of Poverty." Annals of Applied Statistics **10**(1): 451-476.

Jakiela, P. and O. Ozier (2015). The Impact of Violence on Individual Risk Preferences: Evidence from a Natural Experiment. World Bank Policy Research Paper **9870**.

Jakubowski, A., D. Egger, C. Nekesa, L. Lowe, M. Walker and E. Miguel (2021). Self-Reported Mask Wearing Greatly Exceeds Directly Observed Use: Urgent Need for Policy Intervention in Kenya. medRxiv

James, P. (2017). Creating capacities for human flourishing: an alternative approach to human development. Cultures of Sustainability and Wellbeing, Routledge**:** 23-45.

Janssens, W., M. Pradhan, R. de Groot, E. Sidze, H. P. P. Donfouet and A. Abajobir (2021). "The short-term economic effects of COVID-19 on low-income households in rural Kenya: An analysis using weekly financial household data." World Development **138**: 105280.

Jarvis, C. I., K. Van Zandvoort, A. Gimma, K. Prem, CMMID COVID-19 working group, P. Klepac, G. J. Rubin and W. J. Edmunds (2020). "Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK." BMC Med **18**(124).

Josephson, A., T. Kilic and J. D. Michler (2020). Socioeconomic Impacts of COVID-19 in Four African Countries. World Bank Policy Research Working Paper. **9466**.

Justino, P. (2009). "Poverty and Violent Conflict: A Micro-Level Perspective on the Causes and Duration of Warfare." Journal of Peace Research **46**(3): 315-333.

Justino, P. (2012). "War and Poverty." IDS Working Papers **2012**(391): 1-29.

Justino, P., C. Ivan, M. Becky and M. Catherine (2012). Women Working for Recovery: The Impact of Female Employment on Family and Community Welfare after Conflict, in UN Women, UN Women Sourcebook on Women, Peace and Security. New York, United Nations Entity for Gender Equality and the Empowerment of Women.

Justino, P., M. Leone and P. Salardi (2014). "Short and Long-Term Impact of Violence on Education: The Case of Timor Leste." World Bank Economic Review **28**(2): 320 - 353.

Justino, P. and P. Verwimp (2013). "Poverty Dynamics, Violent Conflict and Convergence in Rwanda." The Review of Income and Wealth **59**(1).

Kalsbeek, W. D. (1986). Nomad sampling: an analytic study of alternative design strategies. Proceedings of the Section on Survey Research Methods.

Kansiime, M. K., J. A. Tambo, I. Mugambi, M. Bundi, A. Kara and C. Owuor (2021). "COVID-19 implications on household income and food security in Kenya and Uganda: Findings from a rapid assessment." World development **137**: 105199.

Kaplan, L., U. Pape and J. Walsh (2018). "A Light-Touch Method to Improve Accurate Reporting of IDP's Food Consumption." World Bank Policy Research Working Paper(8539).

Kassié, D., A. Roudot, N. Dessay, J.-L. Piermay, G. Salem and F. Fournet (2017). "Development of a spatial sampling protocol using GIS to measure health disparities in Bobo-Dioulasso, Burkina Faso, a medium-sized African city." International Journal of Health Geographics **16**(1): 14.

Keairns, Y. E. (2003). The Voices of Girl Soldiers, New York: Quaker United Nations Office.

Kenya National Bureau of Statistics (2018). Basic Report on Well-Being in Kenya.

Khamis, M., D. Prinz, D. Newhouse, A. Palacios-Lopez, U. Pape and M. Weber (2021). "The Early Labor Market Impacts of COVID-19 in Developing Countries : Evidence from High-Frequency Phone Surveys." World Bank Policy Research Working Paper(9510).

Kijima, Y. and P. Lanjouw (2003). "Poverty in India During the 1990s: A Regional Perspective." Policy Research Working Paper, World Bank.

Kish, L. (1965). Survey Sampling. Wiley Classics Library Edition. New York, John Wiley & Sons.

Koenker, R. and G. Bassett Jr (1978). "Regression quantiles." Econometrica **46**(1).

Kolbe, A. R. and R. A. Hutson (2006). "Human Rights Abuse and Other Criminal Violations in Port-Au-Prince, Haiti: A Random Survey of Households." The Lancet **369**(9538): 864 - 873.

Kondo, M. C., K. D. W. Bream, F. K. Barg and C. C. Branas (2014). "A Random Spatial Sampling Method in a Rural Developing Nation." BMC Public Health **14**(1): 338.

Korinek, A., J. A. Mistiaen and M. Ravallion (2006). "Survey Nonresponse and the Distribution of Income." The Journal of Economic Inequality 4 (1): 33–55.

Kreuter, F., S. McCullock, S. Presser and R. Tourangeau (2011). "The effects of asking filter questions in interleafed versus grouped format." Sociological Methods and Research(1): 88-104.

Kriechbaum-Vitellozzi, E. and R. Kreuzbauer (2006). "Poverty consumption: consumer behaviour of refugees in industrialized countries."

Krizhevsky, A., I. Sutskever and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks.

Krosnick, J. A. (1991). "Response strategies for coping with the cognitive demands of attitude measures in surveys." Applied Cognitive Psychology **5**(3): 213-236.

Krueger, A. B. and J. Malečková (2003). "Education, Poverty and Terrorism: Is There a Causal Connection?" Journal of Economic Perspectives **17**(4): 119 - 144.

Kuklinski, J. H. (1997). "Racial Prejudice and Attitudes Toward Affirmative Action." American Journal of Political Science **41**(2): 402 - 419.

Kumar, N. (2007). Spatial Sampling Design for a Demographic and Health Survey. Population Research and Policy Review. **26:** 581 - 599.

Kurasha, F. M. N. (2021). Wealth, health, and inequality: a study of hyperinflation and households in Zimbabwe. PhD, University of Cape Town.

Kydd, A. H. and B. F. Walter (2006). "The Strategies of Terrorism." International Security **31**(1): 49 - 80.

LaFree, G., L. Dugan, M. Xie and P. Singh (2012). "Spatial and Temporal Patterns of Terrorist Attacks by ETA 1970 to 2007." Journal of Quantitative Criminology **28**(1): 7 - 29.

Lange, S., U. Pape and P. Putz (2022). "Small area estimation of poverty under structural change." Review of Income and Wealth.

Lanjouw, J. O. (1999). Demystifying Poverty Lines. New York City, NYC, United Nations Development Programme.

Lanjouw, P. and M. Ravallion (1995). "Poverty and household size." The Economic Journal: 1415 - 1434.

Larochelle, C., J. Alwang and N. Taruvinga (2014). "Inter-temporal Changes in Well-being During Conditions of Hyperinflation: Evidence from Zimbabwe." Journal of African Economies **23**(2).

Lin, W. and D. P. Green (2016). "Standard Operating Procedures: A Safety Net for Pre-Analysis Plans." Political Science & Politic: 495 - 500.

Lin, Y. and D. P. Kuwayama (2016). "Using satellite imagery and GPS technology to create random sampling frames in high risk environments." International Journal of Surgery **32**: 123 -128.

Linard, C., V. A. Alegana, A. M. Noor, R. W. Snow and A. J. Tatem (2010). "A high resolution spatial population database of Somalia for disease risk mapping." International Journal of Health Geographics **9**(13).

Little, R. J. and D. B. Rubin (2019). Statistical analysis with missing data, John Wiley & Sons.

Liu, J., V. Iannacchione and M. Byron (2002). Decomposing design effects for stratified sampling. Proceedings of the survey research methods American Statistical Association**:** 2124 - 2126.

Luitel, N., M. Jordans, R. Sapkota, W. Tol, B. Kohrt, S. Thapa, I. Komproe and B. Sharma (2013). "Conflict and mental health: a cross-sectional epidemiological study in Nepal." Psychiatry Psychiatr Epidemiol **48**(2): 183–193.

Lupu, N. and K. Michelitch (2018). Advances in Survey Methods for the Developing World. Annual Review of Political Science. **21:** 195 - 214.

Luttmer, E. F. P. (2005). "Neighbor as Negatives: Relative Earnings and Well-Being." The Quarterly Journal of Economics **120**(3): 963 - 1002.

Makoka, D. (2008). "The impact of drought on household vulnerability: The case of rural Malawi." MPRA(MPRA Paper No. 15399).

Mancini, G. and G. Vecchi (2022). On the Construction of a Consumption Aggregate for Inequality and Poverty Analysis. Washington, D.C., World Bank.

Manley, J., S. Gitter and V. Slavchevska (2013). "How Effective Are Cash Transfers at Improving Nutritional Status?" World Development **48**: 133 - 155.

Mano, Y., A. Iddrisu, Y. Yoshino and T. Sonobe (2012). "How Can Micro and Small Enterprises in Sub-Saharan Africa Become More Productive? The Impacts of Experimental Basic Managerial Training." World Development **40**(3): 458 - 468.

Marhuenda, Y., I. Molina, D. Morales and J. N. K. Rao (2017). "Poverty Mapping in Small Areas Under a Twofold Nested Error Regression Model." Journal of the Royal Statistical Society. Series A: Statistics in Society **180**(4): 1111-1136.

Martin-Shields, C. P. and W. Stojetz (2019). "Food security and conflict: Empirical challenges and future opportunities for research and policy making on food security and conflict." World Development **119**: 150 - 164.

Mazar, N., O. Amir and D. Ariely (2008). "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." Journal of Marketing Research **45**(6): 633-644.

Mazar, N. and D. Ariely (2006). "Dishonesty in Everyday Life and Its Policy Implications." Journal of Public Policy & Marketing **25**(1): 117-126.

McKay, S., M. Robinson, M. Gonsalves and M. Worthen (2006). Girls Formerly Associated with Fighting Forces and their Children: Returned and Neglected. London, Coalition to Stop the Use of Child Soldiers.

McKenzie, D. and C. Woodruff (2014). "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?" World Bank Research Observer **29**(1): 48 - 82.

McKenzie, D. J. (2005). "Measuring inequality with asset indicators." Population Economics **18**(2): 229 - 260.

Mellander, C., J. Lobo, K. Stolarick and Z. Matheson (2015). "Night-Time Light Data: A Good Proxy Measure for Economic Activity?" PLOS ONE **10**(10): e0139779.

Menon, N. and Y. v. d. M. Rodgers (2013). "War and Women's Work: Evidence from the Conflict in Nepal." Journal of Conflict Resolution.

Mercier, M., R. Ngenzebuke and V. Philip (2016). The Long-Term Effects of Conflict on Welfare: Evidence from Burundi, Households in Conflict Network.

Meyer, B. (1995). "Natural and Quasi-Experiments in Economics." Journal of Business and Economic Statistics **13**(2): 151 - 161.

Miguel, E. and G. Roland (2011). "The long-run impact of bombing Vietnam." Journal of Development Economics **96**(1): 1-15.

Miller, J. (1984). "A New Survey Technique for Studying Deviant Behavior." Phd thesis.

Milner, H. V., D. L. Nielson and M. G. Findley (2016). "Citizen Preferences and Public Goods: Comparing Preferences for Foreign Aid and Government Programs in Uganda." Review of International Organizations **11**(2): 219 - 245.

Minasny, B., A. B. McBratney and D. J. Walvoort (2007). "The variance quadtree algorithm: Use for spatial sampling design." Computers & Geosciences **33**(383 - 392).

Ministry of General Education and Instruction (MoGEI) (2016). "Education Management and Information systems ".

Minnesota Population Center (2017). Integrated Public Use Microdata Series, International: Version 6.5 [dataset]. Minneapolis, MN: University of Minnesota.

Minoiu, C. and O. N. Shemyakina (2014). "Armed conflict, household victimization, and child health in Côte d'Ivoire." Journal of Development Economics **108**: 237 - 255.

Miranda, A. (2003). "Cogs in the Wheel? Women in the Liberation Tigers of Tamil Eelam." Civil Wars **6**(4): 37 - 54.

Mneimneh, Z. N., B.-E. Pennell, J. Kelley and K. C. Hibben (2016). Surveys in societies in turmoil. The SAGE Handbook of Survey Methodology 178.

Molina, I. and J. N. K. Rao (2010). "Small Area Estimation of Poverty Indicators." Canadian Journal of Statistics **38**(3): 369-385.

Montalvo, J. G. and M. Reynal-Querol (2007). "Fighting against Malaria: Prevent Wars while Waiting for the "Miraculous" Vaccine." The Review of Economics and Statistics **89**(1): 165 - 177.

Moya, A. (2015). Violence, mental trauma, and induced changes in risk attitudes among the internally displaced population in Colombia. Working Paper, Agricultural and Resource Economics, UC Davis. Sacramento, University of Southern California-Davis.

Müller, A., U. Pape and L. Ralston (2019). "Broken Promises: Evaluating an incomplete Cash Transfer Program." World Bank Policy Research Working Paper **no 9016**.

Muñoz, J. and E. Langeraar (2013). "A census independent sampling strategy for a household survey in Myanmar." Available at: bit. ly/TU94rr.

Murthy, S. and R. Lakshminarayani (2006). "Mental health consequences of war: a brief review of research findings." World Psychiatry **5**(1): 25-30.

Mvukiyehe, E. and C. Samii (2008/9). Laying a foundation for peace? A quantitative impact evaluation of united nations operations in Cote d'Ivoire (with technical appendix). Basis of UN-OIOS report to the UN General Assembly. A/63/713, UN Office of Internal Oversight Services

Nasir, M., M. Rockmore and C. Ming Tan (2015). It's No Spring Break in Cancun: The Effects of Exposure to Violence on Risk Preferences, Pro-Social Behavior, and Mental Health in Mexico. Rimini, the Rimini Centre for Economic Analysis.

Neal, J., M. Burke, M. Xie, W. M. Davis, D. B. Lobell and S. Ermon (2016). "Combining satellite imagery and machine learning to predict poverty." Science **353**(6301): 790–794.

Nechifor, V., E. Ferrari, E. Kihiu, J. Laichena, D. Omanyo, R. Musamali and B. Kiriga (2020). COVID-19 impacts and short-term economic recovery in Kenya. J. R. C. (JRC). Luxembourg, the European Commission's science and knowledge service.

Newhouse, D., R. Swindle, S. Wang, J. D. Merfeld, U. Pape, K. Tafere and M. Weber (2022). How Well Did Real-time Indicators Track Household Welfare Changes in Developing Countries During the Covid-19 Crisis? World Bank Policy Research Working Paper. Washington, D.C., World Bank.

Neyman, J. (1934). "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection." Journal of the Royal Statistical Society. Series A: Statistics in Society **97**(558 - 625).

Nguyen, M. C., H. Wu, C. Lakner and M. Schoch (2021). Update to the Multidimensional Poverty Measure : What's New. Global Poverty Monitoring Technical Note. Washington, DC, World Bank. **17**.

Nitsch, V. and D. Schumacher (2004). "Terrorism and International Trade: An Empirical Investigation." European Journal of Political Economy, The Economic Consequences of Terror **20**(2): 423 - 433.

Nordhaus, W. D. (1973). "The Effect of Inflation on the Distribution of Economic Welfare." Journal of Money, Credit and Banking **5**(1).

Nunez-Chaim, G. and U. Pape (2022). Poverty and violence: The immediate impact of terrorist attacks against civilians in Somalia. World Bank Policy Research Working Paper. Washington, D.C., World Bank.

Nussbaum, M. C. (2011). Creating Capabilities: The Human Development Approach, Harvard University Press.

O'Muircheartaigh, C., E. English and S. Eckman (2007). Predicting the relative quality of alternative sampling frames. Proceedings of the Survey Research Methods Section, American Statistical Association.

Olson-Lanjouw, J. and P. Lanjouw (2001). "How to compare apples and oranges: Poverty measurement based on different definitions of consumption." Review of Income and Wealth **47**(1): 25–42.

OPHI (2018). Global Multidimensional Poverty Index 2018: The Most Detailed Picture to Date of the World's Poorest People. University of Oxford, Oxford Poverty and Human Development Initiative.

Oseni, G., J. Durazo and K. McGee (2017). "The Use of Non-Standard Units for the Collection of Food Quantity."

Osier, G. (2016). Unit Non-Response in Household Wealth Surveys, European Central Bank. **15**.

Ostby, G. (2016). "Violence Begets Violence: Armed conflict and domestic sexual violence in Sub-Saharan Africa." HiCN Working Papers.

Ozer, E. J., L. C. H. Fernald, A. Weber, E. P. Flynn and T. J. VanderWeele (2011). "Does Alleviating Poverty Affect Mothers' Depressive Symptoms ? A Quasi-Experimental Investigation of Mexico's Oportunidades Programme." International Journal of Epidemiology: 1565 - 1576.

Pape, U. (2020). Rapid Emergency Response Survey. Data Collection in Fragile States: Innovations from Africa and Beyond. J. Hoogeveen and U. Pape, Springer International Publishing**:** 15-32.

Pape, U. (2020). Using Video Testimonials to Give a Voice to the Poor. Data Collection in Fragile States. J. H. a. U. Pape, World Bank.

Pape, U. (2021). Measuring Poverty Rapidly Using Within-Survey Imputations. Policy Research Working Paper. Washington, D.C., World Bank.

Pape, U., and J Mistiaen (2015). Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu. Washington DC: World Bank. Proceedings of ABCA. Washington DC: World Bank.

Pape, U., J. Baraibar Molina, A. S. Delius, C. L. Gitau and L. A. Rios Rivera (2020). Socioeconomic Impacts of COVID-19 in Kenya on Households: Rapid Response Phone Survey Round One. Washington, D.C., World Bank.

Pape, U., T. Beltramo, F. K. Appler, J. Fix, F. Nimoh, P. Nan and L. A. Rios Rivera (2019b). Understanding the Socio-Economic Conditions of Refugees in Kenya Kakuma Camp : Results from the 2019 Kakuma Socioeconomic Survey. Washington, D.C, World Bank. **B**.

Pape, U., M. Benson, M. Ebrahim and J. Lole (2017). Reducing poverty through improved agro-logistics in a fragile country: findings from a trader survey in South Sudan. Washington, D.C., World Bank.

Pape, U., A. Delius, A. Haynes, E. Cameron, R. Gupta, R. Khandelwal, S. Siewers and Y. Xu (2021). How COVID-19 Continues to Affect Livelihoods in Kenya : Rapid Response Phone Survey Rounds 1 to 5. Washington, D.C., World Bank.

Pape, U., A. Delius, R. Khandelwal and R. Gupta (2021). Socioeconomic Impacts of COVID-19 in Kenya. Washington, D.C, World Bank.

Pape, U. and N. C. Dihel (2017). South Sudan - economic update : taming the tides of high inflation - policy options. Washington, D.C., World Bank.

Pape, U., K. Huckvale, J. Car, A. Majeed and C. Millett (2015). "Impact of 'Stretch' Targets for Cardiovascular Disease Management within a Local Pay-for-Performance Programme." PLOS ONE **10**(3): e0119185.

Pape, U. and R. W. Karamba (2019). Findings from Wave 2 of the Somali High Frequency Survey. Washington, D.C., World Bank**:** 1-206.

Pape, U., C. Millett, J. T. Lee, J. Car and A. Majeed (2013). "Disentangling secular trends and policy impacts in health studies: use of interrupted time series analysis." Journal of the Royal Society of Medicine **106**(4): 124-129.

Pape, U. and J. Mistiaen (2014). Utilizing Mobile Technology to Innovate CAPI Data Collection in Fragile Contexts. Washington, D.C., World Bank.

Pape, U. and J. Mistiaen (2018). Household expenditure and poverty measures in 60 minutes : a new approach with results from Mogadishu. Policy Research Working Paper. Washington, D.C., World Bank.

Pape, U. and L. Parisotto (2019). Estimating Poverty in a Fragile Context - The High Frequency Survey in South Sudan. Policy Research Working Paper. Washington, D.C., World Bank**:** 1-42.

Pape, U., L. Parisotto, V. Phipps-Ebeler, A. J. M. Mueller, L. R. Ralston, T. Nezam and A. Sharma (2018). Impact of Conflict and Shocks on Poverty : South Sudan Poverty Assessment 2017. Washington, D.C., World Bank**:** 1-288.

Pape, U. and V. Phipps (2018). Impact of conflict on adolescent girls in South Sudan. Policy Research Working Paper. Washington, D.C., World Bank.

Pape, U. and A. Sharma (2019). Using Micro-Data to Inform Durable Solutions for IDPs. Washington, D.C., World Bank.

Pape, U. and P. Verme (2023). Measuring Poverty in Forced Displacement Contexts. Policy Research Working Paper. Washington, D.C., World Bank.

Pape, U. and P. R. Wollburg (2019). Estimation of Poverty in Somalia Using Innovative Methodologies. Policy Research Working Paper. Washington, D.C., World Bank**:** 1-62.

Pape, U. and P. R. Wollburg (2019). Impact of Drought on Poverty in Somalia. Policy Research Working Paper. Washington, D.C., World Bank.

Pape, U. J. and L. Parisotto (2019). "Estimating Poverty in a Fragile Context -- The High Frequency Survey in South Sudan." World Bank Policy Research Working Paper **no 8722**: 1-42.

Pape, U. J. S., A.; Nezam, T.; Petrini, B.; Ebrahim, M.F.; Udo-Udo, J.; Appler, F.K.; Woodhouse, A.F.; Phipps-Ebeler, V.; Meckelburg, A.B.; Iqbal, S.A. (2019a). "Using Micro-Data to Inform Durable Solutions for IDPs."

Parisotto, L. and U. Pape (forthcoming). Impact of Conflict on Livelihoods in South Sudan. Policy Research Working Paper. Washington, D.C., World Bank.

Pasquale, A. D., R. S. McCann and N. Maire (2017). "Assessing the population coverage of a health demographic surveillance system using satellite imagery and crowd-sourcing." PLOS ONE **12**.

Pinheiro, J. C. and D. M. Bates (2000). Mixed Effects Models in S and S-PLUS, Springer.

Pinkovskiy, M. and X. Sala-i-Martin (2016). "Lights, Camera… Income! Illuminating the national accounts-household surveys debate." The Quarterly Journal of Economics **131**(2): 579–631.

Porter, C. (2012). "Shocks, Consumption and Income Diversification in Rural Ethiopia." Journal of Development Studies **48**(9): 1209-1222.

Poterba, J. M. and L. H. Summers (1986). "Reporting Errors and Labor Market Dynamics." Econometrica **54**(6): 1319-1338.

Powell, G. N. and D. A. Butterfield (2003). Gender, gender identity, and aspirations to top management. Women in Management Review. **18:** 88 - 96.

Qader, S., V. Lefebvre, A. Ninneman, K. Himelein, U. Pape, L. Bengtsson, A. Tatem and T. Bird (2019). "A novel approach to the automatic designation of pre-defined census Enumeration Areas and population sampling frames: A case study in Somalia." World Bank Policy Research Working Paper(8972).

Qader, S., V. Lefebvre, A. Tatem, U. Pape, K. Himelein, A. Ninneman, L. Bengtsson and T. Bird (2021). "Semi-automatic mapping of pre-census enumeration areas and population sampling frames." Humanities and Social Sciences Communications **8**(1): 1-14.

Qader, S. H., V. Lefebvre, A. J. Tatem, U. Pape, W. Jochem, K. Himelein, A. Ninneman, P. Wolburg, G. Nunez-Chaim, L. Bengtsson and T. Bird (2020). "Using gridded population and quadtree sampling units to support survey sample design in low-income settings." International Journal of Health Geographics **19**(1): 10.

Raleigh, C. and C. Dowd (2015). Armed Conflict Location and Event Data Project (ACLED) Codebook.

Rasinski, K. A., P. S. Visser, M. Zagatsky and E. M. Rickett (2005). "Using implicit goal priming to improve the quality of self-report data." Journal of Experimental Social Psychology **41**(3): 321 - 327.

Ravallion, M. (1998). Poverty Lines in Theory and Practice. Washington, D.C., The World Bank.

Ravallion, M. (2008). Poverty lines. The New Palgrave Dictionary of Economics 2.

Ravallion, M. (2017). "Poverty comparisons." Routledge.

Ravallion, M. (2020). "On Measuring Global Poverty." Annural Review of Economics **12**: 167 - 188.

Ravallion, M. and B. Benu (1994). How Robust Is a Poverty Profile? The World Bank Economic Review. **8:** 75 - 102.

Ravallion, M., S. Chen and P. Sangraula (2009). "Dollar a Day Revisited." The World Bank Economic Review **23**(2): 163-184.

Ravallion, M. and M. Lokshin (2006). "Testing Poverty Lines." Review of Income and Wealth **52**: 399 - 421.

Reinhart, C. and M. Savastano (2003). "The Realities of Modern Hyperinflation." Finance & Development **40**(2).

Riegle-Crumb, C., C. Moore and A. Ramos-Wada (2011). "Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity." Science Education **95**(3): 458 - 476.

Rigby, R. A. and D. M. Stasinopoulos (2005). "Generalized Additive Models for Location, Scale and Shape." Applied Statistics **54**: 507-554.

Roberts, B., S. Guy, E. Sondorp and L. Lee-Jones (2008). "A basic package of health services for post-conflict countries: implications for sexual and reproductive health services." Reprod Health Matters **16**(31): 57 - 64.

Roberts, B., K. Ocaka, J. Browne, T. Oyok and E. Sondorp (2008). "Factors associated with post-traumatic stress disorder and depression amongst internally displaced persons in northern Uganda." BMC Psychiatry **8**(38).

Rojas-Perilla, N., S. Pannier, T. Schmid and N. Tzavidis (2017). Data-driven Transformations in Small Area Estimation.

Rolstad, S., J. Adler and A. Rydén (2011). "Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis." Value in Health **14**(8): 1101-1108.

Rosenfeld, B., K. Imai and J. N. Shapiro (2016). "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." American Journal of Political Science **60**(3): 783-802.

Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys, John Wiley & Sons.

Ruiz Estrada, M. Arturo, D. Park and A. Khan (2018). "The Impact of Terrorism on Economic Performance: The Case of Turkey." Economic Analysis and Policy **60**: 78 - 88.

Saha, J., B. Barman and P. Chouhan (2020). Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, Children and Youth Services Review. **116**.

Samuelson, P. A. (1974). "Complementarity - An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory." Journal of Economic Literature **15**: 24 - 55.

Särndal, C. and J. Wretman (2003). Model Assisted Survey Sampling. New York.

Schaaf, Z. (2015). MCD43A3 MODIS/Terra+ Aqua BRDF/Albedo Daily L3 Global–500m V006, NASA EOSDIS Land Processes DAAC.

Schräpler, J.-P., Schupp, Jürgen , and Gert G. Wagner (2010). "Changing from PAPI to CAPI: Introducing CAPI in a Longitudinal Study." Journal of Official Statistics **26**(2): 239 - 269.

Sedda, L., A. J. Tatem, D. W. Morley, P. M. Atkinson, N. A. Wardrop, C. Pezzulo, A. Sorichetta, J. Kuleszo and D. J. Rogers (2015). "Poverty, health and satellite-derived vegetation indices: their inter-spatial relationship in West Africa." International Health **7**: 99 - 106.

Sen, A. (1985). Commodities and Capabilities. Amsterdam, North-Holland.

Sen, A. (1987). The Standard of Living. Cambridge, Cambridge University Press.

Shankleman, J. (2011). Oil and State Building in South Sudan, United States Institute of Peace.

Shannon, H. S., R. Hutson, A. Kolbe, B. Stringer and T. Haines (2012). Choosing a survey sample when data on the population are limited: A method using Global Positioning Systems and aerial and satellite photographs. Emerging Themes in Epidemiology. **9**.

Shao, J. (1997). An Asymptotic Theory for Linear Model Selection. Statistica Sinica. **7:** 221 - 264.

Shaver, A. and J. N. Shapiro (2016). The Effect of Civilian Casualties on Wartime Informing Evidence from the Iraq War. HiCN WP 210. Brighton, Households in Conflict Network, The Institute of Development Studies - at the University of Sussex.

Shemyakina, O. (2011). "The effect of armed conflict on accumulation of schooling: Results from Tajikistan." Journal of Development Economics **95**(2): 186-200.

Siegelman, L. (1981). "Question-Order Effects on Presidential Popularity." Public Opinion Qarterly: 199 - 207.

Slesnick, D. T. (2001). Consumption and Social Welfare: Living Standards and their Distribution in the United States. Cambridge, Cambridge University Press.

Smith, L. C., O. Dupriez and N. Troubat (2014). "Assessment of the reliability and relevance of the food data collected in national household consumption and expenditure surveys." International Household Survey Network.

Snow, J. (1855). On the Mode of Communication of Cholera. London, John Churchill.

Snyder, C. F., M. E. Watson, J. D. Jackson, D. Cella and M. Y. Halyard (2007). "Patient-Reported Outcome Instrument Selection: Designing a Measurement Strategy." Value in Health **10**: S76-S85.

Solon, G., S. J. Haider and J. Wooldridge (2013). What Are We Weighting For? Working Paper 18859, National Bureau of Economic Research.

Soumare, B., S. Tempia, V. Cagnolati, A. Mohamoud, G. Van Huylenbroeck and D. Berkvens (2007). "Screening for Rift Valley fever infection in northern Somalia: a GIS based survey method to overcome the lack of sampling frame." Veterinary Microbiology **121**: 249 - 256.

Soundararajan, V., S. Soubeiga, D. Newhouse, A. Palacios-Lopez, U. Pape and M. Weber (forthcoming). "How well do Internet-based Surveys Track Labor Market Indicators in Middle-Income Countries?" World Bank Policy Research Working Paper.

Southern Sudan Center for Census, S., and Evaluation, (2010). Southern Sudan Counts: Tables from the 5th Sudan Population and Housing Census, 2008.

Southwick SM., Morgan CA III., Nicolaou AL. and C. DS. (1997). "Consistency of memory for combat-related traumatic events in veterans of operation desert storm." Am J Psychiatr **154**: 173 - 177.

Stecklov, G., A. Weinreb and C. Carletto (2017). Can incentives improve survey data quality in developing countries? Results from a field experiment in India. Series A - Statistics in Society**:** 1 - 24.

Stevens, F. R., A. E. Gaughan, C. Linard and A. J. Tatem (2015). "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." PLoS ONE **10**(2).

Stone, M. (1977). "An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion." Journal of the Royal Statistical Society **39**(Series B): 44 - 47.

Strachan, A. L. and H. Haider (2015). "Gender and conflict: Topic guide."

Sundberg, R. and E. Melander (2013). "Introducing the UCDP Georeferenced Event Dataset." Journal of Peace Research **50**(4): 523 - 532.

Tabakis, C., G. K. Ten, D. Newhouse, U. J. Pape and M. Weber (2022). "The Welfare Implications of COVID-19 for Fragile and Conflict-Affected Areas." World Bank Policy Research Working Paper(10081).

Talwar, V., C. Arruda and S. Yachison (2015). "The effects of punishment and appeals for honesty on children's truth-telling behavior." Journal of Experimental Child Psychology: 209- 217.

Tarozzi, A. and A. Deaton (2009). "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas." Review of Economics and Statistics **91**(4): 773-792.

Ten, G. K., J. Merfeld, D. Newhouse, U. J. Pape and K. Tafere Hirfrtot (2022). "How Well Can Real-Time Indicators Track the Economic Impacts of a Crisis Like COVID-19?" World Bank Policy Research Working Paper(10080).

Tourangeau, R., L. J. Rips and K. Rasinski (2000). The Psychology of Survey Response. Cambridge, Cambridge University Press.

Turkstra, J. and M. Raithelhuber (2004). Urban slum monitoring. 24th Annual ESRI International User Conference 9th-13th August. Citeseer.

UN OCHA (2017). United Nations Office for the Coordination of Humanitarian Affairs from South Sudan.

UNFPA (2002). The Impact of Conflict on Women and Girls. Geneva: United Nations Population Fund, UNFPA.

UNFPA (2014). Population Estimation Survey 2014 for the 18 pre-war regions of Somalia.

UNHCR (2018). Camp Coordination and Camp Management Cluster, United Nations High Commissioner for Refugees.

UNHCR (2018). "South Sudan Regional Refugee Response Plan."

UNHCR (2018). South Sudan Situation (31 December 2017), UNHCR.

UNICEF (2005). "The Impact of Conflict on Women and Girls in West and Central Africa and the UNICEF Response."

UNICEF (2013). Multiple Indicator Cluster Surveys. MICS5 tools: manual for mapping and household listing. New York, NY USA, UNICEF.

United Nations Statistical Division (2005). "Household surveys in developing and transition countries."

Usman, I. M., F. Ssempijja, R. Ssebuufu, A. M. Lemuel, V. B. Archibong, E. T. Ayikobua, J. O. Aruwa, S. Kembabazi, E. S. Kegoye and J. T. Ayuba (2020). "Community drivers affecting adherence to WHO guidelines against covid-19 amongst rural Ugandan market vendors." Frontiers in public health **8**: 340.

Van Buuren, S. (2007). "Multiple imputation of discrete and continuous data by fully conditional specification." Statistical methods in medical research **16**(3): 219-242.

Van Den Broeck, G. and M. Maertens (2017). Moving Up or Moving Out? Insights into Rural Development and Poverty Reduction in Senegal. World Development**:** 95 - 109.

van Ommeren M., Saxena S. and S. B. (2005). "Mental and social health during and after acute emergencies: emerging consensus?" Bull World Health Organ **83**(1): 75 - 76.

Vecchi, G. (2017). Measuring Wellbeing. New York, Oxford University Press.

Verme, P. (2016). "The Economics Of Forced Displacement: An Introduction." Region et Developpement **44**: 141-163.

Verme, P., Gigliarano, C., Wieser, C., Hedlund, K., Petzoldt, M. & Santacroce, M. (2016). "The Welfare of Syrian Refugees."

Verme, P. G., C. (2019). "Optimal targeting under budget constraints in a humanitarian context." World Development **119**(C): 224 - 233.

Verme, P. S., K. (2021). "The impact of forced displacement on host communities: A review of the empirical literature in economics." Journal of Development Economics **150**.

Vinceti, M., T. Filippini, K. Rothman, F. Ferrari, A. Goffi and G. Maffeis, et al. (2020). "Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking." EClinicalMedicine **25**.

Vinski, M. and S. Watter (2012). "Priming Honesty Reduces Subjective Bias in Self-Report Measures of Mind Wandering." Consciousness & Cognition **21**(1): 451 - 455.

Vintar, M., T. Beltramo, A. Delius, D. Egger and U. J. Pape (2022). "Impact of COVID-19 on Labor Market Outcomes of Refugees and Nationals in Kenya." World Bank Policy Research Working Paper(9960).

Voors, M. J., E. E. Nillesen, P. Verwimp, E. H. Bulte, R. Lensink and D. P. Soest (2012). "Violent Conflict and Behavior: A Field Experiment in Burundi." American Economic Review **102**(2): 942 - 964.

Wampler, P. J., Rediske, R. R., & Molla, A. R. (2013). "Using ArcMap, Google Earth, and Global Positioning Systems to select and locate random households in rural Haiti." International Journal of Health Geographics **12**(1): 3.

Wang, S. and Z. Weina (2016). "The Unintended Long-term Consequences of Mao's Mass Send-Down Movement: Marriage, Social Network, and Happiness." **15-03**.

Wardrop, N. A., W. C. Jochem, T. J. Bird, H. R. Chamberlain, D. Clarke, D. Kerr, L. Bengtsson, S. Juran, V. Seaman and A. J. Tatem (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. **PNAS 201715305**.

Warner, S. L. (1965). " A Survey Technique for Eliminating Evasive Answer Bias. Randomized Response." Journal of the American Statistical Association **60**(309): 63 - 69.

Warner, S. L. (2012). "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." Journal of the American Statistical Association.

Wessells, M. (2006). "Child Soldiers: From Violence to Protection."

White, H. (1980). " A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." Econometrica: 817 - 838.

Wieser, C. D., N.K., Ambel, A.A.; Tsegay, A.H., Mugera, H., Tanner, J. (2021). Monitoring COVID-19 Impact on Refugees in Ethiopia. Results from a High-Frequency Phone Survey of Refugees. Washington, DC., World Bank. **2**.

World Bank (2014). "South Sudan Poverty Note: Impact of a Continued Internal Conflict on Food Security and Poverty."

World Bank (2014). South Sudan Poverty Note: Impact of a Continued Internal Conflict on Food Security and Poverty. Washington, D.C., World Bank.

World Bank (2015). The Fiscal Impact of Declining Oil Prices on South Sudan. Washington, D.C., World Bank.

World Bank (2015). Monitoring Welfare and Perceptions in South Sudan 2012 – 2014. Washington, D.C., World Bank.

World Bank (2015). Transition amid Risks with a Special Focus on Intergovernmental Fiscal Relations. Somalia Economic Update. Washington, D.C., World Bank.

World Bank (2015). World Development Report: Mind, Society, and Behavior. Washington, D.C., The World Bank.

World Bank (2016). Monitoring Global Poverty: A Cover Note to the Report of the Commission on Global Poverty Chaired by Prof. Sir Anthony B. Atkinson. Washington, DC, The World Bank.

World Bank (2016). Somali Poverty Profile 2016. Findings from Wave 1 of the Somali High Frequency Survey. Washington, D.C., World Bank.

World Bank (2016). Somali Poverty Profile. Findings from Wave 1 of the Somali High Frequency Survey. Washington, D.C., World Bank.

World Bank (2016). South Sudan Poverty Profile 2015. Washington, D.C., World Bank.

World Bank (2017). Mobilizing Domestic Revenue to Rebuild Somalia. Somalia Economic Update. Washington, D.C., World Bank. **2**.

World Bank (2017). South Sudan - Governance analysis : building sustainable public sector capacity in a challenging context. Washington, D.C., World Bank.

World Bank (2018). Somali Poverty and Vulnerability Assessment. Washington, D.C, World Bank.

World Bank (2018). Somalia - Systematic Country Diagnostic. Washington, D.C., World Bank. **123807**.

World Bank (2018). Somalia drought impact and needs assessment: synthesis report Washington, D.C., World Bank.

World Bank (2020). Poverty and Shared Prosperity 2020 : Reversals of Fortune. Washington, D.C., World Bank.

World Bank (2021). Monitoring Social and Economic Impacts of COVID-19 on Refugees in Uganda: Results from the High-Frequency Phone Survey - Second Round. Washington, D.c., World Bank.

World Bank (2021). Sampling Refugees and Host Communities for Integration into National Household Surveys. Washington, DC, World Bank.

World Bank (2021). Socioeconomic Impacts of COVID-19 in Kenya on Households. Rapid Response Phone Survey Round 1, January 2021. Washington, D.C., World Bank.

Xu, Y., A. Delius and U. J. Pape (2022). "Gender Differences in Household Coping Strategies for COVID-19 in Kenya." World Bank Policy Research Working Paper(9959).

Yilmazkuday, H. (2021). "Stay-at-home works to fight against COVID-19: International evidence from Google mobility data." Journal of Human Behavior in the Social Environment.

Young, A. (2016). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. London, London School of Economics.

Zezza, A., C. Carletto, J. L. Fiedler, P. Gennari and D. Jolliffe (2017). Food counts. Measuring food consumption and expenditures in household consumption and expenditure surveys (HCES) Introduction to the special issue. Food Policy. **72:** 1 - 6.