ESSAYS IN DEVELOPMENT ECONOMICS: HUMAN CAPITAL ACCUMULATION, SOCIAL PROTECTION, AND HAWTHORNE EFFECTS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Andrew Martin Simons

August 2016

ESSAYS IN DEVELOPMENT ECONOMICS: HUMAN CAPITAL ACCUMULATION, SOCIAL PROTECTION, AND HAWTHORNE EFFECTS

Andrew Martin Simons, Ph.D.

Cornell University 2016

Understanding how the poor respond to programs, policies, and shocks in the developing world is central to the study of development economics. In this dissertation, I examine how households and communities respond to various stressors at three scales and in three different contexts. First, I examine how changes in rainfall and temperature affect household's investment in schooling for their children in Indonesia. Secondly, I examine how communities alter the distribution of aid in a large social safety net program in rural Ethiopia when communities are not given sufficient resources to implement the program as designed. Thirdly, I examine the behavioral response to the presence of researchers during the data collection process of a field evaluation of fuel-efficient cookstoves in Uganda. By examining the varied responses to stressors, these essays contribute to understanding the processes of development around the world.

BIOGRAPHICAL SKETCH

Andrew Simons worked for seven years as a development practitioner in Ethiopia and Honduras in various senior-level NGO management positions prior to starting his Ph.D. Before that he received a Master's of Public Administration in International Development from the Kennedy School of Government at Harvard University in 2004 and a Bachelor's of Arts in Biology from Taylor University in 2001. Starting in the fall of 2016, Andrew will be an assistant professor in the Economics Department at Fordham University.

To Laura and Isabella

ACKNOWLEDGMENTS

I would like to thank my wife, Laura, and my daughter Isabella for your support over the past five years. It is an amazing feeling to come home each night and feel unconditionally loved.

Thank you to Chris Barrett for your confidence in me, and generosity with your time. Thank you for always making me feel important, making my questions seem valuable, and encouraging me as a person. I hope to emulate your example of how to excel at being a professor, advisor, and balanced human being. To Garrick Blalock, thank you for taking me under your wing early on in my time at Cornell. I learned more than you probably know about how to manage a complex field-based experiment, how to organize and analyze data, and how to write academic papers from our cookstove collaboration in Uganda. To Shanjun Li, thank you for your guidance in teaching our first econometrics course and conversations about the job market process and what journals I should submit my work to. Your even-keeled example of how to be very productive while still having time for your students is to be emulated.

Thank you to Alemayehu Seyoum Taffesse, Bart Minten and the team at the International Food Policy Research Institute's Ethiopia Strategy Support Program for many useful discussions and logistical support during my time in Ethiopia. I am grateful to the staff of Ethiopia's Central Statistical Agency (CSA) who conducted the household surveys, and to the CSA's Director General, Samia Zekaria for sharing the data. Thank you to the field staff of the Center for Integrated Research and Community Development (CIRCODU) for their expertise in collecting the data for the cookstove evaluation in Uganda. Special thanks to Juliet Kyaesimira, Stephen Harrell

and Joseph Ndemere Arineitwe for overseeing field operations in Uganda.

I gratefully acknowledge support from the National Science Foundation through Cornell University's Food Systems and Poverty Reduction Integrative Graduate Education and Research Traineeship program. The cookstove study was funded by the United States Agency for International Development under Translating Research into Action, Cooperative Agreement No. GHS-A-00-09-00015-00. I would also like to thank the Atkinson Center for a Sustainable Future at Cornell University, the Institute for the Social Sciences at Cornell University and the Cornell Population Center for additional funding of related expenses for the cookstove study.

Most writing processes are collaborative, and this dissertation is no exception. I would like to thank my co-authors for their patience and insights as we developed these essays. Thank you to Brian Theide and Clark Gray for introducing me to and including me in your work in Indonesia. For the cookstove project, thank you to David Levine for your guidance and offering direct but extremely useful advice to help me clearly articulate my ideas while writing. I have become a better writer over the last four years due to your direction and blunt edits. Thank you to Theresa Beltramo for the many thankless tasks you performed in your role overseeing the Uganda cookstove grant, organizing the data collection process, and as a co-author. And thank you to Garrick Blalock (mentioned above) as a co-author on the cookstove project.

Lastly I would like to thank my parents, Gary and Linda Simons, for instilling in their children a desire live and work internationally and in devoting themselves to serve their neighbors both near and far.

TABLE OF CONTENTS

1. Introduction **1**

2. Temperature, Rainfall and Children's Education **4**

2.1 Introduction 4

2.2 Links Between Climate and Schooling 7

2.3 Data and Methods 8

2.4 Results 13

2.5 Discussion 19

2.6 References 21

2.7 Tables and Figures 27

3. What is the Optimal Locus of Control for Social Assistance Programs **40**

3.1 Introduction 40

3.2 Background and Targeting of Productive Safety Net Program 43

3.3 Estimating Communities' Revealed Preferences 47

3.4 Data and Descriptive Statistics 60

3.5 Results 64

3.6 Conclusion 70

3.7 Appendices 74

3.8 References 82

3.9 Tables and Figures 88

4. Using Unobtrustive Sensors to Measure and Minimize Hawthorne Effects **101**

4.1 Introduction 101

4.2 Data and Methodology 103

4.3 Specification 107

4.4 Results 108

4.5 Adjusting for the Hawthorne Effect 109

4.6 Conclusion 112

4.7 References 114

4.8 Tables and Figures 120

CHAPTER 1

INTRODUCTION

In my dissertation I examine how the poor respond to programs, policies and shocks in the developing world. This theme—the poor's response to various stressors—is a central one within the study of development economics. While my dissertation essays are distinct pieces of work, this unifying theme runs through each essay. I examine this theme in three different geographic contexts and at three different scales.

In "Temperature, Rainfall and Children's Education: Evidence from Indonesia" I, along with my co-authors, begin at a large scale with a national dataset designed to represent more than 80% of the world's fourth largest country (Indonesia) and examine how changes in rainfall and temperature over time affect household investment in education. Because weather variability is expected to increase as a result of global climate change, studying how weather variations impact human development is essential to develop evidence-based interventions to mitigate expected climate change. This has important implications for the long-term trajectory for the development of Indonesia as building human capital is an essential part of a country's development process. I find that primary students face higher and increasing risks (versus secondary students) for dropout as temperatures increase, and that the presence of irrigation lessens the negative effects of monsoon onset delay.

Next, I change geographic locations and narrow the scale of work to focus on a social protection program intended to assist about eight million rural Ethiopians annually. In "What is the Optimal Locus of Control for Social Assistance Programs? Evidence from the Productive Safety Net Program in Ethiopia" I examine how communities choose to allocate aid. The centralized implementation mandates of Ethiopia's Productive Safety Net Program (PSNP) require a full and uniform payment to each person in an eligible household. However, communities do not receive enough funding to fully implement the program as designed so I examine how they allocate aid in this constrained setting.

To do this, I recover the preferences revealed by local communities' actual aid allocations. I find that communities allocate more aid to underprivileged groups with lower wage earning potential (e.g., teenage girls vs. teenage boys, adult women vs. adult men, elderly vs. working age adults) than they would have if they followed the federal implementation mandates. However, despite communities' pro-poor implementation, the program with constrained funding does not significantly lower overall poverty rates. In simulations with full funding, the program reduces poverty in both cases of centralized and decentralized program control, even while using different criteria for how the funds are allocated. The major policy implication of this essay is that the financial scale of the safety net program is much more important to poverty reduction than whether implementation decisions are made centrally by the federal government or locally by the communities themselves.

In my final essay "Using Unobtrusive Sensors to Quantify and Minimize Hawthorne Effects: Evidence from Cookstoves" I further narrow the scope and focus of my unifying theme to examine the stress that we as researchers potentially put on study participants during the data gathering process. To do this, I, along with my co-authors, deployed a detailed study of daily cooking behaviors of 168 rural households in the southwestern region of Mbarara, Uganda. We introduced fuel-efficient cookstoves as a substitute for traditional three stone fires and tracked cooking behaviors before and after the introduction of the new cookstove.

People act differently when they know they are being observed. This phenomenon—the Hawthorne effect—can bias estimates of program impacts. We designed our study to examine the ability of unobtrusive sensors (temperature loggers) to substitute for human observation to alleviate this bias. We find a large Hawthorne effect: when in-person measurement begins, participants increase fuel-efficient stove use approximately three hours/day (54%) and reduce three-stone fire use by approximately two hours/day (30%). When in-person measurement ends, participants reverse those changes.

This Hawthorne effect induced by the presence of researchers biases estimates of fuel use and particulate matter concentrations. A naïve examination of our data shows a modest success for the intervention, that firewood usage and indoor air particulates both declined by 11% between baseline and end line. However, when we correct for the induced Hawthorne effect, we find these results are reversed. Firewood usage and

indoor particulate matter actually increased by 4%. Our results showing that the presence of a Hawthorne effect not only alters the magnitude, but also (potentially) the sign of the impact of an intervention has broad implications for any social science data gathering process where researchers interact with participants. This final essay reinforces the importance of accounting for Hawthorne effects, especially for policy-relevant impact evaluations.

In these essays I have examined how the poor react to various stressors—changing climatic conditions, underfunding of social protection programs, and the presence of researchers—in three different developing country contexts. These essays represent my effort to enhance our understanding of development processes and contribute to the broader development economics literature.

CHAPTER 2

TEMPERATURE, RAINFALL AND CHILDREN'S EDUCATION: EVIDENCE FROM INDONESIA

Introduction

Given that weather variability is expected to increase as a result of global climate change, studying how weather variations impact human development is essential to develop evidence-based interventions to adapt to expected climate change. Perhaps unsurprisingly, the number of studies of how individuals, households, and communities are affected by and respond to weather variability and environmental shocks has grown significantly in recent years (as reviewed by Dell, Jones, and Olken (2014)).

Scholars have documented the effects of weather and climate variability on a number of important economic and demographic outcomes including: economic output, agriculture, labor productivity, health, conflict, and migration.[1] These studies demonstrate that variation in weather can cause notable behavioral responses among

[1] A sample of notable papers touching on each of these topics are: economic output (Barrios, Bertinelli, and Strobl 2010; Dell, Jones, and Olken 2012; Hsiang 2010; Hsiang and Jina 2014), agriculture (Auffhammer and Schlenker 2014; Burke and Emerick 2015; Feng, Krueger, and Oppenheimer 2010; Lobell et al. 2008; Lobell et al. 2011), labor productivity (Connolly 2008; Dunne, Stouffer, and John 2013; Graff Zivin and Neidell 2014; Zander et al. 2015), health (Barreca 2012; Barreca et al. 2013; Deschênes 2014; Deschênes and Greenstone 2011), conflict (Burke et al. 2009; Couttenier and Soubeyran 2014; Fjelde and Uexkull 2012; Harari and La Ferrara 2014; Hsiang, Burke, and Miguel 2013), and migration (Bohra-Mishra, Oppenheimer, and Hsiang 2014; Dillon, Mueller, and Salau 2011; Fussell, Hunter, and Gray 2014; Gray and Bilsborrow 2013; Gray and Mueller 2012; Marchiori, Maystadt, and Schumacher 2012) (Bohra-Mishra, Oppenheimer, & Hsiang, 2014; Dillon, Mueller, & Salau, 2011; Fussell, Hunter, & Gray, 2014; Gray & Bilsborrow, 2013; Gray & Mueller, 2012; Hunter, Luna, & Norton, 2015; Marchiori, Maystadt, & Schumacher, 2012).

affected individuals and households. We fill one important gap in the climate-economy literature—the link between weather variability and schooling—that has important long-term development implications. Understanding the educational consequences of weather variability is particularly important because adverse outcomes not only reflect short-term resource constraints, but also may negatively affect the life chances and future productivity of affected youth. To the extent that the educational impacts of weather variability vary systematically between groups, they may also shape future patterns of social inequality.

In this paper, we examine temperature and rainfall data over a long period of time to understand how small deviations from long-term weather averages affect schooling outcomes in the near-term. Previous studies that have focused on educational outcomes have largely focused on responses to large shocks (Cameron and Worswick 2001; Gertler, Levine, and Ames 2004; Levine and Ames 2003; Thomas et al. 2004) or how early-life or *in utero* shocks affect developmental outcomes much later in life (Almond and Currie 2011; Maccini and Yang 2009). Our study is different in that instead of looking at major shocks *per se*, we focus on small deviations similar to what would be expected due to gradual climate change. We focus on Indonesia due to its importance as a large, growing developing country and because high resolution historic weather data and multiple rounds of panel data covering 15+ years are available. While the effects of future climate change will likely include increased major environmental shocks (e.g., cyclones, hurricanes, etc.), our examination focuses

on how communities in the tropics respond to gradual changes in temperatures and more variable rainfall.

In this analysis we use the first four rounds of the Indonesia Family Life Survey dataset. We plan on expanding the current analysis by incorporating the fifth round of the IFLS dataset and also focusing on growing degree days during the agricultural season (instead of annual deviations from long temperature means). However, some tentative findings using the existing analysis appear to be that primary students face higher and increasing risks (versus secondary students) for dropout as temperatures increase, and that the presence of irrigation lessens the negative effects of monsoon onset delay.

Links between climate and schooling

We hypothesize that the weather variability-agricultural income link drives changes in schooling investments in Indonesia. This has been documented in other developing country settings, for example, agriculture-focused households can pull their children out of school; stop making payments for school-related activities, or limit health and medical check-ups in response to changes in rainfall (Beegle, Dehejia, and Gatti 2006; Björkman-Nyqvist 2013; Jacoby and Skoufias 1997; Jensen 2000; Shah and Steinberg 2013). However, this link has not been explicitly studied in the context of schooling in Indonesia. Other shocks examined in the context of how they affect educational outcomes in Indonesia include: shocks due to El Niño (Cameron and Worswick 2001), the Asian financial crisis of the late 1990s (Thomas et al. 2004; Levine and Ames

2003), parental death (Gertler, Levine, and Ames 2004), and rainfall shocks in the year of birth (Maccini and Yang 2009). We build on this body of work but focus instead on smaller, more gradual changes (as opposed to major shocks) and the weather variability-agricultural income link. Given the expectations for a gradually warming and more variable climate, understanding the response to variations in temperature and rainfall will give insight to potential future impacts of climate change.

Data and Methods

Our analyses draw upon the first through fourth rounds of the Indonesian Family Life Survey (IFLS). The IFLS collected a comprehensive set of social and demographic data on over 30,000 individuals between 1993 and 2008. The original sample of survey respondents was drawn from the population in 13 of Indonesia's 27 provinces, and is representative of approximately 83% of the national population. The IFLS has a remarkably high rate of follow-up from one panel to the next.[2]

For our purposes, we use the four IFLS panels to construct a longitudinal dataset of Indonesian school aged children. Each observation includes characteristics measured at the first survey of each period (i.e., period baseline or t_0) and variables (described in detail below) indicating end-of-period educational investments in that child and whether the child dropped out of school during that period. We examine school dropout and educational investments in children attending school in primary (grades 1-

[2] For example, 88% of eligible participants interviewed in IFLS1 were also interviewed in IFLS4 (Thomas et al., 2012).

6), lower secondary (grades 7-9), and upper secondary (grades 10-12). In order to link the period baseline covariates with inter-survey and end-of-period educational outcomes, we restrict our sample to individuals who were observed during two consecutive surveys (i.e., at least one inter-survey period).

In constructing the sample, there are two important possible sources of attrition: death and migration. Death rates of school-aged children is very low with less than 1% of school aged children passing away between rounds (e.g., 0.6% of children (55/8566) aged 0-18 passed away between IFLS1 and IFLS2, and 0.8% (63/7428) between IFLS2 and IFLS3) therefore such small levels of attrition are unlikely to be a source of bias. However, we do observe some migration. Geo-coordinates for IFLS observations are only available for the 304 communities in which the first round of the IFLS was fielded (1993/4), therefore we lose any migrant household that moved to a community outside of the original 304 communities. Across the three-pooled waves (1993-1997, 1997-2000, 2000-2007) approximately 10.5% of the person-period observations are dropped due to not having geo-codes. However, most of these observations are adults, for example a single adult could migrate (to look for work) while the family and children are left behind in the original geo-coded location and therefore included in the data we use for the analysis. We also retain any household that moved as long as the move was to one of the original 304 areas with geo-codes.

Using geo-coordinates for household location at period baseline, we link observations from the IFLS to daily rainfall and temperature estimates for 1984-2011. Weather data

were produced by NASA's Modern-Era Retrospective-Analysis for Research and Applications (MERRA) (Rienecker et al. 2011).

Outcomes

We estimate the effect of weather variation on (a) the risk of school dropout and (b) the level of educational investments. We define school dropout as children who were in school at the period baseline or entered school after the period baseline; but were not enrolled in school at the end of the period and had not graduated from the level of school they last attended (e.g., primary school). Children who did not continue their education after graduation from the level of school they last attended due to resource constraints were not identified as dropouts in this analysis, although this arguably represents an adverse outcome as well. We define the population at risk of school dropout to include any child who was in school at the period baseline or entered school after the period baseline.

We measure household educational investments in children using reported expenditures on education during the year prior to the end-of-period survey. This measure includes the following categories of expenditures: school registration, exam fees, books, writing supplies, uniforms, sports, transportation, housing and food, special courses, other scheduled fees, and other school-related expenses. Due to the possibility that poor weather could induce inflationary pressure on commonly purchased items, for example housing, food and transportation, we also create a more focused educational expenses variable that excludes the costs of room, board, and

transportation. We run the analysis using both of these definitions of educational expenses. Expenditures were reported separately for each school-aged child for the twelve-month period prior to the end-of-period survey, so they represent child-specific levels rather than household averages per child.

Measuring weather variability

Our analyses focus on the effect of weather conditions on educational outcomes. We focus on rainfall and temperature, which we simultaneously control for in each model (Auffhammer et al. 2013). We use a measure of temperature based on the deviation of the average temperature for each community c_n in each year y_n from the long-term (1984-2011) mean for that community. The final temperature measure represents the average deviation over the four-year periods corresponding to each IFLS inter-survey period included in our dataset. We calculate an identical measure with respect to annual rainfall levels. We also generate a count variable when monthly temperatures or monthly rainfall is one standard deviation or two standard deviations above or below the long term mean for that calendar month. As an example the count would be the number of calendar months during the inter-survey period (*e.g.*, 1993-1997) that the mean monthly temperature for that calendar month was above the long-term mean monthly temperature for that calendar month across the years 1984-2011. We also model the effect of the onset of the annual monsoon, which plays a key role in agricultural production (Naylor et al. 2007; Naylor et al. 2002). We base this measure on previous research (Naylor et al. 2007; Skoufias, Katayama, and Essama-Nssah

2012) and measure monsoon onset as the number of days after August 1st when cumulative rainfall reaches 20cm. To measure delays in monsoon onset, we calculate the deviation of monsoon onset from the long-term (1984-2011) mean for each community and the average deviation over the four-year spans corresponding to the periods in our data.

Methods

To assess weather effects on school dropout and grade repeats we estimate a series of discrete-time event-history models. These models estimate the effect of rainfall and temperature deviations on the probability of school dropout or the probability of repeating a grade. We estimate a logit model that takes the form:

$$D_{it} = F(a_0 + \delta W_{it} + \beta X_{it} + \phi(W_{it} * X_{it}) + \chi S_{it} + a_t + a_d + e_{it} \qquad (2.1)$$

where *F( )* is the logistic functional form, D_{it} is observed dropout for individual i in period t, a_0 is a constant, W_{it} is a vector of weather variables, X_{it} is a vector of demographic variables, $W_{it} * X_{it}$ represent a vector of interaction terms, and S_{it} is a vector of infrastructure access variables. The vectors δ, β, ϕ, and χ are parameters associated with each control variable or interaction term; a_t and a_d are time and regional fixed effects, and e_{it} is an error term. We use district fixed effects to account

for the level of government that sets educational policy as the district manages and funds the majority of educational expenditures in Indonesia (Al-Samarrai 2013; OECD and Asian Development Bank 2015). Standard errors (SEs) are adjusted for clustering at the level of the 0.5°-by-0.5° areas for which weather conditions are measured.

To examine educational investments, we estimate a linear regression model in the form:

$$y_{it} = a_0 + \gamma W_{it} + \theta X_{it} + \rho(W_{it} * X_{it}) + \psi S_{it} + a_t + a_d + e_{it} \quad (2.2)$$

where y_{it} is the total educational expenditure on child *i* during the last year of period *t*, W_{it}, X_{it}, $W_{it} * X_{it}$, S_{it}, a_0, a_t, a_d are as above. The vectors γ, θ, ρ and ψ are parameters associated with each control variable or interaction term and u_{it} is an error term. Standard errors (SEs) are adjusted in the same manner as above.

Results

First, we present descriptive statistics concerning the observed dropout rates, educational expenditures, weather conditions, and community infrastructure variables during the study period. Then we present estimates of the effect of weather shocks on school dropout and educational investments.

Descriptive statistics

There are 21,053 student-period observations where students were at risk for dropout. Table 2.1 presents basic educational and demographic statistics. Approximately 14% of these student-periods resulted in the student dropping out of school. The average per student total educational expenditure in the twelve months prior to the period's end was 1,983,000 rupiah per student (about USD 215),[3] which was about 12% of total annual household income. The sample is evenly split by gender of the student (49% female), about one fourth of the sample (25%) are the only child in the household. The average student is 9.6 years old, the average household has 5.5 people, and the average educational attainment of the household head is 6.0 years. About one tenth of the sample is from female-headed households (10%). The average household has about USD 11,610 worth of assets (106,812,000 rupiah), an average annual household income of USD 1,872 (17,218,000 rupiah), and minimal sales of assets in the preceding twelve months. See Figure 2.1 for a graphical display of the conditional frequency of dropout across completed years of school. The highest dropout rates are just after completing primary school (grades 1-6) and after completing lower secondary (grades 7-9).

The mean temperature (Table 2.2) during the inter-survey periods observed was 27.3 °C. The data indicates warming temperatures with the deviation of the annual mean temperature from the long-term mean (1984-2011) of 0.03 °C, or 0.12 standard deviations. During the inter-survey periods there are more months of hotter than

[3] All rupiah figures are converted to 2007/08 rupiah equivalents using the consumer price index from the World Development Indicators published by the World Bank. The approximate USD to rupiah exchange rate in 2007/08 was 9,200 rupiah to 1 USD.

normal temperatures than lower than normal temperatures (an average of 10.1 months above 1SD of that calendar month's long term mean vs. 8.5 months below 1SD of that calendar month's long term mean temperature). The mean monthly rainfall over the inter-survey period was 265 mm. The data indicates less rainfall over time, and more monsoon onset delay. The deviation of the mean monthly rainfall from the long-term mean was -6.2 mm, or -0.20 standard deviations, while the inter-survey period average monsoon onset delay was 1.2 days. There are about an equal number of months with higher rainfall than normal rainfall as lower than normal rainfall in each inter-survey period (an average of 11.1 months above 1SD of that calendar month's long term mean rainfall vs. 11.4 months below 1SD of that calendar month's long term mean rainfall).

A slight majority of the children represented in the data come from rural households (Table 2.3, 53%). There was more access in terms of elementary schools per community (6.1), followed by junior high schools (2.9), then high schools (2.5). On average, children lived in households with about 22.0 hours of electricity access per day. A majority of students (60%) were in households with no irrigation, some had basic irrigation (21%) and others had mechanized irrigation (19%). Large majorities had access to paved roads as their main transportation infrastructure (74%), with fewer using paved stone roads (16%), dirt roads (9%) or waterways (1%). Almost the entire sample (93%) stated that the primary roads to their household are accessible by a vehicle for twelve months a year.

School dropout and educational investments

We present estimates from a total of nine specifications for each of the dependent variables of interest. Table 2.4 examines the estimated effect of climate variables on the probability of school dropout. In the first specification (Specification A, Table 2.4) we include a quadratic function of temperature and rainfall. Our estimates indicate that temperature and rainfall have no significant effect on dropout (and no evidence of non-linear effects). However, since the timing of rainfall may be more important than the level of rainfall we include monsoon onset (Specification B) instead of rainfall and show that monsoon onset delay has a significant and positive effect on dropout. In Specification C we remove the squared terms and retain only the temperature and monsoon onset delay variables. This will be the basis of weather variables for the to examine variation in climate effects across subpopulations of interest.

We interact the student being in a primary school with temperature deviations and monsoon onset delay (Specification D, Table 2.5) and find the effect of temperature is statistically significant and negative for secondary school students, the difference between the baseline group (secondary students) and primary students is positive and statistically significant. The net effect for primary students is positive. This is depicted graphically in Figure 2.2 which shows declining rates of dropout for secondary students as the temperature deviations from the long term mean increase, but increasing rates of dropout for primary students along the same gradient. The effect on monsoon onset delay is positive and statistically significant for secondary students, the difference between the baseline group (secondary students) and primary students is

negative and statistically significant, with a net effect that is negative for primary students (graphical depiction in Figure 2.3).

Specification E (Table 2.5) interacts gender with climatic conditions. The effects of temperature deviations are negative but non-significant for both boys and girls, and the between-sex differences in the magnitude of this effect are non-significant. The effects of monsoon onset delay are slightly negative for boys; the difference between boys and girls is positive and statistically significant.

In Specification F (Table 2.5) we examine membership in a farm household (defined as generating some farm income) as farm households may be disproportionately affected by changes in climatic conditions. Our estimates show that temperature deviations have a negative (but not statistically significant) effect on the probability of dropout for non-farm households. The difference between the baseline group (non-farm households) and farming households is positive and statistically significant. The net effect for farm households is positive. Monsoon onset delay is associated with increases (though not statistically significant) in dropout rates for non-farming households; there are no statistically significant differences in rates of dropout when interacted with monsoon onset delay for farming and non-farming households.

We continue the examination of interaction effects with climatic conditions in Table 2.6. Because the simple variable of whether a household generates farm income (Specification F above) may not adequately describe the agrarian lifestyle, we

examine the interaction effects of rural location and using irrigation with climatic effects (Specification G and H, respectively) and a consolidated regression (Specification I). We interact the student being in a rural location with temperature deviations and monsoon onset delay (Specification G, Table 2.6) and find the effect of temperature is statistically significant and negative for urban students, the difference between the baseline group (urban students) and rural students is positive and statistically significant, however, the net effect for rural students is approximately zero. The effect on monsoon onset delay is positive and statistically significant for urban students, the difference between the baseline group (urban students) and rural students is negative and statistically significant, with a net effect that is about zero for rural students.

Next we examine the interaction effects of using irrigation with climatic effects. Because irrigation should mitigate the negative effects of poor climatic conditions this is important to analyze. We interact the student being in a household without with irrigation with temperature deviations and monsoon onset delay (Specification H, Table 2.6) and find the effect of temperature is (weakly) statistically significant and negative for students without irrigation, the difference between the baseline group (no irrigation) and those with irrigation is positive and statistically significant. The net effect for students who have irrigation is positive. The effect on monsoon onset delay is slightly positive (not statistically significant) for students with no irrigation, the difference between the baseline group (no irrigation) and students who live in a household with irrigation is negative and statistically significant, with a net effect that

is negative for students in households with irrigation in the presence of monsoon onset delay.

In Specification I, we consolidate the interaction effects from specifications D though H and see that only primary school and irrigation have statistically significant effects when interacted with the climatic variables.

In Tables 2.7, 2.8, and 2.99 we examine the same regression specification but with educational expenditures per child rather than on the probability of dropout. In the consolidated regression (Specification I, Table 2.9) we see that being in primary school and being female have negative interaction effects with increasing temperatures; being in primary school has a negative interaction with monsoon onset delay and having irrigation has a positive interaction with monsoon onset delay.

Discussion

We contribute to the growing field that quantifies the effects of climate on important human activities. To date, this research has focused on demographic and economic outcomes such as economic output, agriculture, labor productivity, health, conflict, and migration. We will expand the current analysis by incorporating the fifth round of the IFLS dataset and also focusing on growing degree days during the agricultural season (instead of annual deviations from long temperature means). However, some tentative findings using the existing analysis appear to be that primary students face higher and increasing risks (versus secondary students) for dropout as temperatures

increase, and that the presence of irrigation lessens the negative effects of monsoon onset delay.

REFERENCES

Almond, Douglas, and Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25 (3): 153–72. doi:10.1257/jep.25.3.153.

Al-Samarrai, Samer. 2013. "Local Governance and Education Performance: A Survey of the Quality of Local Education Governance in 50 Indonesian Districts." Jakarta, Indonesia.

Auffhammer, Maximillian, Solomon M Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change." *Review of Environmental Economics and Policy* 7 (2): 181–98. doi:10.1093/reep/ret016.

Auffhammer, Maximillian, and Wolfram Schlenker. 2014. "Empirical Studies on Agricultural Impacts and Adaptation." *Energy Economics* 46: 555–61. doi:10.1016/j.eneco.2014.09.010.

Barreca, Alan I. 2012. "Climate Change, Humidity, and Mortality in the United States." *Journal of Environmental Economics and Management* 63 (1): 19–34. doi:10.1016/j.jeem.2011.07.004.

Barreca, Alan I., Karen Clay, Olivier Deschênes, Michael Greenstone, and Joseph S. Shapiro. 2013. "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the 20th Century." Cambridge, MA. http://www.nber.org/papers/w18692.

Barrios, Salvador, Luisito Bertinelli, and Eric Strobl. 2010. "Trends in Rainfall and Economic Growth in Africa: A Neglected Cause of the African Growth Tragedy." *Review of Economics and Statistics* 92 (2): 350–66. doi:10.1162/rest.2010.11212.

Beegle, Kathleen, Rajeev H. Dehejia, and Roberta Gatti. 2006. "Child Labor and Agricultural Shocks." *Journal of Development Economics* 81 (1): 80–96. doi:10.1016/j.jdeveco.2005.05.003.

Björkman-Nyqvist, Martina. 2013. "Income Shocks and Gender Gaps in Education: Evidence from Uganda." *Journal of Development Economics* 105. The Author: 237–53. doi:10.1016/j.jdeveco.2013.07.013.

Bohra-Mishra, Pratikshya, Michael Oppenheimer, and Solomon M Hsiang. 2014. "Nonlinear Permanent Migration Response to Climatic Variations but Minimal Response to Disasters." *Proceedings of the National Academy of Sciences of the United States of America* 111 (27): 9780–85. doi:10.1073/pnas.1317166111.

Burke, Marshall, and Kyle Emerick. 2015. "Adaptation to Climate Change: Evidence from US Agriculture." Palo Alto, CA. http://economics.cornell.edu/seminars/Marshall.Burke_JMP.pdf.

Burke, Marshall, Edward Miguel, Shanker Satyanath, John A Dykema, and David B Lobell. 2009. "Warming Increases the Risk of Civil War in Africa." *Proceedings of the National Academy of Sciences of the United States of America* 106 (49): 20670–74. doi:10.1073/pnas.0907998106.

Cameron, Lisa A, and Christopher Worswick. 2001. "Educational Expenditure Responses to Crop Loss in Indonesia: A Gender Bias." *Economic Development and Cultural Change* 49 (2): 351–63.

Connolly, Marie. 2008. "Here Comes the Rain Again: Weather and the Intertemporal Substitution of Leisure." *Journal of Labor Economics* 26 (1): 73–100. doi:10.1086/522067.

Couttenier, Mathieu, and Raphael Soubeyran. 2014. "Drought and Civil War in Sub-Saharan Africa." *Economic Journal* 124 (575): 201–44. doi:10.1111/ecoj.12042.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics* 4 (3): 66–95. doi:10.1257/mac.4.3.66.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken. 2014. "What Do We Learn from the Weather? The New Climate-Economy Literature." *Journal of Economic Literature* 52 (3): 740–98. doi:10.3386/w19578.

Deschênes, Olivier. 2014. "Temperature, Human Health, and Adaptation: A Review of the Empirical Literature." *Energy Economics* 46: 606–19. doi:10.1016/j.eneco.2013.10.013.

Deschênes, Olivier, and Michael Greenstone. 2011. "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US." *American Economic Journal: Applied Economics* 3 (4): 152–85. doi:10.1257/app.3.4.152.

Dillon, Andrew, Valerie Mueller, and Sheu Salau. 2011. "Migratory Responses to Agricultural Risk in Northern Nigeria." *American Journal of Agricultural Economics* 93 (4): 1048–61.

Dunne, John P., Ronald J. Stouffer, and Jasmin G. John. 2013. "Reductions in Labour Capacity from Heat Stress under Climate Warming." *Nature Climate Change* 3 (6): 563–66. doi:10.1038/nclimate1827.

Feng, Shuaizhang, Alan B Krueger, and Michael Oppenheimer. 2010. "Linkages among Climate Change, Crop Yields and Mexico-US Cross-Border Migration." *Proceedings of the National Academy of Sciences of the United States of America* 107 (32): 14257–62. doi:10.1073/pnas.1002632107.

Fjelde, Hanne, and Nina von Uexkull. 2012. "Climate Triggers: Rainfall Anomalies, Vulnerability and Communal Conflict in Sub-Saharan Africa." *Political Geography* 31 (7): 444–53. doi:10.1016/j.polgeo.2012.08.004.

Fussell, Elizabeth, Lori M. Hunter, and Clark L. Gray. 2014. "Measuring the Environmental Dimensions of Human Migration: The Demographer's Toolkit." *Global Environmental Change* 28: 182–91. doi:10.1016/j.gloenvcha.2014.07.001.

Gertler, Paul, David I. Levine, and Minnie Ames. 2004. "Schooling and Parental Death." *Review of Economics and Statistics* 86 (1): 211–25. doi:10.1162/003465304323023769.

Graff Zivin, Joshua, and Matthew Neidell. 2014. "Temperature and the Allocation of Time: Implications for Climate Change." *Journal of Labor Economics* 32 (1): 1–26. doi:10.1086/671766.

Gray, Clark, and Richard Bilsborrow. 2013. "Environmental Influences on Human Migration in Rural Ecuador." *Demography* 50 (4): 1217–41. doi:10.1007/s13524-012-0192-y.

Gray, Clark, and Valerie Mueller. 2012. "Drought and Population Mobility in Rural Ethiopia." *World Development* 40 (1): 134–45. doi:10.1016/j.worlddev.2011.05.023.

Harari, Mariaflavia, and Eliana La Ferrara. 2014. "Confict, Climate and Cells: A Disaggregated Analysis." Bocconi, Italy.

Hsiang, Solomon M. 2010. "Temperatures and Cyclones Strongly Associated with Economic Production in the Caribbean and Central America." *Proceedings of the National Academy of Sciences of the United States of America* 107 (35): 15367–72. doi:10.1073/pnas.1009510107.

Hsiang, Solomon M, Marshall Burke, and Edward Miguel. 2013. "Quantifying the Influence of Climate on Human Conflict." *Science* 341 (6151): 1235367. doi:10.1126/science.1235367.

Hsiang, Solomon M, and Amir S Jina. 2014. "The Causal Effect of Environmental Catastrophe on Long-Run Economic Growth: Evidence from 6,700 Cyclones." Cambridge, MA.

Jacoby, H, and E Skoufias. 1997. "Risk, Financial Markets, and Human Capital in a Developing Country." *Review of Economic Studies* 64: 311–35.

Jensen, Robert. 2000. "Agricultural Volatility and Investments in Children." *American Economic Review* 90 (2): 399–404. doi:10.1257/aer.90.2.399.

Levine, David I, and Minnie Ames. 2003. "Gender Bias and The Indonesian Financial Crisis: Were Girls Hit Hardest?" C03-130. Berkeley, CA.

Lobell, David B., Marianne Bänziger, Cosmos Magorokosho, and Bindiganavile Vivek. 2011. "Nonlinear Heat Effects on African Maize as Evidenced by Historical Yield Trials." *Nature Climate Change* 1 (1): 42–45. doi:10.1038/nclimate1043.

Lobell, David B., Marshall B. Burke, Claudia Tebaldi, Michael D. Mastrandrea, Walter P. Falcon, and Rosamond L. Naylor. 2008. "Prioritizing Climate Change Adaptation Need for Food Security in 2030." *Science* 319 (5863): 607–10.

Maccini, Sharon, and Dean Yang. 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review* 99 (3): 1006–26.

Marchiori, Luca, Jean François Maystadt, and Ingmar Schumacher. 2012. "The Impact of Weather Anomalies on Migration in Sub-Saharan Africa." *Journal of Environmental Economics and Management* 63 (3): 355–74. doi:10.1016/j.jeem.2012.02.001.

Naylor, Rosamond L., David S. Battisti, Daniel J. Vimont, Walter P. Falcon, and Marshall B. Burke. 2007. "Assessing Risks of Climate Variability and Climate Change for Indonesian Rice Agriculture." *Proceedings of the National Academy of Sciences* 104 (19): 7752–57. doi:10.1073/pnas.0701825104.

Naylor, Rosamond L., Walter Falcon, Nikolas Wada, and Daniel Rochberg. 2002. "Using El Niño-Southern Oscillation Climate Data to Improve Food Policy Planning in Indonesia." *Bulletin of Indonesian Economic Studies* 38 (1): 75–91.

OECD, and Asian Development Bank. 2015. "Education in Indonesia: Rising to the Challenge." Vol. 20. Paris, France. doi:10.1525/as.1951.20.15.01p0699q.

Rienecker, Michele M., Max J. Suarez, Ronald Gelaro, Ricardo Todling, Julio Bacmeister, Emily Liu, Michael G. Bosilovich, et al. 2011. "MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications." *Journal of Climate* 24 (14): 3624–48. doi:10.1175/JCLI-D-11-00015.1.

Shah, Manisha, and Bryce Millett Steinberg. 2013. "Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital." *National Bureau of Economic Research Working Paper Series* No. 19140. doi:10.3386/w19140.

Skoufias, Emmanuel, Roy S. Katayama, and B. Essama-Nssah. 2012. "Too Little Too Late: Welfare Impacts of Rainfall Shocks in Rural Indonesia." *Bulletin of Indonesian Economic Studies* 48 (3): 351–68.

Thomas, Duncan, Kathleen Beegle, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, and Graciela Teruel. 2004. "Education in a Crisis." *Journal of Development Economics* 74 (1): 53–85.

Zander, Kerstin K., Wouter J. W. Botzen, Elspeth Oppermann, Tord Kjellstrom, and Stephen T. Garnett. 2015. "Heat Stress Causes Substantial Labour Productivity Loss in Australia." *Nature Climate Change* 5: 647–52. doi:10.1038/nclimate2623.

Table 2.1
Demographic and wealth variables: summary statistics

Dropped out during inter-survey period, share	0.14 (0.35)
Repeated grade during inter-survey period, share	0.06 (0.24)
Total educational expenditures (1000s 07/08 rupiah)	1,983 (12,048)
Focused educational expenditures (1000s 07/08 rupiah)	1,409 (10,888)
Female, share	0.49 (0.50)
Only child, share	0.25 (0.43)
Age, years	9.58 (4.47)
Household size	5.47 (1.90)
Female headed household, share	0.10 (0.30)
Household generates farm income, share	0.39 (0.49)
In primary school, share	0.76 (0.43)
Household head educational attainment (years)	6.02 (4.25)
Household's current assets (1000s 07/08 rupiah)	106,812 (398,679)
Household's total assets sold last 12 months (1000s 07/08 rupiah)	26 (441)
Total household income: all sources last 12 months (1000s 07/08 rupiah)	17,183 (190,278)
Observations	21,053

Note: These are the demographic and wealth variables that line up with the inter-survey period unit of observation for school-aged children who were in school and therefore at risk of dropping out. The focused educational expenditures includes: school registration, exam fees, books, writing supplies, uniforms, sports, other scheduled fees, and special courses. The total educational expenditures variables also includes the cost of room and board and the transportation costs related to getting to school. Average exchange rate in 2007/2008 was approximately 9,200 rupiah per USD. As an example, the average total household income of 17,218,000 rupiah equals approximately 1,872 USD. Means presented with standard deviations in parenthesis.

Table 2.2
Weather variables: summary statistics

Temperature, inter-survey period, celsius	27.27
	(1.62)
Deviation of annual temperature from long term mean	0.03
	(0.10)
Standardized deviation of temp from long term mean	0.12
	(0.41)
Mean monthly temperature above 1SD of mean for that calendar month (count)	10.07
	(4.45)
Mean monthly temperature below 1SD of mean for that calendar month (count)	8.46
	(6.36)
Mean monthly temperature above 2SD of mean for that calendar month (count)	2.65
	(2.46)
Mean monthly temperature below 2SD of mean for that calendar month (count)	0.49
	(1.02)
Rainfall, monthly, inter-survey period, mm	264.91
	(60.63)
Deviation of annual rainfall from long term mean	-6.23
	(22.63)
Standardized deviation of rain from long term mean	-0.20
	(0.60)
Mean monthly rainfall above 1SD of mean for that calendar month (count)	11.11
	(7.50)
Mean monthly rainfall below 1SD of mean for that calendar month (count)	11.39
	(4.21)
Mean monthly rainfall above 2SD of mean for that calendar month (count)	1.98
	(1.64)
Mean monthly rainfall below 2SD of mean for that calendar month (count)	0.68
	(0.94)
Monsoon onset delay from long term mean, days	1.20
	(3.35)
Observations	21,053

Note: These are the weather variables that line up with the inter-survey period unit of observation for school-aged children who were in school and therefore at risk of dropping out. Monsoon onset delay is calculated as the number of days after August 1st when cumulative rainfall reaches 20cm. Means presented with standard deviations in parenthesis.

Table 2.3
Community infrastructure variables: summary statistics

Rural community, share	0.53
	(0.50)
Elementary schools in community	6.10
	(3.50)
Junior high schools in community	2.93
	(1.76)
High schools in community	2.46
	(2.59)
Hours per day with electricity	22.04
	(6.31)
No irrigation, share	0.60
	(0.49)
Non-mechanized irrigation, share	0.21
	(0.40)
Mechanized irrigation, share	0.19
	(0.40)
Paved asphalt roads, share	0.74
	(0.44)
Paved stone roads, share	0.16
	(0.36)
Dirt roads, share	0.09
	(0.28)
Waterway, share	0.01
	(0.11)
Road accessible by vehicle twelve months per year, share	0.93
	(0.25)
Observations	21,053

Note: These are the infrastructure variables that line up with the inter-survey period unit of observation for school-aged children who were in school and therefore at risk of dropping out. Means presented with standard deviations in parenthesis. Values presented are rounded to two decimal places.

Table 2.4

Estimated effects of climate variables on probability of dropout

	(1) Dropout Spec A	(2) Dropout Spec B	(3) Dropout Spec C
Deviation of annual temperature from long term mean	0.42	-1.12	-0.68
	(0.91)	(0.99)	(0.78)
Deviation of annual temperature from long term mean, squared	-3.58	3.27	
	(4.36)	(4.00)	
Deviation of annual rainfall from long term mean	0.00		
	(0.00)		
Deviation of annual rainfall from long term mean, squared	-0.00		
	(0.00)		
Monsoon onset delay, days		0.04*	0.01
		(0.02)	(0.02)
Monsoon onset delay, days, squared		-0.00***	
		(0.00)	
Female	0.02	0.02	0.02
	(0.05)	(0.05)	(0.05)
Only child	-0.15**	-0.15**	-0.15**
	(0.07)	(0.07)	(0.07)
Age	0.44***	0.44***	0.44***
	(0.03)	(0.03)	(0.03)
Household size	0.01	0.01	0.01
	(0.02)	(0.02)	(0.02)
Female headed household	-0.01	-0.00	-0.00
	(0.09)	(0.09)	(0.09)
Household head education (years)	-0.12***	-0.12***	-0.12***
	(0.01)	(0.01)	(0.01)
Household's current assets (1000s 07/08 rupiah)	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)
Household income: all sources last 12 months (1000s 07/08 rupiah)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Rural location	0.19	0.25	0.24
	(0.41)	(0.38)	(0.40)
Household has farm income	-0.20***	-0.20***	-0.20***
	(0.07)	(0.07)	(0.07)
In primary school	0.41***	0.41***	0.41***
	(0.11)	(0.11)	(0.11)
Uses irrigation	-0.33**	-0.35**	-0.32**
	(0.15)	(0.15)	(0.15)
Hours per day with electricity	0.01**	0.01**	0.01**
	(0.01)	(0.01)	(0.01)
Constant	-6.27***	-6.51***	-6.47***
	(0.55)	(0.54)	(0.54)
Period and community fixed effects	Yes	Yes	Yes
Observations (person-periods)	20,443	20,443	20,443
Prob > chi^2	0.000	0.000	0.000
Pseudo R^2	0.325	0.325	0.325

Standard errors clustered at 0.5 x 0.5 degree pixel; *** p<0.01, ** p<0.05, * p<0.1

Note: The variables for assets sold in the last 12 months, roads accessible for last 12 months, type of road, and the count of elementary, junior high, and high schools in each community are included in the regressions, but the coefficients were not statistically significant. Therefore they are removed due to space considerations.

Table 2.5

Estimated effects of climate variables with interactions on probability of dropout

	(1) Dropout Spec D	(2) Dropout Spec E	(3) Dropout Spec F
Deviation of annual temperature from long term mean	-3.10***	-0.86	-1.31
	(0.86)	(0.79)	(0.81)
Monsoon onset delay, days	0.12***	-0.00	0.03
	(0.03)	(0.02)	(0.02)
Primary school x Deviation of annual temperature	4.16***		
	(1.18)		
Primary school x Monsoon onset delay	-0.16***		
	(0.04)		
Female x Deviation of annual temperature		0.42	
		(0.59)	
Female x Monsoon onset delay		0.03**	
		(0.02)	
Household owns farm x Deviation of annual temperature			1.61**
			(0.62)
Household owns farm x Monsoon onset delay			-0.02
			(0.02)
Female	0.03	-0.05	0.02
	(0.05)	(0.06)	(0.05)
Only child	-0.16**	-0.15**	-0.15**
	(0.07)	(0.07)	(0.07)
Age	0.44***	0.44***	0.44***
	(0.03)	(0.03)	(0.03)
Household size	0.01	0.01	0.01
	(0.02)	(0.02)	(0.02)
Female headed household	-0.00	-0.00	-0.00
	(0.09)	(0.09)	(0.09)
Household head education (years)	-0.12***	-0.12***	-0.12***
	(0.01)	(0.01)	(0.01)
Household's current assets (1000s 07/08 rupiah)	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)
Rural location	0.19	0.25	0.24
	(0.38)	(0.40)	(0.39)
Household has farm income	-0.20***	-0.20***	-0.21***
	(0.07)	(0.07)	(0.08)
In primary school	0.51***	0.41***	0.41***
	(0.13)	(0.11)	(0.11)
Uses irrigation	-0.32**	-0.32**	-0.33**
	(0.14)	(0.15)	(0.15)
Hours per day with electricity	0.01	0.01*	0.01*
	(0.01)	(0.01)	(0.01)
Constant	-6.39***	-6.44***	-6.39***
	(0.53)	(0.54)	(0.54)
Period and community fixed effects	Yes	Yes	Yes
Observations (person-periods)	20,443	20,443	20,443
Prob > chi^2	0.000	0.000	0.000
Pseudo R^2	0.330	0.325	0.325

Standard errors clustered at 0.5 x 0.5 degree pixel; *** p<0.01, ** p<0.05, * p<0.1

Note: The variables for assets sold, income, and roads accessible over the last 12 months, type of road, and the count of elementary, junior high, and high schools in each community are included in the regressions, but the coefficients were not statistically significant. Therefore they are removed due to space considerations.

Table 2.6

Estimated effects of climate variables with interactions on probability of dropout

	(1) Dropout Spec G	(2) Dropout Spec H	(3) Dropout Spec I
Deviation of annual temperature from long term mean	-1.87**	-1.50*	-3.86***
	(0.82)	(0.77)	(0.98)
Monsoon onset delay, days	0.06**	0.02	0.12***
	(0.03)	(0.02)	(0.04)
Primary school x Deviation of annual temperature			4.05***
			(1.12)
Primary school x Monsoon onset delay			-0.16***
			(0.04)
Female x Deviation of annual temperature			0.34
			(0.58)
Female x Monsoon onset delay			0.03*
			(0.02)
Household owns farm x Deviation of annual temperature			1.04
			(0.72)
Household owns farm x Monsoon onset delay			0.01
			(0.02)
Rural location x Deviation of annual temperature	1.78**		-1.00
	(0.79)		(0.81)
Rural location x Monsoon onset delay	-0.07***		-0.02
	(0.03)		(0.03)
Uses irrigation x Deviation of annual temperature		2.78***	2.45**
		(0.91)	(0.97)
Uses irrigation x Monsoon onset delay		-0.06**	-0.05**
		(0.02)	(0.02)
Female	0.02	0.02	-0.04
	(0.05)	(0.05)	(0.06)
Rural location	0.24	0.26	0.23
	(0.41)	(0.41)	(0.38)
Household has farm income	-0.20***	-0.20***	-0.26***
	(0.07)	(0.07)	(0.09)
In primary school	0.41***	0.41***	0.52***
	(0.11)	(0.11)	(0.13)
Uses irrigation	-0.27*	-0.34**	-0.30**
	(0.15)	(0.14)	(0.13)

Note: See continuation of table for other covariates.

Table 2.6, continued

Estimated effects of climate variables with interactions on probability of dropout

	(1) Dropout Spec G	(2) Dropout Spec H	(3) Dropout Spec I
Only child	-0.16**	-0.15**	-0.16**
	(0.07)	(0.07)	(0.07)
Age	0.44***	0.44***	0.44***
	(0.03)	(0.03)	(0.03)
Household size	0.01	0.01	0.01
	(0.02)	(0.02)	(0.02)
Female headed household	-0.00	-0.00	-0.00
	(0.09)	(0.09)	(0.09)
Household head education (years)	-0.12***	-0.12***	-0.12***
	(0.01)	(0.01)	(0.01)
Household's current assets (1000s 07/08 rupiah)	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)
Household's total assets sold last 12 months (1000s 07/08 rupiah)	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Household income: all sources last 12 months (1000s 07/08 rupiah)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Elementary schools in community	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)
Junior high schools in community	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)
High schools in community	-0.02	-0.02	-0.02
	(0.01)	(0.01)	(0.01)
Paved stone roads	0.04	0.03	0.05
	(0.10)	(0.11)	(0.10)
Dirt roads	-0.01	-0.05	0.08
	(0.15)	(0.16)	(0.16)
Waterways	0.23	0.38	0.27
	(0.40)	(0.41)	(0.39)
Road accessible by vehicle 12 months per year	0.08	0.13	0.08
	(0.17)	(0.16)	(0.15)
Hours per day with electricity	0.01	0.01**	0.01
	(0.01)	(0.01)	(0.01)
Constant	-6.47***	-6.60***	-6.50***
	(0.54)	(0.56)	(0.55)
Period and community fixed effects	Yes	Yes	Yes
Observations (person-periods)	20,443	20,443	20,443
Prob > chi^2	0.000	0.000	0.000
Pseudo R^2	0.325	0.326	0.332

Standard errors clustered at 0.5 x 0.5 degree pixel; *** p<0.01, ** p<0.05, * p<0.1

Note: Road type or waterways is a categorical variable with paved asphalt roads as the omitted category.

Table 2.7

Estimated effects of climate variables on educational expenditures (1000s 07/08 rupiah)

	(1) Expend Spec A	(2) Expend Spec B	(3) Expend Spec C
Deviation of annual temperature from long term mean	4,463	2,552	5,091
	(6,021)	(5,181)	(4,802)
Deviation of annual temperature from long term mean, squared	8,852	14,781	
	(28,671)	(28,361)	
Deviation of annual rainfall from long term mean	-32*		
	(18)		
Deviation of annual rainfall from long term mean, squared	-0		
	(0)		
Monsoon onset delay, days		323***	194**
		(123)	(90)
Monsoon onset delay, days, squared		-30**	
		(12)	
Female	100	104	101
	(174)	(173)	(174)
Only child	-378	-376	-386
	(257)	(258)	(259)
Age	-98**	-95**	-96**
	(40)	(40)	(40)
Household size	-46	-52	-51
	(41)	(42)	(42)
Female headed household	-127	-151	-156
	(240)	(235)	(235)
Household head education (years)	82***	79***	80***
	(23)	(23)	(23)
Household's current assets (1000s 07/08 rupiah)	0	0	0
	(0)	(0)	(0)
Household income: all sources last 12 months (1000s 07/08 rupiah)	0	0	0
	(0)	(0)	(0)
Rural location	6,160*	5,850*	5,604
	(3,541)	(3,411)	(3,444)
Household has farm income	-29	-53	-77
	(167)	(163)	(162)
In primary school	-94	-81	-83
	(241)	(241)	(240)
Uses irrigation	-515	-731	-598
	(579)	(563)	(577)
Hours per day with electricity	-81	-92	-88
	(60)	(59)	(60)
Constant	-3,306	-3,088	-2,670
	(4,277)	(4,173)	(4,042)
Period and community fixed effects	Yes	Yes	Yes
Observations (person-periods)	20,649	20,649	20,649
R^2	0.082	0.083	0.083
F-test	2.756	2.737	2.936
Prob > F	0.001	0.001	0.001

Standard errors clustered at 0.5 x 0.5 degree pixel; *** p<0.01, ** p<0.05, * p<0.1

Note: The variables for assets sold in the last 12 months, roads accessible for last 12 months, type of road, and the count of elementary, junior high, and high schools in each community are included in the regressions, but the coefficients were not statistically significant. Therefore they are removed due to space considerations.

Table 2.8

Estimated effects of climate variables on educational expenditures (1000s 07/08 rupiah)

	(1) Expend Spec D	(2) Expend Spec E	(3) Expend Spec F
Deviation of annual temperature from long term mean	10,843**	6,372	4,818
	(4,595)	(4,967)	(4,938)
Monsoon onset delay, days	323***	193**	161
	(105)	(92)	(118)
Primary school x Deviation of annual temperature	-8,466**		
	(3,304)		
Primary school x Monsoon onset delay	-156**		
	(64)		
Female x Deviation of annual temperature		-2,693*	
		(1,501)	
Female x Monsoon onset delay		5	
		(24)	
Household owns farm x Deviation of annual temperature			2,023
			(2,515)
Household owns farm x Monsoon onset delay			49
			(83)
Female	99	164	100
	(174)	(190)	(174)
Only child	-403	-383	-382
	(261)	(259)	(259)
Age	-76**	-96**	-96**
	(42)	(42)	(42)
Household size	-54	-51	-49
	(42)	(42)	(42)
Female headed household	-192	-149	-155
	(236)	(235)	(235)
Household head education (years)	79***	80***	80***
	(23)	(23)	(23)
Household's current assets (1000s 07/08 rupiah)	0	0	0
	(0)	(0)	(0)
Rural location	5,621	5,606	5,671
	(3,446)	(3,446)	(3,453)
Household has farm income	-105	-76	-198
	(162)	(162)	(239)
In primary school	478	-87	-82
	(333)	(241)	(241)
Uses irrigation	-509	-603	-653
	(570)	(577)	(571)
High schools in community	-71	-72*	-72*
	(43)	(43)	(43)
Constant	-3,597	-2,702	-2,489
	(4,083)	(4,046)	(4,043)
Period and community fixed effects	Yes	Yes	Yes
Observations (person-periods)	20,649	20,649	20,649
R^2	0.083	0.082	0.082
Prob > F	0.000	0.001	0.001

Standard errors clustered at 0.5 x 0.5 degree pixel; *** p<0.01, ** p<0.05, * p<0.1

Note: The variables for assets sold, income, road accessibility, type of road, count of elementary and junior high schools in each community, and hours per day with electricity are included in the regressions, but the coefficients were not statistically significant. Therefore they are removed due to space considerations.

Table 2.9

Estimated effects of climate variables on educational expenditures (1000s 07/08 rupiah)

	(1) Expend Spec G	(2) Expend Spec H	(3) Expend Spec I
Deviation of annual temperature from long term mean	3,173	2,940	9,527**
	(5,121)	(4,880)	(4,556)
Monsoon onset delay, days	203	111	274*
	(131)	(89)	(147)
Primary school x Deviation of annual temperature			-9,525***
			(3,515)
Primary school x Monsoon onset delay			-147**
			(65)
Female x Deviation of annual temperature			-2,801*
			(1,519)
Female x Monsoon onset delay			3
			(23)
Household owns farm x Deviation of annual temperature			-1,615
			(2,163)
Household owns farm x Monsoon onset delay			100
			(66)
Rural location x Deviation of annual temperature	4,623		4,436
	(3,185)		(3,524)
Rural location x Monsoon onset delay	-22		-192
	(126)		(124)
Uses irrigation x Deviation of annual temperature		4,331	2,850
		(3,242)	(3,841)
Uses irrigation x Monsoon onset delay		262*	342**
		(139)	(150)
Female	98	97	160
	(172)	(172)	(188)
Rural location	5,579	5,709*	5,811*
	(3,419)	(3,353)	(3,314)
Household has farm income	-48	-37	-151
	(165)	(165)	(236)
In primary school	-103	-114	465
	(244)	(242)	(338)
Uses irrigation	-721	-772	-601
	(594)	(589)	(582)

Note: See continuation of table for other covariates.

Table 2.9, continued

Estimated effects of climate variables on educational expenditures (1000s 07/08 rupiah)

	(1) Expend Spec G	(2) Expend Spec H	(3) Expend Spec I
Household size	-50	-51	-53
	(42)	(42)	(42)
Female headed household	-148	-164	-192
	(235)	(237)	(238)
Household head education (years)	81***	80***	80***
	(23)	(23)	(23)
Household's current assets (1000s 07/08 rupiah)	0	0*	0*
	(0)	(0)	(0)
Only child	-384	-372	-390
	(258)	(257)	(260)
Age	-98**	-100**	-79**
	(41)	(41)	(37)
Household's total assets sold last 12 months (1000s 07/08 rupiah)	0	0	0
	(0)	(0)	(0)
Household income: all sources last 12 months (1000s 07/08 rupiah)	-0	-0	-0
	(0)	(0)	(0)
Elementary schools in community	83	70	77
	(63)	(63)	(63)
Junior high schools in community	164	166	168
	(124)	(122)	(122)
High schools in community	-71	-69	-68
	(44)	(44)	(45)
Hours per day with electricity	-91	-92	-96
	(61)	(59)	(63)
Paved stone roads	-139	33	64
	(412)	(422)	(414)
Dirt roads	541	679*	769*
	(400)	(389)	(394)
Waterways	-2,003*	-2,023*	-2,085*
	(1,193)	(1,138)	(1,102)
Road accessible by vehicle 12 months per year	1,117	1,096	1,023
	(737)	(725)	(730)
Constant	-2,327	-2,896	-3,822
	(4,010)	(4,031)	(4,087)
Period and community fixed effects	Yes	Yes	Yes
Observations (person-periods)	20,649	20,649	20,649
R^2	0.082	0.083	0.085
Prob > F	0.001	0.000	0.001

Standard errors clustered at 0.5 x 0.5 degree pixel; *** p<0.01, ** p<0.05, * p<0.1

Note: Road type or waterways is a categorical variable with paved asphalt roads as the omitted category.

Figure 2.1

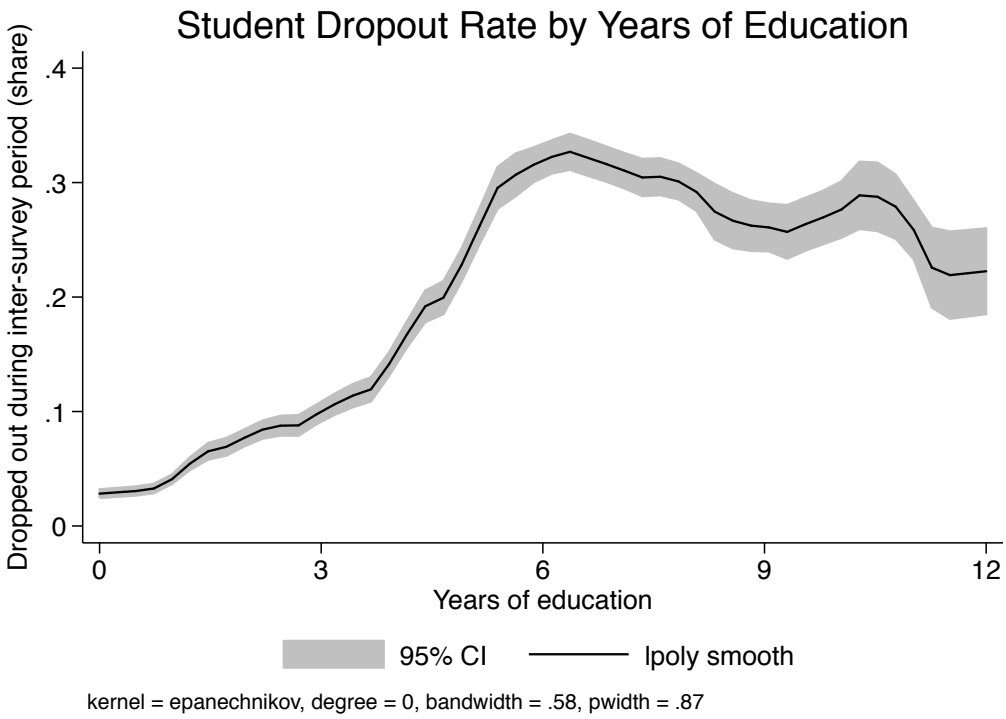

Student Dropout Rate by Years of Education
Dropped out during inter-survey period (share)
.4
.3
.2
.1
0
0
3
6
9
12
Years of education
95% CI
lpoly smooth
kernel = epanechnikov, degree = 0, bandwidth = .58, pwidth = .87

Figure 2.2

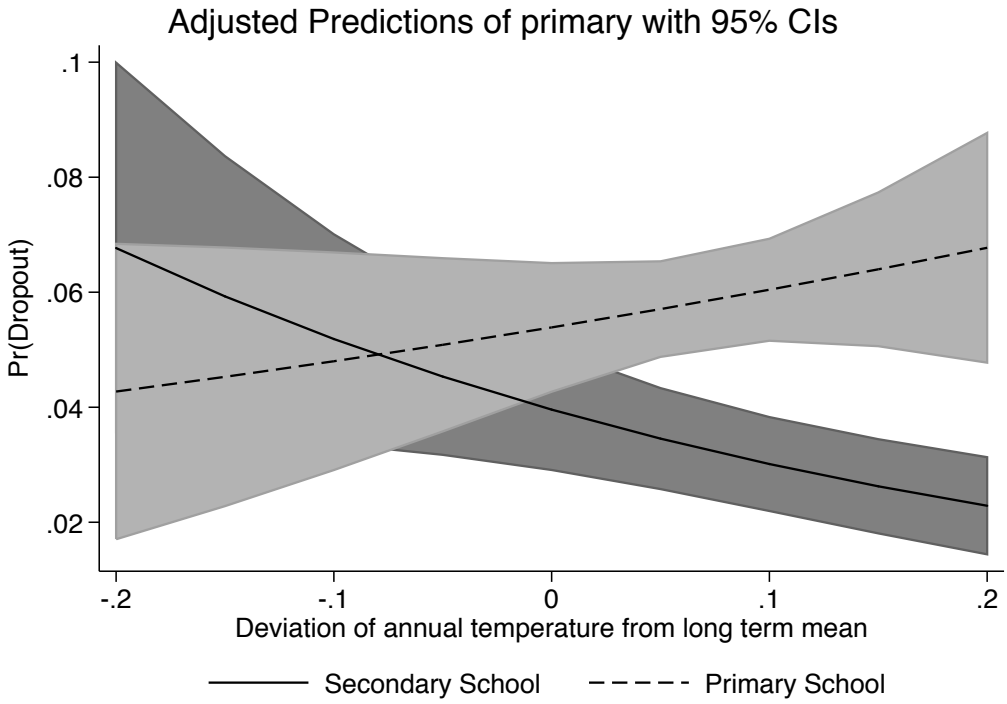
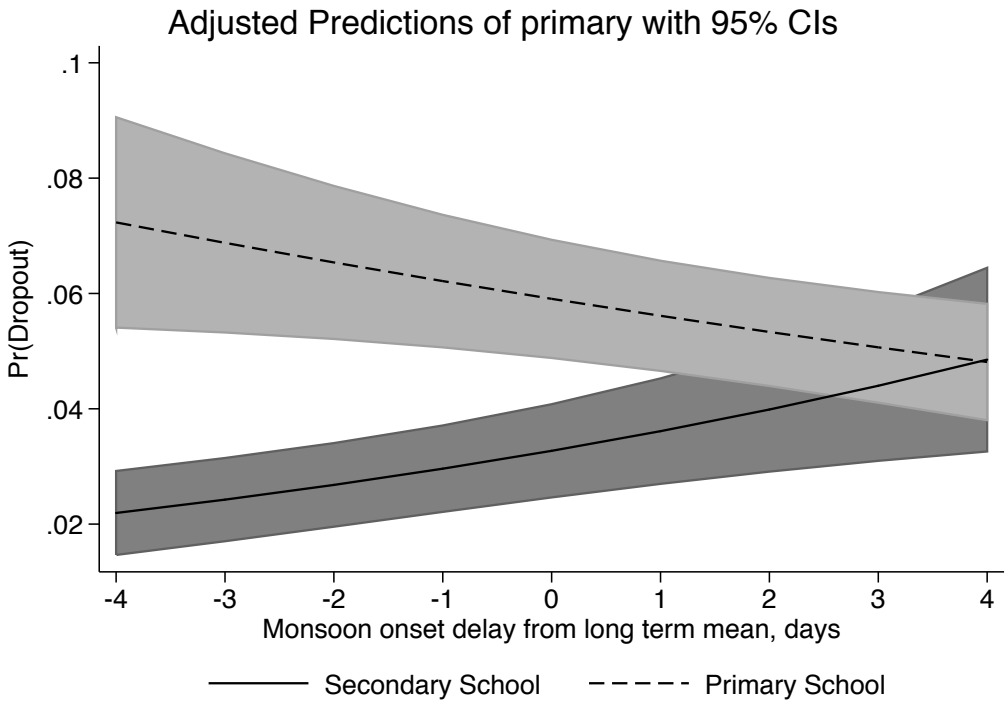


Figure 2.3



CHAPTER 3

WHAT IS THE OPTIMAL LOCUS OF CONTROL FOR SOCIAL ASSISTANCE PROGRAMS? EVIDENCE FROM THE PRODUCTIVE SAFETY NET PROGRAM IN ETHIOPIA

Introduction

The debate over the optimal locus of government program control dates at least to Oates (1972) who argued that the main rationale for decentralization is bringing decision-makers closer to the people, thereby increasing the chances that leaders' choices reflect the preferences of the people. Since Oates, a broad and robust literature—theoretical and empirical—debates whether a centralized or decentralized approach is more appropriate for the delivery of social assistance programs.1 Further, when focusing on implementation of social assistance programs in developing countries, a significant body of research argues that decentralization makes things worse for the poor. For example, decentralized education spending in Uganda was unintentionally regressive (Reinikka and Svensson 2004); local governments selected

[1] Arguments favoring decentralized implementation include lower costs to acquire detailed and accurate information (Alderman 2002), increased responsiveness to local needs (Faguet 2004), increased knowledge of what is politically and socially feasible in the local context (Pritchett 2005), increased accountability (Agrawal and Ribot 1999), less corruption (Fisman and Gatti 2002), and encouragement of local participation (Véron et al. 2006). However, others argue that decentralization worsens service delivery, for example when there is political capture by local elites (Bardhan and Mookherjee 2000; Bardhan and Mookherjee 2005), when preferences of local decision makers are not pro-poor (Conning and Kevane 2002), when local governments have weak implementation capacity (Smith 1985), when monitoring mechanisms such as free press are weak (Lessmann and Markwardt 2010), or in ethnically heterogeneous or sparsely populated areas (Olken 2006).

local work projects with less employment of the poor in West Bengal (Bardhan and Mookherjee 2006); and test scores in poor municipalities did not change despite improvements to test scores in wealthier municipalities after decentralization of schools in Argentina (Galiani, Gertler, and Schargrodsky 2008).

In this paper we examine winners and losers from the unintended decentralized implementation of Ethiopia's Productive Safety Net Program (PSNP). The central government dictates a precise allocation rule, with full and uniform transfers per person for participating households. In practice, however, communities do not receive enough funding to implement the program according to the central government mandate. Therefore, communities must use local discretion to distribute aid. In order to examine the distributional impacts of this decentralized approach, we need a rule to compare the welfare of households of differing sizes and compositions; however, this rule must be flexible enough to allow for community specific criteria to influence which households are selected into the program and how much aid they receive. The literature does not provide an adequate rule, so we derive one theoretically based on the principle of revealed preference and develop an empirical method to implement it, building on Olken (2005). We then use the estimated allocation rule to examine the distributional consequences of decentralized implementation.

We find that decentralization can lead to more pro-poor allocation rules, and that Ethiopian communities allocate more aid to underprivileged groups with lower wage earnings potential (e.g., teenage girls vs. teenage boys, adult women vs. adult men,

elderly vs. working age adults). However, when comparing the communities' pro-poor approach with the central government's uniform allocation rule—in which the same reduced budget gets spread evenly among fewer beneficiaries, each of whom receives the full prescribed payment—we find that overall poverty rates do not vary significantly between the two approaches. Furthermore, the poverty rates in both limited budget scenarios is not statistically significantly different than a counterfactual of no PSNP payment. By contrast, when we simulate poverty levels with a fully funded program, we find that both the decentralized approach and the centralized approach reduce overall poverty rates by approximately the same amount.

The contribution of this paper is twofold. The first is methodological. We extend Olken's (2005) technique of estimating locally determined equivalence scales to settings where communities determine not only the extensive margin (i.e., whether a household receives aid or not) but also the intensive margin (i.e., how much aid each beneficiary household receives). This adds to the thin literature on "socially adequate" consumption levels as described by Pollak and Wales (1979). They argue that equivalence scales based on demand systems aptly deal with the creation of cost of living indexes, where it is appropriate for the social scientist to specify the base preference ordering against which all households are measured. They also argue, however, that traditional demand based techniques of calculating equivalence scales do not deal adequately with the broader question of what communities or societies feel is the "socially adequate" consumption level. This paper presents a method to arrive at "socially adequate" consumption levels.

The second is a policy contribution. The overall program funding level matters far more for poverty reduction than whether the central government or local communities control program targeting and implementation. This paper provides operational advice to the government of Ethiopia that whatever funds are used to monitor and enforce the full and uniform payment levels, could be redeployed to enlarge the overall transfer amount of the program and reduce poverty more. Additionally, this paper contributes to the broader policy debate about how best to target the poor in social assistance programs. It adds to growing evidence that scale matters more for poverty reduction than whether central or local governments implement the targeting procedure. For example, Alatas et al. (2012) randomize targeting approaches in Indonesia and find that centralized targeting efforts (proxy means tests) perform slightly better than community-based targeting in identifying those below the poverty line. However, community-based methods create higher community satisfaction, and the difference between these methods does not yield statistically significantly different effects on reducing overall poverty rates. The case of Ethiopia's PSNP is much the same in that the resultant poverty levels are statistically indistinguishable between the decentralized and centralized approaches. In sum, the overall funding level drives performance in reducing poverty more than the locus of program control does.

Background and Targeting of the Productive Safety Net Program

More than 80% of Ethiopia's population lives in rural areas and relies on rain-fed agriculture as its main livelihood. Historically, insufficient and variable rainfall caused

cycles of food shortage and famine, and the government of Ethiopia requested international assistance when help was needed. In the early 2000s, the government and a consortium of international donors moved towards a model to address underlying chronic food insecurity instead of the repeated *ad hoc* emergency appeals for acute food shortages caused by drought. Therefore, the government of Ethiopia launched the PSNP in January 2005.

The PSNP is designed to assist approximately 7-8 million people per year and has an approximate annual budget of USD$350 million (Development Assistance Group 2010). The PSNP has two major parts: 1) a large public works (PW) program in which food insecure households provide daily labor to public works projects in exchange for food or cash, [2] and 2) a smaller direct support (DS) component in which households without available labor (generally the elderly or disabled) receive a transfer with no work requirement.

Targeting of program participants

A combined administrative and community targeting approach is used in the PSNP. The amount of aid allocated to each district is determined at the federal level (based on need and historic receipt of food aid). Once district aid levels are determined, the

[2] One of the government of Ethiopia's initial stated goals of the PSNP was to move away from food aid and towards cash payments as aid. However, some donors, particularly the United States, would only give their contribution to the PSNP in the form of food aid, so the areas supported by US government resources are generally chosen to be the most remote and those with the least market access where food aid is perhaps a better option than cash transfers.

districts work with all of the villages within that district to determine exact beneficiary lists.[3] The Program Implementation Manual (PIM) mentions key criteria for participant selection including: the household is a member of the community, the household has faced continuous food shortages, and the household has faced sudden serious shock, and/or the household lacks adequate family support or other means of social protection. These criteria are broad and allow for significant local level discretion in determining who participates and who does not (as documented by Caeyers and Dercon (2012) in a similar Ethiopian program).

The exact administrative process for determining which households are included in the PSNP beneficiary list is an iterative process between the district and villages. The village level committees, their composition, and responsibilities are as follows:

(i) The Village Council is the elected leadership of the village and is tasked with approving the beneficiary list that is passed to the district and ensuring that the PSNP client list, along with program plans and budgets, are posted in a public place.

(ii) The Village Food Security Task Force (VFSTF) is comprised of village administrators, agriculture extension workers, health extension workers, volunteer community health workers, teachers, and community members.

[3] In Ethiopia, a district is known as a *woreda* (20,000-250,000 population), and a village is known as a *kebele* (2,000-4,000 population). The village is the lowest administrative unit of the government. We use the English names in this paper.

The VFSTF determines which households are eligible for the public works program versus the direct support program.

(iii) The Community Food Security Task Force (CFSTF) is comprised of one representative of the VFSTF, one agriculture extension worker, one health extension worker or volunteer community health worker, 2-3 elected female representatives, 2-3 elected male representatives, and one elected youth representative. Depending on the size of the village, there may be one or more CFSTFs created per village. They mobilize the community for the actual participatory planning exercise to determine households with the highest need. They also organize a public meeting to discuss the proposed list of PSNP participants and give community members the opportunity to suggest the addition or removal of names.

The District Food Security Task Force (DFSTF) approves the plans it receives from the village councils, and if there is some disagreement, it gives additional guidance and direction to the village council and other committees. While the program's design allows local discretion to determine which households are in or out of the PSNP (the extensive margin), the instructions in the PIM are explicit that a uniform payment per household member is required conditional on PSNP participation (the intensive margin). However, the data on actual payment levels shows large deviations from the prescribed uniform payment levels. In fact, contrary to design, between 60-80% of the variation in individual level payments is associated with the lowest administrative level of government (the village). See Appendix A for a full description of the variance decomposition exercise showing this.

Estimating Communities' Revealed Preferences Based on Aid Allocations

Typically equivalence scales are used to compare welfare across households of differing sizes and compositions. Historically this is accomplished by assigning some aspect of a household's demand decisions to be indicative of the household's welfare. For example, the food share of expenditures (Engel 1895) or the total expenditures on adult goods (Rothbarth 1943). Then the social scientist infers the amount of additional expenditures required to compensate a household with a different demographic composition, so that it has the same welfare as a reference household (Deaton 1997; Lewbel and Pendakur 2008). The drawback of traditional consumption based equivalence scales is the strong *a priori* assumption that the social scientist has selected the indicator(s) that correctly proxy for household welfare. For example, the Engel method has been shown to overstate the cost of children (Nicholson 1976), and the Rothbarth method understates the cost of children (Barten 1964). To deal with these biases, the literature developed more and more complex demand systems to account for the substitution effects between adult and children goods (e.g., Deaton & Muellbauer 1986; Lewbel 1997; Browning et al. 2013).

Pollak and Wales (1979) describe a meaningful disagreement within the broader economics field in how comparisons of consumption needs are calculated. They lament that researchers have not been able to develop a method to recover "socially adequate" consumption levels in the typical demand system approaches because the researcher assigns which aspect of household demand is indicative of actual welfare.

Olken (2005) proposes an innovative alternative to traditional demand-based equivalence scales based on the revealed preferences of communities when they allocate aid themselves. This approach would satisfy Pollak and Wales' critique, in that it removes the discretion of the social scientist in deciding what aspect of household demand is most indicative of welfare as it simply observes how communities make the inter-household comparisons for themselves.

Olken's technique is powerful, but it only examines decisions made at the extensive margin (whether or not the household is included in the aid program) rather than also the intensive margin (how much aid the household receives once included in the program). We extend his method and include the intensive margin in estimating the community's revealed preferences concerning the receipt of aid. In doing so, we add to the thin literature on methods for defining a "socially adequate" consumption level.

Estimating revealed community equivalence scales

Conceptually Olken's model is as follows. Each household's indirect utility function, as evaluated by the community is:

$$v(y,n,k,x,p,a) \tag{3.1}$$

where y represents total household expenditures (not including aid receipts), n represents total number of people in the household, k represents the number of children in the household, x represents other household characteristics, p represents a

vector of prices, and a represents the amount of aid received by the household. Assume v is concave in y and the community maximizes a social welfare function:

$$\max \sum_{i=1}^{I} \beta(y_i, n_i, k_i, x_i, p) v(y_i, n_i, k_i, x_i, p, a_i) \qquad \text{s.t.} \sum_{i=1}^{I} a_i = A \tag{3.2}$$

where β represents welfare weights on each household, I is the total number of households in the community, and A represents total amount of aid to be distributed. There are important distinctions between β and v. For example, many aspects of a household's welfare might affect the community's decisions such as vulnerability to shocks or increased medical expenditures for the sick. These are captured in v. However, it is possible that other factors besides pure welfare maximization affect a village's decision of how to allocate aid, for example, the political connectedness of a household or a desire to provide social insurance to those suffering a recent unexpected shock. These are captured by β. Because the weights of β may also be related to household composition (through n or k) we cannot separately identify the community welfare weights β and the indirect utility function v in this context (Olken 2005). We can, however, identify the product of the two (called the overall community benefit function), which is denoted:

$$B(y_i, n_i, k_i, x_i, p, a_i) = \beta(y_i, n_i, k_i, x_i, p) v(y_i, n_i, k_i, x_i, p, a_i) \tag{3.3}$$

Then the community maximization problem becomes:

$$\max_{a_i} \sum_{i=1}^{I} B(y_i, n_i, k_i, x_i, p, a_i) \qquad \text{s.t.} \sum_{i=1}^{I} a_i = A \tag{3.4}$$

To differentiate the cost of children relative to adults and introduce household economies of scale, we parameterize these effects following Deaton (1997) and Olken (2005). For a given set of prices, let α be the cost of children relative to adults, so that each child costs as much as α adults. Define total number of effective adults to be $(n-(1-\alpha)k)^{\theta}$, where θ captures household economies of scale. As θ increases from 0, economies of scale within the household decline; constant returns to scale in household size corresponds to $\theta = 1$. (The federal uniform benefit schedule of the PSNP corresponds to $\alpha = 1$ and $\theta = 1$). Rewrite *B* so that it depends on household composition only through the effect of household composition on household expenditure per effective adult (Olken 2005). Expenditure per equivalent adult is defined as:

$$\tilde{y} = \frac{y}{(n-(1-\alpha)k)^{\theta}} \tag{3.5}$$

and then rewrite *B* so that it depends on *n* and *k*, only through $\tilde{y}$:

$$B(\tilde{y}, x_i, a_i) \tag{3.6}$$

Following Olken (2005), assume that prices in a local context are constant and remove the price vector p from the community benefit function.[4] Assume B is quasi-concave in income per equivalent adult $\tilde{y}$. Additionally assume that aid's only effect on welfare is through its value as an income supplement. Therefore it follows that:

$$\frac{\partial^2 B}{\partial \tilde{y} \partial a_i} < 0 \tag{3.7}$$

meaning that conditional on all other household characteristics x, the marginal utility of aid is higher for households with lower effective consumption (i.e., the marginal utility of aid is higher for the poor).

Based on the community benefit function and the assumptions presented, conditional on household characteristics x, the households with the lowest consumption per equivalent adult will receive aid. In theory, this means there is a threshold where all the households above the threshold do not receive aid and all the households below the threshold do receive aid. This threshold will vary by community based on how much aid the community has to distribute, the distribution of household utilities in the community, and the community's preference for targeting aid among the very poor, captured by the magnitude of $(\partial^2 B / \partial \tilde{y} \partial a_i)$.

[4] This assumption means that communities perceive that all households within the community face the same prices at a given time.

Next introduce an error term, and the probability that a household receives aid is equal to the probability that a household's consumption per effective adult, as evaluated by the community equivalence scales, is lower than some threshold. This threshold varies by community, so it can be modeled as a binary choice model with community fixed effects. This is equivalent to an equation in the form:

$$\Pr(Receive\_aid_{ij}) = F\left[\gamma_j + \gamma_2 B\left(\frac{y_{ij}}{(n_{ij} - (1-\alpha)k_{ij})^{\theta}}, x_{ij}\right)\right] \tag{3.8}$$

Where γ_j is the community fixed effect that captures different thresholds in each community and F is the distribution function for the error term.

Empirical specification of revealed community equivalence scales

Empirical estimation of the community benefit function (3.8) requires a functional form for B and the distribution of the error term F. Following Olken we use the log indirect utility function. Therefore the probability a household i in community j receives aid is:

$$\Pr(Receive\_aid_{ij}) = F\left[\gamma_j + \gamma_2 \log(y_{ij}) - \gamma_2 \theta \log(n_{ij} - (1-\alpha)k_{ij}) + \gamma_3 x_{ij}\right] \tag{3.9}$$

Because this is nonlinear, we estimate a linear approximation:[5]

$$\Pr(Receive\_aid_{ij}) = F\left[\gamma_j + \gamma_2 \log(y_{ij}) - \gamma_2\theta \log(n_{ij}) + \gamma_2\theta(1-\alpha)\left(\frac{k_{ij}}{n_{ij}}\right) + \gamma_3 x_{ij}\right] \qquad (3.10)$$

This can be extended to include different child age or gender categories to separately estimate equivalence scales for different groupings of children or to examine if communities exhibit a sex bias when distributing aid. To do that, include the percentage of household members in each child age or gender grouping rather than just the overall percentage of children.

Following Olken, we assume that the error term takes the logistic form, which allows us to use the conditional fixed-effects logit model. Rewriting equation (3.10) to incorporate this functional form requires additional notation. Let r_{ij} be a binary dependent variable equal to 1 if household i in village j received PSNP aid, and 0 otherwise. Let N_j be the number of households in village j and T_j be the number of households in village j that received PSNP aid. Denote d_{ij} to be a dummy variable equal to 1 if household i in village j received PSNP aid or 0 if the household did not receive aid, and denote by S_j the set of all possible vectors $d_j = \{d_1,...,d_{Nj}\}$ such that $\sum_{i=1}^{N_j} d_{ij} = T_j$. Define $\lambda_1 \equiv \gamma_2$, $\lambda_2 \equiv -\gamma_2\theta$, $\lambda_3 \equiv \gamma_2\theta(1-\alpha)$, and $\lambda_4 \equiv \gamma_3$. Substituting

[5] This is similar to the Working-Leser (Working 1943; Leser 1963) functional form used by Deaton and Muellbauer (1986).

the logistic CDF for F in equation 3.10 and conditioning out the fixed effects yields an empirical specification of the form:

$$\Pr\left(r_{ij} = 1 \mid \sum_{i=1}^{N_j} y_{ij} = T_j \right) = \frac{\exp\left[\sum_{i=1}^{N_j} y_{ij} \left(\lambda_1 \log(y_{ij}) + \lambda_2 \log(n_{ij}) + \lambda_3 \left(\frac{k_{ij}}{n_{ij}} \right) + \lambda_4 x_{ij} \right) \right]}{\sum_{d_j \in S_j} \exp\left[\sum_{i=1}^{N_j} d_{ij} \left(\lambda_1 \log(y_{ij}) + \lambda_2 \log(n_{ij}) + \lambda_3 \left(\frac{k_{ij}}{n_{ij}} \right) + \lambda_4 x_{ij} \right) \right]} \tag{3.11}$$

Equation (3.11) is estimated with maximum likelihood. Then using the estimated coefficients λ_1, λ_2 and λ_3 we recover estimates of θ and α. To compute the revealed community equivalence scale, which is the ratio of the income of the household with a given composition to that of a reference household, set the welfare levels for the reference and comparison household equal, and solve. As per Olken (2005), define a reference household with income y^R, size n^R, and number of children k^R, and comparison household with income y^C, size n^C, and number of children k^C. Setting equation (3.10) for the reference and comparison households equal yields:

$$\lambda_1 \log\left(\frac{y_{ij}^C}{y_{ij}^R} \right) = \lambda_3 \left(\frac{k_{ij}^R}{n_{ij}^R} - \frac{k_{ij}^C}{n_{ij}^C} \right) - \lambda_2 \log\left(\frac{n_{ij}^C}{n_{ij}^R} \right) \tag{3.12}$$

Dividing the right hand side by λ_1 and taking exponents yields the equivalence scales. In this model, the equivalence scale is independent of the income of the reference household (Olken 2005). To extend to multi-year data, we use the conditional fixed-

effects logit model, but instead of conditioning out community level fixed effects, we condition out community-year fixed effects. This allows the threshold point for program participation to change for a given community in each time period, however, it assumes that the parameter coefficients are the same for a given community over time (i.e., $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ do not change over time).

Empirical approach to calculating the intensive margin of participation

After estimating how household demographic structure affects selection into the PSNP (extensive margin of participation), we extend the method to consider whether household demographic structure affects the levels of payments once a family is included in the PSNP (intensive margin of participation). In a sequential model we use the results from (3.11) as the first stage of a two-stage selection model. We capture the predicted probability of PSNP participation, then convert that predicted probability to an inverse Mill's ratio (IMR) and include the IMR as a control variable in a pooled OLS regression with village-year fixed effects in the second stage. The IMR is calculated as:

$$\Lambda_{ijt} = \frac{\phi(\hat{r}_{ijt})}{\Phi(\hat{r}_{ijt})} \tag{3.13}$$

where $\phi(\hat{r}_{ijt})$ is the probability distribution function, and $\Phi(\hat{r}_{ijt})$ is the cumulative distribution function of $\hat{r}_{ijt}$, the predicted probability of PSNP participation from (3.11).[6] The second stage is a pooled OLS regression modeled as:

$$P_{ijt} = \alpha + Family\_Structure_{ijt}{}'\beta + X_{ijt}{}'\gamma + J_{tj} + \theta\Lambda_{ijt} + \varepsilon_{ijt} \qquad (3.14)$$

where P_{ijt} is the payment to household i at year t at village j and $Family\_Structure_{ijt}$ is a vector of household characteristics such as number of household members in age categories (ages 0-6, 7-15, 16-60, and 61+), and X_{ijt} is a vector of household characteristics that might affect payments such as annual expenditures, gender and age of household head, marital status, education level, asset holdings, local political connectedness, and household level shocks. J_{tj} is the village-year fixed effect, Λ_{ijt} is the IMR converted from the predicted probability of PSNP participation from the first stage equation, and ε_{ijt} is the error. To account for the non-negative censoring of P_{ijt} the Λ_{ijt} term serves as an estimate of the (otherwise) omitted variable of the probability of selection into the PSNP (Heckman 1979). The variables for local political participation are excluded from the second stage, the coefficients on these variables are statistically significant in the first stage, but if included they are not significant in the second stage. In essence, this means that local political participation

[6] Note the additional t subscript in $\hat{r}_{ijt}$ to denote time since our predicted probabilities come from equation (3.11) after it is taken to the multi-year extension.

can help a household enter the PSNP, but once in the PSNP, local political connections do not alter payment amounts made to households, so those variables serve as an effective instrument to identify the first stage.

The vector β is interpreted as the additional payout per household holding all else constant for one additional person in each of the age categories, it assumes that each person within a given age bracket is assigned the same value for β. The vector γ is interpreted as the additional household payment holding all else constant for an additional unit of each household characteristic, and θ is the coefficient on the IMR.

However, it is important to note that there is no guarantee that communities allocate aid to households using a two-stage sequential process. For example, the very real possibility that local communities receive less funding than the necessary amount to fully fund all qualified participants (combined with local communities' authority to select participants) means communities could simultaneously decide what households are included in the PSNP and their level of payment. In that case the selection into the PSNP and the selection of payment amount would be modeled simultaneously. To model the decision as simultaneous we use a standard tobit model in the form:

$$P_{ijt} = \begin{cases} P^*_{ijt} & if \quad P^*_{ijt} > 0 \\ 0 & if \quad P^*_{ijt} \leq 0 \end{cases} \tag{3.15}$$

where P_{ijt}^{*} is the latent variable:

$$P_{ijt}^{*} = \alpha + Family\_Structure_{ijt}{}'\eta + X_{ijt}{}'\psi + K_{jt} + \varepsilon_{ijt} \tag{3.16}$$

Where P_{ijt}^{*} is the latent variable of payment to household *I* at year *t* at village *j* and $Family\_Structure_{ijt}$ and X_{ijt} are vectors of household characteristics as above, K_{jt} is the village-year fixed effect, and μ_{ijt} is the error. We use a J-test (Davidson and MacKinnon 1993) to test which model better fits the data statistically. See Appendix B for a broader discussion of a sequential versus simultaneous model.

Simulating poverty reduction of decentralized versus centralized implementation

Because the PSNP is part of the government of Ethiopia's overall poverty reduction plan,[7] we simulate four main cases, which can be conceptualized in a 2x2 matrix with the level of funding (limited or full) on one axis and the locus of program control (decentralized vs. centralized) on the other axis. The four cases are: constrained funding and decentralized implementation (what we see in practice), constrained funding and centralized implementation (we give beneficiaries with the highest predicted probability of PSNP inclusion from equation (3.11) a full payment amount until the district budget is exhausted, then others receive no payment), full funding and decentralized implementation (the community allocation rules with full funding), and

[7] The PASDEP (Plan for Accelerated and Sustained Development to End Poverty) launched in 2006 was the Ethiopian government's over-arching poverty reduction strategy. The PSNP is a central pillar of the food security plan, which is a key element of the PASDEP.

full funding and centralized implementation (the program as designed). See Figure 3.1 for a visual representation.

We calculate the Foster, Greer, and Thorbecke (1984) poverty metrics under each of these simulated scenarios to understand how the actual allocation decisions at the local level affect poverty levels. The FGT metrics are calculated as follows:

$$FGT_{\alpha} = \frac{1}{N}\sum_{i=1}^{H}\left(\frac{z - y_i}{z}\right)^{\alpha} \tag{3.17}$$

where z is the Ethiopian government's poverty line, N is the total number of people in the economy, H is the number of poor (those at or below the poverty line), y_i is individual income (or, expenditures) and α is a sensitivity parameter. As α increases from zero the individuals farther away from the poverty line are given more weight. A higher FGT metric means more poverty in the economy. Expenditures (y_i) include the amount of aid (α_i) received if households received aid.

For any of these calculations we must recover the budget constraint for each district. Unfortunately, we do not have administrative records of the actual resources transferred from the federal level to the district government. We can, nonetheless, construct an estimate of the budget constraint faced by each district using the planning records for the PSNP caseload per district (Ethiopian Ministry of Agriculture and Rural Development 2010a). We generate an estimate of the budget constraint faced in

each district by multiplying the planned PSNP caseload with the per capita payment observed in each district. Since surveyed participants were selected randomly from the roster of beneficiaries, the average per capita payment received should be equivalent to the average per capita payment across the district. While not perfect, this measure should approximate the budget constraint faced by each district.

Data and Descriptive Statistics

The data are from the Ethiopian Food Security Survey (EFSS), a panel survey collected every two years in the four largest regions[8] of Ethiopia. The Central Statistical Agency (CSA) of Ethiopia collected the data with the support of the International Food Policy Research Institute (IFPRI). The dataset focuses on PSNP implementation areas and comprised 3,689 households in 2006; it expanded to 4,654 households in 2008 and beyond. Starting in 2008, the additional households were given the same questionnaire as the rest of the sample. The surveys take place in the traditional hungry season (June-August), which immediately precedes harvest time (September-October). In subsequent rounds, the same households were re-surveyed regardless of whether they had joined or left the PSNP.[9]

[8] The four largest regions, which comprise about 84% of Ethiopia's population, are Amhara, Tigray, Oromia, and Southern Nations and Nationalities People's Region (SNNPR).

[9] See Gilligan et al. (2007) and Gilligan et al. (2009) for a detailed description of the sampling methodology.

The survey asked households about their PSNP payments for the previous 17 months,[10] leaving a gap of 7 months with no data every other year. Therefore, the analysis of payments uses a yearly panel from 2006-2009 with the total January-May PSNP payments as the key dependent variable (see Figure 3.2).[11] While this is recall data, there is evidence that the recall data of monthly household aid receipts is likely to be accurate. The PSNP began the process of adding client cards for each beneficiary in 2009/2010. Once client cards were distributed, enumerators were instructed to ask households to produce their PSNP client card during the interview. The client card lists months and payments received (see sample client card in Figure 3). In the 2012 round of data collection, about half of the respondents could produce their client card and about half could not. The beneficiaries that could not produce their client card were asked to list payments from recall following the same procedure as in past rounds of data collection. The recall payment amounts and payments copied from the client cards had almost identical distributions for the two groups, with the mean payments only 2% different between groups (Berhane et al. 2013).

[10] The 2006 survey (performed approximately in July) only recounted PSNP payment data for 12 of the 18 months since the program start in January 2005.

[11] PSNP work is designed to occur between January-June (avoiding the primary agricultural season July-December) so the January-May data covers almost all of the scheduled PSNP payments. However, it is noted that arrears in payments occurred in some years (Berhane et al. 2011).

Descriptive statistics of program versus non-program participants

Table 1 presents the basic descriptive statistics of the variables used in the analysis that follows.[12] The average non-participant household has about 20% higher expenditures than a PSNP household, not including PSNP payments (12,458 ETB vs. 10,407 ETB). A PSNP household is much more likely to be female headed (25.7% vs. 16.8%), less educated (1.20 vs. 1.06 years completed), and slightly smaller in size (5.31 vs. 5.15). PSNP households have about one third of a hectare lower land holdings (1.43 vs. 1.13), fewer livestock (4.88 vs. 3.30 tropical livestock units), and less productive equipment (270 ETB vs. 244 ETB). PSNP households have more direct local political connections (12.9% vs. 9.5%) where an immediate family member is a member of the village government. PSNP households also have more extended local political connections (21.7% vs. 18.8%) where a friend or extended relative is a member of the village government. PSNP participants are more likely to have experienced the death of a spouse (2.5% vs. 1.7%) but the likelihood of facing a drought or illness shock is not statistically significantly different between groups. PSNP households have fewer working age adults (46.2% vs. 47.9%) and more elderly (7.9% vs. 6.2%), there is no statistical difference in the demographic composition of older children (24.9% vs. 24.6%) or young children (21.1% vs. 21.2%). A hypothetical "average" PSNP family has 5.15 members and is entitled to an annual payout[13] of

[12] Expenditures and value of productive assets are adjusted to 2009 currency units using World Bank consumer price index data for Ethiopia downloaded at: http://data.worldbank.org/indicator/FP.CPI.TOTL

[13] Because the PSNP payment rate changes throughout the dataset, PSNP payment amounts are normalized to 2009 payment levels. For example this adjustment makes a 65% payment in 2007 (117

1,545 ETB (5.15 people * 300 ETB/person/year), which is approximately 15% of annual household expenditures. Figure 4 displays a box plot of the PSNP payments received for each household size. While some households do receive their full entitlement, in general as household size increases the median household payment lags considerably compared to the full entitlement.

Descriptive statistics of district level budget constraints

On average, communities only received 62% of the full amount required to implement the program as per the district planning documents (Ethiopian Ministry of Agriculture and Rural Development 2010a); 89% of communities (69 of 78 districts in our data) received less than the necessary amount, while 11% received more than necessary to implement the program (Figure 5). While there is no publically available administrative records that document the amount of money sent to each district, the PSNP is structured in such a way that the level of international donations into the program determines the overall budget. An assessment document of the PSNP by the Independent Evaluation Group of the World Bank (2006) finds that the program anticipates a funding gap of USD$194.6 million in the 2007-09 period, which is equivalent to more than 20% of the total three year planned budget of USD$915.3 million. It is worth noting, however, that the report states that anticipated gaps can increase or decrease over time as implementation priorities change or more donors contribute funding.

ETB when the pay schedule is 180ETB/year) equal to a 65% payment in 2009 (195 ETB when the pay schedule is 300ETB/year).

Results

We begin by examining the community's selection of participants at the extensive margin, then the intensive margin, followed by poverty simulations comparing locus of control and level of funding.

Revealed community equivalence scales at the extensive margin of participation

The estimated odds ratios of the revealed community equivalence scales at the extensive margin (based on equation 3.11) are presented in Table 3.2 (the logistic regression results are presented in Appendix Table A1). The specifications are run with (col. 2 and 4) and without household controls (col. 1 and 3). Because communities are likely to take into account observable characteristics, the preferred specification is with household controls.

Larger households are associated with a higher probability of inclusion into the program as the odds ratio is statistically significantly higher than one in each specification. However, when examining the age structure of households (col. 2), the probability of inclusion in the PSNP is not different between age cohorts (none of the odds ratios are statistically significantly different from one), meaning communities treat all people as equivalent when assigning PSNP status. Because none of the coefficients on the age structure of households are statistically significant, it is not meaningful to calculate an actual set of equivalence scales as per equation 3.12.

When examining gender differences on the extensive margin (col. 4), there is no sex bias; boys and girls are treated equally (odds ratios are not statistically significantly different from one). However, households with more adult men are less likely to be selected into the PSNP than households with more adult women. Households with higher expenditures are less likely to be included in the PSNP; households with a marital status of only one spouse (single, divorced, widowed) have higher probabilities of inclusion than married couples (the omitted category). Local political connectedness (having a family member or friend with a position in the village) is associated with a higher probability of inclusion in the program, while higher asset holdings (land, livestock) are associated with lower probabilities of program inclusion. Suffering a household shock neither increases nor decreases the probability of inclusion.

These results are markedly different from those found in a similarly structured program in Indonesia. Olken (2005) finds that Indonesian communities allocate aid as if adding an additional child to a household only requires 76% of the amount spent on each of the first two adults. Our results suggest that, unlike in Indonesia, children and adults are treated as equivalent in determining household aid eligibility in Ethiopia. The one exception to that rule, the fact that having more working age men reduces the likelihood of PSNP participation in spite of working age men's relatively greater consumption requirements, is the first signal that Ethiopian communities use a different concept of poverty (one based on earnings potential rather than consumption needs) to define neediness.

Revealed community equivalence scales at the intensive margin of participation

Do communities decide how much households receive as a one stage or two-stage process? Based on the project implementation manual, we might expect a sequential two-stage process where households are initially chosen for inclusion to the PSNP and then households receive the federally mandated fixed per capita payment in a second stage. However, this is unlikely the case if communities have less funding than necessary to implement as per the PIM. Households, on average, only received 62% of the funds necessary to implement as per the instructions in the PIM. Furthermore, World Bank appraisal documents highlight international donor funding shortfalls of more than 20% for the PSNP (World Bank 2006). In a constrained setting like this, it seems highly probable that communities make decisions about the extensive and intensive margins of PSNP participation simultaneously to account for the shortfall in budget (e.g., the community includes a household in the program, but does so knowing they will assign the household a lower level of payment). The results of the simultaneous model (tobit) are presented here and the results of the sequential model are presented in Table A2 and discussed in Appendix B.

The simultaneous model (Table 3.3) specifications are run with (col. 2 and 4) and without household control variables (col. 1 and 3). The preferred specification is with household controls. A child aged 0-6 (col. 2) receives an estimated 15% more than an adult aged 16-60 (95.2 ETB vs. 83.1 ETB) but this difference is not statistically significantly different from zero. Older children aged 7-15 receive 27% more ($p<0.05$)

than adults (105.4 ETB vs. 83.1 ETB). The elderly receive 41% more than working aged adults (116.8 ETB vs. 83.1 ETB), however this difference is not statistically significantly different from zero (likely due to the small sample size of elderly).

When allowing for gender differences (col. 4), girls and boys aged 0-6 receive about the same payment (89.9 ETB vs. 95.0 ETB), girls aged 7-15 receive 34% more ($p<0.05$) than boys the same age (120.4 ETB vs. 90.1 ETB). Adult working aged women are paid 74% more ($p<0.01$) than working aged men (109.5 ETB vs. 62.8 ETB). Elderly women receive 64% more than elderly men (143.7 ETB vs. 87.7 ETB) but this difference is not statistically significantly different from zero. Further evidence that the "socially adequate" equivalence scale in place favors those with the lowest external wage earnings potential is that adult working aged men (those most likely to be able to find work in the marketplace) have the smallest coefficient estimate (62.8 ETB) of any age gender grouping.[14] The coefficient on adult working aged men is statistically significantly smaller than the coefficient estimate of all other age and gender categories ($p<0.05$), with the exception of young boys aged 0-6 ($p<0.10$) and elderly men.

[14] For evidence that women are disadvantaged in labor markets in Africa see Glick and Sahn (1997) who document gender wage gaps where women are paid less than men for the same job in Guinea, and Appleton, Hoddinott, and Krishnan (1999) who document this phenomena for three African countries including Ethiopia. When comparing education levels of the husband and wife within the PSNP data, adult women have much lower education levels than adult men (1.47 years vs. 3.64 years) in households that are not single headed households. See Schultz (2004) for evidence of the wage rate returns to schooling in Africa.

Comparing poverty levels using head count, poverty gap, and poverty gap squared

Now we generate policy simulations to understand the poverty reducing effects of the allocation rules chosen by the communities versus those designed by the central government. The first row of Table 4 presents a counterfactual of poverty levels without the PSNP by examining expenditure data less PSNP payments. Because of a very low propensity to invest PSNP proceeds in productive assets (Gilligan, Hoddinott, and Taffesse 2009; Hoddinott et al. 2012) it is unlikely that PSNP proceeds created a return for households outside of their consumption value, therefore simply subtracting the PSNP proceeds from household expenditures seems a feasible counterfactual of what poverty levels in communities would have looked like without the program. The headcount poverty rate is 0.5879 in this counterfactual scenario, the poverty gap is 0.2181 and the gap-squared measure is 0.1077.

Next, we keep the program budget equal to what we observed in the field and generate two scenarios. One is based on the communities' actual allocation rules and the other mimics the federal implementation mandates by allocating a full and uniform payment to each household member, but allows fewer households into the program based on the observed district budget constraint. We select these households by choosing households that had the highest probability of inclusion in the PSNP according to equation 3.11. Once the entire district budget is exhausted the remaining households receive zero payments.

The actual implementation scenario (row 2) has FGT metrics of 0.5705, 0.2025, and 0.0957 for alphas 0,1,2 respectively. The limited budget and centralized allocation approach (row 3) has FGT metrics of 0.5719, 0.2001, and 0.0934. Using the statistical inference tests developed by Kakwani (1993) neither the head count poverty nor the poverty gap measures are statistically significantly different from the no PSNP program scenario and only the gap-squared measure in the limited budget and centralized allocation approach (row 3) is weakly statistically significantly different ($p<0.10$) from the counterfactual of no PSNP program payments. None of the differences in any metrics ($\alpha=0,1,2$) are statistically significantly different between either of the limited funding scenarios.

The fully funded scenarios are presented in the fourth and fifth rows of Table 3.4. When following the community allocation rules with a full budget (row 4) the FGT metrics are 0.5627, 0.1932 (difference with no PSNP payment scenario (row 1) significant at $p<0.10$), and 0.0893 (difference with no PSNP payment scenario (row 1) significant at $p<0.05$). Following the centralized implementation plan of a full and uniform payment and a full budget (row 5) the FGT metrics are 0.5573, 0.1906 (difference with no PSNP payment scenario (row 1) significant at $p<0.05$), and 0.0876 (difference with no PSNP payment scenario (row 1) significant at $p<0.05$). While the point estimates in the fully funded central allocation scenario (row 5) show slightly lower poverty rates than each estimate in the fully funded but local allocation rules (row 4), none of the differences of any metrics ($\alpha=0,1,2$) between either of the fully funded scenarios are statistically significantly different from each other.

The broad take away is that under limited funding there is essentially no change in poverty rates, but with full funding, poverty rates are statistically significantly reduced. It does not matter which allocation rules (central or local) are used, the scale of the program drives poverty reduction. However, the finding that the local communities' allocations are pro-poor, but that these allocation decisions do not result in lower poverty levels (in the limited funding scenario) is an apparent contradiction that merits further discussion. Essentially, households receiving PSNP aid are so far below the poverty line that a low level of payments does not bring enough households above (closer to) the poverty threshold to statistically significantly alter the economy wide FGT poverty metrics. The PSNP has been noted as well-targeted (Coll-Black et al. 2012). It is a program that is successfully targeting the poor, in that the community's decision to favor the poorest groups means giving households under (but closer to) the poverty threshold relatively less aid, while giving households further below the threshold more aid. This would be consistent with seeing no change in headcount poverty metrics (α=0), no change in poverty gap metrics (α=1), but perhaps some effect on the poverty gap squared (α=2) measure. According to rows 2 and 3 of Table 4, this is what we observe.

Conclusion

Because communities do not receive sufficient funds to follow the federal implementation mandates of a full and uniform payment to each PSNP beneficiary, communities must instead exercise local discretion in allocating aid. This gives us the

chance to examine "socially adequate" consumption levels as determined by the local communities themselves. In order to recover the preferences revealed by communities in how they allocate aid we extend the technique developed by Olken (2005) to include the intensive margin of participation (how much aid paid out) in addition to the extensive margin of participation (whether or not a household is included in the aid program). The first key finding of this paper is that the preferences revealed by Ethiopian communities show that they allocate aid in a pro-poor fashion by allocating more aid to underprivileged groups with lower wage earnings potential (e.g., teenage girls vs. teenage boys, adult women vs. adult men, elderly vs. working age adults). This first finding is distinctive in regards to at least two strands of literature.

First, a majority of the literature on equivalence scales (e.g., Engel 1895; Rothbarth 1943; Deaton & Muellbauer 1986; Deaton 1997; Lewbel 1997; OECD 2008; Browning et al. 2013) is based on the underlying assumption that adult equivalence should be calculated based on consumption needs. While basing equivalence scales on consumption is certainly logical (e.g., fixed costs of running a household decline on a per capita basis as the household gets larger due to shared housing, shared electricity, etc.), doing so ignores the preferences of communities in how they define poverty for themselves. Understanding community preferences in the context of a definition of poverty may assist in the long standing challenge for researchers—raised by Pollak and Wales (1979)—of not just equalizing consumption but in actually defining a "socially adequate" consumption level.

One way to recover this "socially adequate" level is to observe how communities make the decision for themselves, such as in the Ethiopian context discussed in this paper or in the Indonesian context described by Olken (2005). When communities are given the opportunity to allocate aid, they may signal that other characteristics are more important than consumption in their definition of poverty. We find that Ethiopian communities appear to prioritize factors like limited future earnings potential more than current period consumption. Our finding is consistent with Alatas et al. (2012) despite using a different methodological approach. Alatas et al. experimentally compare a community-based targeting method with central government led proxy means testing within a large social assistance program in Indonesia. They find that communities have different preferences than the central government as to whom should receive social assistance and that communities prioritize neediness in terms of earnings capacity rather than just consumption needs.

Second, our finding of pro-poor aid allocations is distinctive in that it shows that decentralized implementation of social assistance programs can lead to pro-poor allocation rules. The Ethiopian case presented in this paper runs contrary to a large body of research that describes how decentralized program implementation are often distributionally regressive (e.g., Reinikka & Svensson 2004; Bardhan & Mookherjee 2006; Galiani et al. 2008; Ricker-Gilbert & Jayne 2012; Lunduka et al. 2013; Kilic et al. 2015).

The second key finding of this paper is that despite the communities' pro-poor implementation, the program does not significantly lower poverty rates due to constrained funding. However, in simulations with full funding the program significantly reduces the poverty gap and poverty gap-squared measures in *both* the decentralized and centralized implementation, even though they use different allocation criteria. This finding is meaningful for at least two reasons.

First, it provides practical operational advice for the Government of Ethiopia and other program implementers that fewer (not additional) resources should be deployed to enforce the full and uniform payment mandate. This insight is potentially quite useful as the Government of Ethiopia highlighted the requirement of local communities to make a full and uniform payment to each member of eligible households in its revised Program Implementation Manual published at the beginning of 2010 (Ethiopian Ministry of Agriculture and Rural Development 2010b). Any funding allocated towards ensuring compliance of this mandate may be more effectively re-deployed towards enlarging the overall program budget, rather than regulating how communities allocate the aid they do receive.

Second, the finding of no difference in poverty rates when comparing centralized versus decentralized implementation builds on an emerging trend in recent research. Alatas et al. (2012) found that central government led proxy means tests performed slightly better at identifying the poor than a local community-based approach, but that these differences in targeting were not large enough to make a difference in overall

poverty rates. In that study—as in this one—the major policy takeaway is that the overall level of funding is more important for resultant poverty reduction than the locus of control over implementation.

Appendix A: Variance decomposition of payments to participants

Because the per-capita payment is designed to be uniform, there should be no variation in the marginal transfer when increasing household size by one person (conditional on PSNP participation). Therefore, we decompose the variance in marginal payments across the administrative levels of government to determine if (and to what degree) local governments deviate from central implementation mandates.

Construction of marginal payment variable

The marginal payment for an additional household member is the difference in payment between what a household actually received and the mean payment (conditional on being in the PSNP) in that same *woreda* for a household whose size is larger by one member. The probability density function of the incremental differences in marginal payouts for additional household members is the distribution of:

$$MP_{itw} = (\overline{T_{tw}} \mid H_{tw} = m+1) - (T_{itw} \mid H_{itw} = m) \tag{3.18}$$

Where $\overline{T_{tw}}$ is the mean transfer for households with size $H_{itw} = m+1$ in year t and *woreda* w and T_{itw} is the amount of transfer and $H_{itw} = m$ is household size for

household *i* specifically.[15] We symmetrically trim the 1% of outliers from each tail of the marginal payment sample to reduce the effect of outliers.[16]

Decomposing the variance of marginal payment

We adapt the nonparametric variance decomposition approach of Barrett and Luseno (2004) to decompose the variance in marginal payments at differing levels of the government structure within the PSNP. The decomposition works as follows. Let *i* index individual households, *k* is the *kebele* (village) location, *w* is the *woreda* (district) location, *z* is the zone location, *r* is the region location, and *f* is the federal level.[17] Simply begin with the obvious statement that marginal payment of a given household equals the marginal payment of that same household.[18]

---

[15] This is a one-step-ahead estimator. We also construct a one-step-behind estimator (i.e. $MP_{ikw} = (T_{ikw}|H_{ikw} = m) - (T_{iw}|H_{iw} = m - 1)$. Results using the one-step-ahead estimator are presented in the main paper; results using the one-step-behind estimator do not materially change and are presented in Appendix Table A3.

[16] We also calculate additional marginal payment measures using the *kebele* rather than the *woreda* as the reference point. The *kebele* is a lower administrative unit and calculating the marginal payment this way would be advantageous if the *kebele* is the locus of determination in marginal payments. It has the disadvantage, however, of data loss, as there are more boundary problems and more potential gaps in the data when creating the distribution of marginal payments. Additionally, the marginal payment variable is calculated with and without simple non-parametric smoothing to reduce the impact of any outliers in a given *woreda* or *kebele*. Irrespective of the way the marginal payment variable is generated, the qualitative variance decomposition results are largely the same.

[17] The administrative levels of Ethiopian government from the least to the most central are: *kebele* (village), *woreda* (district), zone, region, federal.

[18] The decomposition is executed only on data points from the same year; therefore the year subscript is dropped in the model specification.

$$MP_{ikwzrf} = MP_{ikwzrf} \tag{3.19}$$

Then repeatedly add and subtract the same term to the right hand side of equation (3.19) and regroup with parentheses.

$$MP_{ikwzrf} = (MP_{ikwzrf} - \overline{MP_k}) + (\overline{MP_k} - \overline{MP_w}) + (\overline{MP_w} - \overline{MP_z}) + (\overline{MP_z} - \overline{MP_r}) + (\overline{MP_r} - \overline{MP_f}) + \overline{MP_f} \tag{3.20}$$

Equivalently this can be rewritten as:

$$MP_{ikwzrf} = K + W + Z + R + F + \overline{MP_f} \tag{3.21}$$

where $K \equiv (MP_{ikwzrf} - \overline{MP_k})$ is the deviation of household marginal payment from the *kebele* mean marginal payment in the same *kebele*; $W \equiv (\overline{MP_k} - \overline{MP_w})$ is the deviation of *kebele* mean marginal payment from *woreda* mean marginal payment in the same *woreda*; $Z \equiv (\overline{MP_w} - \overline{MP_z})$ is the deviation of *woreda* mean marginal payment from zonal mean marginal payment in the same zone; $R \equiv (\overline{MP_z} - \overline{MP_r})$ is the deviation of zonal mean marginal payment from regional mean marginal payment in the same region; and, lastly, $F \equiv (\overline{MP_r} - \overline{MP_f})$ is the deviation of regional mean marginal payment from federal mean marginal payment. Taking the variance of equation 3.21 gives the following decomposition:

$$Var(MP_{ikwzrf}) = Var(K) + Var(W) + Var(Z) + Var(R) + Var(F) + \\ 2[Cov(K,W) + Cov(K,Z) + Cov(K,R) + Cov(K,F) + Cov(W,Z) + \\ Covs(W,R) + Cov(W,F) + Cov(Z,R) + Cov(Z,F) + Cov(R,F)] \quad (3.22)$$

Simplification and splitting the covariance shares equally between the two components leads to the following five sources of variation in marginal payments:

$$KS \equiv Var(K) + Cov(K,W) + Cov(K,Z) + Cov(K,R) + Cov(K,F) \quad (3.23)$$

is the *kebele* (village) source variation,

$$WS \equiv Var(W) + Cov(K,W) + Cov(W,Z) + Cov(W,R) + Cov(W,F) \quad (3.24)$$

is the *woreda* (district) source variation,

$$ZS \equiv Var(Z) + Cov(K,Z) + Cov(W,Z) + Cov(Z,R) + Cov(Z,F) \quad (3.25)$$

is the zonal source variation,

$$RS \equiv Var(R) + Cov(K,R) + Cov(W,R) + Cov(Z,R) + Cov(R,F) \quad (3.26)$$

is the regional source variation, and

$$FS \equiv Var(F) + Cov(K,F) + Cov(W,F) + Cov(Z,F) + Cov(R,F) \qquad (3.27)$$

is the federal source variation. Substituting these five variables into equation (3.22) and dividing both sides by $Var(MP_{ikwzrf})$ gives a decomposition of the sources of variation of marginal payment:

$$1 = ks + ws + zs + rs + fs \qquad (3.26)$$

where the lower case variables are shares of variation from each source.

Variance decomposition of marginal payment

The federally mandated uniform benefit schedule implies zero variance across the sample because marginal annual wage payment is uniform in a given year.[19] However, the variance decomposition (Appendix Table A3) shows that the largest share of variance in marginal payment is concentrated at the *kebele* (village) (ks) level (61.6% to 79.8%), followed by the *woreda* (district) (ws) (16.7% to 35.6%). In short, there is considerable variation in marginal PSNP payments and local governments account for most of that variation. Results are robust to whether the marginal payment variable is calculated with the *woreda* or *kebele* as the reference point. It appears that the actual

[19] Recall the daily wage rage in 2006-07 was 6 ETB/day (6 ETB/day*5 days/month*6 months = 180 ETB/person/year), 2008 wages were 8 ETB/day (8 ETB/day*5 days/month*6 months = 240 ETB/person/year) and in 2009 was 10 ETB/day (10 ETB/day*5 days/month*6 months = 300 ETB/person/year). Direct support beneficiaries receive the same payment allocation with no work requirement.

payment schedule is largely determined in a decentralized manner at the most local level of government and does not follow the uniform payment schedule stipulated by the central government.

Appendix B: Simultaneous vs. Sequential Model

Because households that receive PSNP payments are not randomly selected, a selection model with censoring at zero is necessary to estimate the value of parameters associated with household demographic structure at the intensive margin of participation. Deciding on an appropriate selection model, however, raises an ancillary but important question; do rural Ethiopian communities make decisions about the extensive and intensive margins of participation in the PSNP sequentially or simultaneously? It seems likely that in a situation where communities face funding constraints that a simultaneous approach may be more likely (e.g., allow a needy household into the program, all the while knowing that they will not receive a full payment).

We test the hypothesis that local communities simultaneously determine participation in the PSNP and PSNP payment amounts. To do this, we adapt the technique proposed by Bellemare and Barrett (2006) and model payment levels at the intensive margin both sequentially (equation 3.14) and simultaneously (equation 3.16) and then use a sequential J-test (Davidson and MacKinnon 1993) to see if we can reject the simultaneous hypothesis.

Results of J-test

We obtain the predicted values for the sequential (2 step) model and include them as regressors in the simultaneous (tobit) model. We obtain the predicted values for the simultaneous (tobit) model and include them as regressors in the sequential (2 step) model. The null hypotheses are: (1) the estimated coefficient for the predicted value of the sequential model is not statistically significantly different from zero in the simultaneous model and (2) the estimated coefficient for the predicted value of the simultaneous model is not statistically significantly different from zero in the sequential model. These test, respectively, that (1) the sequential model has no explanatory power with respect to the simultaneous model, and (2) the simultaneous model has no explanatory power with respect to the sequential model. The coefficient estimate for (1) is 5.72 with a t-statistic of 9.24 ($p<0.000$), and the coefficient estimate for (2) is -0.48 with a t-statistic of -4.39 ($p<0.000$). Therefore we cannot reject the hypothesis that local communities simultaneously determine participation in the PSNP and PSNP payment amounts. As such, we focus on the results of the simultaneous model (tobit) in the body of the paper and present the results of the sequential model below.

Results of sequential model

The sequential model results are presented in Table A2. Without accounting for any control variables (recall that the PIM states that once selected into the program all participants should receive the same payment), younger children aged 0-6 (col. 1)

receive about the same as children aged 7-15 (81.7 ETB vs. 77.8 ETB) and this is about 9% larger than what working aged adults receive (74.7 ETB). However, none of these estimates are statistically significantly different from each other. Elderly receive a lower payment (54.5 ETB), which is marginally statistically different from children aged 0-6 ($p<0.10$), but not statistically significantly different from other age groups. When including controls (col. 2), young children receive the same as working aged adults (59.0 ETB vs. 60.2 ETB), but older children receive 12% more than working aged adults (67.6 ETB vs. 60.2 ETB), but this difference is not statistically significant either. When splitting age cohorts by gender (col. 4) there are no statistically significant differences between sexes within the same age cohort, nor are their statistically significant differences across any of the age and gender groups.

REFERENCES

Agrawal, Arun, and Jesse Ribot. 1999. "Accountability In Decentralization: A Framework With South Asian and West African Cases." *Journal of Developing Areas* 33 (4): 473–502. doi:10.2307/4192885.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 102 (4): 1206–1240. doi:10.1257/aer.102.4.1206.

Alderman, Harold. 2002. "Do Local Officials Know Something We Don't? Decentralization of Targeted Transfers in Albania." *Journal of Public Economics* 83 (3): 375–404. doi:10.1016/S0047-2727(00)00145-6.

Appleton, Simon, John Hoddinott, and Pramila Krishnan. 1999. "The Gender Wage Gap in Three African Countries." *Economic Development and Cultural Change* 47 (2): 289–312.

Bardhan, Pranab, and Dilip Mookherjee. 2000. "Capture and Governance at Local and National Levels." *American Economic Review, Papers and Proceedings* 90 (2): 135–139.

———. 2005. "Decentralizing Antipoverty Program Delivery in Developing Countries." *Journal of Public Economics* 89 (4): 675–704. doi:10.1016/j.jpubeco.2003.01.001.

———. 2006. "Pro-Poor Targeting and Accountability of Local Governments in West Bengal." *Journal of Development Economics* 79 (2): 303–327. doi:10.1016/j.jdeveco.2006.01.004.

Barrett, Christopher B., and Winnie K. Luseno. 2004. "Decomposing Producer Price Risk: A Policy Analysis Tool with an Application to Northern Kenyan Livestock Markets." *Food Policy* 29 (4): 393–405. doi:10.1016/j.foodpol.2004.07.008.

Barten, Antonius Petrus. 1964. "Family Composition, Prices and Expenditure Patterns." In *Econometric Analysis for National Economic Planning*, edited by Peter Edward Hart, Gordon Mills, and John King Whitaker, Volume 16. London: Butterworths.

Bellemare, Marc F, and Christopher B Barrett. 2006. "An Ordered Tobit Model of Market Participation: Evidence from Kenya and Ethiopia." *American Journal of Agricultural Economics* 88 (2): 324–337.

Berhane, Guush, Samuel Gidney, John Hoddinott, Neha Kumar, Alemayehu Seyoum Taffesse, Yisehac Yohannes, Jeremy Lind, et al. 2013. "The Implementation of the Productive Safety Nets Programme and the Household Asset Building Programme in the Ethiopian Highlands, 2012: Program Performance Report." Washington, D.C.

Berhane, Guush, John Hoddinot, Neha Kumar, Alemayehu Seyoum Taffesse, Michael Tedla Diressie, Yisehac Yohannes, Rachel Sabates-Wheeler, et al. 2011. "Evaluation of Ethiopia's Food Security Program: Documenting Progress in the Implementation of the Productive Safety Nets Programme and the Household Asset Building Programme." Washington, DC.

Browning, M., P. Chiappori, and A. Lewbel. 2013. "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power." *Review of Economic Studies* 80 (4): 1267–1303. doi:10.1093/restud/rdt019.

Caeyers, Bet, and Stefan Dercon. 2012. "Political Connections and Social Networks in Targeted Transfer Programs: Evidence from Rural Ethiopia." *Economic Development and Cultural Change* 60 (4): 639–675.

Coll-Black, Sarah, Daniel O. Gilligan, John Hoddinott, Neha Kumar, Alemayehu Seyoum Taffesse, and William Wiseman. 2012. "Targeting Food Security Interventions in Ethiopia: The Productive Safety Net Programme." In *Food and Agriculture in Ethiopia: Progress and Policy Challenges*, edited by Paul A. Dorosh and Shahidur Rashid, 280–317. Philadelphia, PA: University of Pennsylvania Press.

Conning, Jonathan, and Michael Kevane. 2002. "Community-Based Targeting Mechanisms for Social Safety Nets: A Critical Review." *World Development* 30 (3): 375–394. doi:10.1016/S0305-750X(01)00119-X.

Davidson, R., and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. Oxford, U.K.: Oxford University Press.

Deaton, Angus S. 1997. *The Analysis of Household Surveys: A Microeconometric*

Approach to Development Policy. Baltimore: Johns Hopkins University Press.

Deaton, Angus S, and John Muellbauer. 1986. "On Measuring Child Costs: With Applications to Poor Countries." *Journal of Political Economy* 94 (4): 720–744.

Development Assistance Group. 2010. "Aid Management and Utilisation in Ethiopia." Addis Ababa, Ethiopia.

Engel, Ernst. 1895. "Die Libenskosten Belgischer Arbieter — Familien Fruher and Jetzt." *International Statistical Institute Bulletin* 9: 1–74.

Ethiopian Ministry of Agriculture and Rural Development. 2010a. "PSNP Annual Plan by Region and Woredas." Addis Ababa, Ethiopia.

———. 2010b. "Productive Safety Net Programme: Programme Implementation Manual (Revised)." Addis Ababa, Ethiopia.

Faguet, Jean-Paul. 2004. "Does Decentralization Increase Government Responsiveness to Local Needs?" *Journal of Public Economics* 88 (3-4): 867–893. doi:10.1016/S0047-2727(02)00185-8.

Fisman, Raymond, and Roberta Gatti. 2002. "Decentralization and Corruption: Evidences across Countries." *Journal of Public Economics* 83 (3): 325–345. doi:10.1016/S0047-2727(00)00158-4.

Foster, James, Joel Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52 (3): 761–766.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. 2008. "School Decentralization: Helping the Good Get Better, but Leaving the Poor behind." *Journal of Public Economics* 92 (10-11): 2106–2120. doi:10.1016/j.jpubeco.2008.05.004.

Gilligan, Daniel O, John Hoddinott, and Alemayehu Seyoum Taffesse. 2009. "The Impact of Ethiopia's Productive Safety Net Programme and Its Linkages." *Journal of Development Studies* 45 (10): 1684–1706. doi:10.1080/00220380902935907.

Gilligan, Daniel O., John Hoddinott, Alemayehu Seyoum Taffesse, S. Dejene, N. Tefera, and Yisehac Yohannes. 2007. "Ethiopia Food Security Programme Report on 2006 Baseline Survey." IFPRI, Washington, DC.

Glick, Peter, and David E Sahn. 1997. "Gender and Education Impacts on Employment and Earnings in West Africa: Evidence from Guinea." *Economic Development and Cultural Change* 45 (4): 793–823.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161.

Hoddinott, John, Guush Berhane, Daniel O. Gilligan, Neha Kumar, and Alemayehu Seyoum Taffesse. 2012. "The Impact of Ethiopia's Productive Safety Net Programme and Related Transfers on Agricultural Productivity." *Journal of African Economies* 21 (5): 761–786. doi:10.1093/jae/ejs023.

Kakwani, Nanak. 1993. "Statistical Inference in the Measurement of Poverty." *Review of Economics and Statistics* 75 (4): 632–639.

Kilic, Talip, Edward Whitney, and Paul Winters. 2015. "Decentralised Beneficiary Targeting in Large-Scale Development Programmes: Insights from the Malawi Farm Input Subsidy Programme." *Journal of African Economies* 24 (1): 26–56. doi:10.1093/jae/eju021.

Leser, C. E. V. 1963. "Forms of Engel Functions." *Econometrica* 31 (4): 694–703.

Lessmann, Christian, and Gunther Markwardt. 2010. "One Size Fits All? Decentralization, Corruption, and the Monitoring of Bureaucrats." *World Development* 38 (4): 631–646. doi:10.1016/j.worlddev.2009.11.003.

Lewbel, Arthur. 1997. "Consumer Demand Systems and Household Equivalence Scales." *Handbook of Applied Econometrics* 2: 167–201.

Lewbel, Arthur, and Krishna Pendakur. 2008. "Equivalence Scales." In *New Palgrave Dictionary of Economics*, edited by Steven N Durlauf and Lawrence E Blume, 2nd ed. Palgrave Macmillan.

Lunduka, Rodney, Jacob Ricker-Gilbert, and Monica Fisher. 2013. "What Are the

Farm-Level Impacts of Malawi's Farm Input Subsidy Program? A Critical Review." *Agricultural Economics* 44 (6): 563–579. doi:10.1111/agec.12074.

Nicholson, J L. 1976. "Appraisal of Different Methods of Estimating Equivalence Scales and Their Results." *Review of Income and Wealth* 22 (1): 1–11.

Oates, Wallace E. 1972. *Fiscal Federalism*. New York, NY: Harcourt, Brace, Jovanovich.

OECD. 2008. "What Are Equivalence Scales?" Paris, France.

Olken, Benjamin A. 2005. "Revealed Community Equivalence Scales." *Journal of Public Economics* 89 (2-3): 545–566. doi:10.1016/j.jpubeco.2004.03.003.

———. 2006. "Corruption and the Costs of Redistribution: Micro Evidence from Indonesia." *Journal of Public Economics* 90 (4): 853–870. doi:10.1016/j.jpubeco.2005.05.004.

Pollak, Robert A, and Terence J Wales. 1979. "Welfare Comparisons and Equivalence Scales." *American Economic Review, Papers and Proceedings* 69 (2): 216–221.

Pritchett, Lant. 2005. "A Lecture on the Political Economy of Targeted Safety Nets." Social Protection Discussion Paper Series No. 0501. The World Bank, Washington, DC.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *Quarterly Journal of Economics* 119 (2): 679–705.

Ricker-Gilbert, Jacob, and Thomas S. Jayne. 2012. "Do Fertilizer Subsidies Boost Staple Crop Production and Reduce Poverty Across the Distribution of Smallholders in Africa? Quantile Regression Results from Malawi." In *International Association of Agricultural Economists (IAAE) Triennial Conference*. Foz do Iguacu, Brazil.

Rothbarth, Erwin. 1943. "Note on Method of Determining Equivalent Income for Families of Different Composition." In *War-Time Pattern of Saving and Expenditure*, edited by Charles Madge, Volume 4. Cambridge, U.K.: Cambridge

University Press.

Schultz, T Paul. 2004. "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education." *Journal of African Economies* 13 (suppl 2): ii95–ii148. doi:10.1093/jae/ejh044.

Smith, Brian C. 1985. *Decentralization: The Terriorial Dimension of the State*. London, UK: George Allen & Unwin.

Véron, René, Glyn Williams, Stuart Corbridge, and Manoj Srivastava. 2006. "Decentralized Corruption or Corrupt Decentralization? Community Monitoring of Poverty-Alleviation Schemes in Eastern India." *World Development* 34 (11): 1922–1941. doi:10.1016/j.worlddev.2005.11.024.

Working, Holbrook. 1943. "Statistical Laws of Family Expenditure." *Journal of the American Statistical Association* 38 (221): 43–56.

World Bank. 2006. "Project Appraisal Document for Ethiopia Productive Safety Net APL II Project." Washington D.C.

Table 3.1
Descriptive Statistics

	PSNP Status		
	Non-Participant	Participant	Difference
Total household expenditures, birr/year	12457.9	10406.8	2051.1***
	(8093.8)	(7247.2)	(125.3)
Age of household head	46.18	46.77	-0.589*
	(14.87)	(15.05)	(0.243)
Female headed household	0.168	0.257	-0.0895***
	(0.374)	(0.437)	(0.00660)
Household head highest grade attained	1.200	1.062	0.139***
	(2.299)	(2.145)	(0.0362)
Household size	5.312	5.147	0.164***
	(2.111)	(2.151)	(0.0347)
Percent children aged 0-6	0.212	0.211	0.00183
	(0.188)	(0.187)	(0.00305)
Percent children aged 7-15	0.246	0.249	-0.00283
	(0.188)	(0.193)	(0.00310)
Percent adults aged 16-60	0.479	0.462	0.0175***
	(0.210)	(0.217)	(0.00347)
Percent adults aged 61+	0.0622	0.0787	-0.0165***
	(0.159)	(0.191)	(0.00285)
Landholdings in hectares	1.430	1.158	0.272***
	(1.258)	(1.002)	(0.0186)
Livestock in tropical livestock units	4.897	3.298	1.599***
	(5.474)	(3.550)	(0.0757)
Household member has position in *kebele*	0.0953	0.129	-0.0340***
	(0.294)	(0.336)	(0.00512)
Friend or relative has position in *kebele*	0.188	0.217	-0.0282***
	(0.391)	(0.412)	(0.00653)
Value of productive equipment, birr	269.9	244.2	25.68***
	(319.3)	(309.0)	(5.118)
Drought mentioned as most important shock	0.492	0.494	-0.00242
	(0.500)	(0.500)	(0.00814)
Death of a spouse	0.0169	0.0248	-0.00792***
	(0.129)	(0.156)	(0.00232)
Crops suffered from illness of household member	0.102	0.104	-0.00271
	(0.302)	(0.306)	(0.00495)
Observations	7,867	7,250	

Note: Data is pooled from 2006-2009. Currency measures are adjusted according to consumer price index and listed in 2009 equivalent units. Differences significant at *** p<0.01, ** p<0.05, * p<0.1.

Table 3.2

Extensive Margin of PSNP participation (2006-2009), presented as odds ratios

	(1) FE logit	(2) FE logit	(3) FE logit	(4) FE logit
Log annual household expenditures	0.41***	0.51***	0.41***	0.51***
	(0.03)	(0.03)	(0.03)	(0.03)
Log household size	1.22*	2.30***	1.27**	2.36***
	(0.13)	(0.27)	(0.13)	(0.28)
Percent children aged 0-6	1.06	0.63		
	(0.27)	(0.21)		
Percent children aged 7-15	1.29	0.82		
	(0.32)	(0.25)		
Percent adults aged 16-60	0.73*	0.72		
	(0.13)	(0.16)		
Percent boys aged 0-6			0.89	0.58
			(0.26)	(0.21)
Percent girls aged 0-6			1.12	0.61
			(0.33)	(0.21)
Percent boys aged 7-15			1.00	0.69
			(0.27)	(0.24)
Percent girls aged 7-15			1.48	0.87
			(0.40)	(0.29)
Percent men aged 16-60			0.55***	0.58**
			(0.12)	(0.16)
Percent women aged 16-60			0.93	0.85
			(0.19)	(0.21)
Household head highest grade attained		1.00		1.00
		(0.00)		(0.00)
Marital Status: Single		1.99***		1.99***
		(0.51)		(0.50)
Marital Status: Divorced		2.08***		2.07***
		(0.36)		(0.36)
Marital Status: Widowed		1.82***		1.81***
		(0.20)		(0.20)
Household member has position in *kebele*		1.95***		1.95***
		(0.21)		(0.21)
Friend or relative has position in *kebele*		1.38***		1.39***
		(0.11)		(0.11)
Landholdings in hectares		0.88***		0.87***
		(0.03)		(0.03)
Livestock in tropical livestock units		0.85***		0.85***
		(0.02)		(0.02)
Kebele-year fixed effects	Yes	Yes	Yes	Yes
Observations	15,548	13,645	15,548	13,645
Chi-square test	224.5	294.1	237.8	320.3
Prob > chi^2	0.000	0.000	0.000	0.000
Pseudo R^2	0.041	0.101	0.042	0.101

Standard errors clustered at *kebele* level, presented in exponentiated form

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

Note: The data is pooled from Jan.-May of years 2006-2009. Currency variables adjusted according to consumer price index and listed in 2009 equivalent units. Marital status is categorical with married as omitted category. Age of household head, value of productive equipment, drought mentioned as most important shock, death of a spouse, and crops suffered from illness of household member were all included in the regressions, but with statistically insignificant coefficients, so they were removed from the table due to space constraints.

Table 3.3

Simultaneous Model of Intensive Margin of PSNP participation

	(1) Tobit	(2) Tobit	(3) Tobit	(4) Tobit
Annual household expenditures (100's birr)	-2.60***	-2.02***	-2.55***	-2.01***
	(0.17)	(0.16)	(0.17)	(0.16)
Number children aged 0-6	79.30***	95.20***		
	(7.85)	(8.32)		
Number children aged 7-15	77.15***	105.41***		
	(6.84)	(7.20)		
Number of adults aged 16-60	31.94***	83.13***		
	(7.35)	(7.91)		
Number adults aged 61+	33.69**	116.82***		
	(16.34)	(22.21)		
Number boys aged 0-6			75.97***	89.91***
			(10.44)	(10.77)
Number girls aged 0-6			79.49***	94.97***
			(10.15)	(10.38)
Number of boys aged 7-15			57.25***	90.14***
			(9.36)	(9.65)
Number of girls aged 7-15			99.69***	120.36***
			(9.74)	(9.83)
Number men aged 16-60			4.33	62.82***
			(9.85)	(10.39)
Number women aged 16-60			79.19***	109.45***
			(11.97)	(11.77)
Number men aged 61+			-35.60	87.74***
			(22.45)	(28.77)
Number women aged 61+			122.22***	143.67***
			(26.40)	(29.48)
Household head highest grade attained		-9.43**		-8.40**
		(3.68)		(3.69)
Marital Status: Single		200.26***		194.97***
		(60.11)		(59.75)
Marital Status: Divorced		120.00***		109.13***
		(30.86)		(30.80)
Marital Status: Widowed		116.32***		97.62***
		(23.32)		(24.48)
Household member has position in *kebele*		221.13***		219.85***
		(24.07)		(24.09)
Friend or relative has position in *kebele*		99.60***		99.71***
		(22.00)		(21.98)
Livestock in tropical livestock units		-55.72***		-55.18***
		(5.02)		(5.00)
Death of a spouse		142.24***		150.29***
		(52.42)		(52.95)
Kebele-year fixed effects	Yes	Yes	Yes	Yes
Observations	13,645	13,645	13,645	13,645
F-test	4.280	4.650	4.281	4.660
Prob > F	0.000	0.000	0.000	0.000
Pseudo R^2	0.030	0.037	0.031	0.037

Robust standard errors in parentheses

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

Note: Landholdings, value of productive equipment, drought, and illness of family member mentioned as important shocks were included in regressions, but results were statistically insignificant, therefore they are removed for space.

Table 3.4

Poverty metrics comparing various simulated scenarios of PSNP implementation

	α=0	**α=1**	**α=2**
Counterfactual: No PSNP payments	0.5879	0.2181	0.1077
	(0.0211)	(0.0122)	(0.0077)
Decentralized and Limited Budget: program as implemented	0.5705	0.2025	0.0957
	(0.0217)	(0.0116)	(0.0069)
Centralized and Limited Budget: fewer people, full and uniform payment	0.5719	0.2001	0.0934*
	(0.0217)	(0.0113)	(0.0067)
Decentralized and Full Budget: community allocation rules, full budget	0.5627	0.1932*	0.0893**
	(0.0218)	(0.0112)	(0.0065)
Centralized and Full Budget: program as envisioned	0.5573	0.1906**	0.0876**
	(0.0217)	(0.0112)	(0.0065)

Note: The table examines simulated scenarios of limited and full funding, and centralized versus decentralized program control. Poverty metrics are calculated based on Foster, Greer, Thorbecke (1984) and official Ethiopian government poverty lines. The statistical inference between poverty measures and the counterfactual of no PSNP payments is calculated with a one-sided test based on Kakwani (1993), significance levels for this test are: ** $p<0.05$, * $p<0.10$. The estimates and standard errors take into account sampling weights. When $\alpha=0$ the FGT metric is share of population under the head count poverty line. When $\alpha=1$ the FGT metric is the poverty gap, and when $\alpha=2$ the FGT metric is the poverty gap squared. When comparing poverty metrics between the two limited budget scenarios, neither of the metrics is statistically significantly different from each other. This is also the case when comparing poverty metrics between the two full budget scenarios.

Figure 3.1
Overview of simulated policy scenarios

Locus of Program Control	Level of Funding: **Limited Budget**	Level of Funding: **Full Budget**
Decentralized	Observed in Practice: Community Allocation Rules and Limited Budget	Simulated Scenario: Community Allocation Rules with a Full Budget
Centralized	Simulated Scenario: Fewer People Get Full and Uniform Payment	Simulated Scenario: Central Design, All Participants Get Full and Uniform Payment

Figure 3.2
Data used in analysis

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2005						X	X	X	X	X	X	X
2006	X	X	X	X	X	Survey						
2007	X	X	X	X	X	X	X	X	X	X	X	X
2008	X	X	X	X	X	Survey						
2009	X	X	X	X	X	X	X	X	X	X	X	X
2010	X	X	X	X	X	Survey						

Designed six months per year of PSNP implementation

The X's represent months in which PSNP payment data was collected. There was a policy change affecting full family targeting in January 2010. Therefore we use the annual panel (January - May) of monthly payments prior to that date.

Figure 3.3
PSNP Client Card

ANNEX 6: PSNP CLIENT CARD

Months	PW days	DS days	Transfer Type/Amount			Date received	Signature
Mar							
Apr							
May							
Jun							

Year 2014 _____ No of Months of Entitlement ☐

Wage rate _____

No PW_____ No DS_____ TOTAL _____

Months	PW days	DS days	Transfer Type/Amount			Date received	Signature
Jul							
Aug							
Sep							
Oct							
Nov							
Dec							
Jan							
Feb							
Mar							
Apr							
May							
Jun							

CHARTER OF RIGHTS AND RESPONSIBILITIES

RIGHTS

~ If you have been selected as a PSNP beneficiary you must be issued with a Client Card free of charge.

~ You have the right to receive your transfer on time. You should receive your transfer no later than 45 days after the month to which the payment relates.

~ You have the right to receive your full transfer. You will be informed of the transfer rates at the beginning of the year. No one should deduct any money for any reason from your transfer.

~ If you are more than four months pregnant, in your first 10 months breastfeeding your child, or weakened through age, illness or disability you should not participate in public works. If your status changes in the course of the year due to sickness or pregnancy, you have the right to shift between public works and direct support.

~ Your household should not provide more than five days of labour per household member per month. Furthermore, no one person should work for more than 20 days a month.

~ You have the right to appeal if you have been incorrectly excluded or have not been categorised correctly as direct support or public works.

~ You have the right to know the criteria for graduation and to remain in the programme if you do not meet these criteria.

RESPONSIBILITIES

~ You must provide accurate and complete information to targeting committees.

~ Households with able bodied members must provide labour for public works and be committed to complete works to an acceptable standard.

~ You must not send a child under 16 to contribute their labour to public works

~ You must present your Client Card at the transfer site to record the receipt of payment.

~ Should you lose your card you must report its loss immediately to the Keble Administration.

~ You have a responsibility to build your assets and work towards graduation

~ You must report any abuses of these rights whether affecting yourself or your neighbour to the Kebele Appeal Committee. If you are not satisfied with the response you may pursue your complaint up to the Woreda Council.

PSNP
Productive Safety Net Programme

PRODUCTIVE SAFETY NET PROGRAM
CLIENT CARD

PASS ID No: _____

Name of HH head: _____ Sex:☐ Female ☐Male

Name of Spouse: _____

Region:_____ Zone:_____

Wereda: _____ Kebele:_____

Mender: _____

HH Size: _____

HH Categorisation: PW☐ DS ☐

Client's Signature:_____ Spouse's: Signature ----------

Issuing Authority:_____ Position:_____

Signature:_____ Date Issued:_____

Serial Number: _____

HH Head PIC

Spouse PIC

Year 2010 _____ No of Months of Entitlement ☐

Wage rate _____

No PW_____ No DS_____ TOTAL _____

Months	PW days worked	DS days eligible	Transfer Type/Amount			Date received	Signature
Jul							
Aug							
Sep							
Oct							
Nov							
Dec							
Jan							
Feb							
Mar							
Apr							
May							
Jun							

Year 2011 _____ No of Months of Entitlement ☐

Wage rate _____

No PW_____ No DS_____ TOTAL _____

Months	PW days	DS days	Transfer Type/Amount			Date received	Signature
Jul							
Aug							
Sep							
Oct							
Nov							
Dec							
Jan							
Feb							
Mar							
Apr							
May							
Jun							

Year 2012 _____ No of Months of Entitlement ☐

Wage rate _____

No PW_____ No DS_____ TOTAL _____

Months	PW days	DS days	Transfer Type/Amount			Date received	Signature
Jul							
Aug							
Sep							

Months	PW days	DS days	Transfer Type/Amount			Date received	Signature
Oct							
Nov							
Dec							
Jan							
Feb							
Mar							
Apr							
May							
Jun							

Year 2013 _____ No of Months of Entitlement ☐

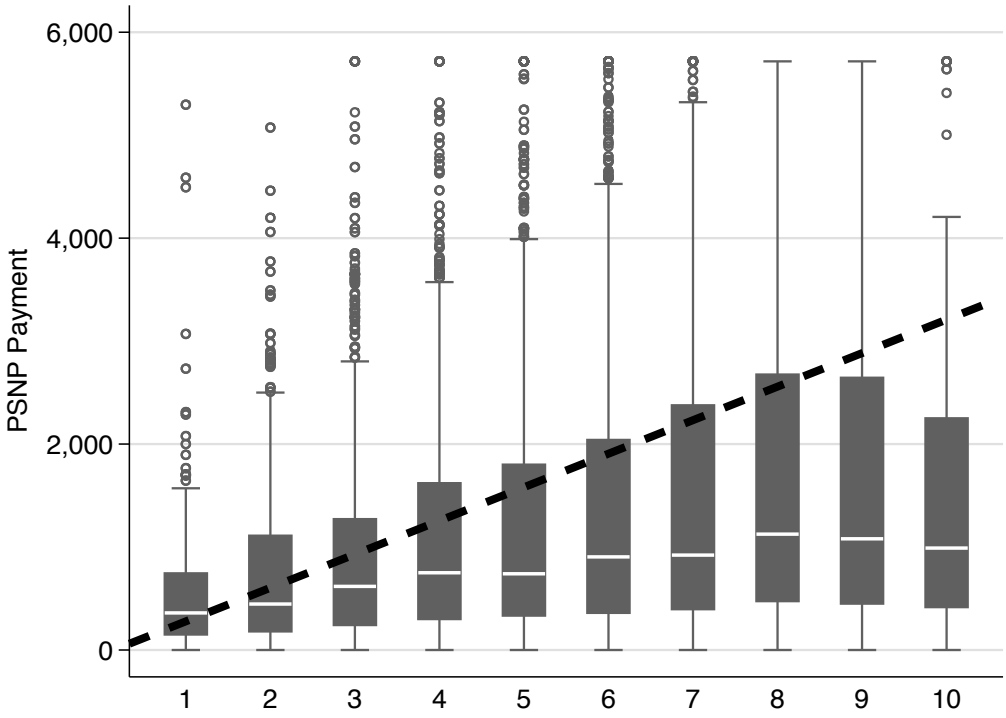
Wage rate _____

No PW_____ No DS_____ TOTAL _____

Months	PW days	DS days	Transfer Type/Amount			Date received	Signature
Jul							
Aug							
Sep							
Oct							
Nov							
Dec							
Jan							
Feb							

Figure 3.4

Box plot showing distribution of PSNP payment by household size compared to household entitlement



Data includes all payments from 2006-2009. The dashed line is the entitlement of 300 ETB per person. Payments are scaled to be in 2009 payment equivalents (for example, a payment of 90 ETB when the payment schedule calls for 180 ETB is scaled to 150 ETB out of 300 ETB, which is the full payment schedule for 2009). Payments are top and bottom coded at 2%. The while vertical line in the box plot represents the median value, the ends of the solid rectangles represent the 75th and 25th percentiles of the distribution for that household size. The lines extending from the solid rectangles are 1.5 times the interquartile range, and the dots outside of those lines are extreme values.

Figure 3.5

Histogram of share of distict budget received versus needed according to planning documents. Approximately 89% of districts did not receive sufficient funds to implement as per the planning documents.

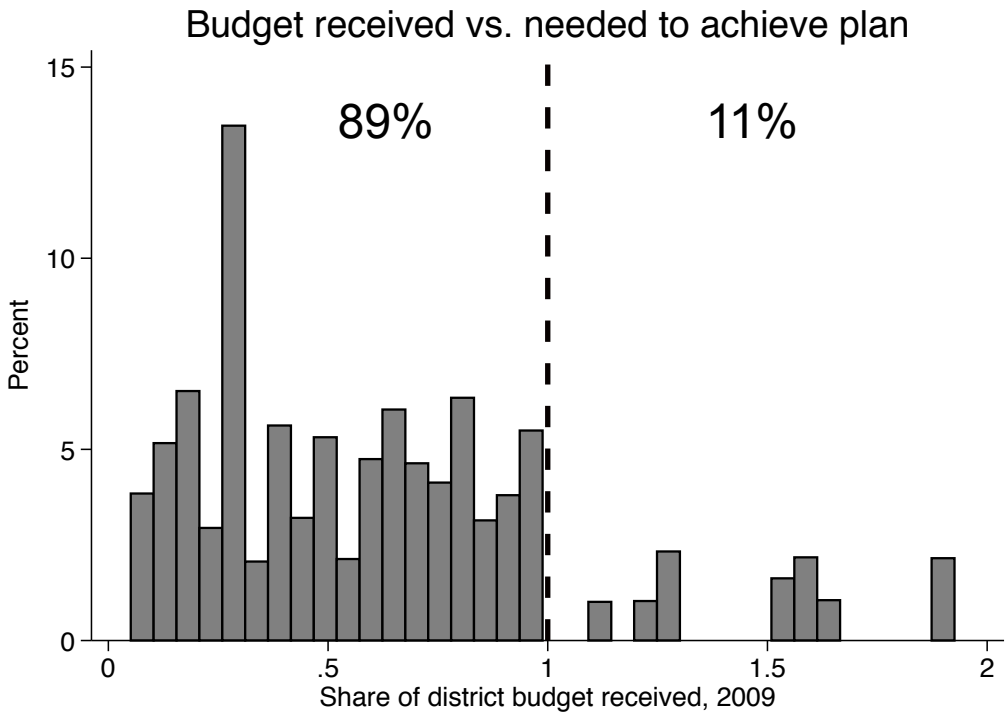


Table A3.1

Extensive Margin of PSNP participation (2006-2009), logistic regression results

	(1) FE logit	(2) FE logit	(3) FE logit	(4) FE logit
Log annual household expenditures	-0.90***	-0.67***	-0.90***	-0.67***
	(0.06)	(0.06)	(0.06)	(0.06)
Log household size	0.20*	0.83***	0.24**	0.86***
	(0.10)	(0.12)	(0.10)	(0.12)
Percent children aged 0-6	0.06	-0.45		
	(0.26)	(0.32)		
Percent children aged 7-15	0.26	-0.20		
	(0.25)	(0.31)		
Percent adults aged 16-60	-0.32*	-0.33		
	(0.18)	(0.23)		
Percent boys aged 0-6			-0.12	-0.54
			(0.29)	(0.36)
Percent girls aged 0-6			0.12	-0.50
			(0.29)	(0.34)
Percent boys aged 7-15			-0.00	-0.37
			(0.27)	(0.34)
Percent girls aged 7-15			0.39	-0.14
			(0.27)	(0.33)
Percent men aged 16-60			-0.61***	-0.55**
			(0.23)	(0.28)
Percent women aged 16-60			-0.07	-0.17
			(0.21)	(0.25)
Household head highest grade attained		-0.00		-0.00
		(0.02)		(0.02)
Marital Status: Single		0.69***		0.69***
		(0.25)		(0.25)
Marital Status: Divorced		0.73***		0.73***
		(0.17)		(0.17)
Marital Status: Widowed		0.60***		0.59***
		(0.11)		(0.11)
Household member has position in *kebele*		0.67***		0.67***
		(0.11)		(0.11)
Friend or relative has position in *kebele*		0.32***		0.33***
		(0.08)		(0.08)
Landholdings in hectares		-0.13***		-0.13***
		(0.04)		(0.04)
Livestock in tropical livestock units		-0.17***		-0.17***
		(0.03)		(0.03)
Kebele-year fixed effects	Yes	Yes	Yes	Yes
Observations	15,548	13,645	15,548	13,645
Chi-square test	224.5	294.1	237.8	320.3
Prob > chi^2	0.000	0.000	0.000	0.000
Pseudo R^2	0.041	0.101	0.042	0.101

Standard errors clustered at *kebele* level

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

Note: The data is pooled from Jan.-May of years 2006-2009. Currency variables adjusted according to consumer price index and listed in 2009 equivalent units. Marital status is categorical with married as omitted category. Age of household head, value of productive equipment, drought mentioned as most important shock, death of a spouse, and crops suffered from illness of household member were all included in the regressions, but with statistically insignificant coefficients, so they were removed from the table due to space constraints.

Table A3.2

Sequential Model of Intensive Margin of PSNP participation

	(1) 2nd Stage	(2) 2nd Stage	(3) 2nd Stage	(4) 2nd Stage
Number children aged 0-6	81.74***	58.99***		
	(10.83)	(9.34)		
Number children aged 7-15	77.77***	67.62***		
	(7.84)	(8.16)		
Number of adults aged 16-60	74.65***	60.17***		
	(7.59)	(8.13)		
Number adults aged 61+	54.48***	55.65***		
	(15.48)	(19.85)		
Number boys aged 0-6			80.51***	56.99***
			(12.64)	(10.81)
Number girls aged 0-6			80.90***	58.81***
			(11.98)	(11.18)
Number of boys aged 7-15			79.36***	70.18***
			(10.12)	(9.89)
Number of girls aged 7-15			74.78***	63.48***
			(9.96)	(10.75)
Number men aged 16-60			84.00***	69.68***
			(10.23)	(10.55)
Number women aged 16-60			59.64***	46.59***
			(10.41)	(10.94)
Number men aged 61+			73.06***	63.06***
			(20.18)	(24.01)
Number women aged 61+			30.43	47.72
			(28.82)	(29.54)
Annual household expenditures (100's birr)		-0.01		0.01
		(0.17)		(0.16)
Marital Status: Single		-35.33		-38.46
		(47.40)		(47.45)
Marital Status: Divorced		-180.88***		-180.41***
		(44.12)		(44.41)
Marital Status: Widowed		-137.05***		-134.02***
		(27.38)		(29.42)
Landholdings in hectares		37.03*		37.64*
		(20.64)		(20.64)
Livestock in tropical livestock units		10.07***		10.39***
		(3.55)		(3.43)
Inverse Mills Ratio	-87.51***	-295.83***	-96.98***	-303.91***
	(29.99)	(69.87)	(32.76)	(67.47)
Kebele-year fixed effects	Yes	Yes	Yes	Yes
Observations	6,728	6,728	6,728	6,728
R-squared	0.65	0.66	0.65	0.66

Standard errors clustered at *kebele* level

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

Note: Dependent variable is PSNP payments made between Jan.-May in years 2006-2009. PSNP payments and currency-based variables are adjusted according to consumer price index and are listed in 2009 equivalent currency units. Marital status is a categorical variable with married as the omitted category. Age of household head, education of household head, value of productive equipment, drought mentioned as most important shock, death of a spouse, and crops suffered from illness of household member were all included in the regressions, but with statistically insignificant coefficients, so they were removed from the table due to space constraints.The local political participation variables are used in the first stage, but are excluded in the second stage.

Table A3.3
Decomposition of Source Variation in Marginal Payments: Productive Safety Net Program (PSNP)

	***kebele* source (ks)**	***woreda* source (ws)**	**zonal source (zs)**	**regional source (rs)**	**federal source (fs)**	**sample size (N)**
Kebele as reference, step ahead	0.9160	0.0301	0.0346	0.0173	0.0020	1327
Kebele as reference, step behind	0.9281	0.0286	0.0207	0.0202	0.0023	1327
Kebele as reference, step ahead, smoothed	0.7786	0.0797	0.0873	0.0494	0.0050	1262
Kebele as reference, step behind, smoothed	0.7829	0.0898	0.0551	0.0690	0.0033	1259
Woreda as reference, step ahead	0.7976	0.1667	0.0208	0.0125	0.0023	1560
Woreda as reference, step behind	0.7150	0.2455	0.0210	0.0174	0.0011	1540
Woreda as reference, step ahead, smoothed	0.7949	0.0297	0.0903	0.0721	0.0129	1552
Woreda as reference, step behind, smoothed	0.7315	0.0546	0.1119	0.0981	0.0038	1532
Average Source Variation (2006)	**0.8056**	**0.0906**	**0.0552**	**0.0445**	**0.0041**	**1420**

	***kebele* source (ks)**	***woreda* source (ws)**	**zonal source (zs)**	**regional source (rs)**	**federal source (fs)**	**sample size (N)**
Kebele as reference, step ahead	0.8739	0.0500	0.0518	0.0138	0.0104	1690
Kebele as reference, step behind	0.8719	0.0535	0.0509	0.0131	0.0106	1679
Kebele as reference, step ahead, smoothed	0.7480	0.0901	0.1179	0.0271	0.0168	1570
Kebele as reference, step behind, smoothed	0.6863	0.1167	0.1438	0.0326	0.0207	1569
Woreda as reference, step ahead	0.7148	0.2304	0.0238	0.0084	0.0226	2065
Woreda as reference, step behind	0.6935	0.2613	0.0254	0.0111	0.0086	1994
Woreda as reference, step ahead, smoothed	0.6859	0.0712	0.1028	0.0329	0.1071	2059
Woreda as reference, step behind, smoothed	0.6865	0.0396	0.1546	0.0621	0.0572	1987
Average Source Variation (2007)	**0.7451**	**0.1141**	**0.0839**	**0.0251**	**0.0318**	**1827**

Table A3.3, continuation

Decomposition of Source Variation in Marginal Payments: Productive Safety Net Program (PSNP)

	kebele **source (ks)**	***woreda*** **source (ws)**	**zonal source (zs)**	**regional source (rs)**	**federal source (fs)**	**sample size (N)**
Kebele as reference, step ahead	0.8869	0.0670	0.0197	0.0150	0.0114	1702
Kebele as reference, step behind	0.8742	0.0699	0.0325	0.0106	0.0129	1694
Kebele as reference, step ahead, smoothed	0.7952	0.1178	0.0388	0.0307	0.0174	1588
Kebele as reference, step behind, smoothed	0.8021	0.0982	0.0675	0.0157	0.0165	1584
Woreda as reference, step ahead	0.7336	0.2389	0.0139	0.0048	0.0087	2091
Woreda as reference, step behind	0.6690	0.3031	0.0127	0.0050	0.0102	2001
Woreda as reference, step ahead, smoothed	0.7855	0.0451	0.0919	0.0273	0.0502	2083
Woreda as reference, step behind, smoothed	0.6902	0.0615	0.1238	0.0404	0.0841	1996
Average Source Variation (2008)	**0.7796**	**0.1252**	**0.0501**	**0.0187**	**0.0264**	**1842**

	kebele **source (ks)**	***woreda*** **source (ws)**	**zonal source (zs)**	**regional source (rs)**	**federal source (fs)**	**sample size (N)**
Kebele as reference, step ahead	0.8519	0.0762	0.0361	0.0289	0.0069	1553
Kebele as reference, step behind	0.8681	0.0631	0.0389	0.0224	0.0075	1511
Kebele as reference, step ahead, smoothed	0.7272	0.1155	0.0781	0.0667	0.0124	1414
Kebele as reference, step behind, smoothed	0.7605	0.1011	0.0770	0.0500	0.0115	1385
Woreda as reference, step ahead	0.6162	0.3562	0.0190	0.0052	0.0033	1938
Woreda as reference, step behind	0.6007	0.3688	0.0103	0.0136	0.0065	1840
Woreda as reference, step ahead, smoothed	0.7769	0.0575	0.1188	0.0287	0.0180	1922
Woreda as reference, step behind, smoothed	0.6782	0.0606	0.0969	0.1130	0.0513	1824
Average Source Variation (2009)	**0.7350**	**0.1499**	**0.0594**	**0.0411**	**0.0147**	**1673**

Note: For a given household the marginal PSNP payment is calculated by finding the difference between that household's payment and the mean payment of households in the same location that differed in size by one member. A one-step-ahead (one-step-behind) estimator compares the actual payment of a participant household of size 4 with the mean payment received of participant households of size 5 (size 3) in the same geographic location (either *kebele* (rows 1,2,3,4) or *woreda* (rows 5,6,7,8)). Simple non-parametric local smoothing is used to reduce the effect of outliers (rows 3,4,7,8) while no smoothing is used in rows 1,2,5,6. The sample includes all payments to households for the five-month period (Jan.-May) each year and removes outliers (the top 1% and bottom 1% of marginal payments).

CHAPTER 4

USING UNOBTRUSIVE SENSORS TO MEASURE AND MINIMIZE HAWTHORNE EFFECTS: EVIDENCE FROM COOKSTOVES

Introduction

The validity of empirical research depends on data quality. Unlike the physical sciences, for which data often is generated in controlled laboratory settings, the social sciences typically measure variables involving human behaviors, which make data quality a challenge. Respondents often do not answer surveys candidly (Bertrand and Mullainathan 2001) and the act of surveying can change later behaviors of those being surveyed (Zwane et al. 2011). These drawbacks to surveys have been one factor contributing to a push for more experiments in environmental economics (Greenstone and Gayer 2009) and social science research more generally (Falk and Heckman 2009; Banerjee and Duflo 2009; Duflo, Glennerster, and Kremer 2008). While much has been learned from experiments in environmental economics, these types of experiments measuring human behaviors are susceptible to issues such as observation bias, or Hawthorne effect.

We explore an emerging class of technology—small, inexpensive, and unobtrusive sensors—as a remedy to the Hawthorne effect. A growing variety of sensors have become available to researchers. GPS trackers and motion detectors, for example, allow non-obtrusive measurement of subject location and body movements (Ermes et al. 2008). Medical doctors wear sensors that detect the scent of alcohol used in hand

sanitizers to alert the doctor and/or patient if the doctor has not washed his or her hands recently (E. Smith 2014; Srigley et al. 2014). Loop detectors installed in the lanes of freeways allow monitoring of congestion and driver behavior (Bento et al. 2014).

The degree to which these sensors interfere with subjects' behavior can vary widely. In some cases, individuals may choose to be observed to motivate their own behavioral response. For example, long-distance bikers and runners can opt into programs that will report the location, time, and speed of excursions to a website that others can monitor (Mueller et al. 2010). Users of such schemes typically hope peer observation will increase their motivation. In other cases, such as room occupancy detectors that control lighting and climate control, the sensor may be far harder to notice (Buchanan, Russo, and Anderson 2014).

A major challenge for direct observational studies is that they alter participants' behavior. The effects of observers have been noted in cookstove studies (Ezzati, Saleh, and Kammen 2000; Smith-Sivertsen et al. 2009), in energy consumption (Schwartz et al. 2013), in public health (Clasen et al. 2012; Das, Hammer, and Leonard 2008; Leonard and Masatu 2006; Srigley et al. 2014) in development economics (Leonard 2008; Leonard and Masatu 2010; Muralidharan and Sundararaman 2010) and in social sciences more broadly (Levitt and List 2007; Levitt and List 2011). We demonstrate a technique to remedy the Hawthorne effect that uses unobtrusive temperature sensors in an evaluation of fuel-efficient cookstoves in

Uganda. We use minimally invasive temperature sensors to measure usage of the fuel-efficient cookstoves and of the traditional three-stone fires.[1] We then compare usage of each stove in periods when observers visit the households with periods when no observers are present. We find a large Hawthorne effect: households increase the use of the fuel-efficient stove and decrease the use of three-stone fires on days they expect observers.

The observers visited homes to measure wood use and household exposure to particulate matter. Unfortunately, changes in cooking practices due to the observers will bias measures of wood use and of exposure to particulates. Fortunately, once the magnitude of this Hawthorne effect is known, we can estimate unbiased impacts of how fuel-efficient stoves affect wood use and exposure to particulate matter.

Data and Methodology

We implemented a series of studies in rural areas of the Mbarara District in southwestern Uganda from February to September 2012, which focused on the adoption, and use of fuel-efficient stoves. At baseline almost all families cooked on a traditional three-stone fire (97%), usually located within a separate enclosed cooking hut. We introduced an Envirofit G-3300 stove. Its manufacturer reports that it uses 50% less fuel and reduces smoke and harmful gasses by 51% compared to a three stone fire (Envirofit Inc. 2011). The study area is characterized by agrarian livelihoods

[1] A three-stone fire is simply three large stones, approximately the same height, on which a cooking pot is balanced over a fire.

including raising livestock and farming *matooke* (starchy cooking banana), potatoes, and millet.

We tracked stove usage before and after the purchase of a fuel-efficient stove at 168 households spread across fourteen rural parishes in Mbarara.[2] Upon arriving in a new parish, staff displayed the fuel-efficient stove (Envirofit G-3300) and offered it for sale to anyone who wanted to purchase at 40,000 Ugandan Shillings (approximately USD $16, see Beltramo et al. (2015) and Levine et al. (2016) for an overview of the sales contract). Households were eligible to participate in the impact study if they mainly used wood as a fuel source, regularly cooked for eight or fewer persons, someone was generally home every day, and cooking was largely in an enclosed kitchen.

Eligible households who wanted to buy the stove were randomly assigned to two groups: early buyers, late buyers. Because it is crucial to measure both the use of the new stove *and* any reduction in use of traditional stoves (Ruiz-Mercado et al. 2011), we asked both early buyers and late buyers if they would agree to have stove use monitors (SUMs) that read stove temperatures placed on their traditional and Envirofit stoves. After giving consent, three stone fires were fitted with SUMs immediately and we collected a baseline round of data with only three stone fires present in homes.

Approximately two to three weeks later the early buyers group received their first Envirofit stove. We did a midline round of data collection that is not used in this

[2] The population of most Ugandan rural parishes ranges from 4,000 to 6,000.

study (but will be the basis of an impact evaluation, when the data are cleaned). Approximately five to six weeks later the late buyers received their first Envirofit stove. About six weeks after late buyers received their Envirofits, both groups were surprised with a second Envirofit stove. Because common cooking practices in the area require two simultaneous cooking pots (for example rice and beans, or *matooke* and some type of sauce), and the Envirofit is sized for one cooking pot, we gave a second Envirofit to permit normal cooking using only fuel-efficient stoves. We then collected our endline data, the core data we use in this study.

We tracked stove temperatures for approximately six months (April–September 2012). To track usage, we used small, inexpensive and unobtrusive sensors: stove use monitors (SUMs) that record stove temperatures without the need for an observer to be present.[3] Using SUMs to log stove temperatures was initially suggested by Ruiz-Mercado et al. (2008) and has been used successfully in various settings (Mukhopadhyay et al. 2012; Ruiz-Mercado et al. 2013; Pillarisetti et al. 2014). We installed SUMs on two Envirofits and two three-stone fires (by the end of the study numerous SUMs had been lost or burned up; therefore, at the end line we measured both Envirofits and the primary three stone fire).

[3] The SUMs used for our project, iButtons™ manufactured by Maxim Integrated Products, Inc., are small stainless steel temperature sensors about the size of a small coin and the thickness of a watch battery which can be affixed to any stove type. Our SUMs record temperatures with an accuracy of +/- 1.3 degrees C up to 85°C. For additional details see the product description website at: http://berkeleyair.com/services/stove-use-monitoring-system-sums/ The SUMs cost approximately USD$16 each and could record temperature data for 24 hours a day for six weeks in a household before needing minimal servicing from a technician to download the data. After the data download they can be reset and re-used.

We also performed standard kitchen performance tests (KPT) (Bailis, Smith, and Edwards 2007) in each household to measure the quantity of fuel wood used, record detailed food diaries, and measure household air pollution. The KPT lasts approximately a week and involves daily visits by a small team of researchers weighing wood, monitoring household air particulate monitors, and collecting survey data on stove usage over the last 24 hours and related topics.

Comparing stove usage calculated from the temperature data collected by the SUMs in the week while KPT measurement teams are present versus stove usage in the week before and after the measurement week provides our test of a Hawthorne effect.

Throughout the study, field staff recorded about 2,400 visual observations of whether a stove was in use (on/off) when they visited homes to exchange stove usage monitors or gather data for the KPT. Then we used a machine-learning algorithm to examine the temperature data immediately before and after the 2,400 visual observations of use. The algorithm analyzed the data to understand how temperature patterns change at times of observed stove use and then predicted cooking behaviors to the wider dataset of 1.7 million temperature readings. This process, detailed in Simons et al. (2014), allowed us to unobtrusively and inexpensively track daily stove usage on a large sample of households for six continuous months.[4]

[4] Overnight, while most participants report sleeping, SUMs record the residual heat absorbed in the large stones of the three stone fires and/or from coals banked overnight. Therefore our algorithm overestimates overnight cooking of three stone fires. We adjust for this in the subsequent analysis. For further discussion and a description of the technical adjustment see Harrell et al. (2016).

Specification

Assign the subscripts *t=-1* to the week prior to measurement week, *t=0* to the measurement week, and *t=1* to the week after the measurement week. Let the coefficient on stove type *s* = *TSF* for three-stone fire or *ENV* for Envirofit, and *Adjacent_Week* be a dummy variable for an adjacent week (*t=-1* or *t*=1). The regression is modeled using Ordinary Least Squares (OLS) as:

$$H_{it}^{S} = B^{S} * Adjacent\_Week + I_i + e_{it} \tag{4.1}$$

where H_{it}^{S} is the total hours cooked per day on stove type *s* for household *i* during the week, I_i is fixed effects for the individual household (which controls for household level characteristics that don't change over these three weeks like family size, income, housing, etc.), and e_{it} is an error term. The coefficient B^{S} is the estimate of how different (in hours cooked per day) the average adjacent week is compared to a measurement week on stove type *s*. Standard errors are clustered at the household.

To test the weeks separately, we use a slightly different specification. Let H_{it-1}^{S} be a dummy variable equal to 1 for the week before the measurement week (when *t=-1*) and 0 otherwise, and let H_{it+1}^{S} be a dummy variable equal to 1 for the week after the measurement week (when *t=1*) and 0 otherwise. Then the regression is modeled using OLS as:

$$H_{it}^{S} = \gamma_1^{S} * H_{it-1}^{S} + \gamma_2^{S} * H_{it+1}^{S} + I_i + e_{it} \tag{4.2}$$

where I_i is household fixed effects and γ_1^S is the estimate of the difference (in hours cooked per day) of the week before the measurement week compared to the measurement week. The coefficient of γ_s^S is the estimate of the difference cooked in the week after the measurement week compared to the measurement week. Standard errors are clustered at the household.

Results

In the week before the observers arrived (when *t=-1*), primary three-stone fires were used an average of 5.99 hours per day (95% CI = [4.77 to 7.21]) and combined usage on Envirofits was an average of 5.53 hours per day (95% CI = [4.36 to 6.71]). We first estimate equation 4.1, where we constrain the effect of the observers arriving to be the same magnitude (but opposite sign) as the effect of the observers leaving. On average, usage of the Envirofit stoves is 2.97 hours higher during the measurement week than during the adjacent weeks (95% CI = [1.79 to 4.15], p<0.01, Table 4.1, column 3). This increase is matched by a reduction of 1.78 hours in usage of the three-stone fire (95% CI = [0.86 to 2.70], p<0.01, col. 1).

In columns 2 and 4 we relax the assumption that stove usage is the same in the week prior to and the week after our measurement period. Contrasted with the measurement week, households use their primary three-stone fire 1.17 hours per day more in the prior week (95% CI = [0.10 to 2.24], p<0.05, col. 2) and 2.37 hours more in the following week (95% CI = [1.12 to 3.62], p<0.01). These coefficients are jointly

significantly different than zero ($p<0.01$), but not statistically significantly different from each other ($p=0.10$).

The total usage of Envirofits follows a mirror image (col. 4), and is 2.58 hours per day lower in the week prior to measurement week than in measurement week (95% CI = [1.21 to 3.94], $p<0.01$) and 3.30 hours per day lower the following week (95% CI = [2.04 to 4.57], $p<0.01$). These coefficients are jointly significantly different from zero ($p<0.01$), but not statistically significantly different from each other ($p=0.20$).

Adjusting for the Hawthorne Effect

Because the kitchen performance test is widely used to measure the effects of new cookstoves on fuel usage and household air pollution (K. Smith et al. 2007; Berrueta, Edwards, and Masera 2008; Johnson et al. 2010)—as well as the basis for the measurement of carbon emissions—estimates of how new stoves affect fuel use and carbon emissions may be substantially biased. The same bias can arise in studies, such as ours, that also measure household air pollution or health effects with repeated household visits. We develop a basic framework for testing for the magnitude of this bias and examine its extent in our setting.

Basic framework

The field of epidemiology has "efficacy trials" that test the effects of an intervention under ideal conditions and "effectiveness trials" that test the effects of an intervention under realistic conditions (Flay 1986). In the context of cookstoves, the kitchen

performance test provides a valid measure of how the new stove affects wood usage during the measurement week (as in an efficacy trial); however, we need to adjust for the gap in usage between measurement weeks and weeks when no observers are influencing behaviors to generalize to weeks without daily visits (that is, to estimate effectiveness). Next we consider various illustrative examples using data from our setting.

Illustrative examples

Table 4.2 presents the daily mean values of firewood consumption, particulate matter concentration and total three stone fire usage prior to the introduction of fuel-efficient stoves. The average household consumes 9.0 kgs of firewood per day (col. 1), has a daily concentration of PM2.5 of 428 $\mu g/m^3$ (col. 2) and cooks for a total of 14.0 daily hours (col. 3) across two three stone fires. To examine the bias introduced by the Hawthorne effect in our setting we need to know the expected biomass and pollution reductions for the new stove. To find the expected reduction we examine the "Emission and Performance Report" for the Envirofit G3300 performed by the Engines and Energy Conversion Lab at Colorado State University. These emissions measurements are based on accepted biomass stove testing protocols in a carefully monitored laboratory setting. The report (Figure 4.1) finds average improvements of 50.1% for fuel use and 51.2% for particulate matter emissions using the Envirofit G3300 versus a three stone fire (Envirofit Inc. 2011).

Using these mean values, we construct illustrative efficacy and effectiveness trails according to the framework above. For the purpose of our illustrative example, we assume a similar sized Hawthorne effect on the usage of the secondary three stone fires as well as what was observed on the first three stone fire (recall that attrition of sensors led us to measure fewer stoves in the endline).

Using the assumptions above, firewood consumed and daily PM2.5 concentrations were 16% lower when observers were present (as in an efficacy trial) than when they were not (as in an effectiveness trial, Table 4.4.3).

Bias introduced by Hawthorne effect

Table 4.4 presents a comparison of the endline to the baseline levels of daily cooking hours (on all stoves combined), daily firewood usage, and PM2.5 daily concentrations. Recall that at baseline no homes had any Envirofits, and at endline homes had two Envirofits. These results are not causal estimates, as seasonal or time effects may influence them.

When we use time periods when observers were present, between baseline and endline: cooking time rose 20% (from 14.0 to 16.8 hours),[5] firewood use declined 11% (9.0 to 8.0 kg/day), and particulate matter also fell 11% (from 428 to 382 $\mu g/m^3$). When examining weeks when observers were not present (as in an effectiveness trial),

[5] While total time cooking increases this is calculated over four stoves (two three stone fires and two Envirofit stoves) during the end line data collection period, but calculated over only two three stone fires during the baseline period. So it is likely that cooks actually spend less of their time cooking at end line because they have more stoves per meal at their disposal.

some important results are reversed. Now cooking time rose 24% (from 14.0 to 17.3 hours), firewood use rose 4% (from 9.0 to 9.3 kg/day) and exposure to particulate matter grew 4% (from 428 to 445 $\mu g/m^3$). That is, adjusting for the Hawthorne effect turned a decline of about 11% in wood use and participate matter into a small increase of about 4%.

This illustrative example shows how important it is to account for Hawthorne effects in impact evaluations. Using the sample means from our data, and the emissions and performance report for the Envirofit G3300 the Hawthorne effect not only biases the magnitude of the change, but (with these assumptions) also reverses the direction of the change over time.

Conclusion

We demonstrate a technique to measure the magnitude of—and adjust for—a Hawthorne effect in a field experiment in the developing world. Given the push for more experiments in environmental economics (Greenstone and Gayer 2009), developing techniques to generate data that does not suffer from observer bias is necessary if the evaluations are to help make unbiased policy recommendations. In our specific setting, the findings of a large Hawthorne effect have implications for the impact of fuel-efficient stoves on fuel use and air particulates. The kitchen performance test is the current "gold standard" for generating Certified Emission Reductions that can be sold into the emissions trading markets of the Clean Development Mechanism. Our findings potentially call into question the veracity of

these CO_2 reductions. More broadly, our results reinforce the importance for observed behaviors to be independently verified with unobtrusive monitoring.

While other forms of unobtrusive objective monitoring exist—such as using administrative records when reliable (Angrist, Bettinger, and Kremer 2006) or tracking take-up at a remote location via redeemed vouchers (Dupas 2009; Dupas 2014)—the recent explosion of small, inexpensive, and unobtrusive sensors expands researchers' ability to quantify and remove observation bias. A wide variety of emerging technologies can be utilized, a partial list includes: smart phones tracking locations through GPS, remote sensors that detect latrine usage (Clasen et al. 2012), sensors to remotely detect the use of water filters (Thomas et al. 2013), medical devices to monitor the hand hygiene of medical professionals (Boyce 2011), smart grid or other energy monitors (Darby 2010), and pedometers or other devices that monitor physical activity (Bravata et al. 2007). Adjusting for Hawthorne effects is essential if the results of impact evaluations are intended to generalize beyond periods of intense in-person observation.

REFERENCES

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96 (3): 847–62.

Bailis, Rob, Kirk R. Smith, and Rufus Edwards. 2007. "Kitchen Performance Test (KPT)." Univerisity of California, Berkeley, CA.

Banerjee, Abhijit, and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1 (1): 151–78. doi:10.1146/annurev.economics.050708.143235.

Beltramo, Theresa, David I. Levine, Garrick Blalock, and Andrew M. Simons. 2015. "The Effect of Marketing Messages and Payment over Time on Willingness to Pay for Fuel-Efficient Cookstoves." *Journal of Economic Behavior & Organization*. doi:10.1016/j.jebo.2015.04.025.

Bento, Antonio, Daniel Kaffine, Kevin Roth, and Matthew Zaragoza-Watkins. 2014. "The Effects of Regulation in the Presence of Multiple Unpriced Externalities: Evidence from the Transportation Sector." *American Economic Journal: Economic Policy* 6 (3): 1–29. doi:10.1257/pol.6.3.1.

Berrueta, Víctor M., Rufus D. Edwards, and Omar R. Masera. 2008. "Energy Performance of Wood-Burning Cookstoves in Michoacan, Mexico." *Renewable Energy* 33 (5): 859–70. doi:10.1016/j.renene.2007.04.016.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review* 91 (2): 67–72.

Boyce, John M. 2011. "Measuring Healthcare Worker Hand Hygiene Activity: Current Practices and Emerging Technologies." *Infection Control and Hospital Epidemiology* 32 (10): 1016–28. doi:10.1086/662015.

Bravata, Dena M, Crystal Smith-Spangler, Vandana Sundaram, Allison L Gienger, Nancy Lin, Robyn Lewis, Christopher D Stave, Ingram Olkin, and John R Sirad. 2007. "Using Pedometers to Increase Physical Activity: A Systematic Review." *Jama* 298 (19).

Buchanan, Kathryn, Riccardo Russo, and Ben Anderson. 2014. "Feeding Back about Eco-Feedback: How Do Consumers Use and Respond to Energy Monitors?" *Energy Policy* 73 (June): 138–46. doi:10.1016/j.enpol.2014.05.008.

Clasen, Thomas, Douglas Fabini, Sophie Boisson, Jay Taneja, Joshua Song, Elisabeth Aichinger, Anthony Bui, et al. 2012. "Making Sanitation Count: Developing and Testing a Device for Assessing Latrine Use in Low-Income Settings." *Environmental Science & Technology* 46 (6): 3295–3303. doi:10.1021/es2036702.

Darby, Sarah. 2010. "Smart Metering: What Potential for Householder Engagement?" *Building Research & Information* 38 (5): 442–57. doi:10.1080/09613218.2010.492660.

Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. 2008. "The Quality of Medical Advice in Low-Income Countries." *Journal of Economic Perspectives* 22 (2): 93–114.

Duflo, Esther, R Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." Edited by T Paul Schultz and John A Strauss. *Handbook of Development Economics* 4 (07): 3895–3962. doi:10.1016/S1573-4471(07)04061-2.

Dupas, Pascaline. 2009. "What Matters (and What Does Not) in Households' Decision to Invest in Malaria Prevention?" *American Economic Review* 99 (2): 224–30. doi:10.1257/aer.99.2.224.

———. 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica* 82 (1): 197–228.

Envirofit Inc. 2011. *G-3300 Woodstove Features*. http://www.envirofit.org/products/?sub=cookstoves&pid=10.

Ermes, Miikka, Juha Pärkkä, Jani Mäntyjärvi, and Ilkka Korhonen. 2008. "Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions." *IEEE Transactions on Information Technology in Biomedicine* 12 (1): 20–26. doi:10.1109/TITB.2007.899496.

Ezzati, Majid, Homayoun Saleh, and Daniel M Kammen. 2000. "The Contributions of Emissions and Spatial Microenvironments to Exposure to Indoor Air Pollution from Biomass Combustion in Kenya." *Environmental Health Perspectives* 108 (9): 833–39.

Falk, Armin, and James J Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326 (5952): 535–38. doi:10.1126/science.1168244.

Flay, Brian R. 1986. "Efficacy and Effectiveness Trials (and Other Phases of Research) in the Development of Health Promotion Programs." *Preventive Medicine* 15 (5): 451–74.

Greenstone, Michael, and Ted Gayer. 2009. "Quasi-Experimental and Experimental Approaches to Environmental Economics." *Journal of Environmental Economics and Management* 57 (1): 21–44. doi:10.1016/j.jeem.2008.02.004.

Harrell, Stephen, Theresa Beltramo, Garrick Blalock, Juliet Kyayesimira, David I. Levine, and Andrew M. Simons. 2016. "What Is a Meal?: Comparing Methods of Auditing Carbon Offset Compliance for Fuel Efficient Cookstoves." *Ecological Economics* forthcoming.

Johnson, Michael, Rufus Edwards, Victor Berrueta, and Omar Masera. 2010. "New Approaches to Performance Testing of Improved Cookstoves." *Environmental Science and Technology* 44 (1): 368–74. doi:10.1021/es9013294.

Leonard, Kenneth L. 2008. "Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect." *Journal of Health Economics* 27 (2): 444–59. doi:10.1016/j.jhealeco.2007.07.004.

Leonard, Kenneth L, and Melkiory C Masatu. 2006. "Outpatient Process Quality Evaluation and the Hawthorne Effect." *Social Science & Medicine* 63 (9): 2330–40. doi:10.1016/j.socscimed.2006.06.003.

Leonard, Kenneth L., and Melkiory C. Masatu. 2010. "Using the Hawthorne Effect to Examine the Gap between a Doctor's Best Possible Practice and Actual Performance." *Journal of Development Economics* 93 (2): 226–34. doi:10.1016/j.jdeveco.2009.11.001.

Levine, David I, Theresa Beltramo, Garrick Blalock, Carolyn Cotterman, and Andrew M. Simons. 2016. "What Impedes Efficient Adoption of Products? Evidence from Randomized Sales Offers for Fuel-Efficient Cookstoves in Uganda." Berkeley, CA. http://escholarship.org/uc/item/2cb9g10z.

Levitt, Steven D, and John A List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21 (2): 153–74.

———. 2011. "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments." *American Economic Journal: Applied Economics* 3 (1): 224–38.

Miller, Grant, and A. Mushfiq Mobarak. 2013. "Gender Differences in Preferences, Intra-Household Externalities, and Low Demand for Improved Cookstoves." National Bureau of Economic Research. http://www.nber.org/papers/w18964.

Mueller, Florian, Frank Vetere, Martin R Gibbs, Stefan Agamanolis, and Jennifer Sheridan. 2010. "Jogging over a Distance: The Influence of Design in Parallel Exertion Games." *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, 63–68.

Mukhopadhyay, Rupak, Sankar Sambandam, Ajay Pillarisetti, Darby Jack, Krishnendu Mukhopadhyay, Kalpana Balakrishnan, Mayur Vaswani, et al. 2012. "Cooking Practices, Air Quality, and the Acceptability of Advanced Cookstoves in Haryana, India: An Exploratory Study to Inform Large-Scale Interventions." *Global Health Action* 5: 19016–19016.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2010. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *Economic Journal* 120 (546): F187–203. doi:10.1111/j.1468-0297.2010.02373.x.

Pillarisetti, Ajay, Mayur Vaswani, Darby Jack, Kalpana Balakrishnan, Michael N Bates, Narendra K Arora, and Kirk R Smith. 2014. "Patterns of Stove Usage after Introduction of an Advanced Cookstove: The Long-Term Application of Household Sensors." *Environmental Science & Technology* 48 (24): 14525–33.

Ruiz-Mercado, Ilse, Eduardo Canuz, Joan L. Walker, and Kirk R. Smith. 2013. "Quantitative Metrics of Stove Adoption Using Stove Use Monitors (SUMs)." *Biomass and Bioenergy* 57 (October): 136–48.

Ruiz-Mercado, Ilse, Nick L Lam, Eduardo Canuz, Gilberto Davila, and Kirk R Smith. 2008. "Low-Cost Temperature Loggers as Stove Use Monitors (SUMs)." *Boiling Point* 55: 16–18.

Ruiz-Mercado, Ilse, Omar Masera, Hilda Zamora, and Kirk R. Smith. 2011. "Adoption and Sustained Use of Improved Cookstoves." *Energy Policy* 39 (12). Elsevier: 7557–66. doi:10.1016/j.enpol.2011.03.028.

Schwartz, Daniel, Baruch Fischhoff, Tamar Krishnamurti, and Fallaw Sowell. 2013. "The Hawthorne Effect and Energy Awareness." *Proceedings of the National Academy of Sciences of the United States of America* 110 (38): 15242–46. doi:10.1073/pnas.1301687110/.

Simons, Andrew M., Theresa Beltramo, Garrick Blalock, and David I. Levine. 2014. "Comparing Methods for Signal Analysis of Temperature Readings from Stove Use Monitors." *Biomass and Bioenergy* 70: 476–88.

Smith, Eleanor. 2014. "Better Hygiene Through Humiliation." *The Atlantic*, August 13.

Smith, Kirk, Karabi Dutta, Chaya Chengappa, P.P.S. Gusain, Omar Masera, Victor Berrueta, Rufus Edwards, Rob Bailis, and Kyra Naumoff Shields. 2007. "Monitoring and Evaluation of Improved Biomass Cookstove Programs for Indoor Air Quality and Stove Performance: Conclusions from the Household Energy and Health Project." *Energy for Sustainable Development* 11 (2): 5–18.

Smith-Sivertsen, Tone, Esperanza Díaz, Dan Pope, Rolv T Lie, Anaite Díaz, John McCracken, Per Bakke, Byron Arana, Kirk R Smith, and Nigel Bruce. 2009. "Effect of Reducing Indoor Air Pollution on Women's Respiratory Symptoms and Lung Function: The RESPIRE Randomized Trial, Guatemala." *American Journal of Epidemiology* 170 (2): 211–20. doi:10.1093/aje/kwp100.

Srigley, Jocelyn A, Colin D Furness, G Ross Baker, and Michael Gardam. 2014. "Quantification of the Hawthorne Effect in Hand Hygiene Compliance Monitoring Using an Electronic Monitoring System: A Retrospective Cohort

Study." *BMJ Quality & Safety* 23 (12): 974–80. doi:10.1136/bmjqs-2014-003080.

Thomas, Evan A, Christina K Barstow, Ghislaine Rosa, Fiona Majorin, and Thomas Clasen. 2013. "Use of Remotely Reporting Electronic Sensors for Assessing Use of Water Filters and Cookstoves in Rwanda." *Environmental Science & Technology* 47 (23): 13602–10. doi:10.1021/es403412x.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, et al. 2011. "Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences of the United States of America* 108 (5): 1821–26. doi:10.1073/pnas.1000776108.

Table 4.1

Regressions testing for Hawthorne effect: estimates of effects of the presence of measurement team in primary three stone fire (TSF) usage and combined Envirofit usage, the coefficients represent the change in hours cooked per day compared to hours cooked per day in the measurement week

	Primary TSF		Combined Envirofit	
	(1)	(2)	(3)	(4)
Week prior to and after measurement week constrained to be equal	1.78***		-2.97***	
	(0.46)		(0.60)	
Week prior to measurement week		1.17**		-2.58***
		(0.54)		(0.69)
Week after measurement week		2.37***		-3.30***
		(0.63)		(0.64)
Household fixed effects	Yes	Yes	Yes	Yes
Observations	316	316	229	229
R-squared	0.82	0.82	0.79	0.79
Household clusters	118	118	89	89

Standard errors clustered at household level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The unit of analysis is a measurement "week" (approximately 72 hours) at a household. The specification in columns 1 and 3 imposes that the weeks prior to and after the measurement week are equal. The specification in columns 2 and 4 tests usage in the week prior to and after the measurement week separately. The coefficient estimates in column 2 are jointly significantly not equal to zero (p<0.01), but not statistically different from each other (p=0.10). The coefficient estimates in column 4 are jointly significantly not equal to zero (p<0.01), but not statistically different from each other (p=0.20).

Table 4.2

Daily mean firewood consumption, particulate matter and three stone fire usage prior to introduction of fuel-efficient stoves

	Wood Consumed (kgs)	PM2.5 (μg/m^3)	Three Stone Fire (hours)
	(1)	(2)	(3)
Mean Values	8.98***	427.79***	13.95***
	(0.33)	(23.18)	(1.03)
Observations	568	609	339
Household clusters	160	159	102

Standard errors clustered at household level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Columns 1, 2, and 3 present the average daily wood consuption, average daily PM2.5 reading, and the total hours of combined cooking on two three stone fires per household prior to receiving a fuel-efficient stove, respectively. Observations are at the household-day level.

Table 4.3

Estimates of biomass usage and indoor air pollution with and without in-person observers

Efficacy Trial (effects of an intervention during week with observers)			
	Daily hours	Biomass per hour	Total biomass (kg)
Total three stone fire usage	8.30	0.64	5.31
Total Envirofit usage	8.50	0.32	2.72
Totals:	16.80		8.03
	Daily hours	PM2.5 per hour	Total PM2.5 (μg/m^3)
Total three stone fire usage	8.30	30.67	254.56
Total Envirofit usage	8.50	14.97	127.25
Totals:	16.80		381.81
Effectiveness Trial (effects of an intervention during week without observers)			
	Daily hours	Biomass per hour	Total biomass (kg)
Total three stone fire usage	11.81	0.64	7.56
Total Envirofit usage	5.53	0.32	1.77
Totals:	17.34		9.33
	Daily hours	PM2.5 per hour	Total PM2.5 (μg/m^3)
Total three stone fire usage	11.81	30.67	362.21
Total Envirofit usage	5.53	14.97	82.78
Totals:	17.34		444.99

Note: Daily hours for the effectiveness trial are taken from the data for the week prior to the KPT. Recall that only about one fifth of the secondary three stone fires had iButtons on them at this point in our experiment. For the purpose of this illustrative table we make the assumption that households with missing values for the secondary three stone fire are equal to the mean value observed for the one fifth of the sample that had an hourly usage reading for the secondary three stone fire. Daily hours for the efficacy trial are based on the Hawthorne effects presented in Table 1. The consumption rates for biomass and PM2.5 with the three stone fires are calculated prior to the introduction of fuel-efficient stoves using the values in Table 2 (8.98 kg/13.95 hours = 0.64 kgs/hour and 427.79 μg/m^3/13.95 hours = 30.67 μg/m^3/hour). The consumption rates for the Envirofit G3300 are calculated using the emissions testing report in Figure 1 (0.64 kgs/hour * 0.499 = 0.32 kgs/hour and 30.67 μg/m^3/hour * 0.488 = 14.97 μg/m^3/hour).

Table 4.4
Bias introduced by the Hawthorne effect

	Daily cooking (hours)	Total biomass (kg)	Total PM2.5 (µg/m^3)
Baseline	13.95	8.98	427.79
Efficacy (observers present)	16.80	8.03	381.81
Effectiveness (no observers)	17.34	9.33	444.99

Note: These calculations are illustrative based on the mean values of data collected in the field and the emissions and performance report performed in a laboratory. These calculations assume a similarly sized Hawthorne effect on the secondary three stone fire as what we observed on the primary three stone fire.

Figure 4.1
Certified Emissions and Performance Report for Envirofit G3300

April 27, 2011

ENGINES & ENERGY CONVERSION LAB

DEPARTMENT OF MECHANICAL ENGINEERING
COLORADO STATE UNIVERSITY

1374 CAMPUS DELIVERY
FORT COLLINS, CO
80523-1374
970.491.4796
970.491.4799 (F)
WWW.EECL.COLOSTATE.EDU

Emissions and Performance Report

The stove listed below has been tested in accordance with the "*Emissions and Performance Test Protocol*", with emissions measurements based on the biomass stove testing protocol developed by Colorado State University (available at www.eecl.colostate.edu). Percent improvements are calculated from three-stone fire performance data collected at Colorado State University.

Stove Manufacturer:	**Envirofit International**
Stove Model:	**G-3300**
Test Dates:	**4/4/2011-4/22/2011**
Average CO emissions (grams):	**18.7**
80% Confidence Interval:	**17.7-19.7**
Percent Improvement:	**65.30%**
Average PM emissions (milligrams):	**995**
80% Confidence Interval:	**944-1046**
Percent Improvement:	**51.20%**
Average Fuel use (grams):	**596.7**
80% Confidence Interval:	**591.6-601.7**
Percent Improvement:	**50.10%**
Average Thermal efficiency:	**32.6**
80% Confidence Interval:	**32.3-32.8**
Percent Improvement:	**105.20%**
High Power (kW):	**3.3**
80% Confidence Interval:	**3.3-3.4**
Low Power (kW):	**1.9**
_____ 80% Confidence Interval:	**1.8-1.9**

The above results are certified by the Engines and Energy Conversion Laboratory at Colorado State University. All claims beyond the above data are the responsibility of the manufacturer.

Morgan DeFoort
EECL Co-Director
Technical Lead, Biomass Stoves Testing Program