Rural electrification: the potential and limitations of solar power
Pre-Analysis Plan

Anna Margret Aevarsdottir[*], Nicholas Barton[†] and Tessa Bold[‡]

December 22, 2015

1 Introduction

This plan outlines the hypotheses to be tested and specifications to be used in the analysis of the data obtained from a randomized controlled trial (RCT) on solar power to be conducted in Mwanza, Tanzania in 2015-2016. Since this pre-analysis plan was completed prior to the authors collecting mid-line and follow up data from the program and also prior to the authors having access to the compiled baseline data the plan can provide a useful reference in evaluating the final results of the study.

This plan is organized in the as follows: Section 2 reviews the motivation for the study and the setting; Section 3 outlines the experimental design including sample selection, treatment assignment and data sources; Section 4 details the hypotheses to be tested as part of the study and Section 5 documents the specifications to be used in analyzing the data.

2 Overview of the Study

2.1 Motivation

There is a general consensus that energy, in particular electricity, is a key input to economic development. A small literature has established a positive relationship between electricity access and a range of household and individual outcomes. These include increases in female employment [Dinkelman, 2011] and improvements in education [Khandker et al., 2013] and health [Barron and Torero, 2014]. The focus of this project is to provide the first experimental evidence on the adoption and impact of solar power at the household level. We will examine how price, liquidity constraints and information influence the decision to adopt a clean and renewable energy source such as solar power and how access to this source of energy influences household outcomes. This study is motivated by three main observations. First, nearly 1.3 billion individuals around the world lack access to electricity, of those, over 600 million reside in sub-Saharan Africa [IEA, 2013, IEA, 2014]. Second, based on current grid expansion plans and high population growth, it is estimated that 530 million people in, primarily rural, sub-Saharan Africa will remain without grid connection for the next 30 years [IEA, 2014]. Finally, due to this slow progress in expanding national grids solar power is being promoted by many as a decentralized and clean solution for rural areas that requires minimal infrastructure investments. Despite the apparent high potential, very little is known about the benefits and limitations of solar power at the household level.

2.2 Setting and Program Description

In partnership with GiveWatts, a NGO working in Kenya and Tanzania, we have designed a randomized field experiment in which households in rural Tanzania will be offered the chance to purchase solar powered lamps with solar panels. The implementation of the project is scheduled to begin late 2015 or early 2016. In our study the lamps will be offered at

[*] anna.aevarsdottir@iies.su.se
[†] barton@econ.uni-frankfurt.de
[‡] tessa.bold@iies.su.se

different levels of subsidization at the household level. The lamps are fitted with a mobile phone charging point and with a daily charge the lamps have enough capacity to provide the household with light for several hours and a small amount of power - enough to, for example, charge a mobile phone.

It may seem that this is a rather limited intervention given that the solar panels and lamps offered in our study are only able to provide a clean source of lighting and a limited amount of power. However, it has been documented that even for rural households in sub-Saharan Africa with grid connections electricity consumption is generally quite low, typically in the range 50 and 100kWh per person per year. To put these numbers in context, annual consumption of 50kWh per person for a five person household would for example power a mobile phone, two compact fluorescent light bulbs and a fan for approximately five hours a day [IEA, 2014]. These low levels of energy consumption despite grid connections lend support to the idea that decentralized energy solutions with minimal infrastructure investments can serve as a short run energy solution in rural areas where demand for energy and willingness-to-pay is low.

Our partner organization has been working with schools in rural Kenya since 2010 to provide solar power energy solutions to households in off-grid areas. Recently they have expanded their operations into the Northern part of Tanzania. The NGO operates based on the following protocol: first the NGO establishes a partnership with each school to facilitate the distribution of the lamps and collection of payments. GiveWatts then organizes a meeting with parents of children in the school to demonstrate how the lamps work and explain the price and payment structure of the lamps. The process of distributing the solar lamps and collecting payments from households is then managed jointly by the Parent Teacher Association in the school and a local agent from the NGO. The current GiveWatts model in Kenya provides the lamps at the recommended retail price of 3500 KSh ($37). Households are offered the lamps on credit and can repay over a number of months. The payment structure is such that households make an initial payment amounting to roughly a third of the total price and then pay the remainder of the cost in instalments over a 4 month period. The roll-out in Tanzania will follow a similar model and prices, with the major difference being that each of the four instalments are equal, at 20,000 TSh ($9.2). Repayment rates in Kenya are quite high with roughly 95% of lamps repaid in full. Despite the lamps being cost-effective when compared with alternative fuel sources, take up at full price is rather low at an average of 10% across GiveWatts program schools.

3 Experimental Design

We designed our experiment to consider two issues. The first is to consider the demand curve for solar lamps. By offering the lamps to households at differing price levels, we can gain a relatively crude impression of the demand curve for solar lamps. We then use the variation in demand for lamps at different prices to induce variation in lamp ownership. Where people receive a high subsidy, we can use this for an intention-to-treat estimation for the effects of lamp ownership.

3.1 Power and Sample Size

To determine the sample size for our experiment, we focus on three core outcome variables: 1) technology adoption (purchase of lamp) 2) educational achievement (test scores) and 3) female employment. Based on data provided by GiveWatts and test score data from the World Bank Service Delivery Initiative in Tanzania (Bold, Gauthier, Maestad, Svensson, 2011) and following a protocol requiring 80% power of detecting a significant difference at the two-sided 10% level, we find that sample of 1800 households would detect a 0.25 standard deviation increase in test scores and female employment and a 0.1 standard deviation increase in technology adoption with 80% power.

The sample is therefore set to 1800 households. For the administrative reasons set out above, we will conduct a stratified randomized controlled trial, by first randomly selecting 60 schools eligible for the GiveWatts programme, randomizing sets of subsidies between them and then randomly selecting 30 households within the school to be allocated – again randomly – a subsidy level for the solar lamp.

Given that there is the possibility of lamp sharing and thus potential spill-over effects in outcomes we decided to add an additional 10 schools to our sample that would serve as pure control schools (without any lamp subsidies). From each of the pure control schools we sampled 30 households. Thus the final study sample consists of 2100 households.

3.2 Sample Selection

The evaluation sample consists of 2100 households in the catchment area of 70 primary schools in the district of Magu in Tanzania. The selection of Magu district was based on the expansion plans of our partner organization and the timing of the project funding. Based on the program structure of our partner organization the sample selection required a two step selection where we first select schools to be part of the study and then select households connected to each selected school.

3.2.1 Selection of schools

The selection of schools was randomized based on a list of all schools in Magu district based on a list provided by the Ministry of Education and Vocational (MoEV) Training in Tanzania. From the list we randomly selected 70 schools to participate in our study. Of the 70 schools 60 were randomly selected to receive the subsidy treatment. The remaining 10 schools were eligible for the standard GiveWatts program.

3.2.2 Selection of students and households

From each of the selected schools we collected student rosters. From the student rosters we randomly selected 30 students per school. All households of students selected for the study will be sent a letter introducing GiveWatts and informing them of the possibility of participating in our study and asking the student's parents to come to the school for a baseline interview. Some students may not be present at the distribution of letters, and it is unlikely that this is purely random. We will attempt to measure the extent of this selection effect by this using grade data from the school, if available, along with attendance data prior to the implementation of the program. In spite of possible sample selection of those present, we feel these are the households who would be likely to purchase lamps in the first place and as such are a representative sample of the households of interest to the study in the Mwanza region, given that households would have to attend school to purchase a lamp any way.

3.3 Treatment Assignment

Our main treatment instrument was a subsidy for a solar lamp, inducing variation in take-up of purchasing solar lamps. In addition to this we also cross randomized and distributed a free children's book to half of the sample. In other words the book randomization was stratified such that half the households for any given subsidy level received a book. We can then consider the complementary effect of owning a book to read on the reading skills of the children.

3.3.1 Assignment of treatment at school level

The 70 schools selected to be part of our study were assigned to one of the three following treatment categories, each with different percentage subsidies available:

A High average subsidy: $S_1 = \{0, 50, 100\}$

B Low average subsidy: $S_2 = \{0, 25, 100\}$

C No subsidy (Control schools)

We randomly assign 30 schools to treatment arm A, with a high average subsidy, and 30 schools to treatment arm B, with low average subsidy. The remaining 10 schools are assigned to control arm C, without any subsidies.

3.3.2 Assignment of treatment at household level

The treatment assignment of households, their level of subsidy, is determined by a random draw from the set of subsidies S assigned to the school. This randomization takes place via a public lottery with the respondents following the baseline interviews. Based on their draw the respondents will be presented with a voucher for their subsidy s_i. They can then redeem the voucher by purchasing a lamp from GiveWatts through the school. The voucher is valid for 2 weeks from the date of the draw. During the entire experiment, households will be able to buy lamps from GiveWatts at the full unsubsidized price.

In all project schools GiveWatts will follow their standard protocol in advertising the information meeting to all households in the school through the teachers and students. During the meeting there will be a demonstration of how the lamps and solar panels work and parents will be given information on the price and payment structure of the lamp. The introductory meeting will be conducted at the school and led by a representative from GiveWatts.

3.4 Data Sources

The primary sources of data are a baseline survey conducted immediately before treatment assignment and a follow-up survey that will be collected approximately 12 months after the baseline survey. Additional sources of data include a brief school survey, administrative data from the schools and administrative data from our partner organization, as well as a midline survey including student testing.

3.4.1 Baseline Survey

To measure the core outcome variables we will, as previously mentioned, conduct a household level survey at baseline before program implementation for all households in the sample. The baseline will consist of a detailed household survey administered at the school. In this survey we will collect information on general household and individual characteristics. The questionnaire will also include questions on fuel consumption and expenditures and time usage of household members. The expenditure questions are intended to measure the household level savings due to reduced fuel expenditures following the repayment period. The time-use questions will include a detailed set of questions on labour market outcomes for household members these include employment status, hours of work and earnings. The lamps may also provide household members with new income generating opportunities. Anecdotal evidence from conversations with our partner organization suggests that some households have used the lamps to sell mobile charging time to other households in the area or to rear animals such as chickens. In addition, the lamps may allow home-run businesses to stay open later into the evenings. At baseline we will also conduct testing of students in both control and treatment schools using a standardized reading test designed by the research team. This will allow us to convincingly compare education outcomes of students across schools. Throughout the program, we will collect administrative data from GiveWatts on take-up and repayment of the lamps. As part of our survey we will also conduct air quality measurements in the respondent's home to measure the impact of cleaner technology on the household indoor environment.

3.4.2 Follow-up Survey

12 months after program implementation we will conduct a second household survey for our initial sample households as well as re-testing the students in our sample. The household survey will take place at the home of each household, allowing us also to obtain GPS coordinates for each household's domicile.

3.5 Other data sources

We collect administrative data from two sources: the study schools and from our partner organization. The data from the schools includes enrolment and attendance data. The data from our partner organization includes take-up data (lamp purchase data) and repayment data.

4 Hypotheses and Estimation Methodology

We will collect a relatively rich data set which will be used to answer the set of questions below by testing the corresponding hypotheses:

4.1 Take-up and spillovers in take-up

The first set of questions and hypotheses will test to what extent variation in subsidy levels has an impact on take up of the solar lamps and whether there are spill-over effects in take up

1. How does demand for a solar powered light source vary with price?

   Hypotheses:

   (a) higher level of subsidy is likely to have (no) positive impact on take-up of solar lamps
   We estimate this hypothesis using the following equation:

   $$L_{i,t} = \alpha + \beta s_i + \epsilon_i$$

   where $L_{i,t}$ is an indicator for whether the household purchased a lamp in the initial baseline experimental phase (t).
   $H_a(H_0) : \beta > 0(\beta \leq 0)$

   (b) It is likely that the demand curve is convex (linear) $H_a(H_0)$ higher levels of subsidy squared is likely to have (no) positive impact on take-up of solar lamps
   We estimate this hypothesis using the following equation:

   $$L_{i,t} = 1 = \alpha + \beta_1 s_i + \beta_2 s_i^2 + \epsilon_i$$

   $H_a(H_0) : \beta_2 > 0(\beta_2 = 0)$

2. Are there spill over effects in adoption of a renewable energy source?

   Hypotheses:

   (a) subsidy treatment of neighbours has (no) positive impact on households take up of solar lamp

   We estimate this hypothesis using the following equation:

   $$L_{i,t} = \alpha + \beta s_i + \gamma TreatmentIntensity_i + \epsilon_i$$

   where $TreatmentIntensity_i$ can be measured in several different ways including the share of high subsidies and average treatment intensity of connected households as follows:

   - $ShareHigh_i = \frac{\sum_{j=1}^{n} d_{ij} \mathbb{1}(s_j = high)}{\sum_{j=1}^{n} d_{ij}}$
   - $Intensity_i = \frac{\sum_{j=1}^{n} d_{ij} \cdot s_j}{\sum_{j=1}^{n} d_{ij}}$ where $s_j \in \{1, 0.5, 0.25, 0\}$ is the subsidy level of household j.

   A high subsidy is defined as a subsidy of 100% or 50% and d_{ij} is an indicator for whether household i and j are connected. We will use two measures of household connectedness

   - geographic proximity of households (measured using GPS coordinates)
   - reported households connections for household questionnaire

   $H_a(H_0) : \beta > 0(\beta \leq 0)$

3. Do learning and experience influence adoption of a renewable and clean energy source?

   Hypotheses:

   (a) subsidy treatment at baseline has (no) positive impact on households take up of solar lamp at follow up
   We estimate this hypothesis using the following equation:

$$L_{i,t+1} = \alpha + \beta s_{i,t} + \epsilon_i$$

   $H_a(H_0) : \beta > 0 \ (\beta \leq 0)$

4.2 Impact on outcomes and spillovers in outcomes

The second set of questions and hypotheses we present mainly test whether our treatment (the solar lamp) had any impact on individual and household level outcomes.

1. $H_a(H_0)$ The solar lamp is likely to have (no) a positive average impact on reading skills (as measured by sections 1-5 on student reading test; the number of correct responses will be aggregated to create a measure of reading skills)
2. $H_a(H_0)$ The solar lamp is likely to have (no) a negative average impact on household expenditures on fuel for lighting (as measured by questions sec8_q2-sec8_q5)
3. $H_a(H_0)$ The solar lamp may have (no) a positive average impact on the likelihood of adult employment outside the home (as measured by question sec1_q1_7 activity 6,8-10,12-26)
4. $H_a(H_0)$ The solar lamp may have (no) negative average impact on adolescent employment outside the home (as measured by question sec1_q1_7 activity 6,8-10,12-26)
5. $H_a(H_0)$ The solar lamp may have (no) positive average impact on household income (as measured by questions sec1earn_act6-sec1earn_act29)
6. $H_a(H_0)$ The solar lamp may have (no) a positive average impact on time spent studying for children (as measured by question sec3_q1_8_2)

For outcomes in which the same question was asked in both baseline and follow-up surveys our main specification will be the following

$$Y_{i,t+1} = \alpha + \beta S_{high,i} + \pi Y_{i,t} + \gamma \mathbf{X_{i,t}} + \epsilon_{i,t+1}$$

where $Y_{i,t+1}$ is the given outcome variable measured at follow-up, $Y_{i,t}$ is its baseline value, $\mathbf{X_{i,t}}$ is a vector of control variables and ϵ_i is the error term. $S_{high,i}$ is an indicator for high subsidy treatment, where high subsidy is defined as a subsidy of 50% or higher. Huber-White standard errors clustered at the school/community level will be used. The estimate of β using this equation will then provide the intention-to-treat effect which is the effect of being assigned a high subsidy. However not all those assigned the high subsidy treatment will buy a lamp and some of those assigned a low subsidy (25% subsidy or lower) will purchase a lamp, as may some receiving no subsidy and paying the full price. Thus we can also estimate the LATE using the following specification

$$Y_{i,t+1} = \alpha + \beta \hat{L}_i + \pi Y_{i,t} + \gamma \mathbf{X_{i,t}} + \epsilon_{i,t+1}$$

where $\hat{L}_i$ is an indicator for purchasing a lamp, which is instrumented by assignment to treatment status $S_{high,i}$ or subsidy level s_i. As an alternate specification to remove the effect of time invariant covariates we can run a regression including household fixed effects ν_i:

$$Y_{i,t} = \nu_i + \gamma_t + \beta \hat{L}_{i,t} + \epsilon_{i,t}$$

where $t \in 0, 1$

Some variables will only be collected in the final survey and so a different specification will be required, not including the initial value of the outcome variable $Y_{i,t}$. This applies to variables such as the air quality in the home, where we must rely on the randomisation having worked well:

$$Y_{i,t+1} = \nu_i + \gamma_t \beta \hat{L}_i + \epsilon_{i,t+1}$$

4.2.1 Complementarities between lamps and books for educational outcomes

In order to test for complementarities between the solar lamps and access to reading material on educational outcomes we adjust the specifications above to include an interaction term between lamp subsidies ($S_{high,i}$) and books (B_i where $B_i = 1$ if household receives book and 0 otherwise)

The main ITT specification becomes:

$$Y_{i,t+1} = \alpha + \beta_1 S_{high,i} + \beta_2 B_i + \beta_3 S_{high,i} \times B_i + \pi Y_{i,t} + \gamma \mathbf{X_{i,t}} + \epsilon_{i,t+1}$$

The corresponding fixed effects regression becomes:

$$Y_{i,t} = \alpha + \beta_1 S_{high,i,t} + \beta_2 B_{i,t} + \beta_3 S_{high,i,t} \times B_{i,t} + \nu_i + \epsilon_{i,t}$$

where $t \in 0, 1$

The main LATE specification becomes:

$$Y_{i,t+1} = \alpha + \beta \hat{L}_i + \beta_2 B_i + \beta_3 \hat{L}_i \times B_i + \pi Y_{i,t} + \gamma \mathbf{X_{i,t}} + \epsilon_{i,t+1}$$

where $\hat{L}_i$ is an indicator for purchasing a lamp, which is instrumented by assignment to treatment status $S_{high,i}$ or subsidy level s_i

In all specifications the outcomes Y_i is reading test scores. The coefficient of interest is β_3 and the null (alternative) hypothesis is $H_0 : \beta_3 = 0$ ($H_a : \beta_3 \geq 0$).

4.3 Heterogeneity in treatment effects

We will examine heterogeneity along the following dimensions:

- gender (for parents employment outcomes and children's educational outcomes)
- age (for school children)
- distance to trading center and infrastructure (main road and/or electrical grid)

- socio-economic status (as measured by asset index)
- initial achievement (high vs. low performance on test at baseline)

4.4 Survey Attrition

Attrition may occur, but will hopefully be tempered by the fact that the enumerators will seek a representative of each household where they live, thus reducing effort to attend the school. We cannot rule out attrition but will check that this is not selective according to the information at baseline.

4.5 Multiple Hypothesis Testing

Given that the treatment has the potential to impact several different outcomes at the individual and household level and that we are interested in exploring heterogenous treatment effects along several dimensions one concern is that of false discovery due to multiple hypothesis testing. In order to address these concerns we will follow the standards emerging in the literature. One way to account for this is to group outcome measures into common domains. In our study the domains could for example be 1) child educational outcomes and 2) household economic activity. One can then sign the the outcomes in each domain such that the hypothesised effect goes in the same direction and construct a standardized treatment effect within each domain. One can also account for multiple inference within a domain by computing and reporting the family-wise error rate adjusted p-values.

References

[Barron and Torero, 2014] Barron, M. and Torero, M. (2014). Short term effects of household electrification:experimental evidence from northern el salvador. *Working paper*.

[Dinkelman, 2011] Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from south africa. *American Economic Review*, 101(7):3078–3108.

[IEA, 2013] IEA (2013). World energy outlook 2013. *International Energy Agency*.

[IEA, 2014] IEA (2014). Africa energy outlook 2014. *International Energy Agency*.

[Khandker et al., 2013] Khandker, S. R., Barnes, D. F., and Samad, H. A. (2013). Welfare impacts of rural electrification: A panel data analysis from vietnam. *Economic Development and Cultural Change*, 61(3):pp. 659–692.