Impact Evaluation of MCC Compact in El Salvador

The Rural Electrification Sub-Activity

Impacts Estimates Using Panel Data from 2009 to 2014

By: Maximo Torero[*] and Manuel Barron[†]

International Food Policy Research Institute
Consortium led by Social Impact
Final Revision: November 2016

[*] Máximo Torero is the Director of the Markets, Trade, and Institutions Division of the International Food Policy Research Institute (IFPRI), m.torero@cgiar.org

[†]Manuel Barron is assistant professor in the Department of Economics at Universidad del Pacífico, MF.BarronA@up.edu.pe

i. List of Acronyms

ARI	Acute Respiratory Infection
CAESS	Compañía de Alúmbrado Eléctrico de San Salvador
CI	Confidence Interval
CIC	Changes in Changes
DIGESTYC	Dirección General de Estadística y Censos
DVD	Digital Versatile Disk
EEO	Empress Eléctrica de Oriente
EHEIPCER	Encuesta de Hogares para Evaluar el Impacto de los Proyectos de Conectividad y Electrificación Rural
EPA	Environmental Protection Agency
EHPM	Encuesta de Hogares de Própositos Múltiples
ERR	Economic Rate of Return
FE	Fixed Effects
FOMILENIO	Fondo del Milenio
GIS	Geographic Information System
GOES	Government of El Salvador
IAP	Indoor Air Pollution
IFPRI	International Food Policy Research Institute
IV	Instrumental Variables
KS	Kolmogorov-Smirnov
MCC	Millennium Challenge Corporation
NHS	National Household Survey
PM10	Coarse Particulate Matter
PM2.5	Fine Particulate Matter
RED	Randomized Encouragement Design
SUR	Seemingly Unrelated Regressions
TV	Television
TVA	Tennessee Valley Authority
UCB-PATS	University of California at Berkeley Particle and Temperature Sensor
USA	United States of America
USD	United States Dollar

ii. Table of Contents

a. Table of Figures

b. List of Tables

iii. Executive Summary

c. Overview of Compact and Electrification Project

This study was conducted during a recent grid extension and intensification program in northern El Salvador, designed to be rolled out in three phases to account for construction costs and accessibility. The Government of El Salvador (GOES) covered all of the grid extension and installation costs up to the electric meter, and households paid for their internal wirings and connection fees (for a safety certification). The fee for the safety certification is around 100 United States Dollars (USD). This is a significant investment, amounting to 18 percent of the average annual per capita income in our sample (550 USD).

The experimental sample consisted of 500 households located in sub-districts that were scheduled to benefit from the project during its first year. We generated experimental variation in the connection fee by offering discount vouchers to a randomly selected subsample. We randomly allocated 200 low-discount vouchers (20 percent discount) and 200 high-discount vouchers (50 percent discount) and left the remaining households as a control group (N=100).

Encuesta de Hogares para Evaluar el Impacto de los Proyectos de Conectividad y Electrificación Rural (EHEIPCER), the household survey implemented for this study, is a standard survey that collects data on demographic characteristics, health, education, housing characteristics, energy use, income, and consumption—among other factors. It includes a detailed module on time allocation for up to four household members: the male head, the female head, and up to two school-age children. Strict training sessions were conducted to ensure high quality in data collection, which was conducted with handheld computers. Enumerators were trained and selected by the authors with the assistance of Dirección General de Estadística y Censos[3] (DIGESTYC), an office of the Ministry of Economy, and International Food Policy Research Institute (IFPRI) staff. The indoor air pollution (IAP) data described was collected by a subset of enumerators who underwent additional specialized training to this end.

The baseline household survey, designed using the 2007 Population Census as the sampling framework, was conducted in November and December 2009. It covered 4,800 households across northern El Salvador. Four follow-up surveys were collected in the same months in 2010, 2011, 2012, and 2013, respectively.

We used four subsamples in the analysis. A non-experimental subsample was formed of all households that were off the grid at baseline and includes 2,014 households. The experimental sample included 500 households in San Miguel and Chalatenango. A subset of these households (N=150) was selected for IAP measurement. The experimental IAP results were based on this subsample. Finally, 207 households from the non-experimental sample in San Miguel and Chalatenango were selected for IAP measurement. These were households that had not connected to the grid by the time the first follow-up survey was administered.

[3] Translated to General Division of Statistics and Census in English.

d. Impact Evaluation Design, Evaluation Questions, and Expected Outcomes

This impact evaluation is based on two main empirical strategies used to identify the effects of electrification on the outcomes of interest (listed below). The first strategy, random encouragement design, is an experimental strategy that exploits random allocation of discount vouchers to a subsample of study households. The second strategy, fixed effects estimation,[4] exploits the longitudinal nature of the data and uses within-household variation in electrification status to estimate the effects of electrification.

In addition, we used novel equipment (University of California at Berkeley Particle Monitors) to study the relationship between electrification and indoor air quality. Reductions in IAP are usually argued by the literature to be the most obvious and ubiquitous benefits from electrification, but there is no evidence to date regarding actual measures of IAP. This study provides the first evidence that access to electricity is linked to large, immediate, and permanent improvements in well-being.

We studied four main types of outcomes:

- **Adoption of formal grid connections**. We were particularly interested in the role of discount vouchers and the role of spillovers (imitation of neighbors' choices). We expected discount vouchers to increase the probability of formal connections. Spillovers could go in either direction; seeing the benefits of electrical connection that their neighbors receive could encourage households to connect themselves, but having more neighbors with formal connections also facilitates informal access to electricity (through the neighbor).
- **Effects on IAP concentration**. As households tend to substitute away from kerosene when they gain electrical connection, indoor air quality should improve. The size of this improvement is unknown and was key to the type of health effects we expected.
- **Effects on time allocation**. Household members will reallocate time across activities as a result of electrical connection, but the sign of these changes is not clear ex-ante. For instance, children will have a home environment that better facilitates studying and thus may decide to study more; on the other hand, electrification makes leisure more enjoyable (for example, access to television [TV] in the home) so children may decide to spend more time on these activities. Similarly, adult household members may decide either to work more and exploit new business opportunities or to enjoy more leisure.
- **Effects on income:** Finally, we also measure the effect on total household income as a result of access to rural electrification.

Each of these outcomes is discussed below—focusing on the experimental estimates that are statistically significant at the 95 percent confidence level (unless otherwise noted).

---

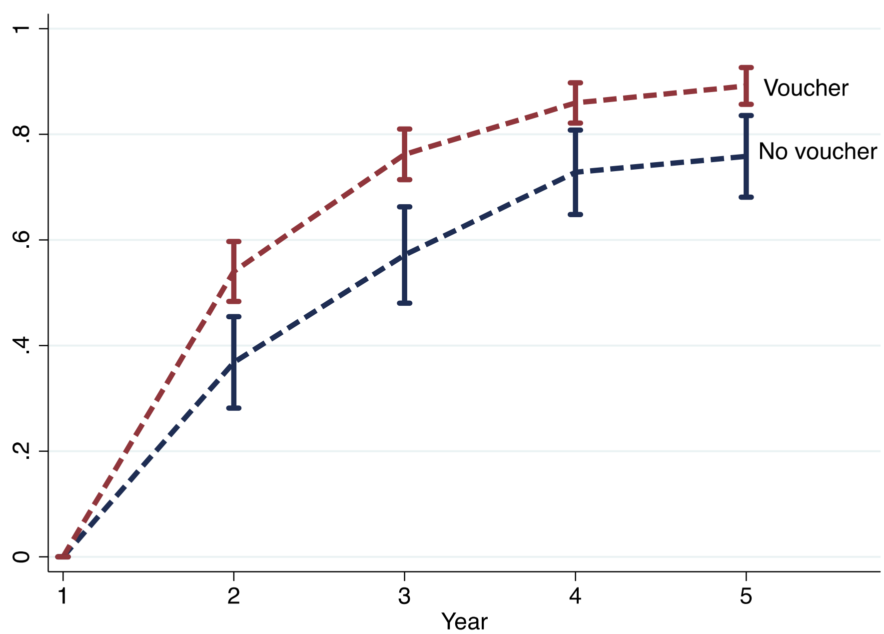
[4] Fixed effects estimation is a generalization of the difference in difference estimator where by the decision to connect to the grid is allowed to be correlated with time-invariant characteristics of the household and one can estimate the impact of connecting to the electric grid after controlling for the selection bias proxy by these characteristics.

e. Impacts of the Electrification Project: Findings

Adoption

- Voucher recipients were between 11 and 19 percentage points more likely to get a formal connection to the grid than control households; this is shown in Figure ES 1. Alternatively, a 10 USD reduction in the connection fee increased the probability of connection by two percentage points, though we find no systematic difference in connection between those receiving the high and low discount vouchers, suggesting diminishing returns to increasing voucher discount.
- There appeared to be an important spillover effect, wherein households are more likely to connect to the electrical grid if others nearby have already connected: a 10 percent increase in the share of eligible neighbors receiving a voucher increased the probability of connection by 1.3 percentage points. An additional connection within 100 meters increased the probability of a single household having a connection by 10 percentage points, almost the same effect as the household itself receiving a voucher.
- Up to this point, households with informal connections are considered off-grid. We studied the type of connection (formal, informal, or none) and found that vouchers increased the probability of having a formal connection and reduced the probability of having informal connections or no connections. We also find that households with an informal connection are significantly more likely to connect.
- There were significant increases in appliance ownership of "leisure" items, like TV sets and Digital Versatile Disk players, but also in ownership of appliances that could be used for home production. Electrification led to increased ownership of refrigerators (54 percentage points), blenders (25 percentage points), and washers (13 percentage points).

Figure ES 1: Voucher Allocation and Connection Rate

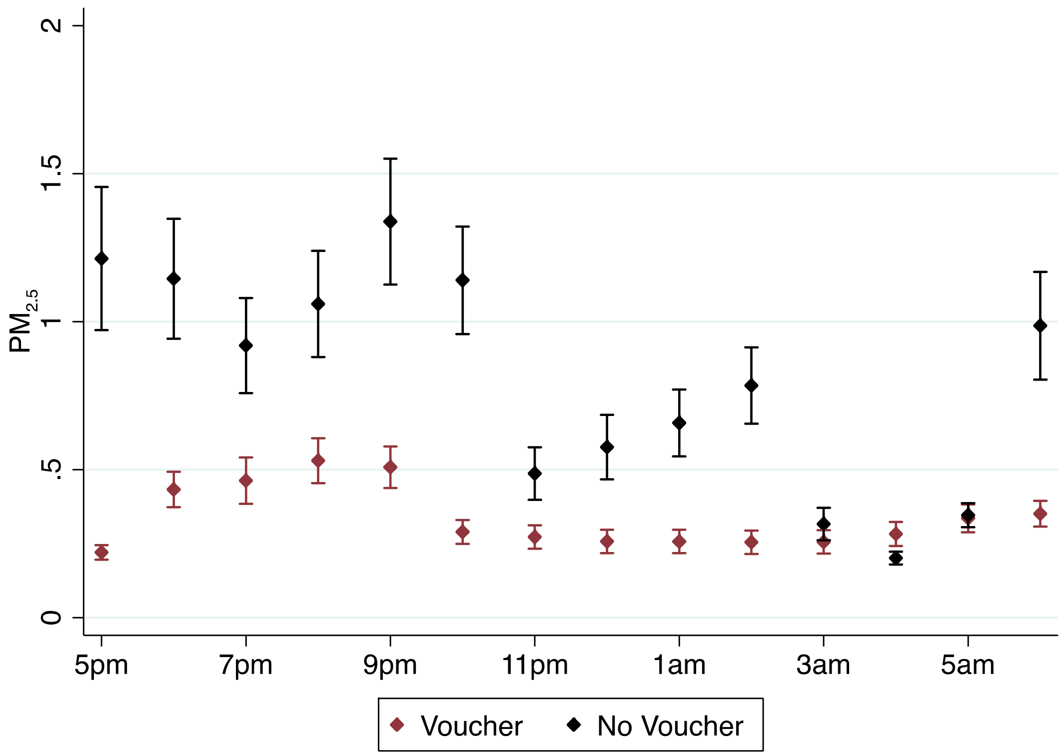


Effects on IAP

- By round three, voucher recipients showed drastic reductions in IAP compared to the non-recipient group, with a 67-73 percent lower fine particulate matter (PM2.5) concentration.[5] When collapsing the data at the household level, the magnitude and significance remained unaltered. In rounds four and five, the coefficients were closer to zero and were not statistically significant. This result may be attributed to the fact that the control group appears to be catching up to the treatment group in terms of their electrification rate. Our sample is simply too small to pick up differences in PM2.5 concentration with differences in electrification rates of around 10 percent.
- Figure ES 2 shows the reduced form results by hour of the day. The effects were larger from 5:00 pm to 10:00 pm, decreasing thereafter as most household members go to sleep around this time, and jump up again from 6:00 am to 7:00 am, when they wake up the next morning. The main mechanism behind the PM2.5 reductions in our study setting was a substitution away from kerosene lighting.
- There were large and statistically significant (only at the 90 percent confidence level) reductions in the incidence of acute respiratory infections (ARI) among children in households with an electrical connection. Vouchers led to a reduction of 16-18 percentage points at round three (significant at the 90 percent confidence level). The experimental estimate suggests electrification reduced ARI incidence by 48 percentage points during the study period.
- The changes in exposure were large for all household members (all above 30 percent), but these gains were unequally distributed across household members. The male head benefitted the most, with a reduction in exposure of 59 percent, while the female head benefitted the least, with a reduction of 33 percent. These differences were due to females spending more time than males in the kitchen, where pollutant concentration is highest, while males spend more time than females outside the home, where pollutant concentration is lowest.

[5] This figure is obtained from the reduced form coefficients: $e^{-1.119}-1=-0.67$; $e^{-1.316}-1=-0.73$.

Figure ES 2: Voucher Allocation and PM2.5 Concentration



Time Allocation – Children

- Electrification increases the probability of children participating in education activities by 78 percentage points. These activities include studying at home, spending time at school, and going to and from school. Separating the participation in different activities, this evaluation finds that the increase is driven by 54 percentage point increase in the probability of spending time studying; and an 84 percentage point increase in the probability of spending time commuting between home and school. This effect on the commute time increase reflects that children were more likely to attend school.
- Children who studied at home did so for an average of two hours a day. Because electrification increases the likelihood of participating in education activities, the treatment and control subgroups who participate in education activities may be different—representing a possible a selection bias discussed in the methodological section. Accordingly, this finding cannot be interpreted as electrification leading to a two hour increase in study time, but it serves an illustrative purpose.

The average time allocated to education (by those who participate in such activities) was 6.1 hours per day. A higher share of children studying at home is an important indicator of improved learning, especially given that this increase was paired with a better study environment (due to electrification).

Time Allocation – Adults

- Workers from connected households were 26 percentage points (at the 90 percent confidence level) more likely to engage in non-farm employment. The effects size was larger among females; the probability of engaging in non-farm employment at some point over the four periods following grid extension was 46 percentage points higher for females among on-grid households.

Effects on Annual Household Income

- Electrification increased the probability of operating a home business by 12 percentage points among the households that connected to the grid because of the voucher. When we split the sample by gender, only the point estimates for females were statistically significant and increased to 25 percentage points.
- We find mixed evidence on the effects of electrification on household income. The experimental estimates of the impact of electrical connection on income suggest that electrification increased annual household income by around 1,600 USD per year, although the point estimate is noisy and not statistically significant. The non-experimental effects are more modest and more precisely estimated. The non-experimental effects suggest an increase of 55 USD in non-labor net income (18 percent increase from baseline) and 208 USD on labor net income (20 percent from baseline)—statistically significant at the 95 percent confidence level. The effect on total net income is 111 USD (8.8 percent of the baseline). The differences across these estimates show that the effects could be very large for the households that connected to grid because of the voucher.

f. Next steps/future analysis

The results presented in this report will be compiled in academic papers to be published in policy and development journals. Presentation dissemination efforts will include: presentation of the report(s) to Millennium Challenge Corporation (MCC) Headquarters staff, presentation in MCC workshops, presentation of findings and key recommendations to local stakeholders, and presentation of findings in other international development conferences.

Table ES 1 Evidence Assessment of Rural Electrification Sub-Activity: Immediate and Short Term

Term	Theme	Impact	Size of Effects	Heterogeneity	Methodology
Immediate	**Coverage and Access**	Both the low and high-discount vouchers increased the probability of adoption of a formal connection.	Individual discount vouchers made households 11 to 19 percentage points more likely to connect to the grid. The effect of low and high-discount vouchers was roughly similar.	No differentiated effect.	Experimental
		Spillover effects were large. A neighbors' connection decision explains one's own connection decision.	An additional connection within 100 meters increased the probability of a single household having a connection by 10 percentage points, almost the same effect as the household itself receiving a voucher.	No differentiated effect.	Experimental
Short-term	**Coping Costs**	Decreased likelihood of using non-electric lighting sources.	Most fuel changes were due to reductions in kerosene use, 24 to 33 percentage points less likely to use it or spend any money on kerosene. Other sources showed economically small and statistically insignificant changes.	No differentiated effect.	Experimental
		Electrification caused large reductions in kerosene expenditures.		No differentiated effect.	Experimental
		No evidence of changes in cooking practices, either in the use of wood for cooking or in the probability of cooking outdoors.	This effect would be unlikely since the use of wood for cooking was around 85 percent at baseline and since cooking with electricity is much more expensive.	No differentiated effect.	Experimental

Table ES 2 Evidence Assessment of Rural Electrification Sub-Activity: Medium Term

Term	Theme	Impact	Size of Effects	Heterogeneity	Methodology
Medium-term	**Health**	Reduction in air pollution due to substitution away from kerosene as a lighting source.	Overnight air pollutant concentration was 67 to 73 percent lower among voucher recipients. The time resilience of the effects strengthened the link between household electrification and health.	No differentiated effect.	Experimental
		Electrification led to reduced incidence of ARI among children under the age of six.	This link is reflected in reductions of 37 to 44 percent in ARI incidence among children under age six.	No differentiated effect.	Experimental
	Education, Leisure and Information	School-age (six to 15 year old) children increased time studying at home. No impact on the probability of school enrollment.	Vouchers increased the probability of spending time in education activities by 78 percentage points.	No differentiated effect.	Experimental
		Increases in appliance ownership, such as TV sets, stereos, refrigerators, and blenders.	Electrification led to increased ownership of refrigerators (54 percentage points), blenders (25 percentage points), and washers (13 percentage points).	No differentiated effect.	Experimental
	Productivity	Beneficiaries of electrification were more likely to engage in	Workers from connected households were 26 percentage points (at the 90		Experimental, Non-Experimental

Term	Theme	Impact	Size of Effects	Heterogeneity	Methodology
		self-employment and in non-agricultural activities.	percent confidence level) more likely to engage in non-farm employment. Electrification increased the probability of operating a home business by 12 percentage points among the households that connected to the grid because of the voucher.	The effects size was larger among females. This increase seemed to come from adults in the 30-40 age range rather than younger workers.	Experimental, Non-Experimental

Table ES 3 Evidence Assessment of Rural Electrification Sub-Activity: Long Term

Term	Theme	Impact	Size of Effects	Heterogeneity
Long-term	**Economic Growth**	Increases in total income and expenditure.	We find mixed evidence on the effects of electrification on household income. The experimental suggest that electrification increased annual household income by around 1,600 USD per year, although the point estimate is noisy and not statistically significant. The non-experimental effects are more modest and more precisely estimated. The non-experimental effects suggest an increase of 55 USD in non-labor net income (18 percent increase from baseline) and 208 USD on labor net income (20 percent from baseline) — statistically significant at the 95 percent confidence level. The effect on total net income is 111 USD (8.8 percent of the baseline).	No differentiated effect.
		Distributional effects and poverty.	Income changes had some distributional consequences, with voucher recipients being 10 percentage points less likely to have income below the median.	No differentiated effect.

1 Introduction

In 2009, 1.3 billion people around the world lacked access to electricity at home (IEA 2011). At night, households with no access to electricity make do mostly with candles or kerosene lamps to satisfy their illumination needs. These sources of light provide poor illumination and, more importantly, emit large amounts of pollutants that are harmful to human health. In addition, these households lack adequate refrigeration technologies and thus face limitations in terms of ability to store food safely. Furthermore, due to the high costs of operating small electronics like radios or cellphones, these households have limited access to information and communication; they also do not have access to power tools or electric water pumps and are thus limited by the constraints of traditional technologies.

Access to electricity could unleash a series of changes in all these dimensions. Some recent evidence suggests that electricity may increase female labor supply (Dinkelman 2011, Grogan and Sadanand 2012) and improve educational outcomes, consumption, and income (e.g. Khandker et al, forthcoming, Van de Walle et al. 2013). However, other studies find no impacts beyond lighting (e.g. Bernard and Torero 2015, Bensch et al 2011). In addition to these mixed results, there is almost no evidence regarding the mechanisms that drive these changes. Some of the observed changes, like improvements in indoor air quality, are expected to be present in most settings where electrification occurs, but in most cases, the effect will depend heavily on household and context characteristics. This is due to the fact that for electricity to impact income or expenditures, households need to invest in resources, acquire new tools and complementary inputs, or build knowledge on how to operate these technologies. On the other side of the market, households also need demand for their goods and services. Constraints in access to credit or inputs, insufficient demand, or lack of knowledge can prevent electrification from affecting economic outcomes like income or expenditures.

To better understand how access to electrification affects the economic lives of rural households, we implemented an experimental study in northern El Salvador which gathered longitudinal data on a sample of households over a five-year period. While most of the literature studies electrification in an entire village or community, our approach allows us to study electrification status at the household level.

2 Overview of the Compact and the Electrification Project

The El Salvador Compact, a Millennium Challenge Corporation (MCC) project, began in September 2007 and ended in September 2012. The Compact consisted of three projects; these projects had the collective goals of stimulating economic growth and reducing poverty through productive development (68 million United States Dollars [USD]), human development (89 million USD), and connectivity (269 million USD). The human development project consisted of an education and training activity and a community development activity. The community development activity consisted of three sub-activities: rural electrification, community infrastructure, and water and sanitation. The goal of the electrification sub-activity was to increase the Northern Zone's electrical coverage from 70 percent to no less than 97 percent, with a total of 235,000 individuals gaining coverage.

2.1 Program Logic - Input, Output, Outcomes, and Ultimate Impact

2.1.1 Compact-level

The overall logic of the Compact was to improve the lives of Salvadorans in the Northern Zone. As such, the Compact combined infrastructural development with technical assistance aimed at connecting Northern El Salvador with the rest of the country. This increased connection intended to create opportunities for the region's residents through improved access to: markets through the east-west highway; electricity through expansions of the electrical grid and distribution of solar panels; water and sanitation facilities to decrease the incidence of disease; and other interventions in education, agriculture, and other productive activities.

The Northern Zone of El Salvador contains half of El Salvador's poorest municipalities and suffered more damage from the country's internal conflict during the 1980s than any other region. Economic and social indicators in the Northern Zone remain worse than the national average. In 2011, 48.4 percent of households in the Northern Zone were poor, compared with the 40.6 percent national estimate, and 18.7 percent of households in the region lived in extreme poverty in 2011, compared with 11.2 percent at the national level. Human capital development is also lower in this region than in any other. The average level of schooling in El Salvador was 6.2 years in 2011, while the average in the Northern Zone was only 4.7 years. The percentage of illiterate people in the Northern Zone was 21.9 percent in 2011 versus a 12.8 national average.[6] The goal of the Compact was thus to reduce rural poverty in the region by increasing regional economic growth through a five-year project consisting of strategic investments and technical assistance in various sectors.

2.1.2 Project-level

The electrification sub-activity consisted of the construction and extension of distribution lines and individual household connections to electrical networks.[7] An estimated 37,000 families now

---

[6] Source: DIGESTYC from national household survey 2011 and 2012 DYGESTYC (2012).

[7] Under this sub-activity, there was also the distribution of off-grid solar systems and technical assistance for

have reliable access to electricity in their homes, thanks to the installation of new power lines and solar power systems.

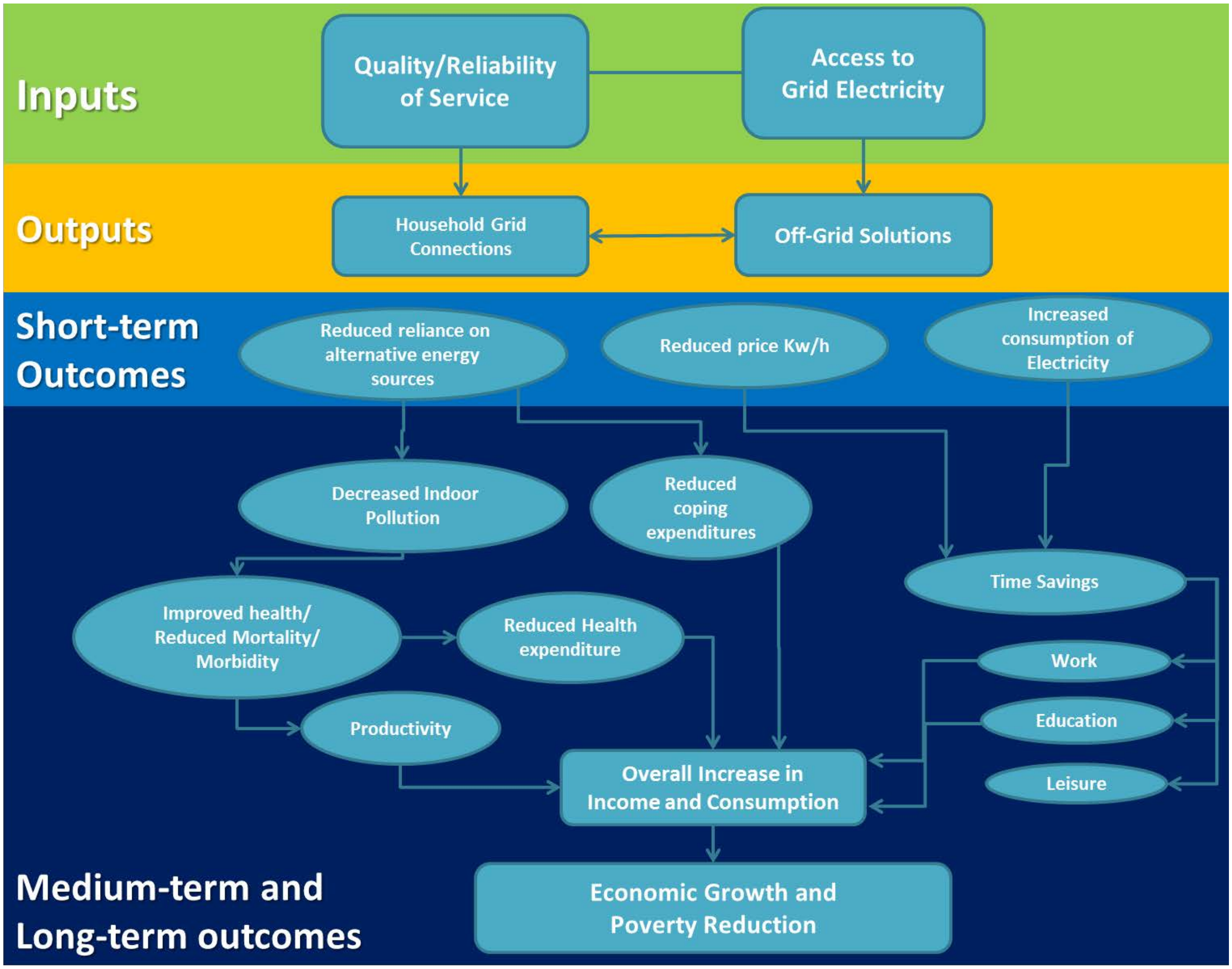
Figure 1 shows the impact pathways that relate inputs, outputs, outcomes, and impacts to achieve the overall objective of the Compact. The principal outcomes of the electrification sub-activity are improved access to the electricity network infrastructure and improved availability and increased quality of electricity services in the Northern Zone. These outcomes are reflected in several indicators, including the proportion of households with electrical connection, the distance to the nearest grid-connection point (which measures access), kilowatt hours of electricity consumed, and hours of electrical service (which measures measure availability and reliability).

As the figure shows, as households switch to electricity instead of traditional sources of energy—like wood and kerosene—this leads to decreased IAP and improved health outcomes. In turn, increased health leads to increases in households' productivity. This is one pathway through which electrification can help meet the overall program target of poverty reduction and economic growth in the Northern Zone. The other pathway occurs through time savings—access to electricity allows household members to allocate more time to productive activities and opens up non-farm business opportunities. These changes imply income flows that are more diverse and perhaps less volatile, promoting resilience and helping households out of poverty.

---

community capacity-building to ensure system maintenance and sustainability. The 1,950 families benefiting from the solar power systems live in geographical areas with difficult grid access and now have a source of electricity free from contamination. This part of the sub-activity is not included in this impact evaluation; rather, we focus on the beneficiaries who were connected to the electrical network.

Figure 1: Electrification Sub-activity Impact Pathways and Program Logic



The expected impacts of the rural electrification sub-activity on the well-being of the program beneficiaries were:

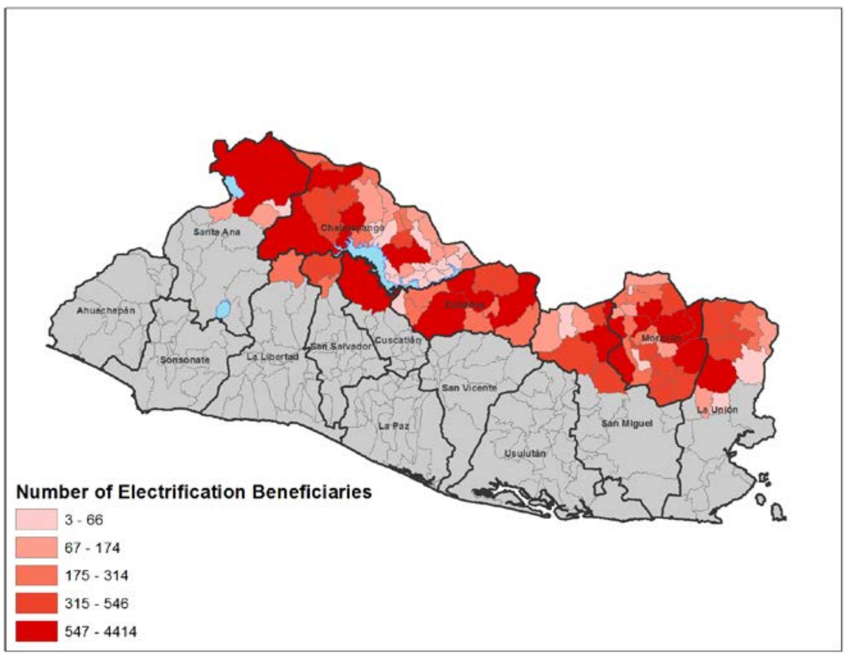
- Increased household income/consumption by at least 13 percent;
- Increased household access to electricity from a 78 percent baseline in 2007 to 90 percent access;
- Increased electricity consumption to 82 kWh per month; and
- Reduced time and money spent on seeking or purchasing electricity
    i. For example, decreasing the cost of electricity from 4.84 USD per kWh to 0.16 USD per kWh..

Other potential impacts of the electrification sub-activity included improvements in education, measured as increased attendance and enrollment due to an increase in the time available for education because of better lighting and increased access to educational information through television (TV)/radio. In addition, impact heterogeneity across gender and socioeconomic status will be explored.

2.2 Summary of the Implementation and Project Costs

The Government of El Salvador (GOES) created “Fondo del Milenio” (FOMILENIO) to be accountable for the Compact. FOMILENIO engaged with public agencies, contractors, and consultants for the direct execution of the projects while also remaining responsible for successful program implementation. The electrification project in Northern El Salvador was financed by MCC and the GOES, benefitting more than 37,000 families through the construction of over 1,500 kilometers of new electrical distribution and transmission lines to create an electrical grid in 94 municipalities in nine departments and connecting over 10,000 households to existing networks via the construction of necessary low-voltage extensions. Figure 2 shows the geographical distribution of the beneficiaries of these rural electrification projects.

Figure 2: Geographic Distribution of Rural Electrification Sub-activity



Project management and supervision were conducted by a team of experts with more than 25 years of experience in the field of electrical grids. The projects were managed and supervised at all stages: formulation (design and budget), construction orders, project execution, inspection of completed projects, labor and materials audits (with their associated costs), and closing of the construction orders.

During the development of the electrification projects, the management and supervision team verified the quality of labor. The team also verified that the materials and equipment used in the projects were new and of high quality, and that they fulfilled all the standards and norms set by El Salvador’s electricity and telecommunications regulator, Superintendencia General de

Electricidad y Telecomunicaciones. FOMILENIO investments were executed by AES El Salvador, a company that agglomerates four electricity distributors in the Northern Zone, and by DELSUR at the end of the Compact. Specifically, the construction of the distribution and transmission lines were as follows:

- Compañía de Alúmbrado Eléctrico de San Salvador (CAESS) with 566 kilometers. This distributor serves the Northern Zone of San Salvador, Chalatenango, and Cuscatlán y Cabañas.
- Empresa Eléctrica de Oriente (EEO) with 823 kilometers. This distributor serves San Miguel, Morazán, La Unión, and parts of Usulután y San Vicente
- CLESA with 116 kilometers. This distributor serves Santa Ana, Sonsonate, Ahuachapán, and parts of the La Libertad department.
- DELSUR 16 kilometers. This distributor serves the departments of La Libertad, San Salvador, La Paz, and San Vicente y Cuscatlán.

In addition, the rural electrification sub-activity built low and medium-voltage distribution lines and miscellaneous grid reinforcement projects in areas where the rural electrification project was expected to have the greatest impact. This included 10 kilometers of transmission lines along the northern Transnational Highway and 17 kilometers of additional transmission lines in the municipalities of Ilobasco, Cabañas, and Lislique, Morazán. The ultimate goal was to maintain high-quality electricity provision in the entire region.

Table 1 and Figure 3 present the development of the rural electrification project, showing monthly evolution in executed grid kilometers from the onset of the project in September 2009 to the close of the project in September 2012. To accomplish the project goal, 2.2 kilometers of distribution lines needed to be built daily; Figure 4 shows that EEO and CAESS had the highest level of distribution line construction.

Table 1: Monthly Evolution of Rural Electrification Project Construction (Kilometers of distribution lines)

Company	Sep-09	Feb-10	Mar-10	Apr-10	May-10	Jun-10	Jul-10	Aug-10	Sep-10
CAESS	31.1	3.4	43.5	12.2	11.1	9.3	12.4	11.9	20
EEO	86.6	3.1	39.7	11	12.5	32.3	40.8	10.4	34.1
CLESA					1.5	7	6.5	4	1.2
TOTAL	117.7	6.5	83.2	23.2	25.1	48.6	59.7	26.3	55.3
CUMM.	117.7	124.2	207.4	230.6	255.7	304.3	364	390.3	445.6

Company	Oct-10	Nov-10	Dec-10	Jan-11	Feb-11	Mar-11	Apr-11	May-11	Jun-11
CAESS	19.2	25.2	43.5	18.1	37.3	8.3	12.7	19.1	13.9
EEO	8.2	15.4	54.23	8.4	25.1	16.4	11.3	16.2	35.6
CLESA	6.4	7.6	3.59	0.4	5.7	0.5	5.5	3.3	9.6
TOTAL	33.8	48.2	101.32	26.9	68.1	25.2	29.5	38.6	59.1
CUMM.	479.4	527.6	628.92	655.82	723.92	749.12	778.62	817.22	876.32

Company	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12
CAESS	24.5	10.8	23.64	19.17	13.69	27.13	18.1	8.22
EEO	30.8	6.5	43.11	23.15	19.96	20.49	28	26.11
CLESA	2.3	0.8	13.19	3.09	3.29	9.38	4.01	0.59
TOTAL	57.6	18.1	79.94	45.41	36.94	57	50.11	34.92
CUMM.	933.92	952.02	1031.96	1077.37	1114.31	1171.31	1221.42	1256.34

Company	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	**TOTAL**
CAESS	19.34	11.03	14.91	9.7	5.1	9.1	0	154.9
EEO	53.21	10.85	40.24	18.31	24	6.36	10.58	270.5
CLESA	6.39	2.39	1.61	0.01	1.38	5.13	0	20.2
DELSUR	1.5	3.5	6	4	1.61			16.61
TOTAL	80.44	27.77	62.76	32.02	32.09	20.59	10.58	1522.59
CUMM.	1336.78	1364.55	1427.31	1459.33	1491.42	1512.01	1522.59	

Figure 3: Monthly Evolution of Rural Electrification Project in Kilometers

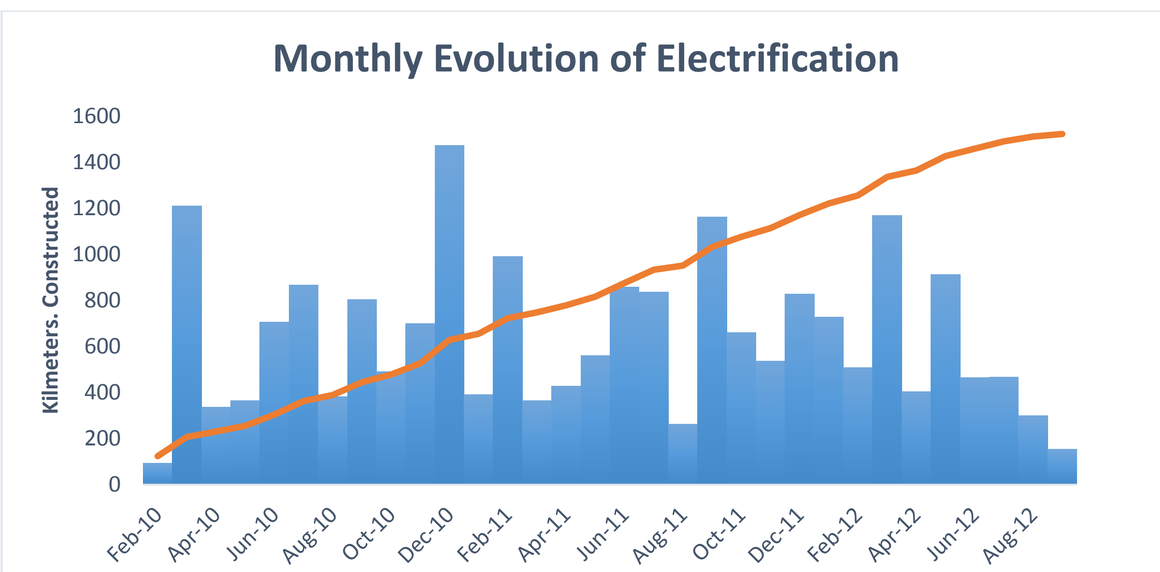


Figure 4: Monthly Evolution of Rural Electrification Projects by Company in Kilometers

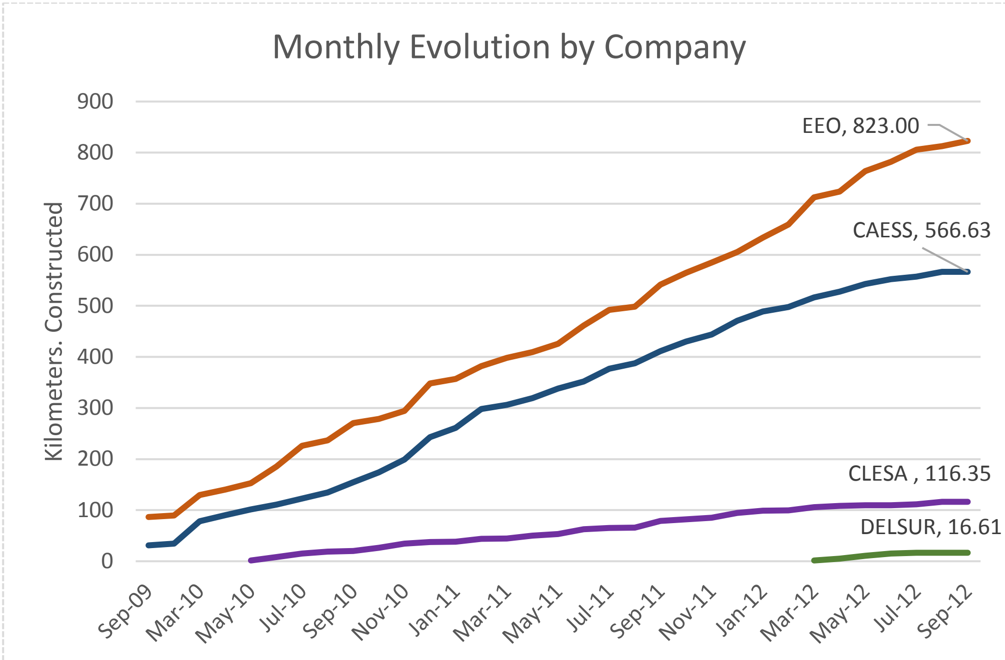


Table 2 shows the number of kilometers of distribution and transmission lines executed by each company and the average costs per kilometer for each distribution company involved in the development of the project. The average costs across each company were similar, with CAESS having the lowest cost per kilometer of electric distribution line constructed.

Table 2: Number of Electrification Projects, Kilometers Executed, and Associated Costs.

Company	Kilometers	Total with Tax	FOMILENIO Contribution	Cost per Kilometer for FOMILENIO
CAESS	567	9,588,745	8,150,433	14,385
CLESA	116	2,208,545	1,877,263	16,132
DELSUR	17	333,089	283,126	17,046
EEO	812	15,426,163	13,112,239	16,141
TOTAL	1,512	27,556,543	23,423,061	15,492

2.3 ERR and Beneficiary Analysis

The Economic Rate of Return (ERR) measures the effectiveness of a program by contrasting the discounted flows of costs and benefits of a specific intervention. The costs are comprised of any initial investment and any required maintenance expenditures throughout the course of the program. The benefits are determined by the gains of the population affected by the project.

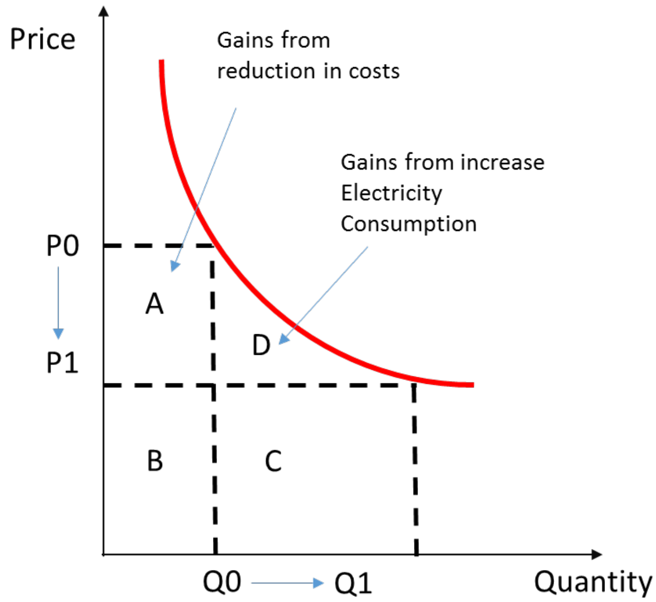
The ERR of the electrification project, including Compact administration costs, was 21.9 percent at Compact closing in 2012 (MCC (2012)). Calculation of the costs of a project is usually straightforward; the only data required is the set of investments required by the project and the selection of an appropriate discount factor to account for the intertemporal nature of the investment flow. However, estimation of benefits is a much more complex task; one of the most common methodologies used to estimate a program's benefits is a surplus approach.

We describe the ERR model used by MCC at Compact closing, using information made available as of August 2012. The closeout ERR for the electrification sub-activity was 19.4 percent over 20 years. The rural electricity sub-activity provided electrical connections to 35,412 households (connecting to new and existing networks). The ERR of 19.4 percent is estimated using beneficiary survey data; benefits include lower costs per kWh consumed and the benefits of consuming more electricity. For the ERR calculations, beneficiaries are the households receiving grid connections.

Access to electricity results in immediate and significant financial savings and increased household productivity. Key benefits of electrification include less time and money spent searching for inferior energy alternatives (e.g., fuel wood, kerosene, and batteries) and reduced negative environmental, health, and safety impacts of using these inferior alternatives in the home. The model uses consumer surplus to estimate the ERR. Consumer surplus is a measure of the total utility derived from consumption less the amount paid for the consumption. The change in consumer surplus can be split into a reduction in costs (including time savings) for the existing level of consumption and the benefits associated with the increased consumption. In Figure 5, the reduction in cost is area A and the increase in consumer surplus due to the increase in consumption is area D. Both of these areas measure the impact on household well-being, but neither one is strictly a measure of an increase in household income. The reduction in costs will allow an equivalent increase in consumption of other goods, as would be the case with an increase in income of the same amount. However, this is not an increase in income.

In the model, the demand curve was estimated by assuming a constant elasticity of substitution (that an x percent change in price induces a y percent change in quantity at all points on the demand curve). The functional form implied by this assumption has two parameters and can be solved for with two "known" points on the curve (current consumption and consumption with project). Once the curve is estimated, the area under the curve is found by integrating and then subtracting off the rectangle to estimate the consumer surplus.

Figure 5: Illustration of Consumer Surplus to Estimate the ERR



The project connected 22,970 households to a new or existing electrical grid and provided 1,950 households with solar power. MCC estimated that without the project, each household consumed the equivalent of 2.11 kWh/month at a price of 4.84 USD per kWh. The reduction in cost for those connected to the grid is 2.11kw*(4.84-0.16 USD) = 9.88 USD/month or 118.51 USD/year per household, shown as area A in Figure 5. The consumer surplus from induced consumption for those connected to the grid (area D, under the constant elasticity of substitution assumption) is 29.85 USD/month or 358.23 USD/year, but as not all of the consumer surplus will be translated into income gains, the benefit is only counted at 50 percent for the ERR. Summing these two yields, the annual benefit is 297.62 USD per year per household connected to the grid; using the same methodology, each household receiving a solar system yields a benefit of 217.33 USD per year. This produced the original ERR of 18.4 percent in August 2010.

At the end of Compact, the model was updated using the 2011 electricity survey data. Rather than using the assumptions about two points on the demand curve to estimate the curve, the curve was derived from the dataset. It was assumed that each survey respondent would, with electricity, consume 84 kWhs per month at 0.13 USD per kWh with a three cent per KWh subsidy. Each respondent therefore had a unique demand curve based on their initial prices and quantities without electricity. This yielded an updated ERR of 19.4 percent, which differs from the previous value in that the updated value accounts for the lower cost of electricity for households with and without a connection, and that the higher consumption of households without a connection that was found in the electricity survey data. This information is presented in Table 3.

Table 3: Rural Electrification Parameters for ERR

Parameter	ERR Value (18.4 percent)	ERR Value (19.4 percent)
Without Connections[1]		
Monthly consumption (kWh/month)	2.11	8.63
Price per kWh	4.84 USD	1.93 USD
Monthly cost	10.21 USD	16.66 USD
Elasticity of demand	-1.08	-1.06
With Connection		
Monthly consumption (kWh/month)	81.75	84
Price per kWh, w/ connection[2]	0.16 USD	0.13 USD
Elasticity of demand	-1.08	-1.06
Monthly cost	13.08 USD	8.40 USD
Solar System		
Monthly consumption (kWh)	11.3	24.47
Price per kWh	2.26 USD	0.44 USD
Monthly cost	5 USD	10.77 USD
Elasticity of demand	-1.08(-2.2)	-0.61

Notes: 1. Closeout values without connections represent the mean of the prices and quantities in the 2011 survey data excluding the top five percent of the distribution. 2. The closeout model accounts for a 0.03 subsidy per KWh.

As MCC points out, there are some caveats to these calculations:

a. One source of energy for households without electricity is car batteries, used by roughly six percent of households without a connection. The cost of recharging a battery is included in the ERR; however, we lacked data regarding the replacement costs of car batteries (when they wear out) and the frequency with which they must be replaced.
b. Due to a lack of data, the model makes certain assumptions. The model used survey data of consumption and prices with electricity and without electricity for different respondents, but did not use any before-and-after data showing how electricity consumption changed for specific individuals once an electric connection was acquired. Therefore, we must either assume that respondents without electricity had the same demand curve (resulting in different quantities consumed with electricity) or that respondents consumed the same amount once they gained a connection (resulting in different demand curves). The final

model assumed that respondents consume the same quantity with connection. As a further update to the ERR, this exercise can be replicated using the follow-up data for different years.

c. Price per kWh without an electrical connection varied widely, from 0.25 USD to 36.00 USD per kWh. The model eliminated the top five percent of observations as outliers; however, exactly which respondents answered accurately and which did not is unknown.

In summary, the households served by the project will realize an estimated savings of almost 100 USD per year when served by a connection to the electrical distribution network (based on a reduction in cost from over 1.93 USD/kWh on average to 0.13 USD/kWh), and roughly 70 USD per year when served by a solar system (based on a reduction in cost from over 1.93 USD/kWh to 0.44 USD/kWh). These savings are equal to more than 11 percent of the average annual household income in the Northern Zone.

The estimated beneficiaries of the rural electrification sub-activity at closeout was 160,770. MCC considers beneficiaries to be those individuals who saw improved standards of living, primarily through higher incomes, as a result of the economic gains generated by the MCC-funded project. The beneficiary estimates account for population growth and exclude accounts for all double counting (within the human development project and between other projects in the Compact).

Electrification: A Literature Review

Energy consumption in developing countries is a pressing matter in the fields of development and environmental economics (Wolfram et al 2012, Greenstone and Jack 2013). The economic effects of electrification play a central role in this topic, but solid empirical evidence regarding these effects is scarce (Bernard 2012).[8] To help fill this gap, we exploited a combination of experimental and non-experimental techniques to explore the mechanisms through which access to electricity affects household behavior and welfare in the short and medium term. As we will see, these mechanisms can have sizable effects on measures of human capital, welfare, and income.

Our research question is situated within the broader area of the effects of electrification, an active area of research in which the debate is far from settled. The massive resources allocated to rural electrification[9] are usually justified on the assumed benefits for health, education, and income, but most of the empirical evidence on which these claims are based is weak (Bernard 2012, IEG 2008) and the more recent literature shows mixed results. Some recent evidence suggests that electricity may save time spent on household chores, thus increasing female labor supply (Dinkelman 2011,Grogan and Sadanand 2012), or that it leads to improvements in educational outcomes, consumption, and income (e.g. Khandker et al, forthcoming, van de Walle et al 2013) and improvements in the human development index (Lipscomb et al 2013). These findings are not the same across the board, however; Chakravorty et al (2016) show large short-term welfare gains from electrification, while others find no such relationships (see e.g. Bernard and Torero 2015; Bensch et al 2011).

To date, the main paper on the effects of electrification on employment was published by Dinkelman (2011), and argues that time saved in fuel collection and other chores due to electrification can be utilized for other income-generating activities. Using IV and fixed effects (FEs), the author finds that although electrification did not alter male labor supply, women

---

[8] As a matter of fact, impact studies are rare in the field of infrastructure in general. There are plenty of difficulties in carrying out an experiment in this setting. Nevertheless, there are several good exceptions in the literature that exploit natural experiments. One such study is the seminal paper by Duflo (2001), which exploits a school building program in Indonesia to study the impact of infrastructure on education and wages. Devoto et al. (2009) study the impact of access to piped water among households in urban Morocco. Despite the fact that access to piped water saved an average of seven hours of time spent fetching water per week, there was no change in the labor supply, nor were there changes in income or educational outcomes. The main effect they found was an increase in happiness measures, mainly arising from increases in leisure, social activities, and social integration. Infrastructure programs may also have distributional effects. Duflo and Pande (2007) study the distributional effects of irrigation dams in India. Using land gradient as an instrument for dam placement, the authors find that dams have adverse distributional effects: poverty declines downstream of the dam but increases in districts where the dam is built. Strobl and Strobl (2011) conduct a similar study in Africa using geographic units instead of districts, which are political units. The authors also find that downstream basins are positively affected, but unlike the findings of Duflo and Pande (2007), upstream geographical units are not adversely affected.

[9] For instance, the World Bank recommends investing 10 USD billion per year between 2010 and 2020 in production and distribution of electricity in rural areas in Africa (World Bank, 2009). Given that the institution aims to provide 250 million people across Africa with modern sources of energy by 2030 (World Bank, 2007), understanding the effects of electrification is of urgent importance.

became 13 percent more likely to participate in the local labor market between 1996 and 2001, a period of rapid electrification in South Africa.

Regarding the effects of electricity on time allocation, Jacobson (2007) shows strong correlations between electricity adoption and the increase of night-time activities like accounting or preparing lectures, but no use of electricity as an input in agricultural production. This finding lines up with the hypothesis that electrification programs may have adverse distributional effects since electricity has a larger impact on non-agricultural activities, which are typically conducted by relatively richer households. Kline and Moretti (2011) study the long-run effects of the Tennessee Valley Authority (TVA) on productivity in the agricultural and manufacturing sectors. The TVA is a major place-based economic development policy that took place in Tennessee, United States of America, roughly between the 1930s and 1960s. The authors study the period between 1900 and 1990 and find that the TVA induced short-run gains in agricultural productivity, but that these were reversed after the program was scaled down. On the other hand, impacts on manufacturing employment continued to intensify even after the program had been scaled down. More recently, Fried and Lagakos (2016) study the macroeconomic effects of electrification in Africa and find that energy investment accounts for one-third of total growth in the region.

Not all recent studies find significant behavioral changes with the arrival of electrification, however. For instance, Madubansi and Shackleton (2007) and Hiemstra-van der Horst and Hovorka (2008) find little changes in energy portfolios with the arrival of electrification in South Africa and Botswana. These studies focus on changes in fuel wood consumption. The main concerns with this approach are that cooking fuels may affect the taste of food, that a shift to electricity would require households to buy new cooking equipment, and that the cost of cooking with electricity is expected to be much higher than that of cooking with fuel wood. All of these issues may prevent households from using electricity for cooking.

In addition to these welfare and income effects, electrification is argued to have effects on indoor air quality. At night, households with no access to electricity make do mostly with candles or kerosene lamps to satisfy their illumination needs. These sources of light provide poor illumination and, more dangerously, emit high amounts of pollutants harmful to human health. In fact, indoor air pollution (IAP) is the third leading risk factor for global disease burden, after high blood pressure and smoking (Lim et al 2013).[10] Given the stylized fact that lighting is one of the first uses of electricity in newly electrified areas (see e.g. Bernard, 2012; Barnes, 2007; IEG, 2008), electrification is expected to decrease IAP levels by replacing traditional sources of lighting such as kerosene, candles, and wood sticks. These reductions and their potential health effects are often argued to be one of the main benefits of electrification, but there is no solid empirical evidence to date.

This report provides the first experimental estimates of the relationship between household electrification and IAP. Within the frame of a clean experimental design, we collected a uniquely rich dataset that pairs minute-by-minute fine particulate matter (PM2.5) concentration with

[10] Ambient pollution also has important negative health effects, as shown by Chay and Greenstone (2003) and more recently by, e.g., Chen (2013) and Hanna and Oliva (2011). This study focuses on IAP only.

detailed data on household members' time allocation. This unique combination of data and experimental framework allows for an accurate estimation of the lower bound of the changes in PM2.5 concentration driven by electrification and, moreover, an assessment of the magnitude of the health effects that these changes imply.

In addition, this study allows for an analysis of the process of adoption of grid connections. There is little evidence regarding what drives household adoption of electricity in the developing world,[11] and since in most contexts households need to pay a large fee to connect to an electrical grid, it is important to understand the process of adoption if we are to understand the effects of electrification. For a variety of reasons, technology adoption studies in developing countries have been typically applied to the agricultural sector (see Sunding and Zilberman (2001) and Foster and Rosenzweig (2010) for reviews on this field), generating interesting evidence on social learning and other factors. For instance, Conley & Udry (2010) study technology adoption among Ghanaian pineapple farmers and show evidence of a special type of social learning: farmers will adopt technologies based on the experiences of their unexpectedly successful peers. In turn, Bandiera and Rasul (2006) study how adoption of sunflowers varied by a farmer's network structure, finding an inverted U-shaped relationship: when few people experiment, adoption by a neighbor promotes adoption, but when many people experiment, adoption by a neighbor deters adoption. More recently, Lee et al. (2015) show that despite the large coverage of the electrical grid in Kenya, most households remain unconnected, mainly due to high connection costs.

Other empirical studies look at technology adoption in the fields of health and information technology. The first randomized design on technology adoption in the presence of externalities in a developing country is the work by Kremer and Miguel (2007) on deworming pills, a follow-up to Miguel and Kremer (2004). The authors find that students who had contacts exposed to deworming pills were less likely to take the pill, mainly because the students had learned about the externalities of getting dewormed. In a study on adoption of antimalarial bed nets, Dupas (2012) also finds evidence of social learning: households that did not receive the initial subsidy were more likely to purchase a bed net if they lived near other households that received favorable subsidies. The same study shows that willingness to pay for a bed net was higher among subsidy recipients than in the control group, evidence of households learning from their own experience.

Thus, a key driver of technology adoption may be the process of learning about the technology. At first glance, the benefits of access to electricity may seem obviously large, but they are not necessarily so. For instance, it is difficult to assess how much a household should value an improvement in indoor air quality arising from substituting away from kerosene to satisfy their artificial illumination needs. In addition, it is necessary to learn about the payment system for an electrical connection. While it is relatively easy to observe how much kerosene is being consumed per day, it is considerably more difficult to estimate how a household's monthly electric bill will respond to an extra hour of using a light bulb. Furthermore, actual income generation due to electricity is far down the causal chain (households need to reallocate resources, invest in electric

[11] See Woolf (1987) for a study of adoption in early twentieth century United States of America.

tools, learn the new activities, and so on.). Hence, households need to learn about electricity before deciding whether to adopt it.

We followed Foster and Rosenzweig (2010) and defined learning as taking place when new information affects behavior and results in outcomes for an individual that are closer to the private optimum. One way in which households learn about electricity is by observing neighbors' behavior. Another way, perhaps less orthodox, is through informal connections. In learning-by-doing agricultural set-ups, individuals experiment with new technologies on small portions of land before expanding to their full plot. In our setting, informal connections are a relatively cheap way of experimenting with access to electricity without investing in the formal connection or committing to monthly electric bills.

An observational study of electricity adoption would likely suffer from placement bias and self-selection. Placement bias would arise if the expansion of the program is designed based on other underlying outcomes, or even correlated with those outcomes. Self-selection arises from the fact that having a connection to the grid is ultimately a household choice and, as such, depends on unobservable characteristics. To avoid placement bias, we studied only the group of households that would benefit earliest from the whole grid extension program ("Phase 1" households: see the third section for more details). To avoid self-selection bias, we generated exogenous variation in the cost of connection by allocating discount vouchers randomly among program households. Discount vouchers increase adoption through several potential mechanisms: reducing connection costs, ameliorating credit constraints, providing incentives to not procrastinate in gaining a connection, and increasing information about the program.

The random allocation of vouchers also created exogenous variation in the number of households that will connect in a given sub-district, so we could study the role of neighbors' choices, thus shedding light on social learning and preferences interactions. On the one hand, observing that neighbors connect to the grid may make households more prone to connect, either because of social learning or because of preferences interactions. Social learning would occur if households observed first-hand the benefits of electricity (better illumination, less smoke at night, better food availability) from their neighbors. Preferences interactions are similar to a "keeping-up with the Jones'" story: a household wants electricity because its neighbors have it. However, there is another side to this story. If connection rates are high in a neighborhood, getting an informal connection is easier, so the number of vouchers around a household may increase the number of informal connections. Thus, the number of formal connections will not necessarily increase with voucher intensity.

To our knowledge, Bernard and Torero (2015) is the first experimental study on the adoption of electrical connections. Using an identification strategy similar to the one used in this study, they find that the probability of connection increased among voucher recipients. In their study, households also responded to the number of vouchers allocated to their neighbors, something the authors attribute to "preferences interactions" (Manski 2000), also known as "bandwagon effects." More recently, Lee et al (2016) provide experimental evidence on the demand for rural electrification in Kenya and find that residential electrification in rural areas may reduce total welfare.

In summary, there are six main differences between this study and the existing literature. First, this study analyzed the electrification status at the household level. Other studies use geographical variables and thus attribute the same electrification status to a cluster of households (e.g. the village). Second, this study used an experimental design component that provided results that allow us to relax the most stringent assumptions in the statistical models required by other studies. Third, the data used in this paper includes detailed information on time allocation as well as data on home-based microenterprises. These two pieces of data tighten the hypotheses regarding time allocation and business generation. Fourth, we also collected data on minute-by-minute IAP concentration to provide evidence on the most immediate benefits of electrification. Fifth, the study design allowed us to study two determinants of adoption: connection cost and spillover effects. Sixth, the multiple follow-up surveys provided an idea of the timing of the changes driven by electrification.

2.4 Evidence Gaps Filled by the Current Evaluation

We provide the first evidence that electricity has large, immediate, and sustained impacts on wellbeing through a reduction in IAP. Electrification has long been argued to improve indoor air quality, but to this point, this had not been proven. Households without electricity rely on traditional fuels to satisfy their household cooking, lighting, and heating needs (Smith et al 2013). These fuels are usually burned inefficiently, which results in substantial emission of air pollutants (Naeher et al 2007). Several studies associate these pollutants with ill-health, both in the case of biofuels for cooking (Lim et al 2012) and in the case of kerosene lighting (Epstein et al 2013, Pokherl et al 2010).

As the main source of IAP, cooking has received the most attention in the literature. Significant efforts have been made to improve cooking practices, including improved cook stoves, ventilation, etc. Laboratory and field tests show that improved cook stoves can in fact reduce PM2.5 emissions. However, with the exception of households in China (Smith et al 1993, Sinton et al 2004), adoption rates by households are low (WHO 2006, PNAS paper). Hana, Duflo, and Greenstone (2012) implemented a randomized trial to study the effectiveness of improved cook stoves in reducing IAP. They show that that one year after introduction of improved stoves, PM2.5 concentration bounced back to its original level because households were not using the stoves adequately or at all.

Kerosene, on the other hand, has received little attention in the literature despite being used to light around 300 million households around the world (Lam et al. 2012a, Lam et al. 2012b). Even though kerosene is usually considered a cleaner alternative to biomass emissions from kerosene devices for cooking and lighting can impair lung function and increase infectious illness, asthma, and cancer risks (Lam et al. 2012). Kerosene emits not only PM2.5, but also carbon monoxide, nitric oxides, and sulfur dioxide (Schare and Smith 1995, Fan and Zhang 2001, Chen et al. 2007). In addition, kerosene lighting is estimated to emit 270,000 metric tons of black carbon per year, which amounts to seven percent of total annual black carbon emissions (Lam et al. 2012).

3 Impact Evaluation Design

3.1 Evaluation Type

This impact evaluation was based on two main empirical strategies to identify the effects of electrification on the outcomes of interest. The first strategy, random encouragement design, is an experimental strategy that exploits random allocation of discount vouchers to a subsample of study households. The second strategy, FEs estimation, exploits the longitudinal nature of the data and uses within-household variation in electrification status to estimate the effects of electrification.

In addition, we used novel equipment (UCB Particle Monitors) to study the relationship between electrification and indoor air quality. Reductions in IAP are argued to be the most obvious and ubiquitous benefits from electrification in the literature, but there is no evidence to date regarding actual measures of IAP. This study provides the first evidence that access to electricity is linked to large and immediate improvements in well-being.

A major challenge in the literature is the identification of causal effects since electrification potentially unleashes a number of changes through a complex chain of causality. The identification of causal effects is further complicated because these changes interact with each other, sometimes increasing the effects and other times attenuating them. Since an electrical grid cannot be expanded randomly, recent studies use time variation and instrumental variables (IV), mainly geographic variables, to deal with the endogeneity of connection. Studies used land gradient (Dinkelman, 2011; Grogan and Sadanand, 2012), distance to hydroelectric dams (Grogan and Sadanand, 2012), distance to electric line (Samad 2009), and distance to power-generating plants and baseline electrification rates in the locality (van de Walle et al 2013).

The first-stage relationship in these studies is clear. Since land gradient affects the cost of grid expansion, it is correlated with the probability of grid connection. The exclusion restriction is more difficult to justify, that is, land gradient might affect the outcomes of interest through other channels, not just through the probability of connecting to the grid. Land gradient, for instance, plausibly affects the cost of building and maintaining other types of infrastructure, such as roads, schools, or hospitals, thus potentially affecting transportation costs and access to markets, as well as education and health outcomes. Land gradient also may affect the crop varieties that can be grown in a region (and their profitability), thus directly influencing economic activity and income flows.[12] As such, the exclusion restriction requires the observed variation in land gradient to be in a range that does not affect other types of infrastructure, crops, or other economic activities. Alternatively, it requires the variation in said variables generated by variation in land gradient to have no effect on the outcomes of interest. This may perhaps be not too far from reality in some settings, especially in studies that interact land gradient with time-varying

[12] The same argument is valid for the other variables: electric lines, hydroelectric dams, and power generating plants tend to be placed in areas with certain characteristics that arguably affect economic outcomes through channels other than electrification. Other papers (e.g. Coen-Pirani et al 2010; Khandker et al 2012) use average electrification and appliance ownership rates in the locality. The same argument applies to these studies.

variables like budget availability (e.g. Lipscomb et al 2013), but it is ultimately an assumption that cannot be directly confronted with the data.

Randomized Encouragement Designs (RED) offer an appealing alternative. This approach, originated by Imbens and Angrist (1994), consists of randomly allocating incentives to connect to the grid and using those incentives as instruments in an instrumental variable estimation. The RED approach has been used extensively in other contexts (e.g. Hirano et al 2000; Devoto et al 2011; Mullally et al 2013; Allcott and Mullainathan 2010), but Bernard and Torero (2015) were the first to implement an RED approach to study electrification in developing countries. We implemented an approach similar to theirs. In our study setting, households were required to pay a 100 USD fee for a security inspection to get an electrical connection. We randomly allocated discount vouchers for 20 percent and 50 percent off the inspection fee, thus generating exogenous variation in the connection cost. Discount vouchers increased adoption through several potential mechanisms: reducing connection costs, ameliorating credit constraints, providing incentives to not procrastinate, and increasing information about the program.

The random allocation of vouchers also creates exogenous variation in the number of households that will connect in a given sub-district, so we were able to study the role of neighbors' choices, thus shedding light on social learning and preferences interactions. On the one hand, observing that neighbors connect to the grid may make households more prone to connect, either because of social learning or by preferences interactions. Social learning would occur if households observed first-hand the benefits of electricity (better illumination, less smoke at night, more food availability) from their neighbors. Preferences interactions are similar to a "keeping-up with the Jones" story: a household wants electricity because its neighbors have it. However, there is another side to this story. If connection rates are high in a neighborhood, getting an informal connection is easier, so the number of vouchers around a household may increase the number of informal connections. Thus, the number of formal connections need not increase with voucher intensity.

3.2 Evaluation Questions and Expected Outcomes

During the design phase of the rural electrification sub-activity, MCC and the GOES developed an ERR model to compare the expected benefits and costs of the project. The main benefits and evaluation indicators are the following:

- Household income/welfare
- Price of electricity per kilowatt-hour
- Consumption of electricity

Other outcomes that are considered relevant are analyzed in this impact evaluation to try to understand the effects of providing electricity to rural Salvadoran households. As discussed in the literature review, rural electrification has been attributed a significant range of benefits;[13] these can be summarized as follows:

[13] As explained in IEG (2009)

- Income benefits due to access to electricity and therefore access to new work opportunities, especially in non-farm activities
- Benefits from lighting and TV/radio, calculated mostly as Willingness to Pay, as shown in IEG (2009)
- Education benefits from higher educational attainment by the children of electrified households, which results in higher future earnings
- Time saved from household chores (additional leisure time), valued at the opportunity cost of labor, that is, the average wage; some evidence for Bangladesh and Peru can be found in Escobal and Torero (2004,2005) and Chowdhury and Torero (2007)
- Increased productivity of home business
- Increased agricultural productivity, calculated as incremental revenue
- Improved health comes from the value of reduced mortality as a result of improved indoor air quality from reduced reliance on kerosene lamps
- Reduced fertility coming from knowledge gained using electricity, valued at the cost of achieving fertility reduction through reproductive health programs
- Public goods benefits, such as increased security (see, for example, Chowdhury and Torero 2007)

It was too ambitious to try to capture all these benefits and clearly identify the causal relationship. In that sense, this impact evaluation of the rural electrification sub-activity attempted to answer questions that look at the overall impact on socioeconomic development:

- What is the impact of electrification on the cost of energy and energy consumption?
- What is the impact of introducing energy-efficient technology (i.e. connection to the grid vs other sources of off-grid energy sources) on uses of electricity?
- What is the impact of electrification on time allocation?
- What is the impact of electrification on indoor air quality?
- What is the impact of electrification on productive activities?
- What is the impact of expanded access to and use of electricity on household economic welfare?
- What are the differential impacts for women vs. men?
- Are results likely to be sustainable?
- Why do we see the impacts we see?

To shed light on these evaluation questions, we will concentrate on the following impact indicators:

i. Indicators of changes in quality of the electricity service:
    - Use of electricity
    - Expenditures on electricity (proportion of total energy sources)
    - Expenditures on electricity (proportion of total expenditure)
    - Number of failures
    - Price
    - Sources of energy

ii. Indicators of changes in welfare:

As shown in Escobal and Torero (2004), household income can be represented as:

$$\text{Y} \quad = L \sum_{i=1}^{n} Sl_i \left(\frac{y_i}{l_i}\right)$$

where Y is income approximated by expenditure, L is total household hours worked, Sl_i is the share of household working hours devoted to the i-th activity (where activities can be farm and non-farm), and $\frac{y_i}{l_i}$ is the hourly wage in the i-th activity. Thus, changes in income can be represented as:

$$\begin{aligned}\Delta Y_i \quad = \quad & L\left[\sum_{i=1}^{n} \Delta Sl_i \left(\frac{y_i}{l_i}\right)\right] + \Delta L\left[\sum_{i=1}^{n} Sl_i \left(\frac{y_i}{l_i}\right)\right] + L\left[\sum_{i=1}^{n} Sl_i \Delta\left(\frac{y_i}{l_i}\right)\right] \\ & + L\left[\sum_{i=1}^{n} \Delta Sl_i \Delta\left(\frac{y_i}{l_i}\right)\right] + \Delta L\left[\sum_{i=1}^{n} \Delta Sl_i \left(\frac{y_i}{l_i}\right)\right] + \Delta L\left[\sum_{i=1}^{n} Sl_i \Delta\left(\frac{y_i}{l_i}\right)\right] + \Delta L\left[\sum_{i=1}^{n} \Delta Sl_i \Delta\left(\frac{y_i}{l_i}\right)\right]\end{aligned}$$

Assuming that interactions in the second row of the equation are negligible, changes in income can be approximated as:

$$\Delta Y_i \quad = \quad L\left[\sum_{i=1}^{n} \Delta Sl_i \left(\frac{y_i}{l_i}\right)\right] + \Delta L\left[\sum_{i=1}^{n} Sl_i \left(\frac{y_i}{l_i}\right)\right] + L\left[\sum_{i=1}^{n} Sl_i \Delta\left(\frac{y_i}{l_i}\right)\right]$$

This equation represents three of the possible channels through which income may be affected by access to electricity. On the one hand, the first component of the equation shows the impact of changes on the proportion of working hours allocated to different activities. In this particular case, we analyzed shifts in labor devoted to agricultural and non-agricultural activities. Our hypothesis is that access to electricity leads to greater opportunities for non-farm work activities. On the other hand, electricity may also create overall employment opportunities. Thus, the second component captures the effect of changes in households' total working hours. Finally, there is scope for increases in rural households' market efficiency through increases in their purchasing power. In this line, the third component captures changes based on returns to labor (that is, hourly wages) allocated to agricultural and non-agricultural activities. In the case of agricultural activities, this will be directly related to the price of agricultural products.

We proxy these impacts through the following indicators:

- Change in total income and expenditure
- Total hours of work–household
- Hours of work–household and individual
- Percent hours of non-agricultural work–household and individual
- Hours spent in chores (especially collecting inputs for energy)
- Hours spent in childcare
- All of the above, by gender

The expected effects from these outcomes are summarized in Table 4.

Table 4: Outcome Indicators and Expected Effects

Term	Theme	Indicator	Expected Impact	Gender heterogeneity
Immediate	**Coverage and Access**	• Percentage of households connected to the grid.	Positive	No differentiated effect.
		• Cost of electricity.	Negative	No differentiated effect.
		• Reliability of electric services.	Positive	No differentiated effect.
Short term	**Coping Costs**	• Number of sources used.	Negative	No differentiated effect.
		• Consumption of electricity.	Positive	No differentiated effect.
		• Energy input collection time use.	Negative	Larger effect for females.
		• Coping expenses in other energy sources.	Negative	No differentiated effect.
	Health	• Indoor pollution.	Negative	No differentiated effect.
		• Incidence of acute respiratory disease among vulnerable groups.	Negative	No differentiated effect.
	Education, Leisure, and Information	• Hours in education or studying in the home.	Positive	No differentiated effect.
		• Hours spent in childcare.	No change	No differentiated effect.
		• Hours spent in entertainment and other leisure activities.	Positive	Larger effect for females.
	Productivity	• Total hours of work.	Positive	Larger effect for females.
		• Percentage of hours of agricultural.	Negative	Larger effect for females.
		• Percentage of hours of non-agricultural work.	Positive	Larger effect for females.
		• In home business productivity/revenue.	Positive	Larger effect for females.
Long term	**Economic Growth**	• Change in total income and expenditure	Positive	Larger effect for females.
		• Percentage of poor households	Negative	Larger effect for females.

3.2.1 Country-specific and International Policy Relevance of Evaluation

Almost two billion people in the world lack access to electricity. According to the 2007 National Census, around 80 percent of the El Salvadorian population had access to electricity in that year. Although this figure is high, there are strong correlations between socioeconomic status, electrification, and use of traditional fuels for lighting or cooking. Figure 6, Figure 7, and Figure 8 show that the poorest municipalities are those with the lowest electrification rates and the highest use of traditional fuels for cooking and lighting. One reason for the lack of electrical access is distance to the electrical grid, but even many households in the vicinity of the grid fail to connect. The questions that this evaluation aims to answer relate to the different effects that having access to electricity can have in the lives of rural households. In doing so, the study allowed us to investigate why households are not connecting to the grid; we look particularly the role of cost and spillover effects. As we have pointed out, there are only a handful of papers that study the effects of electricity to date, and more solid empirical evidence is needed to draw strong conclusions.

Figure 6: District Socioeconomic Status, El Salvador 2007.

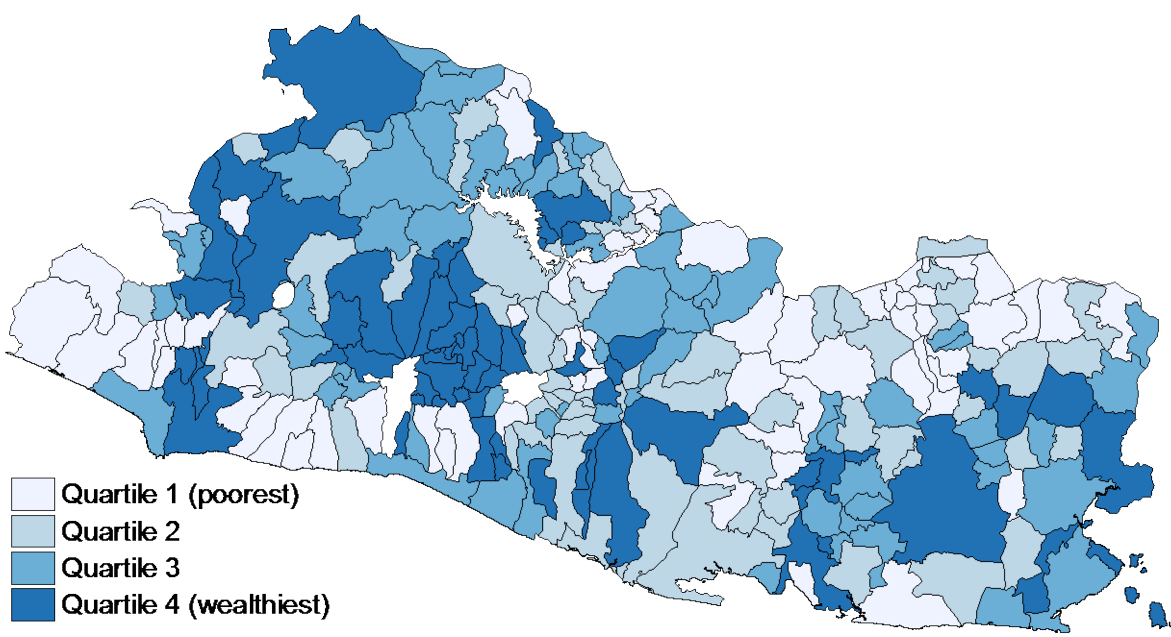


Figure 7: Electrification Rates (Percent), El Salvador 2007.

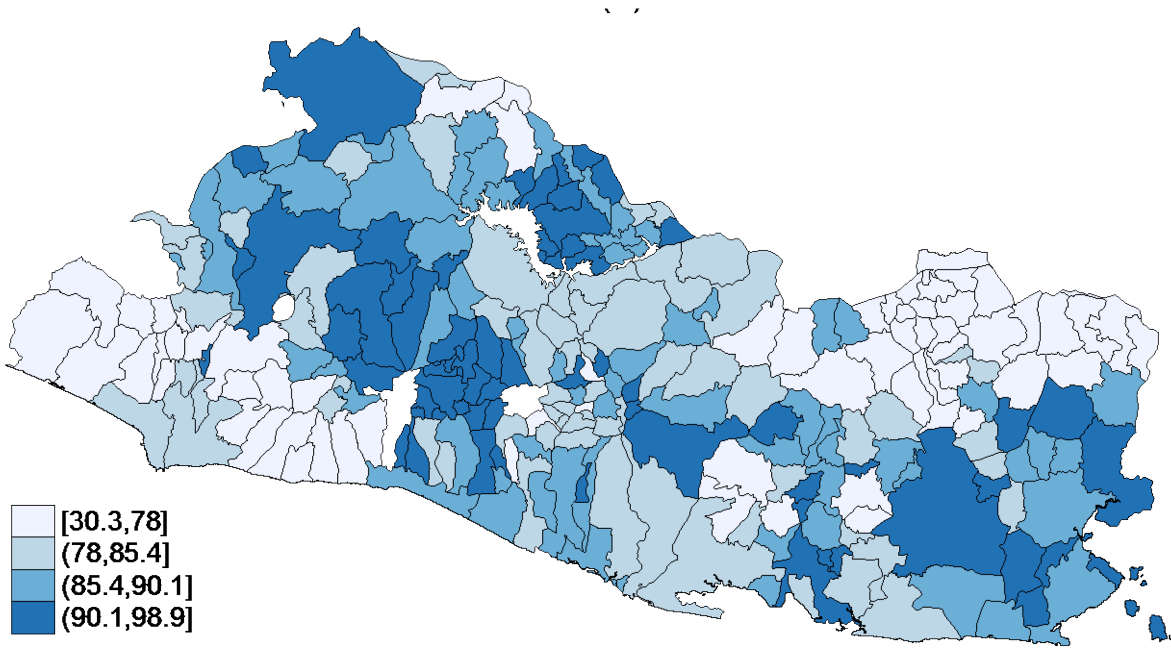
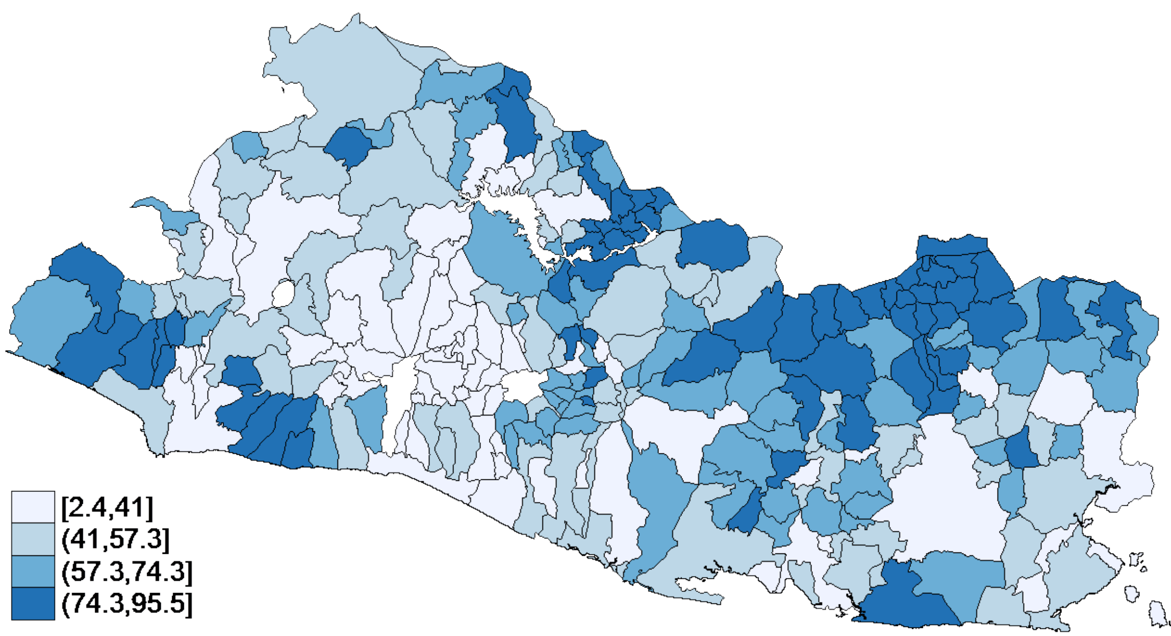


Figure 8: Use of Traditional Fuels for Cooking or Lighting (Percent), El Salvador 2007.



3.2.2 Key Outcomes Linked to Project Logic

It is expected that this project reduced the cost to access high-quality electricity and enabled households to extend their labor activities and diversify their income sources. The overall program goal of the rural electrification sub-activity was to achieve poverty reduction and economic growth in the Northern Zone. One pathway through which this goal could be achieved was increased household productivity as a result of improved health from decreased indoor pollution. The other pathway was through enabling households to allocate time to more productive activities and to engage in off-farm business opportunities. These changes imply income flows that are more diverse and perhaps less volatile, promoting resilience and helping households exit out of poverty.

The first steps in these pathways are related to the project outputs (direct products of the program, such as kilometers of grids constructed, number of households connected, etc.), followed by outcomes or impacts. For example, we can categorize the impacts by the expected time they may take to be realized:

- **Short-term:** changes in access to electricity through increased consumption, lower prices, and use of other energy sources
- **Medium-term**: changes in time allocation, indoor pollution, and productivity
- **Long-term**: changes in income and health outcomes

3.2.2.1 Short-term Impacts

Changes in Access to Electricity: Adoption of Grid Connections

To guide our analysis of the adoption of grid connections, we start from a simple static model with the basic assumption that households adopt an electrical connection if the benefit b_i from connecting is larger than the cost c_i of doing so, subject to their budget constraints. As a result, a household connects if:

$$b > c \text{ , where } b = f(A, E^*, X) \quad \text{and} \quad c = V + F$$

A is the vector of electronic appliances the household expects to acquire once it connects to the grid.[14] This is key in that electricity provides utility gains only through the use of electronic appliances like light bulbs, TV sets, refrigerators, and so on. Put in another way, these appliances are complementary goods to electricity itself. E* is the electrification rate around the household; this is calculated to account for spillover effects as found in Bernard and Torero (2015). These spillover effects may arise from social learning if households learn of the benefits of electrical appliances from their neighbor or from imitation, which could be a way of learning itself or could be a reflection of preferences interactions (“bandwagon effects” or households keeping up with their neighbors; Manski 2000). Finally, X is a vector of household demographic characteristics that may affect the benefits obtained from electrification (e.g. age composition, literacy, etc.).

[14] Some households may already own, for instance, TV sets that are operated with car batteries.

On the other hand, the costs of electrifying a dwelling, once it is within reach of the grid, are a combination of variable and fixed costs. Variable costs cover wiring and installation of light bulbs and sockets, as well as potential upgrades in wall materials. Fixed costs consist of, for example, any fees households must pay as part of the application process, plus costs in time and effort in such a process. We exploited the existence of such fees in our study setting to experimentally vary the cost of connection faced by each household in our sample, de facto creating exogenous variation in F. This exercise also created exogenous variation in F*, the average connection cost faced by the household's neighbors. This cost will change depending on the number of vouchers allocated to the household's neighbors, which in turn generates exogenous variation in E*, the connection rate in the household's vicinity. We examined the role of both sources of variation to study adoption of grid connections.

Before we move forward, we must first consider other variables that may affect adoption of grid connections and their interaction with F and E*. Income is one of the most commonly cited determinants of technology adoption. Richer households are more likely to adopt an electrical connection for two main reasons: first, the marginal disutility of paying F decreases with income, and second, since richer households can expect to buy more electronic appliances that are complementary goods to grid connection, they can extract more utility from a connection. However, the way changes in F and E* interact with income is not straightforward. Let us start with F. There is an income threshold m_1 above which F is trivial, so adoption rates will be unaffected by changes in F if income m is greater than this threshold. Similarly, there is an income threshold m_0 under which virtually any positive F is unaffordable. Hence only households with incomes $m_0<m<m_1$ will respond to changes in F. Similarly, E* will affect households within an income range, but it is not expected to affect either very poor or very rich households.

Second, we discuss the components of V. Wiring costs are an obvious component of V and depend on the number of connections, sockets, light bulbs, etc. as well as the material on the walls. A dwelling cannot be electrified if its walls are made of wood, grass, or metal sheets, for example; wiring can only be done if walls are made of materials like cement, brick, or adobe. Some households will need to upgrade their walling materials in order to pass a safety inspection, which is not a trivial investment.[15] Thus, V includes the time and effort for needed paperwork, hiring contractors, and all activities related to getting an electrical connection. If households exhibit hyperbolic discounting, they can postpone technology adoption indefinitely (Duflo et al 2011). To ameliorate this concern, the validity of the discount had a time limit after the date of issuance, thus providing a salient benefit of taking action before it is too late. V also includes the cost of solving any credit or liquidity constraints faced by the household. If solving liquidity or credit constraints in $t > 1$ periods is cheaper than doing so in one period, V falls with time, so the probability of adoption should increase as time passes, up to its equilibrium level.[16]

[15] On the other hand, a strong desire to get electricity may act as a nudge for households to update their construction materials. With the opportunity for discounted electrification, there is a salient benefit of upgrading while before, there was no such benefit (despite the fact that the actual benefit of upgrading may be high).

[16] The passage of time could also lead to reductions in the number of connections, since some households may

Before the program started, households that wanted a connection had to pay for the grid extension themselves by paying the electric utility for posts, cables, etc. Although there is no official data on these costs, they are presumably high. Once a household became connected to a grid, it could offer informal connections to its neighbors. Informal connections consist of a series of cables connected to a (formally connected) neighbor's outlet.[17] This type of connection is usually enough to reliably operate a couple of light bulbs and, at most, a TV set. It can be argued that households with informal connections value electricity more than those that remain off-grid since they have gone through the trouble of getting this connection in the first place. Additionally, most of these households already own some electronic appliances and thus have higher potential gains from adopting formal electricity than those with no access at all. In the empirical setting, we account for this type of connection with an ordered choice model in which households have the choice between no connection, informal connection, and formal connection, with a formal connection being preferable to an informal connection.

3.2.2.2 Medium-term and Long-term Impacts

Time Allocation

In what follows, we discuss how electrification may lead to changes in time allocation. The basic point is that electricity can increase the marginal return to time spent on any activity, which implies increases in the opportunity cost as well (i.e. the marginal value of time spent on other activities), so the net effect of electrification on time allocated to any given activity is theoretically uncertain. The following sections outline the reasoning behind the changes in the major activities.

Household electrification may open the door to new types of economic activities and shift the status of an activity from non-profitable to profitable, which should result in a reduction in leisure time. However, electrification also facilitates the use of electronic appliances that increase the marginal value of leisure (most notably TV sets, but also stereos and Digital Versatile Disk [DVD] players), which should result in an increase in leisure, making the net effect theoretically uncertain. The effect is further complicated by the fact that, since electronic appliances are not free to buy or operate, households may decide to increase their labor supply because they want to acquire and use these new items, pushing labor supply in the opposite direction and making the net effect on leisure theoretically uncertain.

On the other hand, Banerjee and Duflo (2007) argue that poor households may leave profitable opportunities unexploited because the extra income they could generate would not make a salient impact in their lives, especially after taking into account the effort required to produce this additional income. This suggests that the promise of access to electronic appliances that

---

decide to drop a connection (due to unexpected prices, for example); however, this is not the case in our empirical setting.

[17] Informal connections are not electricity theft, since the utility providing the power is being paid for all the power consumed by the formally connected consumer (who may receive payment by the neighbor who is informally connected).

arguably have salient impacts on well-being, like TV sets and refrigerators, may induce households to pick up some of those pre-existing opportunities by trading leisure time for labor.

Time Studying

In our sample, school-age children (six to14 years old) split most of their time between four types of activities: education, household chores, leisure, and work. Virtually all children have some leisure time at baseline and 67 percent do some sort of household chores. The shares for studying and work are lower, at 22.1 percent and 19.1 percent, respectively.

Education includes time spent studying at home, a key component of educational investment. Electrification radically improves children's study environment, allowing a shift from dimly lit, smoky rooms to well-lit, smokeless rooms. This drastically reduces the effort required to study, encouraging children to spend more time studying. In addition, even if time spent studying is unaffected, a better study environment will likely result in better learning, raising the returns to education. The increase in the returns to education may in turn make parents more willing to incur the cost of sending children to school since they are more likely to perceive their children learning more. In consequence, we could observe improvements in other educational variables like school absenteeism or even enrollment.

On the other hand, electrification also facilitates access to TV, shifting up the marginal value of leisure. This is further complicated by the fact that children may change their participation in chores or labor to compensate for time allocation changes among adult household members. Hence, the net effect on time spent studying is theoretically uncertain. The theoretical ambiguity of the net effect is consistent with respondents' perceptions: at baseline, 90 percent of parents said their children would study more with electricity, but at the same time, 55 percent of parents said their children would waste their time watching TV.

Labor Supply and Household Chores

Electricity can increase the marginal productivity of labor: self-employed workers may now use power tools, farmers may use electric water pumps, and shopkeepers may offer refrigerated goods. This should shift out the labor supply curve, but given the increase in the marginal utility of leisure, the net effect is again theoretically uncertain.[18]

It is important to note that all these changes require access to capital goods (tools, pumps, and refrigerators), which require the use of savings, access to credit, or other sources of financing. They also require complementary inputs, human capital, and demand for the products or services produced. If credit and insurance markets are imperfect, as they seem to be in our empirical setting, better-off households should be more likely to start new activities than poorer households, since the former arguably have better access to resources and are more capable of assuming risks. Even if all households are equally constrained, we may still see some changes in

[18] The net effect of an increase in labor productivity on labor supply is uncertain even without an increase in the marginal value of leisure since it depends on the relative size of the income vs. substitution effect. However, in poor households, the substitution effect is arguably larger than the income effect, so the increase in marginal labor productivity should push out the labor supply curve.

income-generating activities, but at a small scale. For instance, some activities could evolve from being solely household chores to being complementary income-generating activities, such as food preparation, ironing, and washing clothes. Although these complementary income activities may not have sizable impacts on overall income, they might be appealing to households because they do not require a large monetary investment or use of male head's time (in our sample, male heads usually provide a larger share of income than female heads).

3.3 Methodology: General Approach Applied in the Electrification Project

The experimental sample consisted of 500 households located in sub-districts that were scheduled to benefit from the electrification project during its first year. The non-experimental sample was formed by households from the remaining sub-districts. We generated experimental variation in the connection fee by offering discount vouchers to a randomly selected subsample. We randomly allocated 200 low-discount vouchers (20 percent discount) and 200 high-discount vouchers (50 percent discount) and left the remaining households as a control group ($N = 100$). The exogenous variation in the connection fee generated by the random voucher allocation allowed us to deal with self-selection in connection to the grid. Vouchers were valid for a discount toward the safety certification to be reimbursed after households paid the full cost. Vouchers had the name and address of the beneficiary printed on them, were non-transferable, and were valid for nine months.

Random voucher allocation also created exogenous variation in the number of voucher recipients in a given neighborhood of household i (controlling for the number of eligible neighbors). This generated variation in the number of new connections around household i, which allows us to control for the role of spillovers on grid connection. The sign of the effect is theoretically ambiguous. On the one hand, observing their neighbors connect to the grid may make households more prone to connect through a combination of social learning and imitation effects.[19] On the other, higher formal connection rates in a neighborhood reduce the cost of getting an informal connection, so the number of vouchers around a household may increase the number of informal connections. To estimate the role of spillovers on adoption, we used the number of household $i's$ neighbors that received a voucher in a given radius (0-100 meters, 100-200 meters, 200-300 meters), controlling by the number of eligible neighbors in that radius. Eligible households are households with no electricity at baseline.

We estimated the demand for formal connections by exploiting exogenous variation in the connection fee generated by the random allocation of discount vouchers among a subsample of 500 households. We used two variables: a household's individual discount (individual discount) and the share of that household's neighbors that received a discount (density). Individual discount and density could be used as instruments for connection in the first stage of a two-stage least squares (2SLS) model. In the second stage, the outcome variables were regressed against the instrumented value of connection (from the first stage), time FE, and individual FE.[20]

In addition, we took advantage of the longitudinal nature of the data and the multiple survey waves to exploit within-household time variation and to examine the effects of electricity by

---

[19] Social learning would occur if households observed the private benefits of electrification (better illumination, less smoke at night, better food availability, more enjoyable leisure time) from their neighbors. Imitation effects (also known as "preferences interactions" in the literature) are similar to a "keeping-up with the Jones'" story: a household wants electricity because its neighbors have it.

[20] Some studies in the field used geographic variables, like land gradient to instrument for grid placement and assume that all households in the vicinity of the grid will connect. As previously discussed, this has many disadvantages in our setting, since in the context under analysis, many households have the grid available but do not connect to it. In addition, average land gradient turned out to be a weak instrument in this sample.

using FE panel data models. This methodology controlled for any time-invariant unobservable characteristics but would produce biased estimates of the effects of electrification if time-variant unobservable determinants of the outcomes were also correlated with connection to the grid.

3.3.1 Experimental Sample: Reduced Form and Local Average Treatment Effects

Estimation Strategy

Our main estimating equation for the adoption analysis is given by:

$$conn_{it} = \beta_0 + \beta_1 F_1 + \beta_2 E^*_{it} + \beta_3 X_i + \lambda_t + \varepsilon_{it}$$

To include household FE and still be able to estimate β_1, we followed alternative methodologies. (i) We created a *post* dummy that took the value of one in periods two through five, and interacted all the explanatory variables with *post*. (ii) We interacted each explanatory variable with the round FE, thus allowing for different effects in different periods without imposing a linear or quadratic trend. (iii) We interacted the explanatory variables with t and t^2. To allow for non-linear effects of F on connection, we also replaced F by dummies corresponding to the 20 USD and 50 USD discounts.

Since E* is endogenous, we employed two strategies to find its causal effect on connectivity. First, we used the average fee among household i's neighbors as an instrument for E*. Second, we replaced E* by the share of eligible neighbors around household i that received a discount voucher, s*. This estimation gave the "reduced form" coefficient.

$$conn_{it} = \beta_1 Voucher_i + \gamma^{100} V_i^{0-100} + \lambda^{100} N_i^{0-100} + \omega' X_i + v_t + \varepsilon_{it}$$

where *conn* indicates whether household i has a formal connection in year t, *Voucher* indicates whether the individual received a voucher, $V^{0\text{-}100}$ indicates the number of households that received a voucher within a 100-meter radius of household i, while $N^{0\text{-}100}$, the total number of eligible households around a 100-meter radius of household i. As before, X denotes individual baseline characteristics. v_t are year FE included because the follow-up period includes multiple years; we collected up to three yearly measurements per observation in years three through four of the study. ε is the unobserved residual. As is usual in this approach, the standard errors were clustered at the level of treatment, i.e., the household level.

Note that including the share of connections as an explanatory variable for $i's$ decision to connect would be inadequate since, if spillovers play an important role in adoption, this share would also depend upon $i's$ decision. Hence, we only present this "reduced form" type of results, with average cost (or number of encouraged neighbors) as explanatory variables.

In an alternative specification, we allowed for different effects due to size of the discount by estimating:

$$conn_{it} = \beta_{20} V20_i + \beta_{50} V50_i + \gamma^{100} V_i^{0-100} + \lambda^{100} N_i^{0-100} + \omega' X_i + v_t + \varepsilon_{it}$$

Third, we estimated the effect of the amount of the fee, as well as the average fee in the neighborhood of household i. This imposed some structure in the regression, namely, that discounts affect costs linearly and, in turn, costs affect connection rates linearly.

$$conn_{it} = \beta Fee_i + \gamma^{100} Fee_i^{0-100} + \omega' X_i + v_t + \varepsilon_{it}$$

The results from the above regressions give the effects of discount vouchers on adoption averaged over the three follow-up surveys. To analyze the dynamic effects, we interacted the variables on the right-hand side with time dummies.

Informal Connections

A second issue we considered is the existence of informal connections. To take into account the “essential heterogeneity" (Heckman et al 2006) at baseline between households with no connection and those with an informal connection, the above equations could be run separately for households with an informal connection and for those with no connection at baseline, or we could fully interact each variable in the right-hand-side with a dummy for informal connection at baseline.

We studied switching patterns between three connection types: no connection, informal connection, and formal connection. Reductions in F unequivocally increased the probability of household i getting a formal connection to the grid, but reductions in E* generated two opposing forces: on the one hand, imitation-type spillovers would increase the probability of getting a formal connection, while on the other hand, the more neighbors of i are formally connected to the grid, the easier it is for i to get an informal connection.

We studied the probability of switching among different alternatives. For instance, we looked at the probability of a household switching from no electricity to formal electricity at some point over the study period, ignoring households that had informal electricity at baseline. We conducted a similar analysis for the probability of switching from no electricity to informal electricity and from informal to formal connections. While this has the disadvantage of leaving out part of the sample, it provided robust insights without imposing structure on the model.

Alternatively, we exploited the natural ordering among these three types and estimated an ordered probit model. In this setting, formal connections ($y = 2$) are the best type of connection, followed by informal connections ($y = 1$), and finally by no connection ($y = 0$). Ordered choice models start by assuming that households choose the connection type that maximizes their utility level. Subjects will choose their connection type y depending on the value of a latent, unobservable variable y* such that:

$$y^* = \beta X_i + \varepsilon_i$$

where x is a vector of explanatory variables. Although y* is unobservable, we observed y, the subject's choice as:

$$y = 0 \quad \text{if} \quad y^* < \alpha_1$$
$$y = 1 \quad \text{if} \quad \alpha_1 < y^* < \alpha_2$$
$$y = 2 \quad \text{if} \quad y^* > \alpha_2$$

for some (estimable) cut-off parameters α. If we assume that ε follows a normal distribution, we get to the ordered probit model. In this model, the probability of choosing each alternative is given by:

$$\Pr(y = 0) = \Phi(\alpha_1 - x_i\beta)$$

$$\Pr(y = 1) = \Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta)$$

$$\Pr(y = 2) = \Phi(x_i\beta - \alpha_2)$$

In this model, the marginal effects are given by:

$$\frac{\partial \Pr(y = 0|x_i)}{\partial x_k} = \beta_k \phi(\alpha_1 - x_i\beta)$$

$$\frac{\partial \Pr(y = 1|x_i)}{\partial x_k} = \beta_k [\phi(\alpha_1 - x_i\beta) - \phi(\alpha_2 - x_i\beta)]$$

$$\frac{\partial \Pr(y = 2|x_i)}{\partial x_k} = \beta_k (\alpha_2 - x_i\beta)$$

where $P(y = j|x_i)$, j=0,1,2 denote the probability of choosing no connection, informal connection, or formal connection, respectively. Further, note that the direction of the effect of x_k on $P(y = 0|x_i)$ and $P(y = 2|x_i)$ is given by the sign of β_k, but this is not true for $P\ (y = 1|x_i)$ since the effect depends on the sign of $\phi(\alpha_1 - x_i\beta) - \phi(\alpha_2 - x_i\beta)$. Note that the marginal effects for each observation depend on their particular x_i, so they will vary between observations. Thus, the effect may be positive for some households and negative for others. This flexibility is especially interesting for us, since it is theoretically ambiguous whether spillovers will increase or decrease informal connections.

Effects of Electrification on Indoor Air Pollution Concentration

To study the effects of gaining access to electricity, we exploited the experimental variation in connection fees described in the preceding section to instrument for connection to the grid. The second stage is given by:

$$y_{ithsm} = \delta \widehat{conn}_{it} + \psi X_{i0} + hour_h + subdistrict_s + monitor_m + \varepsilon_{it}$$

where y_{it} indicates the outcome of interest (measures of PM2.5 concentration), X_{i0} includes baseline covariates, as well as sub-district, hour of the day, and monitor FE, while ε_{it} is a disturbance term. δ is the main coefficient of interest, as it gives the causal effect of connection on the outcome for the population of compliers. The complier subpopulation may be small because either (i) there is small take-up among the encouraged group or (ii) there is large take-up among the non-encouraged group. In our case, the small complier subpopulation is due to large take-up in the non-encouraged group, especially in rounds four and five.

Due to a small complier subpopulation, the IV point estimates were noisy, which generated large standard errors. To avoid relying on noisy estimates, our main results were based on the reduced form estimates. Effectively, these estimates reported the effect of receiving a voucher on the outcome of interest; as such, these estimates were informative and valuable from a policy perspective. In addition, note that given imperfect compliance, these estimates represented a lower bound of the effects of electrification.

The reduced form is given by:

$$y_{ishm} = \beta voucher_i + \omega X_{i0} + hour_h + subdistrict_s + monitor_m + \varepsilon_i$$

We included hour-of-the-day and sub-district FE. In addition, we added PM2.5 monitor FE to control for potential measurement error in the equipment.

We estimated two variants of the reduced form equation. First, we exploited the minute-by-minute nature of the data. To allow for arbitrary structure in the covariance matrix within a household, we clustered the standard errors at the household level. Second, we collapsed the data at the household level and re-estimate the models (in the spirit of a "between" estimator) to show that the significance of the coefficients of the voucher variables ($\beta's$) was not driven by the large number of observations.[21] Given that the first-stage analysis showed a positive and significant relationship between voucher allocation and grid connection, we found it difficult to argue against electrification being the channel through which vouchers affect PM2.5 concentration.

Effects of Electrification on Other Outcome Variables

In this section, we describe the econometric approach on which our empirical estimates are based. Our main specifications are an IV estimation and the corresponding reduced form estimation.

For the IV estimation, the first stage regression is given, as before, by:

$$conn_{it} = \alpha_0 + \beta_1 voucher_i \times Post_t + \beta_2 s100_i \times Post_t + \lambda_i + t + u_{it}$$

where $conn_{it}$ is the connection status of household i at time t. Our main connection measure was simply a connection indicator that takes the value of one if the household has a formal connection to the grid and 0 otherwise. As robustness checks, we employed time connected to the grid and having a grid connection for at least k years, with k running from one to four (given that we have four follow-up rounds). The results for these variables were strongly consistent with the results for our main connection variable.

[21] The difference between this procedure and the "between" estimator is that the latter requires collapsing the minute-by-minute data from all the rounds to a single point for each household and estimating the effect based on the cross-sectional variation in the resulting sample.

The second stage is given by:

$$y_{it} = \beta_0 + \delta\widehat{conn}_{it} + \lambda_i + \mu_t + \varepsilon_{it}$$

As before, the reduced form provides the intent to treat estimates:

$$y_{it} = \beta_0 + \gamma_1 voucher_i \times Post_t + \gamma_2 s100_i \times Post_t + \lambda_i + \mu_t + \varepsilon_{it}$$

y_{it} is the outcome of interest, *Voucher* takes the value of one if the individual received a discount voucher and 0 otherwise, *Post* takes the value of one for all the follow-up rounds and 0 at baseline, *S100* is the share of eligible neighbors in 100 meters that received a voucher, μ_t are time FE, and λ_i denotes individual FE (or household-level FE in specifications in which the unit of observation is the household). Since vouchers were allocated at the household level, standard errors were clustered at the household level (in the household- and individual-level regressions).

Due to random allocation, *Voucher* and *S100* are uncorrelated with u_{it} in the first stage equation. Under the assumption that vouchers affected the outcome variables only through their effect on the probability of connection, IV should render estimates consistent with the true effects of electrification for the compliers.

Instruments with low correlation between the endogenous regressors are called weak instruments. There is empirical and theoretical evidence that IV estimation with weak instruments may perform poorly, even more poorly than OLS (surveyed in Stock, Wright and Yogo 2002). The relevance of the instruments was tested in the first stage regression. As a rule of thumb, the F-statistic of a joint test, whether all excluded instruments (the variables in the first stage which are not in in the set of regressors of the second stage) were significant, should be bigger than 10 in case of a single endogenous regressor. This F-Test was reported when reporting IV estimates. We note that the adoption regressions discussed above represent a first stage regression and a measure to gauge the suitability of IV estimation. The randomization of the vouchers gave strong evidence of the validity of the instrument, and the adoption regressions show us the relevance of the voucher as an instrument. The additional information provided by the F-test was related to the strength of the correlation in the adoption regression.

There are two features of the data worth highlighting in this setting. First, a key outcome of this paper was time allocation, which is prone to corner solutions. For example, time allocated to activities like studying or working was zero for sizable shares of our sample. If electrification affected the probability of participation in a particular activity and the time spent in such activity, neglecting the selection process will produce biased coefficients even in a randomized control trial,[22] so it would be necessary to deal with selection. Parametric models that deal with selection require establishing strong parametric assumptions about the distribution of the residual term. Non-parametric models, on the other hand, make no such assumptions but require imposing an exclusion restriction: they need a continuous variable that affects participation in an activity but

[22] If an explanatory variable affects the outcome and the probability that the outcome is greater than zero, ignoring selection will render biased estimates even if said explanatory variable is randomly allocated. For instance, in a tobit model: $E[y|x, y > 0] = x\beta + \sigma \frac{\phi(x\beta/\sigma)}{\Phi(x\beta/\sigma)}$

not the amount of time spent in such activity. These types of questions are better suited for a structural approach, which we leave for future research.

Second, in some cases, the outcome variables of interest may be thought of as belonging in a system, for instance, the time allocation to different activities or ownership of different types of appliances. To account for this, we estimated seemingly unrelated regressions (SUR). This methodology took into account the correlation in the disturbance terms across different equations. An additional advantage is that SUR models indicate whether outcomes are pairwise substitutes or complements, thus providing deeper economic insight. The equation-by-equation IV results were qualitatively similar to the results from SUR systems.

3.3.2 Non-Experimental Sample and Fixed Effects Estimation

Indoor Air Pollution

We present non-experimental estimates of the effects of electrification. To do so, we exploited the longitudinal nature of the data to explore the effects of electrification on the outcomes of interest outside the experimental sample. Although there is no exogenous variation in this case, the inclusion of household (or individual) FE allowed us to control for all time-invariant characteristics that affect the outcome variables and that may be correlated with the decision to connect to the grid.

In particular, our non-experimental estimates provided further supportive evidence that the channel through which voucher allocation causally affects PM2.5 concentration is in fact household electrification. For this purpose, we used a sample of approximately 250 *EHEIPCER* households that had not connected to the grid by round two. The first PM2.5 measurement was conducted in round two, and follow-up measurements were taken in rounds three and four.

$$y_{it} = \delta \times connected_{it} + \lambda_t + \mu_i + \varepsilon_{it}$$

where y_{it} is the outcome of interest for household i at time t, $connected_{it}$ indicates whether the household had a connection to the grid, λ_t captures round FE, and μ_i captures household FE. Causal identification of the parameter of interest δ in this setting requires assuming that $connected_{it}$ is uncorrelated with the disturbance term ε_{it} after controlling for time-invariant characteristics μ_i. Since ε_{it} is unobservable, this can never be tested directly, but one way to show supportive, albeit indirect, evidence on the validity of this assumption is showing that the outcome variable of interest followed the same time trend between connected and off-grid households before the former group got a connection.

3.3.3 Validity of the Assumptions

In the experimental sample, the validity of the results required that the random allocation of vouchers was implemented successfully. Table 8 in section 9 shows that average characteristics balanced across treatment arms. The first column shows the means for the control group, the second column shows the means for the households that received a 20 percent discount, and the fourth column shows the means for households that received a 50 percent discount. The third and fifth columns test for differences between each of the treatment arms and the control group.

With a few exceptions, the differences were not statistically significant, indicating that randomization was implemented successfully.

Households were also balanced regarding their ex-ante perceptions of energy sources. The vast majority agreed that electricity illuminates better than kerosene (96 percent) and that wood smoke generates respiratory problems (87 percent). On the other hand, 30-40 percent of respondents said that kerosene is not an expensive source of lighting, and 20-30 percent said it was the best way to illuminate their household.

In the non-experimental sample, the underlying identification assumption is that outcomes among off-grid households are a good counterfactual for outcomes among on-grid households. This can be tested indirectly by comparing trends in outcomes before electrification; this discussion is presented with the results, and Table 7 provides some evidence to this respect.

3.4 Population Studied

This study took place during a recent grid extension and intensification project in Northern El Salvador, designed to be rolled out in three phases in accordance with construction costs and accessibility.

Conscious of budget limitations, we proposed the study of only two departments: Chalatenango and San Miguel. These departments were selected because, according to the electrification project plan, they included the largest numbers of cantons that would benefit from the electrification program. In addition, these districts include a number of cantons that will benefit from the road improvement and the electrification programs.

Table 7 in the Annex presents the main descriptive statistics of our study samples. The first column reports results for a representative sample of beneficiary households. The second column describes the voucher subsample. The third and fourth columns describe the experimental and non-experimental air quality subsamples. Households in the voucher subsample have socioeconomic characteristics that are similar to the whole sample. Household heads are on average 49 years old; 68 percent of them are males and have 2.2 years of schooling on average. Average age in the household is 21.2, average household size is 4.3 people, and the total dependency ratio (the number of non-working-age household members divided by the number of working-age household members) is 0.44. Maximum schooling in the household is around 5.5 years, and roughly half the household heads are literate. Table 10 provides further details on energy consumption, but kerosene is the main energy expenditure (2.5 USD/month) and 70 percent of households use wood for cooking.

There were some differences between the whole sample, which is representative of all the program beneficiaries, and the voucher subsample. For instance, the latter had slightly higher mean annual income, at 770 USD, compared to 650 USD among the whole sample. Households in the experimental air quality subsample are representative of the voucher subsample and, as such, there are no major differences between those groups. However, the non-experimental air quality subsample is formed by households with lower socioeconomic status. This is due to the selection criteria for this sample, which included remaining off-grid during the first two years of the study.

Regarding household perceptions, most households reported that electricity provides better illumination than kerosene, that powering a TV is cheaper with electricity than with a car battery, and that wood smoke generates respiratory problems. Around 60 percent reported that cooking with electricity is not convenient, and roughly half of the households reported that electricity is very expensive. One-third of the households reported that kerosene is not an expensive source of lighting, and 22 percent reported that it is the best way to illuminate their households.

3.4.1 Power calculations and sample size requirements

The baseline household survey was designed using the 2007 Population Census as the sampling framework. This first survey was collected in November and December 2009. It covered 4,800 households all over Northern El Salvador. Four follow-up surveys were collected in the same months in 2010, 2011, 2012, and 2013, respectively.[23]

Following the procedure detailed in Appendix 1, we calculated the minimum sample size for each department (Chalatenango and San Miguel). The power analysis required a total sample of 1,532 households to detect a 20 percent difference in household income between treatment and control groups with 80 percent power and a 95 percent confidence level. The parameters for the sample design were obtained from the 2007 National Household Survey (NHS) (Encuesta de Hogares de Propósitos Múltiples [EHPM] or Multipurpose Household Survey) and various scenarios for the variance of the outcomes were considered to account for the panel nature of the proposed survey. The outcomes that were used as targets were: total income, agricultural and non-agricultural wage and non-wage income, and labor allocated to agricultural and non-agricultural wage and non-wage activities.

The results are presented in Table 5. The scenarios refer to the variance used for the outcomes. First and most conservatively, we simply doubled the variance of the estimates from the 2007 national household survey outcomes; doing so assumed that the primary outcome would not be correlated across the surveys, that each stratum would have exactly the same mean outcome, and that the treatment will not affect the variance of the treatment. In the second scenario, we reduce the doubled variance by 10 percent to simulate a significant decline in sample variance due to stratification. In the third scenario, we simply computed the power calculations using the estimated variance. Finally, we used the estimated variance less 10 percent to account for gains from stratification, but we also assumed a between-period correlation of 0.5 and a within-period correlation of -0.5. Since the baseline variance in outcomes was likely to be smaller than the NHS variance, the fourth estimate was likely to be the most realistic. Thus, that is the one that was used in the sample recommendations with 45 households per cluster.

We present the results of the sample design using other outcome variables in Table 6. Due to the high intracluster correlations observed in variables such as non-agricultural wage income or time allocated to non-agricultural non-wage labor, the power to detect differences in those variables will be lower, although it may still be possible.

---

[23] The surveys use the same instruments as the survey used to evaluate the impact of MCC's investments in the connectivity project.

Table 5: Number of Clusters per Condition and Total Sample Size for Household Income for Each Scenario

		Scenario 1		**Scenario 2**		**Scenario 3**		**Scenario 4**	
		$2\sigma^2$		$1.8\sigma^2$		σ^2		$0.9\sigma^2$	
	Intracluster correlation[6]	**Clusters per condition**	**Total sample size**	**Clusters per condition**	**Total sample size**	**Clusters per condition**	**Total sample size**	**Clusters per condition**	**Total sample size**
Cluster Size=25[5]									
Chalatenango	0.030	41	2027	36	1824	20	1014	15	757
San Miguel	0.073	96	4799	86	4319	48	2399	16	816
Cluster Size=35									
Chalatenango	0.030	31	2147	28	1933	15	1074	11	744
San Miguel	0.073	87	6060	78	5454	43	3030	11	802
Cluster Size=45									
Chalatenango	0.030	25	2281	23	2053	13	1140	8	737
San Miguel	0.073	81	7334	73	6600	41	3667	9	796

Notes: 1 The conditions are “treatment” and “control”. The number of clusters in each condition is equal.
2 Total sample size (treatment + control)
3 The outcome variable is total monthly household income
4 For the specification of each scenario see text, and for the formulae, see Appendix 1
5 Number of observations (households) per cluster
6 Observed in the NHS at the department level
7 $\alpha = 0.05; \beta = 0.20; \Delta = 0.20$

Table 6: Sample Design Results for Other Outcome Variables: Number of Clusters

	25 observations per cluster				35 observations per cluster				45 observations per cluster			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Chalatenango												
Total Income	41	36	20	15	31	28	15	11	25	23	13	8
Agricultural Wage Income	19	17	10	9	13	12	7	6	10	9	5	5
Agricultural Non-wage Income	33	30	17	9	28	25	14	6	25	22	12	5
Non-agricultural Wage Income	10	9	5	5	7	7	4	3	6	5	3	3
Non-agricultural Non-wage Income	36	33	18	8	31	28	16	6	29	26	14	5
Total Labor Hours	27	24	13	10	21	18	10	7	17	15	9	5
Agricultural Wage Hours	108	97	54	51	76	68	38	36	59	53	29	28
Agricultural Non-wage Hours	94	84	47	22	80	72	40	15	73	66	37	12
Non-agricultural Wage Hours	307	276	153	37	288	259	144	26	277	250	139	20
Non-agricultural Non-wage Hours	138	124	69	64	97	87	49	45	75	67	37	35
San Miguel												
Total income	96	86	48	16	87	78	43	11	81	73	41	9
Agricultural Wage Income	25	23	13	9	19	17	10	6	16	15	8	5
Agricultural Non-wage Income	48	43	24	10	41	37	21	7	38	34	19	6
Non-agricultural Wage Income	14	13	7	7	10	9	5	5	8	7	4	4
Non-agricultural Non-wage Income	23	21	12	6	19	17	10	4	17	16	9	3
Total Labor Hours	15	13	7	7	10	9	5	5	8	7	4	4
Agricultural Wage Hours	57	52	29	26	40	36	20	18	31	28	15	14
Agricultural Non-wage Hours	137	123	69	63	96	87	48	44	74	67	37	34
Non-agricultural Wage Hours	107	96	53	18	96	87	48	13	90	81	45	10
Non-agricultural Non-wage Hours	420	378	210	100	359	323	179	70	326	293	163	54

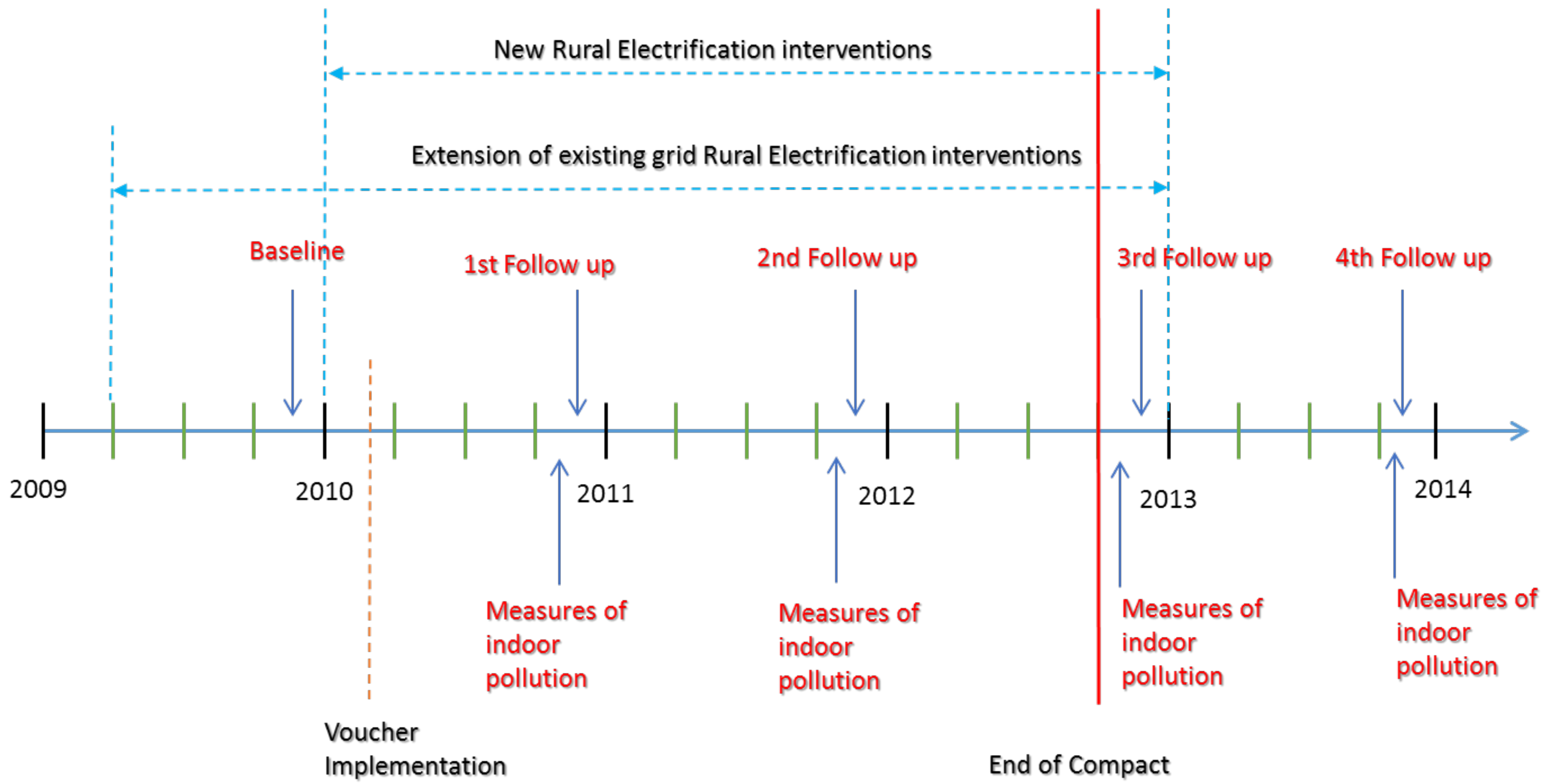
3.5 Timeframe

The timing of the household survey activities was as follows:

- Five household survey rounds[36] in November – December in years 2009 (baseline), 2010, 2011, 2012, and 2013
- 4,800 surveys in the baseline survey that were interviewed in each follow-up survey
- The content and length of the survey was roughly the same in each year, between an hour and an hour and a half.

Concurrent with the household survey, we measured the level of air pollution inside the home in a subsample of households. Figure 9 shows the timeline of activities that led to this impact evaluation.

Figure 9: Timeline of Data Collection and Compact Activities



3.6 Justification for Proposed Exposure Period to Treatment

As discussed in Section 3, very little is known about the effects of electrical access. Even less is known about the timing of these effects or how permanent they are over time. For this reason, it is important to have multiple follow-up surveys. Having multiple follow-up surveys also allows us to shed light on the timing of the effects of electrification. Households need to pay for connection costs as well as for domestic appliances. Saving up for these expenditures will likely take time. In addition, households need to assess information about which appliances to buy; some purchases may be designed to improve leisure, while others will increase productivity. Deciding between opening a shop, starting an ironing service, or enjoying more leisure takes time. Even after the decision is made, as time passes, households may keep changing their resource allocation as their new businesses become more or less profitable.

On the other hand, multiple follow-up surveys also allow us to show that some of the effects of electrification, such as improved indoor air quality, do not fade over time. This is especially important in the case of improved indoor air quality. Some studies, for instance, Hana, Duflo, and Greenstone (2012), show that IAP decreased as a consequence of a stove improvement program, but households bounced back to their original pollution levels soon after.

The distribution over time of the number of beneficiaries connected to the new distribution lines (Figure 10) and the extended electrical system is shown in Figure 11. The beneficiary households were connected to the grid at different dates. By the time we administered each survey, the households had been exposed to the treatment for different lengths of time; some had as much as three years of exposure and others as little as three months. As can be seen in the figure, the majority of the households had been connected to the grid more than two years when the endline survey was administered in November 2013.

Figure 10: Distribution of Beneficiaries of New Distribution Lines

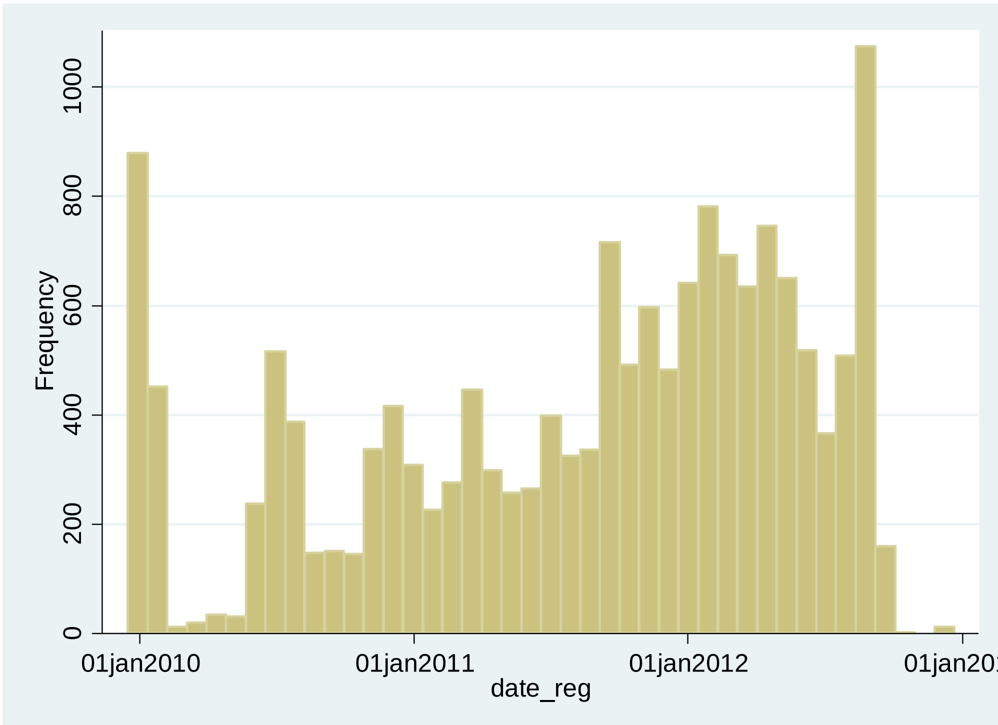
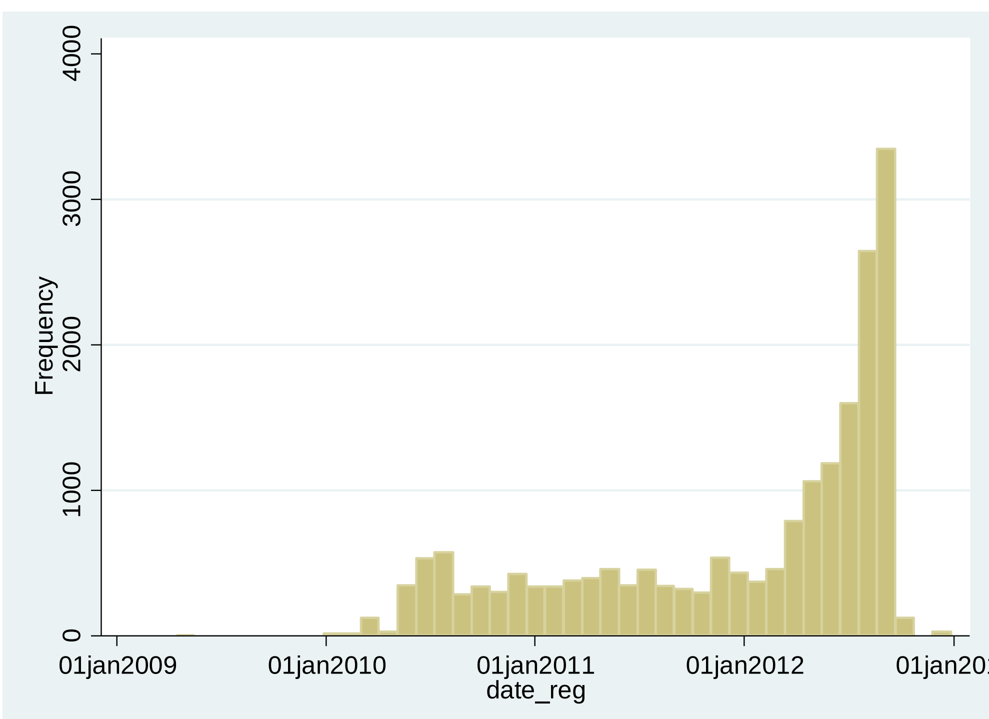


Figure 11: Distribution of Beneficiaries of Extended/Existing Distribution Lines



4 Data

4.1 Description of Databases Collected

The fee for the safety certification is approximately 100 USD. This is non-trivial for households, amounting to roughly 20 percent of annual per capita income in our sample.

EHEIPCER, the household survey implemented for this impact evaluation, is a standard survey that collected data on demographic characteristics, health, education, housing characteristics, energy use, income, and consumption, among other factors. In particular, it includes a detailed module on time allocation for up to four household members: the male head, the female head, and up to two school-age children. Strict training sessions were conducted to ensure high quality in data collection, which was conducted with handheld computers. Enumerators were trained and selected by the authors with the assistance of the Dirección General de Estadística y Censos (DIGESTYC) and International Food Policy Research Institute (IFPRI) staff, in Spanish. The IAP data described below was collected by a subset of enumerators who underwent additional special training to this end.

As described in previous sections, the baseline household survey was designed using the 2007 Population Census as the sampling framework. This survey was collected in November and December 2009 and covered 4,800 households across the Northern Zone of El Salvador. Four follow-up surveys were collected in the same months in 2010, 2011, 2012, and 2013, respectively.

Fine Particulate Matter (PM2.5)

One of the outcomes of interest in this study is PM2.5, which is particulate matter with a diameter of 2.5 microns or less (1 micron = 0.001 millimeters). Particulate matter, also known as particle pollution, is a complex mixture of small particles and liquid droplets composed of potentially hundreds of chemicals. Given the complexity of their composition, particles are mainly classified according to their size. Particles with a diameter between 10 and 2.5 microns are also known as “coarse particulate matter" or “inhalable coarse particulate matter" (PM10), while particulates with a diameter under 2.5 microns are known as “fine particulate matter" (PM2.5). As a reference, a human hair has a diameter of between 50 to 70 microns, 20 to 30 times larger than the cut-off point for PM2.5.

Particle size is inversely linked to its potential for causing health problems. Both PM10 and PM2.5 can pass through the throat and nose and enter the lungs; however, being smaller, PM2.5 can get deeper into the lungs and can also enter the bloodstream, thus causing more health damage than PM10. Both PM10 and PM2.5 have been shown to cause or aggravate heart and lung diseases. Further, there is evidence that states that they weaken the immune system, making the body more vulnerable to disease in general and negatively affecting cognitive ability and productivity.[24]

[24] Other than by size, particles that compose particulate matter are usually classified as primary or secondary

PM2.5 Measurement

A central part of this evaluation consisted of collecting data on overnight PM2.5 concentration. We obtained PM2.5 measurements in two subsamples of households, one experimental and one non-experimental. The experimental subsample included PM2.5 data on 141 randomly selected households from the 500 households that were considered for voucher allocation. The reasons for not selecting the whole sample were logistical and budgetary. Measurements for these households were collected in rounds three and four of the household survey. The non-experimental subsample consisted of 200 EHEIPCER households from neighboring sub-districts in the same departments as the experimental sample (San Miguel and Chalatenango) that had not connected to the grid by September 2010. Measurements in these households were collected with rounds two, three, and four of the household survey. The non-experimental sample was made up of households that had not connected to the grid by round two. Descriptive statistics for both subsamples are reported in Table 7.

In each household, we measured minute-by-minute PM2.5 concentration between 5:00 pm and 7:00 am the next morning in the main evening living area, which is defined as the room in which household members spent most of their time during the evenings while awake. In the majority of cases, this was the living room. Measurements were conducted with the University of California at Berkeley Particle and Temperature Sensor (UCB-PATS). The UCB-PATS is a small, portable, non-intrusive data-logging particle monitor for indoor environments. It uses a photoelectric detector to measure PM2.5 concentrations down to 25 milligrams/m^3. The UCB-PATS records PM2.5 concentration, relative humidity, and temperature at a one-minute time resolution. For details on the development and performance of the UCB-PATS, see Litton et al (2004), Edwards et al (2006), and Chowdhury et al. (2007).

Experienced and meticulously trained enumerators visited the selected households, explained the purpose of the study, and obtained consent to place the UCB-PATS in the home. The protocol implemented to measure PM2.5 concentration is similar to the protocol applied by Northcross et al. (2010) for cook stoves, which is a standard protocol in the cook stove literature. There is no standard protocol in place for measuring IAP emitted by kerosene lanterns. The monitor was placed in the room in which most household members spent most of their time awake during the evenings. For most households, this was the living room but in a handful of cases, it was the master bedroom. The monitor was placed on a wall one meter (horizontally) away from the place where the lamp is usually located in the evenings, at least 1.50 meters away from any working doors or windows, and at a height of 1.50 meters above the ground.[25]

---

particles. Primary particles are emitted directly by a source, like kerosene lamps, cook stoves, unpaved roads, or construction sites. The outcome of interest in this paper falls in this category. Secondary particles, on the other hand, are formed in the atmosphere as a result of sulfur dioxides and nitrogen oxides emitted from power plants, industries, and automobiles. Secondary particles account for most of the particulate matter in developed countries, while the converse is arguably true in developing regions

[25] Studies on the RESPIRE randomized trial in neighboring Guatemala use similar heights, usually ranging from 1.45 to 1.50 meters; see, e.g., Northcross et al. (2010).

In the baseline measurement, enumerators took pictures of the placement to ensure that the monitor would be set up in the same place in the follow-up visits. This reduced the risk of generating artificial variation in PM2.5 concentrations by placing the monitor in different locations.

In follow-up measurements, the enumerators used pictures from previous rounds to place the monitors in the same place as the baseline measurement. The enumerators then filled out a data sheet with exact details of the height, distance, set-up time, and pick-up time, among other indicators. The monitors were placed in the homes before 4:00 pm. If the monitor was placed in a home between Monday and Thursday, it was picked up the next morning starting at around 8:00 am. If it was placed in a home on a Friday, it was picked up the coming Monday starting at around 8:00 am. This procedure was followed to comply with government labor regulations. In a subsample of households, the measurement took place between 5:00 pm on a Friday and 7:00 am the following Monday. Following the standard practice in the environmental health literature, the resulting PM2.5 concentration for those households was averaged across the three days.

5 The Impacts Rural Electrification: Findings

We used four subsamples in the analysis. The non-experimental subsample was formed by all households that were off the grid at baseline and includes 2,014 households. All the FE results were based on this subsample. The experimental sample included 500 households in San Miguel and Chalatenango. A subset of 150 households was selected for IAP measurement. The experimental IAP results were based on this sample. Finally, 207 households from the non-experimental sample in San Miguel and Chalatenango were selected for IAP measurement. These were households that had not connected to the grid by the first follow-up survey. The tables referenced in this section can be found in section 9.

Table 7 shows descriptive statistics by subsample. The first column shows the sample means for the non-experimental sample and the second column shows the sample means for the experimental sample, while the third and second columns report the results for the air quality subsamples (experimental and non-experimental, respectively). Table 7 shows that the experimental air quality subsample was clearly similar to the experimental sample. The non-experimental air quality subsample, on the other hand, was different than the rest: being composed of households that had not connected to the grid by the first follow-up, it was formed of poorer-than-average households. All in all, characteristics in the voucher subsample were roughly similar to the larger non-experimental sample. This is not to say that results from the experimental analysis are representative of the whole sample, but it is at least indicative that the experimental subsample is not too different from the whole sample.

Household heads were on average 50 years old; 69-73 percent of them are male and have 1.5-2.5 years of schooling on average. The average age in the households was between 30-33 years old and households have on average 4.5 members, with a total dependency ratio (the number of non-working-age household members divided by the number of working-age household members) of around 0.45. Annual per capita income was around 650 USD in the whole sample, 770 USD per head in the experimental sample, 620 USD in the air quality experimental subsample, and 430 USD in the air quality non-experimental subsample. Table 8 shows the balancing test by treatment arm; in Table 9, we show the regression adjusted balancing tests for selected indicators, such as age of the head of household, the probability that the head is male, the schooling level of the head, etc. In general, the sample was balanced across groups.

5.1 Adoption of Electric Connections – The Role of Cost and Spillovers

Figure 12 shows the evolution of connection rates by year of the study. This graphic shows that voucher recipients were more likely to be connected to the grid than non-recipients, although the differences seem to decrease toward the final years of the study.

Figure 13 splits vouchers by value and survey year; the results are consistent with Figure 9. Although there is no statistically significant difference in adoption rates between high- and low-discount vouchers, adoption rates are slightly higher in the low-discount group.

Figure 12: Voucher Allocation and Connection Rate

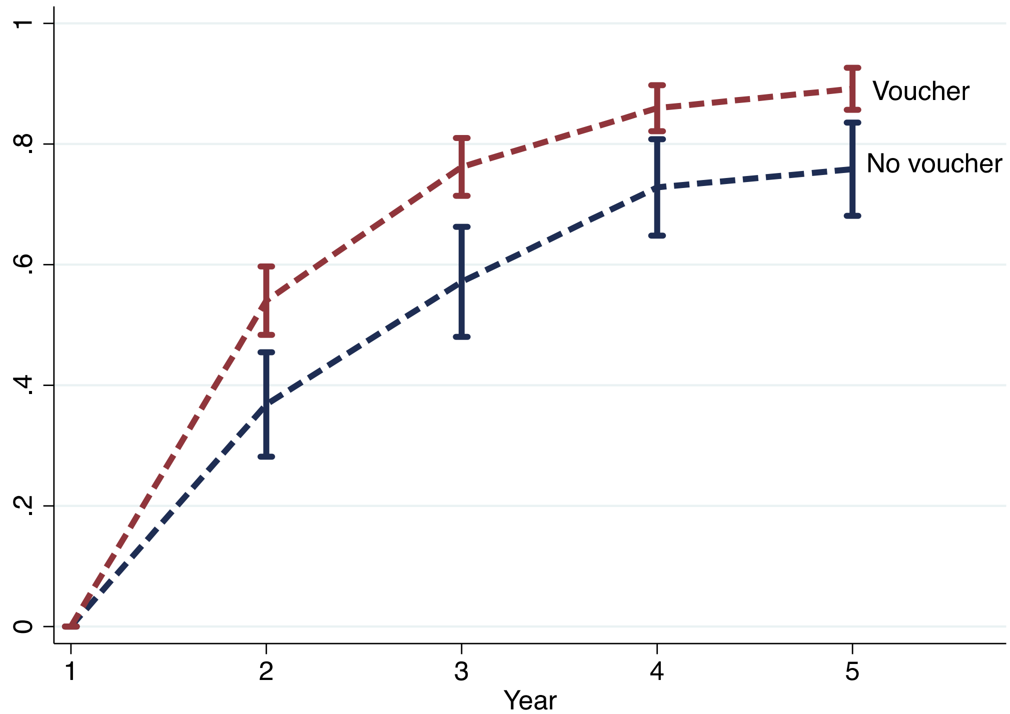


Figure 13: Voucher Allocation and Connection Rate, by Voucher Value

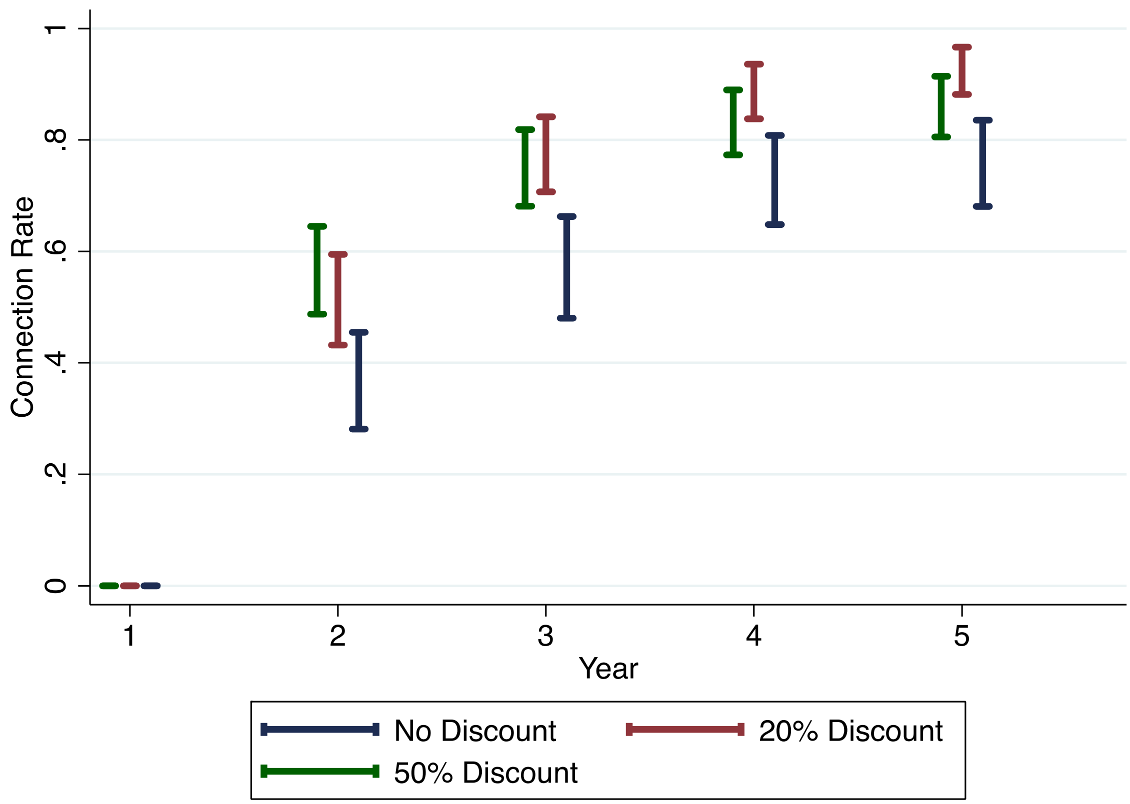


Table 10 shows how changes in F and E*, empirically measured by "voucher" and "s100", affected connection rates. The first column pools the four follow-up surveys (survey rounds two through five), showing that voucher recipients were on average 12 percentage points more likely to get a formal connection in the post-baseline period. There also appear to be important spillovers: a 10 percent increase in s100 increased the probability of connection by 1.3 percentage points. This indicates the existence of potentially large externalities, which are analyzed in greater detail later in this report. The fact that results were identical with and without adding household FE (second column) is further evidence that our randomization worked well.

The third and fourth columns of Table 10 replicate the first two columns interacting voucher and s100 with round dummies, which allows us to explore time patterns. Although the differences were not statistically significant, the coefficients on voucher were larger in rounds two and three (around 15 percentage points) than in rounds four and five (around 10 percentage points). This decline is attributable to the fact that the connection rate in the non-encouraged group started catching up to the encouraged group, as shown in Figure 12. The effect of s100 does not appear to follow such a trend, which would indicate persistence in the externalities.

In Table 11, we estimate specifications equivalent to Table 10, including s200. The coefficients were unchanged and s200 turns out to not be significant. For the sake of parsimony, we did not include s200 in our main specifications. Table 12 analyzed take-up using a different explanatory variable: the inspection fee. The variable takes the value of one if households had to pay the full 100 USD, 0.80 if they received the 20 percent discount, and 0.50 if they received the 50 percent discount. In this specification, spillovers were captured by the average fee paid by eligible neighbors within a 100m radius. The idea is that if the average fee faced by household i's neighbors is lower, household i will have a higher share of neighbors on the grid, and vice versa. Pooling together the three follow-up surveys shows that a 10 USD reduction in the fee would increase the probability of connection by two percentage points (first column). Once again, including household FE leaves the point estimates unchanged (second column).

Next, we analyzed the dynamics in the third and fourth columns of Table 12. The coefficient on the inspection fee at round two is -.32, significant at the one percent level. By round three, the effect size decreases to -0.26, significant at the five percent level. The effect loses significance and approaches zero in rounds four and five (-0.13 and -0.11, respectively). This indicates that a lower fee increased adoption early on, but not later in the process. The average fee (fee100) had significant effects on connection take-up, with a 10 USD reduction increasing the probability of connection by 3.5 percentage points over the study period. The effects are significant in all rounds, with values of.29 (round two), .45 (round three), .30 (round four), and .35 (round five) – all statistically significant. This suggests that peer effects did not decrease with time. In the fifth column, we explore linear and quadratic time trends in the effects of fee and fee100. While this specification imposes more structure than the previous one, said structure is consistent with the more flexible specifications in the previous columns.

We then analyzed how E* affected connections in Table 13. As discussed previously, we used s100 (or fee100) as an instrument for E*. An additional connection within 100 meters increases the probability of household i's connection by 10 percentage points (10.3-11.1, depending on the exact specification), almost the same effect as household i itself receiving a voucher. When expressed in percentage of eligible neighbors, we see that a 10 percentage point increase in the share of eligible neighbors that connect to the grid increased the probability of household i's connection by roughly two percentage points (1.9-2.2, depending on the specification).

The exclusion restriction in this case required the instrument to affect the connection decision only through its effect on E*. This meant that s100 altered E*, which in turn affected E. The exclusion restriction would be violated if, for instance, i's neighbors who received vouchers encouraged household i to connect by providing household i's members with information about the connection process, making them aware of the electrification process, etc. If household i did not receive a voucher and its neighbor did, we would expect household i to refrain from connecting, maybe hoping to get a voucher at a later stage or being upset for not having received said voucher. The information story is not strong either since grid electrification projects are easily visible by all in the community. If it were an information story, we would see the externalities declining with time, but we see no such trend. On the contrary, we found spillover effects that are persistent during the whole study period.

5.2 Connection Choice

In Table 14 and Table 15, we turn to the analysis of connection choice. As discussed previously, households in our sample have the option of getting an informal connection, which typically provides enough power for a couple of light bulbs and maybe a TV. Households with informal connections keep using kerosene for lighting; thus, in our analysis so far, informal connections have been classified as being off the grid. We conducted two types of empirical approaches. First, we looked at the probability of switching between any pair of connection types (from no electricity to formal, from informal to formal, and from no electricity to informal).[26]

The dependent variable takes the value of one if the household switched at some point during the study period, and zero otherwise. The sample only includes households that had no connection at baseline (i.e., it dropped households with informal connection at baseline). This reduced the sample size to 275 observations. Once we controlled for covariates, the effect of voucher and s100 were not significant: households that did not have an informal connection did not respond to vouchers. The coefficient on s100 was too large to reject the existence of positive externalities, but it was not statistically significant. The third and fourth columns reported the results of switching from informal to formal connections. Both voucher and s100 were strongly significant in this subsample. Among this group, voucher recipients were 15 percent more likely to switch to a formal connection; having 40 percent of eligible neighbors receive a voucher (mean s100) increased the probability of connection by four percentage points. Finally, the fifth and sixth columns show that vouchers did not affect the probability of switching from no connection to an informal connection. The results in Table 15 using the fee are consistent with the voucher results.

Next, we present the result of ordered probit estimation for connection choice in Table 16; the first and second columns use the voucher and the third and fourth columns use the fee as excluded instruments. This method imposed more structure in the analysis in two main ways. First, it exploits the difference in electricity quality, which is close to reality since differences in reliability and other characteristics make formal connections better than informal connections. Second, it required imposing the assumption that the disturbances term follows a normal distribution. Results by round are presented in Table 17.

Figure 14, Figure 15, and Figure 16 plot the modeled probabilities of each type of connection by round, according to the ordered probit model in Table 17, the first, third, fifth, and seventh columns. In Figure 14, we estimated the probability of formal connection by round and plotted the kernel densities to give a better sense of the evolution of connection rates. In each round, the mean increased and the variance decreased. Figure 15 shows that the rate of informal connections was low, below 15 percent, and fairly uniform across rounds.

Figure 16 shows that the probability of having an informal connection or of being off-grid, respectively, moved in the opposite direction. Hence, informal connections do not seem to be crowding out formal connections. On the contrary, formal connections seem to be crowding out informal ones.

Figure 14: Households with Formal Connections by Round

---

[26] There are just 16 cases of households switching from informal to off-grid and no cases of households dropping their formal connections.

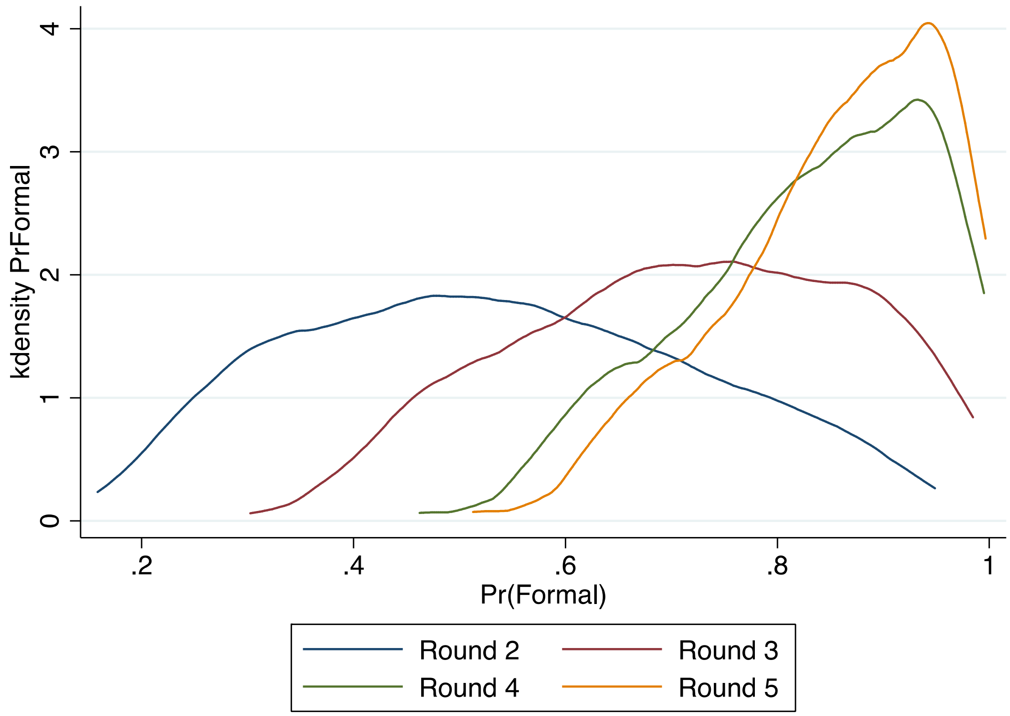

kdensity PrFormal
0
1
2
3
4
.2
.4
.6
.8
1
Pr(Formal)
Round 2
Round 3
Round 4
Round 5

Figure 15: Households with Informal Connections, by Round

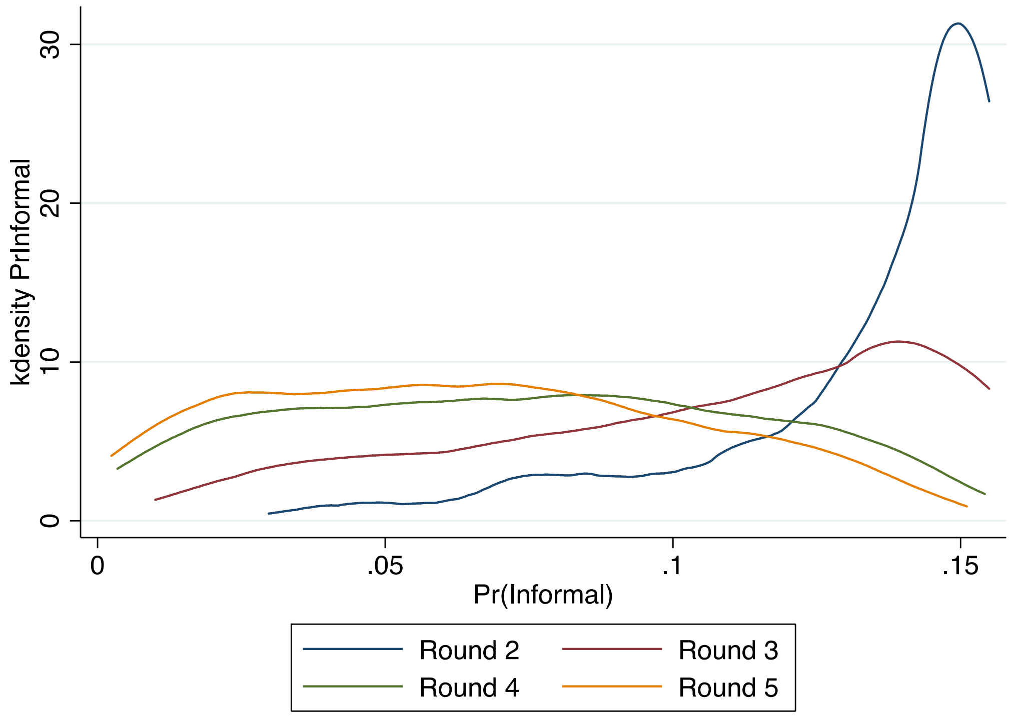


Figure 16: Households with no Connections by Round

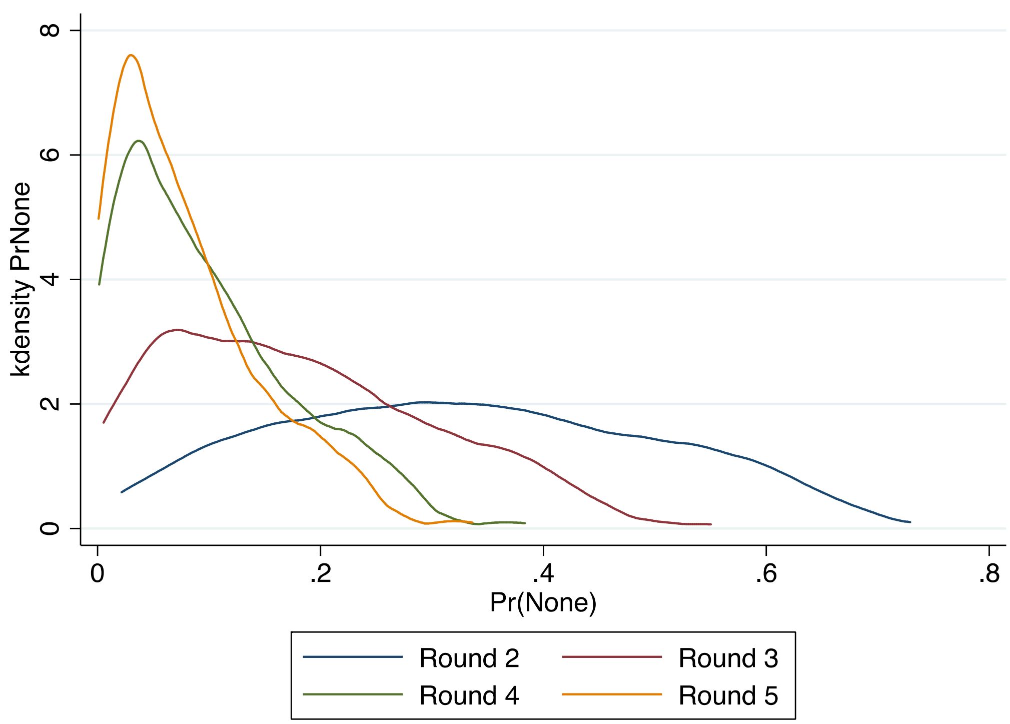


Figure 17, Figure 18, and Figure 19 plot the marginal effects of discount vouchers on the probability of formal (Figure 17) and informal (Figure 18) connections from the ordered probit model in Table 17, the first, third, fifth, and seventh columns. In these plots, the marginal effect for each household was estimated, given each household's characteristics. Figure 17 shows that the effect of receiving a voucher was concentrated around 0.10 at around two, and that it faded away in later rounds, as non-recipients caught up in formal connection rates. In Figure 18, the effect of receiving a voucher on the probability of having an informal connection was initially dispersed with a mean of around zero and a range of -0.02 to +0.20, meaning that receiving a voucher increased the probability of an informal connection among some households by two percentage points and decreased it by 0.2 percentage points among others. In rounds three and later, the effect was negative among almost all households and was concentrated around -0.02. Figure 19 shows that voucher recipients were less likely to remain off-grid by round two, but then the effect became diffused as non-recipients adopted some type of connection.

Figure 17: Vouchers and Probability of Formal Connection by Round, Marginal Effects

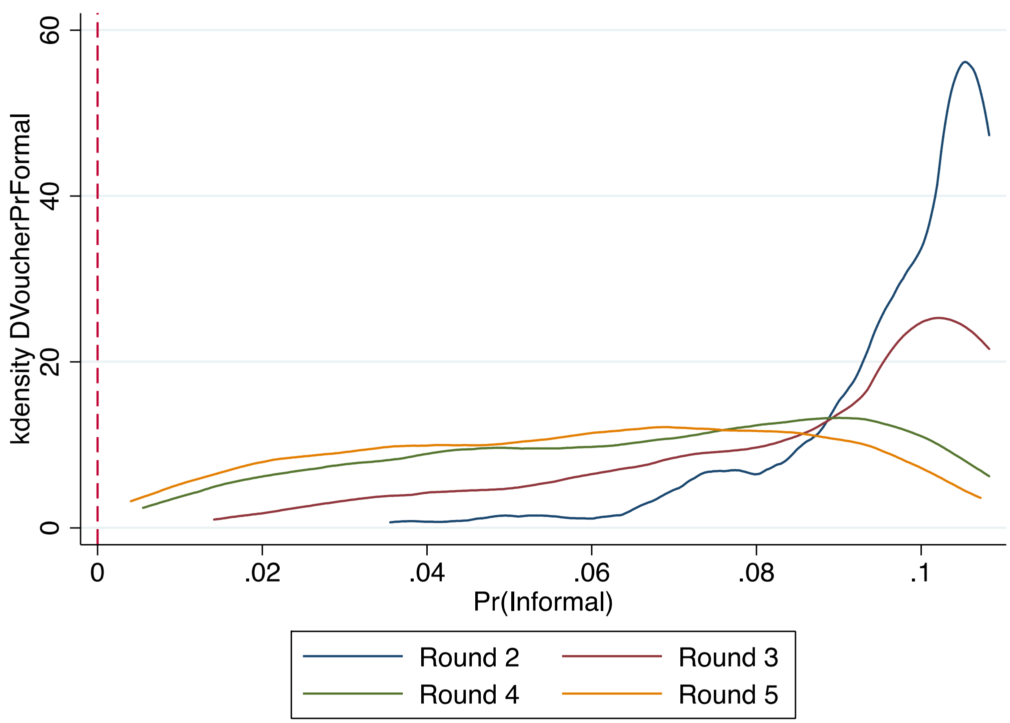


Figure 18: Vouchers and Probability of Informal Connection by Round, Marginal Effects

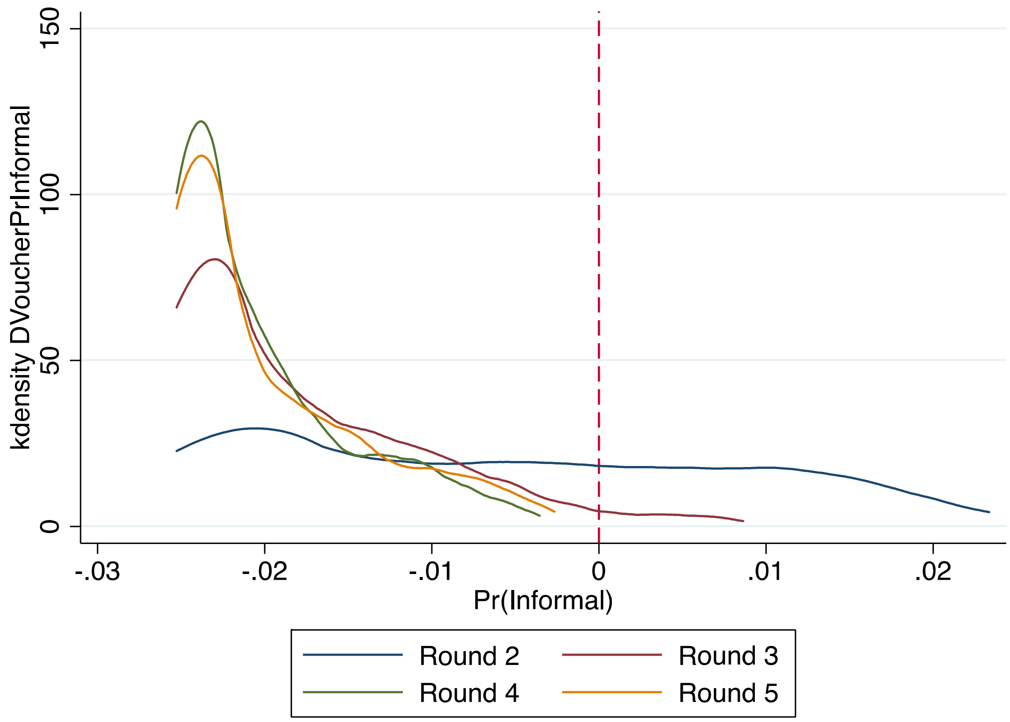
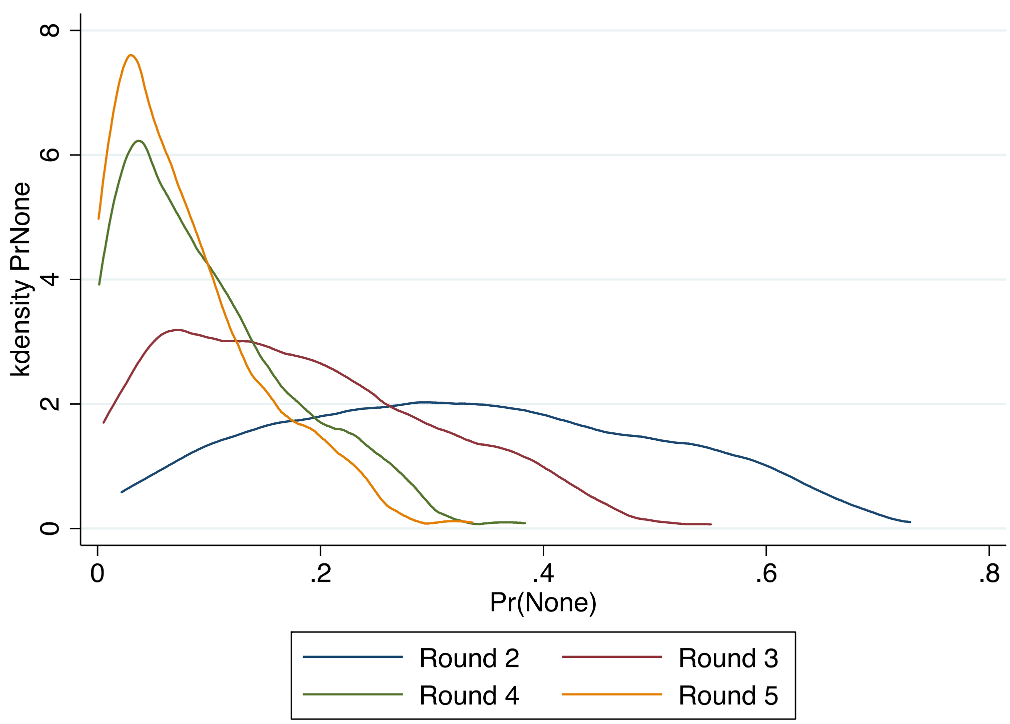


Figure 19: Vouchers and Probability of No Connection by Round, Marginal Effects



Turning to spillover effects, in Figure 20 and Figure 21, we plot the relationship between s100 and connection type. To obtain these figures, we estimated the model in Table 16, the second column and obtained predicted values by simulating different values of s100 while keeping each household's observed characteristics. Both relationships were roughly linear, but the slope on the probability of having a formal connection was positive, while the slope on the probability of having an informal connection was negative. Figure 21 shows that the marginal effect of s100 on the probability of having a formal connection was positive but declines as s100 increases: household i's first connected neighbors had a slightly higher influence on i's decision to get a formal grid connection, and while having more connections increased the probability of formal connection, it does so at a decreasing rate. On the other hand, the marginal effect of s100 on the probability of having an informal connection was small, negative, and fairly constant across values of s100.

Table 15, Table 16, and Table 17 (the second, fourth, sixth, and eighth columns) provide alternative specification using fee and s100, reassuring the robustness of the results presented in this section.

Figure 20: s100 and Connection Type

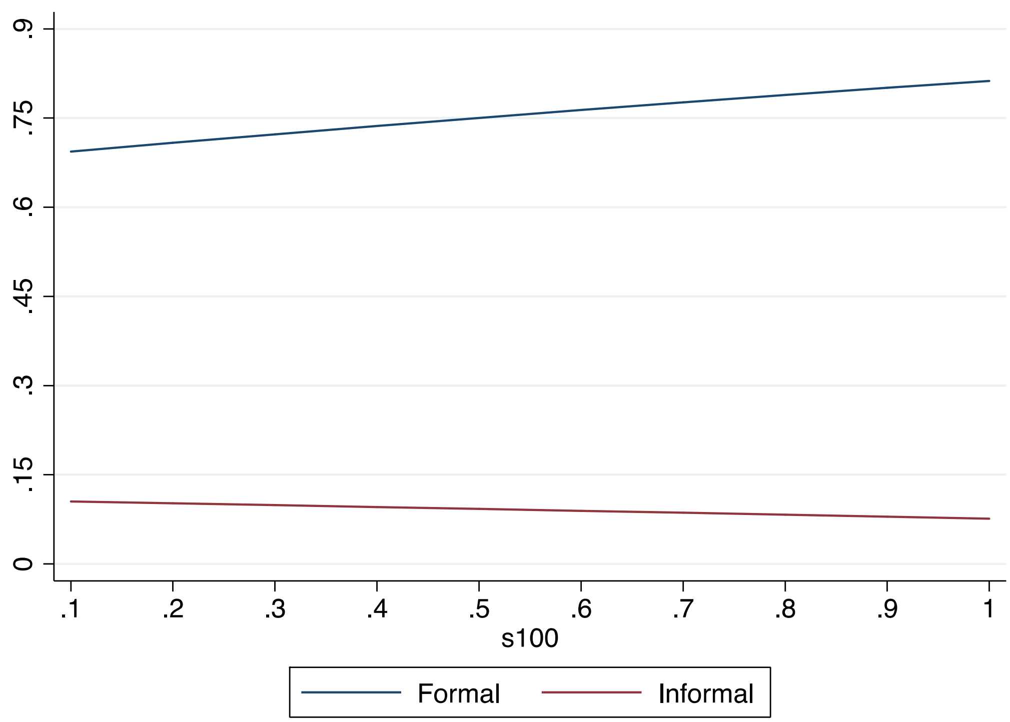
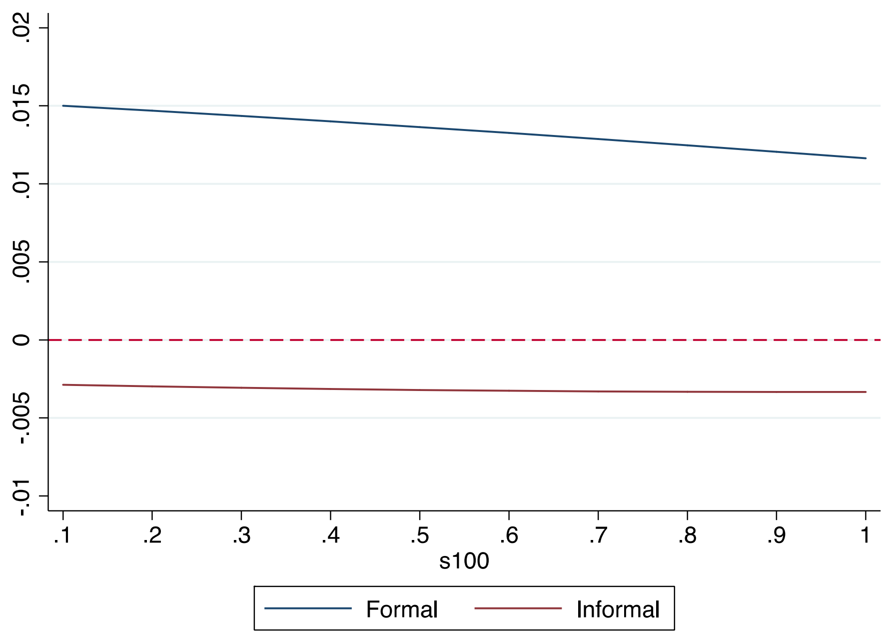


Figure 21: s100 and Connection Type, Marginal Effects



5.3 Characteristics of Adopters

Next, we examine the characteristics of electricity adopters in Table 18 to get better insights into who may benefit from a cost-sharing policy. These were just descriptive regressions, and we claim no causality; they simply provide insights into the characteristics of adopters. The dependent variable took the value of one if the voucher recipient connected at some point over the course of the study and zero otherwise. This regression controlled for sub-district FE.

Households with informal electricity at baseline were 18 percentage points more likely to take up the voucher, as were households with a property title (eight percentage points) and households with floor other than dirt (10 percentage points). Households with more members were also more likely to adopt an electrical connection. Material on the walls had no apparent relation to adoption. Income is positively related to adoption, but the magnitude is economically small: a 1,000 USD increase in annual income is associated with a 1.4 increase in the probability of connection. Age, gender, and literacy status of the household head had no apparent relation to voucher take-up.

Clearly, all the variables are related, but empirically, the main marker with which identify households that will respond to vouchers was informal electrical connection. Those households went through the trouble of getting an informal connection from their neighbors, hanging cables, etc., which reveals that they place a higher value on electricity than the rest.

5.4 Indoor Air Pollution

To illustrate the relationship between kerosene use and indoor air quality in our study setting, Figure 22 shows a non-parametric regression of overnight PM2.5 as a function of monthly kerosene expenditure (more details on the variables and the samples can be found in the next section). There was a strong positive relationship between these two variables, suggesting that reductions in kerosene use could generate important improvements in indoor air quality. Kerosene provides an important source of variation in PM2.5, even with 70 percent of households using wood for cooking.

Figure 22: Monthly Expenditure in Kerosene and Overnight PM2.5 Concentration (with 95 Percent Confidence Bands)

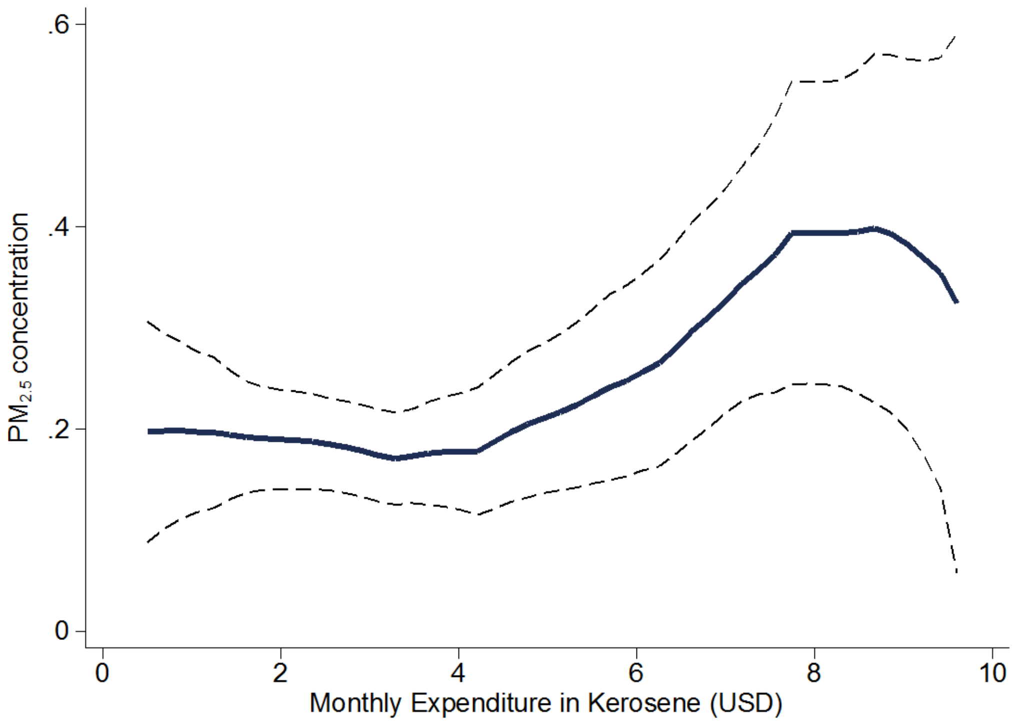


Table 19 reports our main experimental results. The dependent variable is PM2.5 concentration (in logs) between 5:00 pm and 7:00 am.[27] The level of observation is the household-minute in panel A. To allow for the arbitrary structure of the covariance matrix within households, we clustered the standard errors at the household level. By round three, voucher recipients showed drastic reductions in IAP compared to the non-encouraged group, with 67-73 percent lower PM2.5 concentration.[28] When the data is collapsed at the household level (in panel B), the magnitude and significance remain unaltered. In rounds four and five, the coefficients were closer to zero and are not statistically significant. We attributed this result to the control group catching up in electrification rate. Our sample was simply too small to pick up differences in PM2.5 concentration with differences in electrification rates of around 10 percent.

Figure 23 shows the reduced form results by hour of the day. The effects were larger from 5:00 pm to 10:00 pm, decrease thereafter as most household members go to sleep around this time, and jump up again from 6:00 am to 7:00 am when they wake up the next morning.

[27] In this sample, we removed 17 households that reported average overnight PM2.5 concentrations above four milligram/day. Given the small overall number of households, we took the conservative approach of excluding them. When they are included in the sample, the estimated reductions are much larger, but still within the confidence intervals (CIs) in Table 23 (of the order of 90 percent).

[28] This figure is obtained from the reduced form coefficients: $e^{-1.119}-1=-0.67$; $e^{-1.316}-1=-0.73$.

Figure 23: PM2.5 and Voucher Allocation

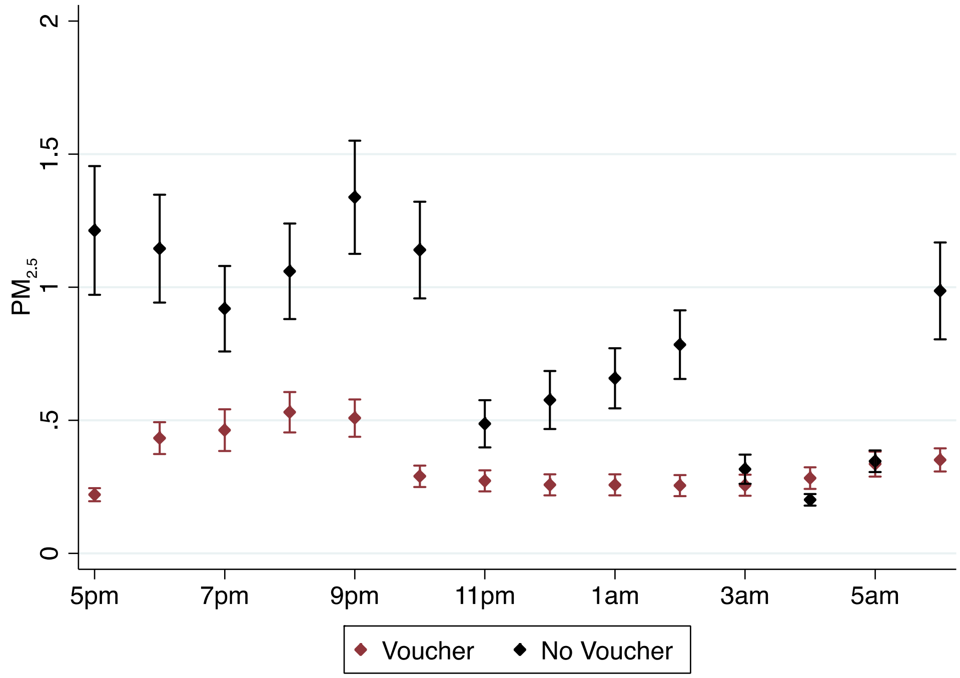


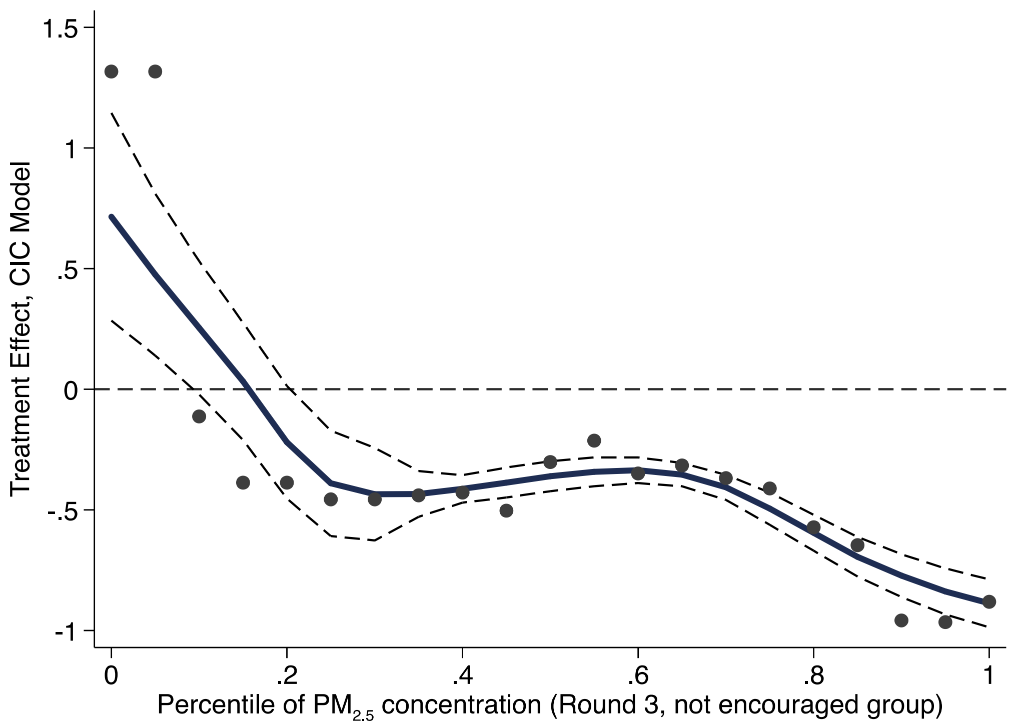
Table 20 presents the IV results. The point estimate on connection was negative and large (implying a 95 percent reduction in overnight PM2.5 concentration), but given the weak first stage, the standard errors were too large to draw any useful inference. The weakening of the first stage with time is consistent with the standard errors of the reductions blowing up in the later rounds, from .10 in the third round to 0.30 in the fourth and 1.8 in the fifth.

To complement these results, we used changes-in-changes (CIC), a non-parametric generalization of differences-in-differences developed by Athey and Imbens (2006). We present the CIC results to examine whether the non-encouraged group caught up with the encouraged group in terms of PM2.5 concentration by round four (given that the electrification rates were similar by then) and to explore the distribution of treatment effects.

The CIC estimator for the average treatment on the treated was -0.70. Given random group assignment, this was also the average treatment effect. The CIC estimator was consistent with the effects found among voucher recipients by round three (-0.67 to -0.73). This strengthens the internal validity of our findings to the extent that eliminating the differences in electrification rates led to an elimination of the differences in overnight PM2.5 concentration.

Figure 24 analyzes the variation in the magnitude of treatment effects along the distribution of overnight PM2.5 concentration. The percentage reductions in the outcome variable were of considerable size all throughout the distribution. However, there seemed to be variability in the treatment effect along the distribution. The reduction was significant, starting roughly from the 20th percentile, and the size of the effect started increasing (becoming more negative) at the 60th percentile. This is consistent with the intuition behind our study setting: treatment effects are significant above a certain threshold of IAP, and higher polluters experience larger reductions.

Figure 24: Treatment Effect Heterogeneity



5.4.1 Non-Experimental Estimates

In this section, we show the results of non-experimental estimates to provide additional support to the experimental estimates from the previous section. We do not claim that these estimates are causal, but we do show that there exists a solid relationship between electrification, kerosene consumption, and overnight PM2.5 concentration in a longitudinal setting.[29] In this section, we show that: (i) households that connect to the grid also reduce kerosene use and IAP, (ii) these changes are not observed before households connect to the grid, (iii) these reductions are similar between groups irrespective of the timing of electrification, and (iv) there is no reversion to pre-electrification levels in PM2.5 concentration.

The non-experimental subsample showed a negative correlation between electrification and PM2.5 (see Table 21, columns one through four). In 2010, the (geometric) mean PM2.5 concentration was 142 [95 percent CI 122-165] mcg/m^3 (N=201). In 2011, PM2.5 concentration was 185 (149-230) mcg/m^3 among non-connected households (N=119) and 128 (86-190) mcg/m^3 among connected households (N=46). In 2012, PM2.5 concentration was 108 (86-136) mcg/m^3

[29] In our study setting, kerosene is mostly used for illumination, not for heating or cooking.

among non-connected households (N=78) and 97 (75-125) $\mu g/m^3$ among connected households (N=90).[30]

In this subsection, we use electrification status at a given point in time to define groups, and we use statistical tools to compare outcomes between those groups. None of the households in this sample had a connection by round two of the *EHEIPCER* survey (2010). T1 is the group of households that connected between rounds two and three of the survey (i.e., between 2010 and 2011). T2 is the group of households that connected between rounds three and four (between 2011 and 2012). T3 is the group of households that remained unconnected by 2012. Just a few households in the IAP sample connected to the grid after 2012. For the sake of readability, we left them in the control group in the following figures.

[30] Kerosene expenditure also shows a negative correlation with electrification in this subsample. In 2010, mean kerosene expenditure was 5.07 (4.57-5.61) USD/month in 2010. In 2011, the figures were 5.54 (4.85-6.32) USD/month for non-connected households and 3.87 (.15-99.42, due to only two non-zero observations) USD/month. In 2012, mean kerosene expenditure was 5.65 (4.62-6.92) USD/month among non-connected households and 2.52 (.93-6.81) USD/month among connected households.

Figure 25: Electrification and PM2.5 Concentration

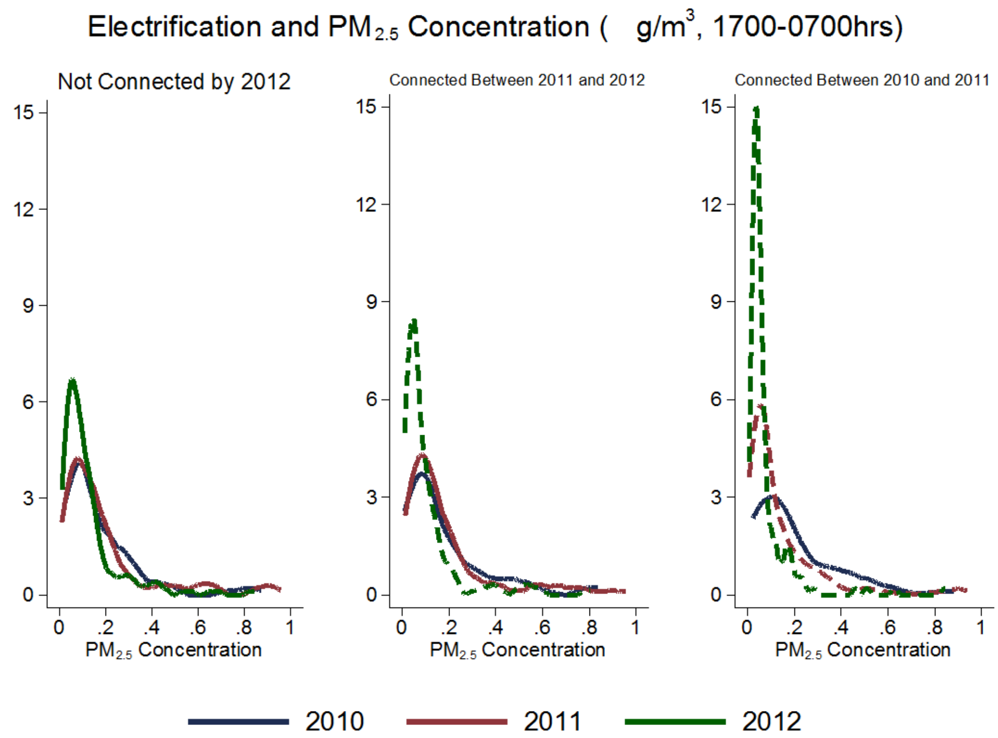
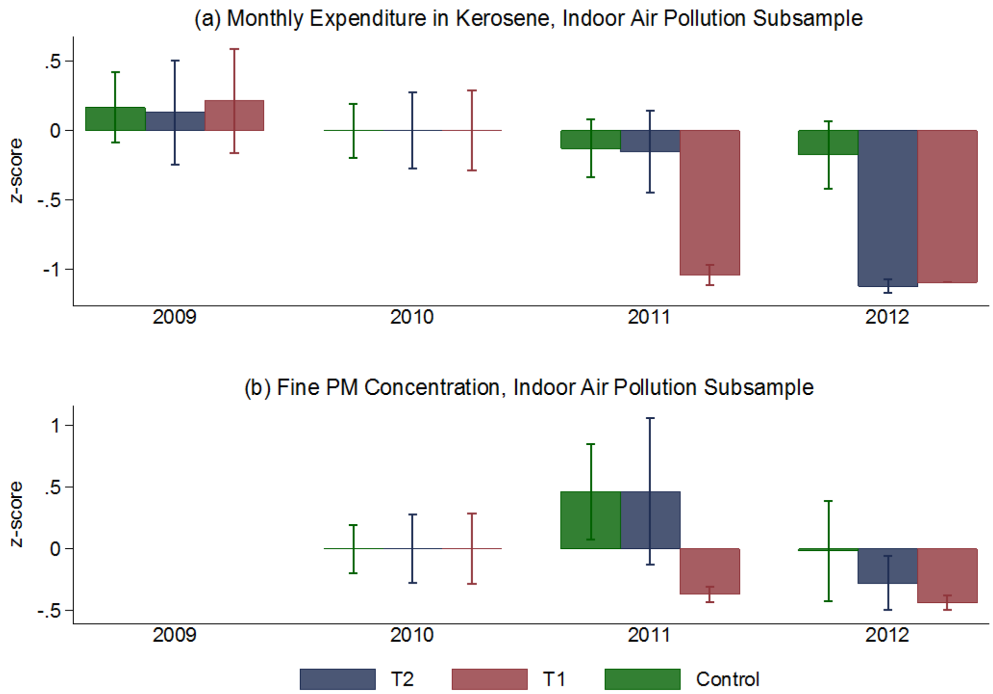


Figure 25 shows the kernel density of PM2.5 concentration between 5:00 pm and 7:00 am by group. Panel (c) shows the kernel density estimates for T3 households. The densities in 2010 and 2011 clearly overlap with each other. The two-sample Kolmogorov-Smirnov (KS) test for equality of distributions generates a p-value of .176, thus, the null hypothesis of equality of distributions cannot be rejected at conventional confidence levels. The density in 2012 shows some differences with respect to the density in 2010; with an associated KS p-value < .001 in this case, thus rejecting the null hypothesis of equality of distributions. Panel (b) shows the same for T2 households. The densities overlap between 2010 and 2011 (KS p-value = .380), but the PM2.5 density estimate shifted to the left 2012 (KS p-value < .001). For T1, the densities corresponding to the 2011 and 2012 measurements fall to the left of the 2010 distribution (KS p-value = .006 and < .001, respectively).

Figure 26: Electrification, Kerosene Expenditure, and Overnight PM2.5 Concentration



To allow for direct comparison with respect to each group's baseline values, the variables in Figure 26 were standardized by subtracting the baseline mean and dividing by the baseline standard deviation of their respective group. Panel (a) shows the change in average monthly expenditures on kerosene in 2011 compared to 2010 levels by treatment arm, with 95 percent CIs. There is no change in mean between the 2009 and the 2010 measurements for any of the groups. T3 does not show any change in mean kerosene expenditures in any of the surveys compared to 2010. T1 shows a large reduction between 2010 and 2011, which is maintained by 2012. T2 shows no change by 2011 (when the group is still off-grid), but a large reduction by 2012 (when the group connects to the grid).

Panel (b) shows the association with PM2.5. PM2.5 concentration did not change significantly in the control group between baseline and follow-up. Households in T1 show a significant reduction in average PM2.5 concentration at the first and second follow-ups. Households in T2 show no reduction in PM2.5 concentration by the first follow-up, but a significant reduction in mean PM2.5 concentration by the second follow-up. This reduction is not statistically different from the average reduction experienced by households in T1.

This graph shows three striking facts. First, both PM2.5 and kerosene expenditure changed when the electrification status changed. Conversely, neither changed if electrification status did not change (except for an *increase* in PM2.5 among the control group in 2011). Second, the average changes among households in T2 were similar to those experienced by households in T1. Third, the new (lower) levels of kerosene consumption and PM2.5 observed for T1 in 2011 were maintained in 2012.

The results are presented in Table 21. Connection to the grid is associated with a 25-33 percent reduction in PM2.5 concentration between 5:00 pm and 7:00 am. This estimate is consistent across specifications and is strongly significant. The first column is a regression PM2.5 on connection and year FE. Adding household FE (second column) did not alter the point estimate, but the standard errors were large, suggesting that in this specification, household FE absorb too much of the variation in connection. In the third and fourth columns, we included sub-district FE and baseline characteristics in lieu of household FE. The resulting point estimates imply a reduction of 22-26 percent in PM2.5 concentration over the course of the study. In the fifth column, we tested for differential treatment effects by round and found that electrification led to substantial reductions in PM2.5 concentration in rounds three and four, but not in the last round. When the sample is restricted to the first four rounds of data, the coefficient on connection in a FE estimation results is -0.49, implying a 39 percent reduction in PM2.5. This suggests something potentially different between households that connected in round five and those that connected in the earlier rounds. In the fifth column, we replaced connection status at round five with connection status at round four and estimate the regression, adding household FE once again. The resulting coefficient is -0.33, significant at the five percent of confidence.

Using the data for T2 and T3 to test for differential pre-treatment trends in PM2.5 and the 2010 *EHEIPCER* wave to test for differential pre-treatment trends in expenditures in kerosene or candles and use of wood or candles, we cannot reject the null hypothesis of parallel pre-treatment trends in any of the tests we performed.

5.5 Discussion on Effect Size

Taken at face value, these effects may seem too large. Given the first stage and reduced form coefficients, the implied IV estimator of the effect of electrification on overnight PM2.5 concentration is given by:

$$\beta_{IV} = \frac{\beta_{RF}}{\beta_{FS}} \approx -\frac{1.12}{0.20} = -5.6$$

This implies a reduction in overnight PM2.5 concentration of 99.6 percent.[31] However, in this section we argue that an effect of this size is not out of the question and present some evidence in support of that claim. Note that the outcome of interest is PM2.5 concentration between 5:00 pm and 7:00 am in the room of main use at night, in most cases the living room. In this room, and during this period, kerosene lamps arguably account for the largest share of PM2.5 emissions.

31 In this case, it is not possible to rely on the approximation $\frac{y_T - y_C}{y_C} \approx \frac{\Delta lny}{\Delta connected} = \beta_1$, since $\beta_1 = -3$ is of considerable magnitude. The exact percentage change is given by the following expression. First, note that:

$$lny_T - lny_C = \ln\left(\frac{y_T}{y_C}\right) = \beta_1$$

, which implies

$$\frac{y_T}{y_C} = e^{\beta_1} = e^{-5.6} - 1 = -0.996$$

Note that Table 28 (discussed below) shows large decreases in the intensive and extensive margin of kerosene use, which should reflect large drops in overnight PM2.5 concentration.[32]

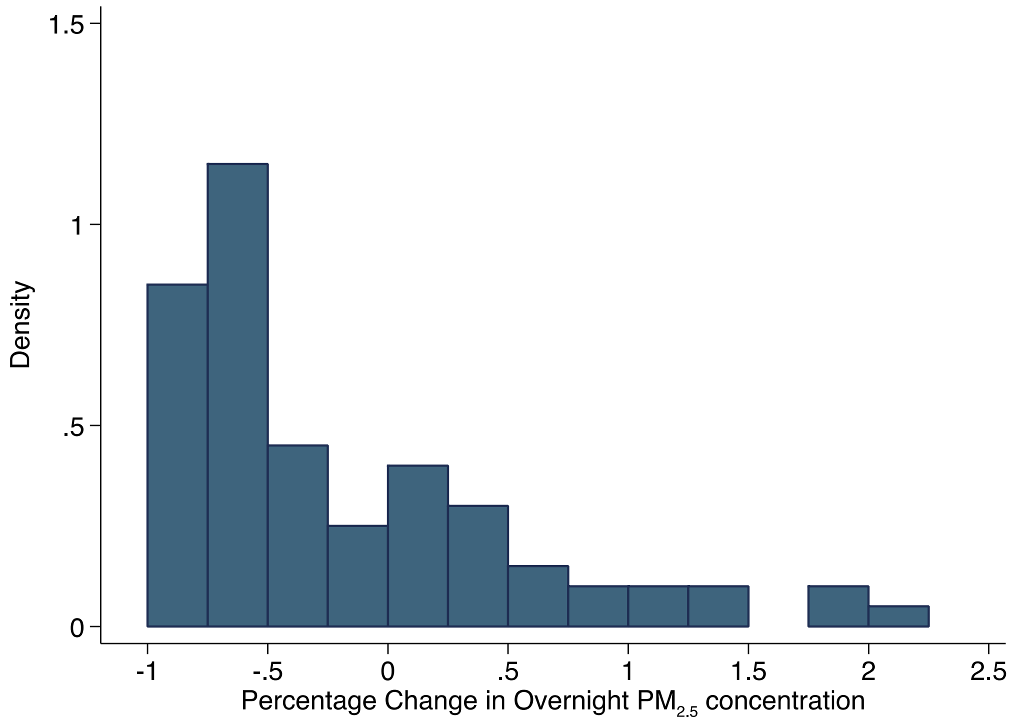
Next, we show that the drops implied by the model were consistent with the raw percentage changes in PM2.5 concentration among households in our sample. We have successful PM2.5 measurements for 85 households in rounds three and four. We calculated the change in PM2.5 concentration for these households.

Figure 27 shows the histogram of these changes.[33] Forty households showed reductions of 50 percent or higher, 14 showed reductions of up to 50 percent, and 30 households showed increases in overnight PM2.5 concentration. A few households showed large percentage *increases*, but these large increases correspond to households with especially low levels of overnight PM2.5 concentration at round three. In terms of level, these increases are rather small. Conditional on showing an increase in PM2.5, the average increase in levels was 0.051 milligrams/m^3, (95 percent CI 0.004 - 0.099 milligram/m^3).

32 Some emissions from cooking may filter through during the day and linger in the living room; as we have seen, cooking practices did not change with electrification. However, there is no reason to believe that PM2.5 concentration in the living room would depend more on filtrations from biomass during the day than from direct use of kerosene lamps in that room during the evenings.

33 Three households show percentage increases higher than 2.5. They are not included in the figure.

Figure 27: Raw Changes in PM2.5 Concentration



5.6 Time Use

5.6.1 Time Use - Children

We first studied how electrification affects the probability of participation in certain activities among school-age children (six to 14 year olds). We considered four categories: education, labor, chores, and leisure. Conditional on participation in each activity, children spent an average of 6.1 hours on educational activities, 3.5 on household chores, 6.7 hours on work, and 8.7 hours on leisure. The dependent variable in each column is an indicator of participation in each activity; the results are reported in Table 22 for the IV estimates and in Table 23 for the FE estimates. We discuss the IV estimates and note that the significance of the FE is lower due to the inclusion of individual effects. Electrification increased the probability of participating in education activities by 78 percentage points in the IV estimates. These include studying at home, spending time at school, and going to and from school. Panel B shows that there were effects on time spent studying (a 54 percentage point increase) as well as time spent on other activities related to education, like spending time in school or commuting between home and school (an 84 percentage point increase). Time spent commuting to school increases because children are less likely to skip classes. This increase in participation is consistent with a perceived increase in the returns to education through better learning due to electrification. Children who studied at home did so for an average of two hours a day. Due to selection issues discussed in the methodological section, this cannot be interpreted as electrification leading to a two-hour increase in study time, but it serves for illustrative purposes. Average time allocated to education by those who participate in such activities was 6.1 hours per day. A higher share of children studying at home is an important indicator of improved learning, especially given that this increase is paired with a better study environment. In addition, there was an interesting increase in computer ownership (14 percentage points, as shown in Table 29, discussed below). Although the literature has not reached a consensus on the effects of computers on learning, this increase may have additional impacts on learning.

The probability of engaging in household chores increased by 96 percentage points. Some of these chores may be home production, which is sometimes difficult to distinguish in the field. The increase in time spent on household chores is important given that, as we will see in the next section, there was an increase in home production, mostly among adult women. Taken together, this suggests that children are taking on some household chores previously undertaken by female heads, who are now allocating some time to other activities.

The point estimate on the probability of children working was negative but not statistically significant.[34] Only 1.7 percent of our sample simultaneously engaged in education and work, which may suggest that a higher participation in educational activities could be accompanied by a reduction in the share of children who work.

5.6.2 Time Use - Adults

Table 24 and Table 25 show that electrification increased adults' participation in non-farm employment and in home business operations. On average, workers from connected households were 26 percentage points more likely to engage in non-farm employment (IV estimates). The third and fourth columns show that the effect was concentrated among women, with women from on-grid households being 46 percentage points more likely to engage in non-farm employment at some point over the four periods, conditional on not having participated in the electrical connection in the year leading to the baseline. Each year on the grid increased this probability by 11 percentage points. These figures can be thought of as an upper bound of the true effect, given that they include people who may have participated for just one month out of the four post-treatment years for which we have data.

In a similar vein, panel (B) shows that electrification increased the probability of adults operating a home business by 12 percentage points. This is more than a 150 percent increase compared to the control group. The effect was concentrated among women, with women from on-grid households being 25 percentage points more likely to operate a home business. For reference, average annual profits are around 1,000 USD a year for females and 1,500 USD for males, suggesting that the new businesses may provide a non-trivial income source to the household.

The point estimates for males were lower than for females and were not statistically significant. Furthermore, the point estimates for home business operations differed statistically by gender and were practically non-existent for males. Changes of this type suggest that electrification has important consequences in women's income that may lead to changes in intra-household

[34] The coefficient on leisure is just to show consistency: since virtually everybody enjoys at least some leisure, the effect of electrification should be null. In fact, the coefficient is close to zero.

bargaining power. Figure 28 and Figure 29 show the point estimates by sex and number of years connected to the grid.

Figure 28: Years on the Grid and Probability of Engaging in Income Generating Activities

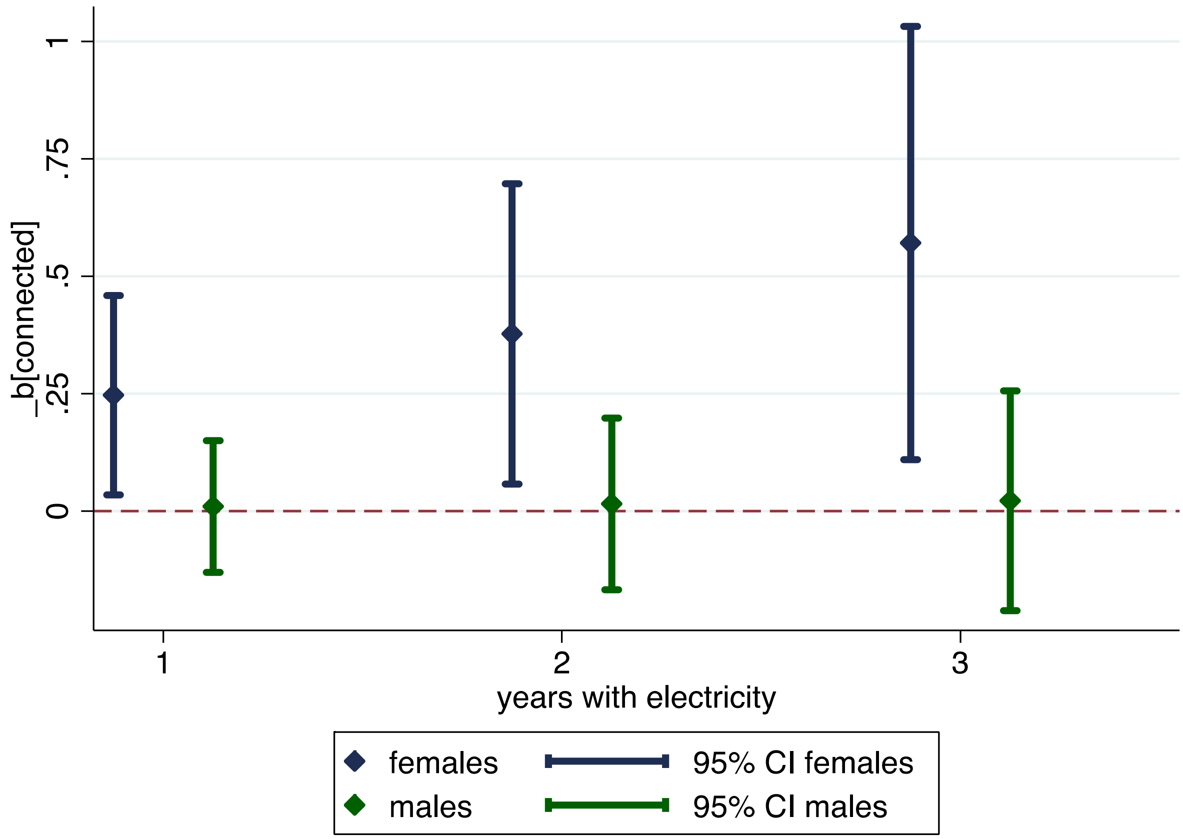
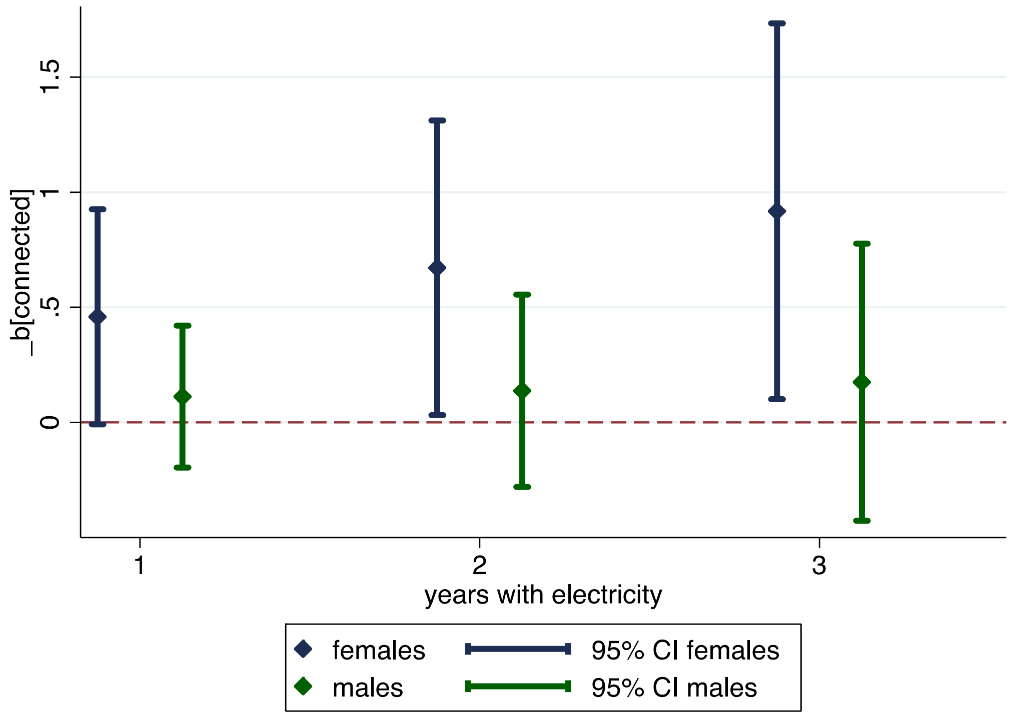


Figure 29: Years on the Grid and Probability of Engaging in Non-farm Employment



5.7 Energy Use

This subsection analyzes changes in traditional fuel use induced by electrification and suggests changes in kerosene use as the main channel through which electrification affected overnight PM2.5 concentration. We report the effects of electrification on energy use in Table 28. Our findings conform to the stylized fact that newly electrified households use electricity first and foremost for illumination.

The dependent variables in Table 28 indicate whether households use the particular energy source or not. The first column shows that voucher recipients were less likely to use kerosene and candles; although consistent with our previous discussion, the coefficients were statistically significant only in round two. However, the IV coefficients were negative and significant, at -0.33 for kerosene ($p < .01$) and -.10 for candles ($p < .05$). There were no significant changes in the use of car batteries, propane, or wood; however, as these sources are less important in households' energy budgets than kerosene, detecting an effect would require larger sample sizes. As mentioned in the preceding section, similar patterns arise in the non-experimental sample.

5.8 Electronic Appliance Ownership and Time Use

Table 29 shows that electrification led to important changes in appliance ownership. Since some households owned some of these appliances at baseline, the sample for each regression was formed by households that did not own that appliance at baseline. This underestimates the effects of electrification on appliance ownership, given that households may buy new appliances to replace old ones.

There were significant increases in appliance ownership of “leisure” items like TV sets and DVD players, but also in ownership of appliances that could be used for home production. Electrification led to increased ownership of refrigerators (54 percentage points), blenders (25 percentage points), and washers (13 percentage points). This is consistent with households starting small businesses based on home production for sale.[35] and is also consistent with non-significant changes in access to credit.

5.9 Income

Table 26 and Table 27 show the effect of electrification on household income. Given that most households had positive income, truncation at zero is not a problem. Note that we are not addressing how electricity would affect potential income (which would require addressing selection into different activities) but rather how income compared between on-grid and off-grid households. The IV estimates of connection on income suggest that electrification increased annual household income by around 1,600 USD per year. However, this estimate is noisy: its standard error is 1,165, so we cannot reject much smaller increases, not even a null increase. Although this is not formal proof, an effect of this magnitude is consistent with the average profits of non-farm businesses.

Some households reported non-labor income only, so selection started to be a more important issue in the labor income regression. Ignoring selection would suggest an increase in labor income of 4,800 USD, but once again, the standard errors are large. The estimate was significant at the 10 percent level only, so much more modest (or even null) results cannot be rejected. The results of non-labor income are presented for completeness only, given that selection is an apparent problem in this income category.

This, along with the estimates in Table 27, shows mixed evidence on the effects of electrification on household income. The non-experimental effects are more modest and more precisely estimated. The non-experimental effects suggest an increase of 55 USD in non-labor net income (18 percent increase from baseline) and 208 USD on labor net income (20 percent from baseline) —statistically significant at the 95 percent confidence level. The effect on total net income is 111 USD (8.8 percent of the baseline). The differences across these estimates show that the effects could be very large for the households that connected to grid because of the voucher.

5.10 Implications for Health Outcomes

This section analyzes the health implications of the observed reductions in PM2.5 concentration. First, we show that acute respiratory infection (ARI) among children under the age of six were lower among voucher recipients.[36] Next, we combined PM2.5 concentration with time allocation data to construct measure of exposure to PM2.5 for four typical household members (adult male,

[35] Although ironing clothes for the neighbors is a common activity in our sample, there is no significant increase in the number of households owning irons, probably because a non-trivial share of households owned charcoal irons at baseline. These households may have switched to electric irons, but the survey did not include information on the type of iron.

[36] We recognize that self-reported ARI carry several disadvantages, but this is the best proxy for health we have at hand.

adult female, male child, and female child), which allowed us to further gauge the health implications of the observed changes in PM2.5 concentration.

5.10.1 Acute Respiratory Infections among Children

Globally, lower respiratory infections caused 2.8 million deaths in 2010 (Lozano et al 2013); thus, they constitute a major public health concern. In this subsection, we show that the reductions in overnight PM2.5 concentration generated by household electrification had sizable effects on respiratory infections among children under six years old. The experimental sample included 192 children in this age range. Despite this relatively small sample size, there were large and statistically significant (at the 90 percent level of confidence) reductions in the incidence of ARI among children.

The dependent variable in Table 30 indicates whether the child had an episode of ARI in the four weeks prior to the survey (self-reported). When the explanatory variables were voucher, round, and their interactions, we found that vouchers led to a reduction of 16-18 percentage points at round three (significant at the 90 percent), depending on whether the regression controlled for baseline characteristics and sub-district FE. The IV estimate on connection was -0.65, meaning that electrification reduced ARI incidence by 48 percentage points. However, it is important to notice that this result was significant at the 90 percent of confidence, and thus we cannot reject much more modest reductions.

The point estimates for rounds four and five were not significant, consistent with the catching up argument. It is worth noting that ARIs did not bounce back up to their original levels, since ARI incidence fell from 44 percent to 10 percent between rounds three and four. Consistent with the analysis of PM2.5 concentration, this shows that the effects of electrification were similar irrespective of whether a household received a voucher.

5.10.2 PM2.5 Exposure and Health Risks

In this subsection, we construct measures of PM2.5 exposure based on PM2.5 concentration and time allocation data from the household survey. The resulting exposure measures were lower for the encouraged group (voucher recipients), but the gains were unequally distributed among household members. Adult males experienced reductions in exposure to PM2.5 nearly twice as large as adult females, mostly due to inequality in time spent in the kitchen, where PM2.5 concentration is highest, and outdoors, where PM2.5 concentration is lowest. It is important to keep in mind that given imperfect compliance with voucher assignment, the reduced form coefficients (the effect of voucher on overnight PM2.5 concentration) constitute lower bounds of the effects of electrification. Compliance with voucher assignment was imperfect because some voucher recipients remained off-grid, while some households that did not receive vouchers got a grid connection.

We start by defining exposure to a pollutant as the amount (in milligrams) of the pollutant that effectively makes its way into a person's respiratory system (as discussed earlier, PM2.5 enters the deep lung and the bloodstream.) For a particular activity conducted for a given amount of time, exposure is estimated as the product of the concentration in the environment where this activity took place multiplied by the inhalation rate while performing said activity. Daily exposure can be estimated by adding together all activities performed during the day, as in the following equation:

$$Exposure = \sum_{j=1}^{J} time_j \times concentration_j \times inhalation_j$$

where $time_j$ is time spent in activity j (hours), $concentration_j$ is PM2.5 concentration in the room while performing activity j (milligrams/m^3), and $inhalation_j$ is the estimated inhalation rate while performing activity j (m^3/hour).

The detailed information regarding the time allocation of the household members gave our study an important advantage over the typical studies found in the literature.[37] It allowed us to estimate fairly accurately the time that individuals spent indoors and, moreover, the type of activity they were performing. Knowing the type of activity and the time allocated to said activity allowed us to input an average inhalation rate adequate for that particular activity during that time, instead of just inputting the average intake per day. For a given pollutant concentration, this leads to better estimates of the amount of pollutants that effectively makes its way into a person's respiratory system.

The time allocation data was collected for up to four household members: the household head, his or her spouse, and up to two school-age children. This allowed us to estimate PM2.5 exposure for four "synthetic individuals": adult female, adult male, female child, and male child. Table 31 presents average time allocation in four type of activities for each of our synthetic individuals. The male and female heads reported 8.9 hours of sleep per day, while the children reported 9.5-10 hours of sleep per day. Time spent at home during the evening (awake) was similar for all members (slightly lower for the children, who sleep 0.5-1.0 hours more than the household heads). The starkest differences were observed in time spent in the kitchen and time spent outside the home. While the female head reported 2.5 hours per day in the kitchen, the male head reported spending an average of just five minutes in the kitchen. On the other hand, the female head reported spending an average of 2.7 hours outside the home, while the male head reported spending 7.8 hours outdoors (consistent with the length of a work day).

The differences in time allocation that arose from this analysis suggest that adult females were more exposed to PM2.5 (as well as other pollutants) since they spent considerably more time in the kitchen than any other household member. On the other hand, males spent almost one-third of their time outside the home; their main activity was farming and walking to and from the farm, in both of which activities it can be safely assumed that exposure to PM2.5 is negligible.

Next, we make explicit the assumptions about the PM2.5 concentration in the environments in which these activities were likely conducted. As shown earlier, average PM2.5 concentration in the living room during the evenings was 0.40 milligrams/m^3. We take this as representative of any room in the house, except the kitchen, between 5:00 pm in the evening and 7:00 am the next

---

[37] This is even more so when compared to outdoor pollution studies, in which time spent outdoors is an unobservable variable

morning. Based on the subsample of households for which we have three-day measurements, we estimated the average PM2.5 concentration in the living room during daytime (from 7:00 am to 5:00 pm) to be 0.26 milligrams/m^3. We took this as representative of the rooms in the household during the daytime, again with the exception of the kitchen. Since we did not collect data on PM2.5 concentration in the kitchen, we used 0.90 milligrams/m^3, which corresponds to average PM2.5 in the kitchen in Guatemalan households (Northcross et al 2010). This figure seemed to be an adequate assumption in our context since it corresponded to a neighboring region where households also rely on fuel and wood for cooking. This made our exposure estimates adequate for households that rely on wood for cooking, and even conservative given that it is not uncommon to find cases where the average concentration is above 2.0 milligrams/m^3. We assumed household members would not be exposed to PM2.5 when they were not home. This assumption did not seem to be too restrictive for the population in our study setting, since most of the time outside the home is spent in outdoor activities, like farming, and very little time is spent conducting activities outside the home that suggest exposure to PM2.5 (e.g. visiting friends at night).

The third and final component in the exposure equation is the inhalation rate. Since inhalation rate depends on age, we estimated it for the sample averages: 43 for the female head, 47 for the male head, 11 for the female child, and 13 for the male child. Air inhalation rates per activity were based on the Environmental Protection Agency (EPA) Exposures Handbook (EPA 2011). Most activities conducted at home are classified as “light activity tasks" by the EPA. Light activities include cooking, washing dishes, ironing, watching TV, doing desk work, writing and typing, and walking at a speed of up to 2.5 mph (2.9 kilometer/h). The average inhalation rate for these activities is 0.78 m^3/hour, while the average air inhalation rate while sleeping is 0.30 m^3/hour, again similar for the four synthetic individuals. The inhalation rate for activities conducted outside the home will vary greatly depending on the intensity of those activities. For instance, walking to work could be classified as light or medium intensity, depending on the speed at which the person is walking. Farming, on the other hand, could be classified as medium to high intensity, but lunch breaks would be light activity. However, the assumption made earlier about PM2.5 concentration being zero outside the home makes the inhalation rates of these activities irrelevant for total exposure.

With these three components, we estimated exposure rates for the four synthetic individuals. Results are shown in Table 31. Estimated exposure measures were highest for the female head, at 5.68 milligrams/day, and lowest for the male head, at 3.20 milligrams/day. The exposure measures for children were in between, with females 4.23 milligrams of PM2.5 per day and males to 3.72 milligrams/day. Taken plainly as units of PM2.5, these concentrations were equivalent to 8.0 cigarettes a month for the male head, 14.2 for the female head, 10.6 for the female child, and 9.3 for the male child.[38] The scientific evidence is as yet inconclusive regarding whether PM2.5 generated by cigarettes is worse than that generated by kerosene combustion.

The changes in exposure were large for all members (all above 30 percent), but these gains were unequally distributed across household members. The male head benefitted the most, with a

[38] One cigarette is estimated to have 12 milligrams of PM2.5.

reduction in exposure of 59 percent, while the female head benefitted the least, with a reduction of 33 percent. As pointed out previously, these differences were due to females spending more time than males in the kitchen, where pollutant concentration is highest, while males spent more time than females outside the home, where pollutant concentration is lowest.

To date, there are no dose-response functions linking exposure to PM2.5 from kerosene combustion to health outcomes. However, Pope et al (2011) present an estimate of a dose-response function linking PM2.5 from first- and second-hand tobacco smoke to lung cancer and cardiovascular diseases. At the bottom of Table 31, we present the relative risks that would be associated with the exposure levels found in our estimations of exposure to PM2.5 if the health effects of PM2.5 from kerosene combustion were similar to those from tobacco smoking. It is worth noting that the dose-response function estimated by Pope et al (2011) is non-linear and has support in exposures of from 0.18 to 0.90 milligrams/day and then above 18 milligrams/day (but not between 0.90 and 18 milligrams/day, while our estimates range from 3.2 to 5.7); thus we need to rely on the linear interpolation of the values up to 18 milligrams/day. This caveat does not seem to be a large weakness since the linearization is highly accurate in this neighborhood, with R-squared values of 0.99 (lung cancer), 0.96 (ischemic heart disease), 0.86 (cardiovascular disease), and 0.80 (cardiopulmonary disease).

The changes in exposure were associated with a decrease in the relative risk of lung cancer (compared to a person with no exposure to PM2.5) from 4.0 to 3.1 for the female head, nearly a 25 percent reduction. The relative risk for the male head fell by 33 percent, while the reduction for the female child is 25 percent and for the male child was almost 30 percent. The estimated reductions in the relative risk of ischemic heart disease, cardiovascular disease, and cardiopulmonary disease were between three and four percent. Consistent with the results for lung cancer, these changes were higher for adult males, but the differences at these levels of exposure were relatively small.

5.11 Implications for Infrastructure Financing

In this section, we study the conditions under which the utility company would benefit by sharing connection costs with project beneficiaries, thus increasing its client base early in the project. Note that the electric company is a natural monopolist that cannot set the price of electricity: this price is determined by the regulator, or costs. However, we argued that the electric company could increase its customer base by paying a fraction of households' connection fees. This practice is not uncommon; for example, there are mobile phone companies in some developing countries that pay for a fraction of the phone in exchange for a commitment of use.

Let household benefits from electrification be given by b, which follows some density function f(b). In this model, households will connect if b>0, and will not connect otherwise.

Electric utility profits can be expressed as:

$$\Pi = \sum_{i=1}^{N} T_i(x)(R(q_i)q_i - x) - C_F$$

N is the total number of households within reach of the grid, T_i=1 if household i is connected to the grid and 0 otherwise, $R(q_i)=p(q_i)-c(q_i)$ is the gross profit obtained from household i (gross of fixed and connection costs), q_i is the quantity of electricity consumed, x is the part of the connection fee paid by the firm, and C_F is the fixed cost of grid extension.

In each round, households can be classified as follows: always-takers, never-takers, and compliers. *Always-takers* are households that would have connected even without the discount. *Never-takers* are households that would not connect even if they received the discount. Finally, *compliers* are households that would connect if and only if they received the discount.

Note that the status of a household as a complier, never-taker, or always-taker is contingent on the period. A household may be a complier in round t and an always-taker in round s > t. For instance, take a household that would have decided to connect in period three without the voucher, but with the voucher decides to connect in period 2. This household is a complier in period 2, but an always-taker in period 3. Despite the fact that it is not possible to know which household is a complier, the size of the complier subpopulation is given by the estimates of the β coefficients in the adoption equations (section 3.1).

Formally, x has an effect on the extensive margin (through new customers) as well as on the intensive margin (more consumption by existing customers). To simplify matters, we assumed the intensive margin to be zero. This simplified the algebra at little cost in terms of insights. This assumption is perhaps more appealing if we think of inter-temporal extensions to this model, where lifetime electricity consumption would vary little by having received a 20 USD or 50 USD discount on the connection fee.

We set the equilibrium such that the marginal cost equals the marginal revenue. Assuming no effect on the intensive margin (i.e. that q_i does not depend on x), this simplifies to:

$$\sum_{i=1}^{N} \frac{\Delta T_i}{\Delta \alpha}(R(q_i)q_i - x) - \sum_{i=1}^{N} T_i = 0$$

Note that $\Delta T_i/\Delta\alpha$ can take only one of two values. It is zero for always-takers and never-takers, and one for compliers. In this context, always-takers are households with b > 0, compliers are households with –x > b > 0, and never-takers are households with b < -x.

$$\frac{\Delta T_i}{\Delta \alpha} = \begin{cases} 1 \; if \; -x < b < 0 \\ 0 \qquad otherwise \end{cases}$$

To keep simplifying matters, assume that q_i is constant among compliers. Thus,

$$R(q_i)q_i = Rq, \forall i\!: -x < b_i < 0$$

With this, we can re-write the first order condition as follows[39]:

[39] We have omitted an intermediate step. To wit:

$$Rq \times NPr\left(\frac{\Delta T_i}{\Delta x} = 1\right) = x \times N \times Pr\left(\frac{\Delta T_i}{\Delta x} = 1\right) + N \times Pr(\Delta T_i = 1)$$

$$Rq\int_{x}^{0} f(b)db = x\int_{x}^{0} f(b)db + \int_{0}^{\infty} f(b)db$$

The above equation means that the additional gains the utility obtained from each complier at this new level of x must be enough to pay x to each of those compliers, plus the marginal subsidy increase to each always-taker (those who would have decided to connect before the incremental change in the subsidy). This confirms the initial insight that if compliers generate enough gains to the utility to pay for the additional subsidy to the always-takers, it is profitable for the utility to offer a strictly positive subsidy.

We show this in the case of normally distributed b. If b ~ N (μ, σ), it can be shown that:

$$\left[\Phi\left(\frac{x^* + \mu}{\sigma}\right) - \Phi\left(\frac{\mu}{\sigma}\right)\right](Rq - x^*) = \Phi\left(\frac{\mu}{\sigma}\right)$$

, which implies that if Rq > x*, then x* > 0.

To find an upper bound to x*, note that the model discussed in section three implies an inverse relationship between q_i and x (Zilberman and Liu 2011). Marginal consumers, those who need a discount to connect, will consume less than the average consumer. The larger the required discount, the smaller q_i. Thus, there is a threshold value for x, call it x_T, such that for any x > x_T,

$$Rq\int_{x}^{0} f(b)db < x\int_{x}^{0} f(b)db + \int_{0}^{\infty} f(b)db$$

This simple model can form the base of a dynamic optimization problem in which the firm either subsidizes in the first period only or subsidizes at every period. In its current form, the model shows some interesting insights. If the compliers can pay for themselves (Rq > x), then the optimal discount is positive. To pin down the optimum discount, we would need to impose some more structure on the problem. Since this paper is not about the optimal subsidy, we leave this issue for future research.

There are many reasons why this type of arrangement may not be implemented widely. The most obvious reason would be due to lack of knowledge of f(b). Next, it could also be the case that the electric company may face the probability of default in payments. Note that the simple model can incorporate the probability of default by multiplying the left-hand side of the equation by the probability of payment. However, more sophisticated modeling of the probability of default and its associated costs could provide more interesting insights. Third, electric companies may simply not have come up with this solution yet.

Next we perform some calculations to simulate revenues per household under different subsidy schemes in the initial years of the electrification program. First, we have the electrification rate per year under each type of voucher, similar to what we obtained in the summary statistics in Table 7. To get at the average revenues perceived by the electric utility, we used the average bills for households with a formal grid connection. The average electric bill in round two was 8.32 USD per month, so the average household generated 99.84 USD of revenues per year.

Multiplying the connection rate times the average bill, we found that the utility received 49.92 USD per household in the area. If the firm shares 20 percent or 50 percent of the cost, the figures are 60.90 USD and 66.92 USD, respectively. To find the most profitable alternative, we subtracted the connection costs that the company shared in each scenario. Note that the company paid 20 USD for all connections in round two and for the increase in connection rates in the future rounds.

With this, we can estimate revenues under each of the subsidy schemes. Although this is not vital for the results, we assumed a five percent interest rate to calculate the revenues in the first three years of operation. If the company provided no subsidy, it would receive roughly 180 USD per household in the first three years. Sharing 20 percent of the connection costs would bump those revenues by 10 percent, up to 200 USD per household. On the other hand, the 50 percent of the cost proves not to be profitable, since revenues would be 175 USD, below the no-subsidy scenario.

6 Impact Mechanisms and Conclusions

This report provides the first experimental evidence on some of the main mechanisms through which household electrification affects human capital formation and income in rural settings.

6.1 Impact Pathways Consistent with the Results Presented

First, we found that vouchers increase adoption of electricity by reducing connection costs, ameliorating credit constraints, providing incentives not to procrastinate in the decision to connect to the grid, and perhaps even increasing awareness about the benefits of electrification. We also analyzed the empirical evidence to try to identify which of these channels likely played a larger role. It does not appear that vouchers worked as commitment devices.[40]

If households needed incentives not to procrastinate, both types of vouchers should have a similar effect in each round. By the end of the study they do, but the timing of adoption is not consistent with this story: high-discount vouchers have a higher adoption rate than low-discount vouchers in the first follow-up surveys. In addition, the non-encouraged group, i.e. households that did not receive vouchers, also had high connection rates (50 percent) starting in the second round; this rate rose to 80 percent by the third round. This behavior is more consistent with the vouchers reducing connection costs than with them acting as commitment devices.

Although the survey did not include questions on information about the electrification program itself, it seems unlikely that vouchers increased program awareness. Projects of this type are easy to observe since they require erecting posts and other types of construction work.

The reimbursement provided by the discount vouchers could have been used to pay the cost of a loan, thus lifting credit constraints. However, access to credit (either formal or informal) did not increase among voucher recipients. This is not to say that these households were not credit-constrained, but rather that vouchers did not increase access to credit markets.

In addition, our findings conform to the stylized fact that newly electrified households use electricity first and foremost for illumination. There were no significant changes in the use of car batteries, propane, or wood. These energy sources were less important in the households' energy budget than kerosene, so detecting an effect for these sources would require larger sample sizes.

[40] In behavioral economics literature, a commitment device is a means of controlling future impulsive behavior, thus aligning choices to the individual's long-term goals. Vouchers may have acted as commitment devices had they incentivized households to not indefinitely postpone their decision to connect to the electric grid.

As mentioned in the preceding section, similar patterns arose in the non-experimental sample. There are significant increases in appliance ownership of "leisure" items like TV sets and DVD players as well as in ownership of appliances that could be used for home production. Electrification led to increased ownership of refrigerators, blenders, and washers. This is consistent with households starting small businesses based on home production for sale.

Second, and with regard to the role that connection costs and spillover effects play in the adoption of formal electric connections, we observed voucher recipients connecting one to two years earlier than the control group, but the control group caught up by the end of the study. Thus, in our setting, households responded to time-limited reimbursements by adopting earlier than the counterfactual. Spillover effects seemed to play an important role in adoption, and their effects did not reduce with time. An additional connection within 100 meters of a household increased the probability of that household connecting formally to the grid by 10 percentage points, roughly the same increase generated by vouchers.

We also analyzed the possibility that households adopted informal connections by modeling the probability of households switching between different alternatives and by estimating an ordered choice model. The first strategy suggests that households with informal connections at baseline were much more responsive to vouchers and spillover effects, while these variables did not affect the adoption rates of households with no connection at baseline. The ordered choice analysis showed that vouchers (allocated directly or to neighbors) increased the probability of formal connections, and decreased the probability of informal ones. This solved any concern about informal connections crowding out formal ones. Among our treatment group, households with informal electricity at baseline were more likely to take up the voucher, as were households with property titles. Better-off households (no drifters, higher income) were also more likely to take up the voucher. Despite gender differences in the benefits of electricity, the gender of the household head was not correlated to voucher take-up. Similarly, other characteristics of the household head like age or literacy status were uncorrelated with take-up.

Finally, we show that the electric utility can actually increase its revenues by providing discount vouchers in a fashion similar to what we used in this study, given that vouchers increase the utility's customer base and revenue flows. In our study, the "optimal" strategy consisted of paying 20 USD of the inspection fee since connection rates among low- and high-discount voucher recipients were very similar.

Third, we found that electrification increases investment in education among school-age children and participation in income-generating activities among adult women. The increases in educational investment materialized through an increase in participation in educational activities. Electrification increased the probability of studying at home by 54 percentage points and of performing other school-related activities (time spent in school, time spent commuting between school and home) by 84 percentage points. One of the main mechanisms for this increase was a dramatic improvement in the study environment, which raises the returns to time studying. A second mechanism may be changes in aspirations: if parents feel that electrification is a sign of progress, which would make their children's schooling more profitable once they reach adulthood, they are more likely to send the kids to school.

Fourth, we found robust results that adult females increased their participation in income-generating activities as a result of electrification. Electrification increased the probability of operating a home business by 12 percentage points. This is more than a 150 percent increase compared to the control group. When splitting the sample by gender, only the point estimates for females were statistically significant. The observed income-generating activities were generally small-scale activities, mostly consisting of offering services like washing and ironing clothes or preparing food for sale. These activities require relatively small investments and do not require the participation of the male head, who typically is the main income earner in the household; they thus imply low risk to the household. However, these activities generate on average 1,000 USD per year, suggesting the effect of electrification on income controlled by women was non-trivial and could unleash important gender dynamics in the household. The literature associates higher income controlled by women with higher intra-household bargaining power among women and with improved welfare outcomes among children (better nutrition, higher expenditure on education).

Fifth, the experimental estimates of electrical connection on income suggests that electrification increased annual household income by around 1,600 USD per year. This is the first time that mid-term effects have been identified in addition to short-term effects.

Finally, the evidence presented also shows that electrification leads to improvements in IAP, which reduced the incidence of ARI among children and lowered exposure to pollutants among adult household members. This is the first experimental evidence that electricity leads to important improvements in welfare through substantial, immediate, and sustained improvements in indoor air quality. Given imperfect compliance with voucher assignment, our experiment produced a lower bound for reductions in average PM2.5 concentration of 67 percent, with the IV estimates suggesting reductions of the order of 95 percent. The most salient mechanism behind the improvements in IAP was a substitution away from kerosene lighting, while there were no discernible changes in the use of other traditional lighting sources or in cooking practices. Given that the mechanism is clear in this context, our results suggest that other clean artificial lighting technologies, like solar lamps, could have similarly strong effects on IAP in households that are too isolated for grid electrification to be feasible. As a result of this drop in overnight PM2.5 concentration, we saw large and significant reductions in ARI among children under the age of 6.

Our IV estimates suggest that connected households had a 65 percent lower incidence of ARI. To gauge further implications on health for populations over the age of six, the observed changes in PM2.5 concentration, together with time-use data and complementary PM2.5 data from Northcross et al. (2010), were used to estimate the change in daily PM2.5 exposure. Our estimation suggests that reductions in average exposure to PM2.5 are large for all household members, but they were distributed unequally among household members. Adult males typically benefitted the most from the estimated reductions in PM2.5 brought about by electrification, with 59 percent lower exposure. Adult females benefitted the least, with estimated reductions of 33 percent. The figures for children are 46 percent (males) and 39 percent (females).

To assess the magnitude of the health effects of this reduction in exposure, we input these figures into the dose-response function estimated by Pope III et al. (2011). The implied risk-ratio for lung cancer fell dramatically, from 4.0 to 3.1 for adult females and from 2.7 to 1.8 for adult males, with the respective figures for children falling from 3.2 to 2.4 (females) and from 3.0 to 2.1 (males). The main caveat in this analysis is that these risk ratios rely on the assumption that the health effects of PM2.5 from kerosene combustion are similar to those of cigarette smoking; thus, the exact figures depend on the health effects of PM2.5 from kerosene combustion relative to those of cigarette smoking. However, given the magnitude of exposure and the reductions we find, we argue that the health effects will be large.

This report also contributes to the environmental health literature by providing evidence of a strong, positive relationship between kerosene use and PM2.5 concentration in a setting with high reliability on biomass for cooking (70 percent of the sample households). This is important given the still scarce evidence regarding the relationship in the field between kerosene use and IAP measures. Despite the large improvements in indoor air quality brought about by electrification, PM2.5 concentration in these households was still high due to the use of fuelwood for cooking; this is also the reason behind the higher exposure levels among females. Thus, it remains important to continue advancing our understanding of the adoption and use of improved cook stoves. We also examined heterogeneity in treatment effect with CIC, a non-parametric approach that allows us to estimate the full distribution of treatment effects. Households up to the 20th percentile (lowest polluters) did not benefit from electrification, but there were large and significant reductions for all households from the 20th percentile onwards. It was also noticeable that the reductions got larger among households in the 60th percentile and are largest for the 15 percent of households with the highest pre-electrification PM2.5 concentration.

6.2 Lessons for Future Interventions and Policy Implications

A rigorous impact evaluation that includes appropriately selected control groups must be a part of rural electrification program designs. Budgeting for evaluation activities and engaging with evaluators at an early stage improves the likelihood of having a high-quality evaluation design. Additionally, if deviations occur after the design stage, the evaluators are better prepared to adjust the design so that the impact results remain informative to policymakers and future program designers.

Another takeaway is the need to specify the expected outcomes and the plausible sizes of impacts with a theoretical model or a program logic based on the existing research. If done at the beginning of the program, this will provide context to the kind of discussion in which policymakers should engage (e.g. if they should focus on health benefits or the potential to diffuse information campaigns to rural households). Moreover, the intervention and the evaluation would have benefited from allowing the evaluation teams to participate in the planning of the timeline for the deployment of transmission and distribution lines. However, this was not feasible given that the electrification deployment designs were already in place when we started the evaluation.

Although our results are consistent with what we found regarding short term impacts in Bernard & Torero (2015), it is important to point out that we also need to focus on external validity when assessing the impact of rural electrification. This can be achieved by evaluating large-scale rural electrification programs, which will provide an opportunity to test whether the results from small-scale impact evaluations translate to other settings. In the majority of impact evaluations, it is commonly assumed that the estimated treatment effects can be generalized to the whole population or to a new location in which no experiment was conducted. However, since individuals in a new location can have different observable and unobservable characteristics, the average treatment effect in that scenario can be significantly different from the average treatment effect obtained from experiments conducted in other locations. Several authors have protested against policy recommendations that they believe are based on implicit extrapolation from a small number of experiments to a wide variety of dissimilar contexts (Deaton 2010; Pritchett and Sandefur 2013). Empirically, a growing body of work shows that identical policies have different effects among individuals with the same observed characteristics living in different contexts (see Allcott 2012; Attanasio, Meghir, and Szekely 2003), because unobserved differences between populations remain. Hence, we need a method that accounts for heterogeneity across locations, or we need to design an evaluation that takes this issue into account from the beginning.

Components which we have not stressed thus far but that are important to keep in mind are the complementarities in the provision of different types of infrastructure. Large projects can provide an opportunity to explore complementarities with other infrastructure programs, such as mobile telephones, road access, and improved water and sanitation access. They can shed light on the most welfare-enhancing policy options when deciding what types of infrastructure to provide in rural areas, especially to poor households.

The benefits of using different methods in the same evaluation cannot be understated. The use of experimental methods and the inclusion of the non-experimental methods taking advantage of the longitudinal nature of the impact evaluation design was extremely important in this evaluation, and the support of MCC was essential. In the future, it will be important to budget for a larger sample size for the experimental design and to explore variations in the discount that consider free connections. This will help us better understand the barriers that households face due to upfront connection costs and to come up with better policy recommendations as to the optimal connection subsidies in different contexts to increase penetration. Moderate discounts toward the connection fee can increase adoption of connections in the early years of a grid expansion program. In addition, electric utility companies can increase their revenues by providing discount vouchers in a fashion similar to that used in this study, given that vouchers increase the utility’s customer base and revenue flows.

To better understand the long-term effects of rural electrification, especially the effects on income through increases in productivity due to health improvements, it is important to measure indoor air quality and health outcomes using objective indicators. Future interventions should explore new ways of measuring indoor pollution as well as measuring objective outcomes that are related to indoor pollution. In addition, electrification includes environmental factors like reduction in kerosene use. Kerosene soot is one of the main sources of black carbon in the atmosphere, so reduced kerosene had positive environmental effects. This target needs to be incorporated as an additional benefit in the cost-benefit analysis of rural electrification projects.

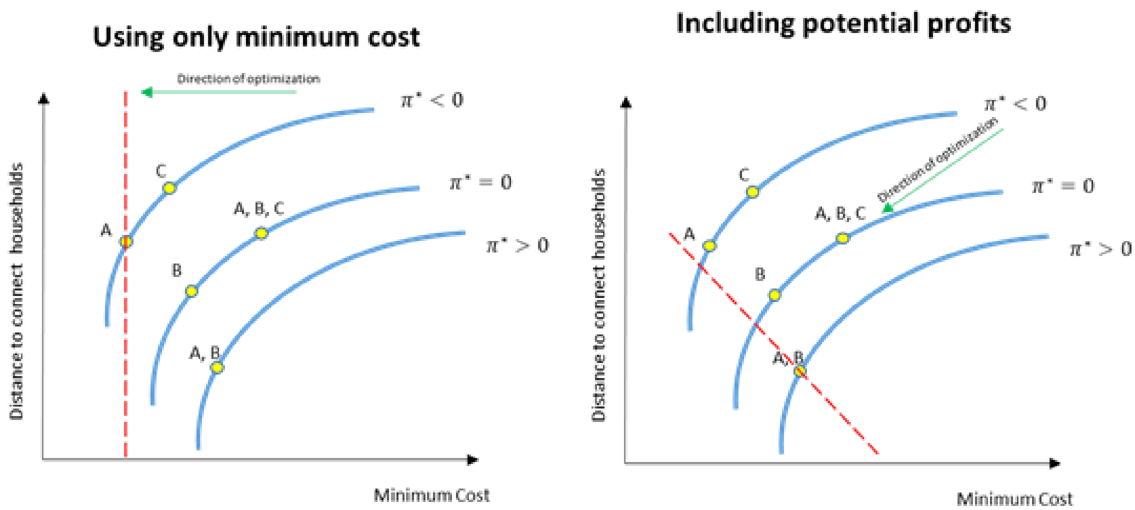
The state has a large role to play in maximizing the effects of electrifications programs. For example, given the increased operation of home businesses and investment in schooling, governments should strive to provide the conditions for these businesses to grow and should increase school investments. In this Compact, these additional and complementary interventions took the form of investments in education, productive activities, community infrastructure, etc. The hope is that the incentives are aligned across these projects and that the impacts could be even bigger than the ones we have shown in the current study.

We conclude this report by highlighting the need for using an objective function that casts a wider net when deciding where to place electrification projects. Focusing solely on cost minimization that is rooted in the difficulty or cost to access a household can result in missed opportunities, considering the spillover effects we found. When deciding where to employ the electric grid in rural areas, it is imperative to take into account the potential profits, specifically a measure of profits that includes the agricultural potential of these areas and the increase in the consumer base due to spillovers.

To illustrate this point, consider that, normally, the implementer solves a cost minimization problem when deciding where to extend an existing grid. There seems to be little attention paid to profit maximization; that is, considering that more remote (and thus more expensive) areas might have high productive potential that would be realized by electrification, thus making the electrification investment ex-post profitable. The duality of cost minimization and profit maximization depends on the quasi-concavity of the production function and complete markets, situations that are not characteristic of the electricity sector—one can easily argue that there are increasing returns to scale in some parts of the production function—and less so in developing countries. This implies that a planner using cost functions or profit functions as objective functions would make different decisions. To illustrate the point, suppose that we have three households, A, B and C, that we want to connect to the electric grid. As shown in Figure 30, if we connect household A at minimum cost, we obtain negative profits, and only connect household A and adjacent households. If we included the potential profits that can be obtained from connecting A to B and C, however, we would arrive at a different conclusion. We would move southwest in the quadrant, to find the allocation that maximizes profit at a minimum cost. We arrive at point (A, B) where profits are positive and households A, B, and adjacent are connected to the grid. Note that it is not always profitable to connect all households, as evidenced by the point (A, B, C) being at the zero isoprofit curve.

By using an objective function that incorporates isoprofits and cost minimization, rural electrification programs have the opportunity to reach more poor households and have larger impacts in the lives of the rural poor by providing new opportunities and enhancing the complementarities between the agricultural and non-agricultural sectors.

Figure 30: Optimization of electric grid using minimum cost and including potential profits



7 Next Steps and/or Future Analysis

7.1 Dissemination Procedures

The results presented in this report will be compiled in an academic paper to be published in policy and development journals. Presentation dissemination efforts will include: presentation of the report(s) to MCC Headquarters staff, presentation in MCC workshops, presentation of findings and key recommendations to local stakeholders, and presentation of the findings in other international development conferences.

8 References

Allcott, H. 2012. Site Selection Bias in Program Evaluation. NBER Working Paper No. 18373. Cambridge, MA, US: the National Bureau of Economic Research.

Attanasio, O., C. Meghir, and M. Szekely. 2003. Using Randomized Experiments and Structural Models for “Scaling Up”: Evidence from the PROGRESA Evaluation. IFS Working Papers WP03/05. London: Institute for Fiscal Studies.

Balke, A. and J. Pearl. 1997. “Bounds on Treatment Effects from Studies with Imperfect Compliance" JASA 92(439) pp 1171-1176.

Bernard, T. 2010. "Impact Analysis of Rural Electrification Projects in Sub-Saharan Africa" The World Bank Research Observer, pages 1-19.

Bernard, T. and M. Torero. 2015. “Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia. Economic Development and Cultural Change, 63(3), 459-484

Chakravorty, U., K. Emerick, and M. Ravago. 2016. “Lighting Up the Last Mile: The Benefits and Costs of Extending Electricity to the Rural Poor” RFF Discussion Paper 16-22.

Chen, L et al. 2007. Emissions from laboratory combustion of wildland fuels: Emission factors and source profiles. *Environ. Sci. Technol.* 41:4317-25

Deaton, A. 2010. “Instruments, Randomization, and Learning about Development.” *Journal of Economic Literature* 48 (2): 424–455.

Devoto, Duflo, Dupas, Pariente, and Pons. 2009. Happiness on Tap: The Demand for and Impact of Piped Water in Urban Morocco. World Bank Working Paper.

Dinkelman, T. 2010. The effects of rural electrification on employment: New evidence from South Africa. AER, forthcoming

Duflo, E., Michael Greenstone and Rema Hanna. 2008. “Indoor Air Pollution, Health and Economic Well-being” Surveys and Perspectives Integrating Environment and Society 1, 1–9.

Duflo, Esther and Rohini Pande. 2007. Dams. Quarterly Journal of Economics, May.

Fan, CW, ad Zhang, N. 2001. Characterization of emissions from portable household combustion devices: Particle size distributions, emission rates and factors, and potential exposures. *Atmos. Environ.* 35: 1281-90.

Fried, S. and D. Lagakos. 2016. “The Role of Energy Investment in Africa’s Recent Growth: A Macroeconomic Analysis” Manuscript.

Haines A, et al. 2007. “Policies for accelerating access to clean energy, improving health, advancing development, and mitigating climate change” *The Lancet* 370: 1264-1281

Hana, Rema, Esther Duflo, and Michael Greenstone. 2012. “Up in Smoke: The Influence of Household Behavior on the Long-Run Impact of Improved Cooking Stoves”. *MIT Department of Economics Working Paper 12-10.*

Hiemstra-van der Horst, G., and A.J. Hovorka. 2008. “Reassessing the ‘Energy Ladder’: Household Energy Use in Maun, Botswana.” Energy Policy, 36(9): 3333–44.

Imbens, G. 2009. Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). NBER Working Paper, April 2009.

International Energy Agency. 2011. *World Energy Outlook*

Jacobson, A. 2007. “Connective Power: Solar Electrification and Social Change in Kenya.” World Development, 35(1): 144–62.

K. R. Smith et al., Energy and human health. Annual review of public health 34, 159 (Mar 18, 2013).

Kline, Patrick, and Enrico Moretti. 2011. Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. Manuscript.

L. P. Naeher et al., Wood smoke health effects: a review. Inhalation toxicology 19, 67 (Jan, 2007).

Lam N, et al. (2012) “Kerosene: A review of household uses and their hazards in low- and middle-income countries” *Journal of Toxicology and Environmental Health, Part B*, 15:396-432.

Lam NL, Chen Y, Weyant C, Venkataraman C, Sadavarte P, Johnson MA, Smith KR, Brem BT, Arineitwe J, Ellis JE, Bond TC. 2012. “Household light makes global heat: high black carbon emissions from kerosene wick lamps” Environ Sci Technol 46(24):13531-13538.

Lee, K., E. Miguel and C. Wolfram (2016) “Experimental Evidence on the Demand for and Costs of Rural Electrification”, NBER Working Paper 22292.

M. B. Epstein et al., Household fuels, low birth weight, and neonatal death in India: The separate impacts of biomass, kerosene, and coal. International journal of hygiene and environmental health, (Jan 21, 2013).

Madubansi, M., and C.M. Shackleton. 2007. “Changes in fuel wood Use and Selection Following Electrification in the Bushbuckridge Lowveld, South Africa.” Journal of Environmental Management, 83(4): 416–26.

McCracken John, et al. (2013) “Longitudinal relationship between personal CO and personal PM2.5 among women cooking with woodfired cookstoves in Guatemala” *PLOS ONE*; 8:1-4

Mills, E. (2005) “The specter of fuel-based lighting” *Science* 308: 1263-64.

Mobarak A, et al (2012) “Low demand for nontraditional cookstove technologies” *PNAS* 109:10815-10820.

Muller, E., R. D. Diabb, M. Binedella and R. Hounsomea. 2003. Health risk assessment of kerosene usage in an informal settlement in Durban, South Africa. Atmospheric Environment 37(15) May 2003, Pages 2015-2022

Pokhrel, Amod, Michael N. Bates, Sharat C. Verma, Hari S. Joshi, Chandrashekhar T. Sreeramareddy, and Kirk R. Smith (2010) “Tuberculosis and indoor biomass and kerosene use in Nepal: A case-control study" Environmental Health Perspectives 2010 April; 118(4): 558-564.

Pope CA, 3rd, Burnett RT, Turner MC, Cohen A, Krewski D, et al. (2011) “Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure-response relationships” *Environmental Health Perspectives* 119: 1616–1621.

Pritchett, L., and J. Sandefur. 2013. Context Matters for Size: Why External Validity Claims and Development Practice Don’t Mix. Working Paper 336. London: Center for Global Development.

Rollin HB, et al. (2004) “Comparison of indoor air quality in electrified and un-electrified dwellings in rural South African villages” *Indoor Air*; 14: 208–216

S. S. Lim *et al.*, A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380, 2224 (Dec 15, 2012).

Schare, S. and Smith, KR, (1995) Particulate emission rates of simple kerosene lamps. *Energy Sutain. Dev.* 2:32-35.

Sinton JE, et al. (2004) An assessment of programs to promote improved household stoves in China. *Energy Sustain Dev* 8(3):33-52.

Smith KR, et al (1993) One hundred million improved cookstoves in China: How was it done? World Development 21(6):941-961.

Smith KR, Peel JL (2010) “Mind the gap” *Environmental Health Perspectives* 118:1643–1645.

Stock, J.H., J.H. Wright, and M. Yogo (2002): “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” Journal of Business and Economic Statistics, 20, 518 – 529.

Strobl, Eric, and Strobl, Robert (2011) “The Distributional Impact of Large Dams: Evidence from Cropland Productivity in Africa” Journal of Development Economics 96 (2011): 432-450.

World Health Organization (2006) *Fuel for life: Household Energy and Health* (World Health Organization, Geneva)

Zhou Z, et al. (2011) “Household and community poverty, biomass use, and air pollution in Accra, Ghana” *PNAS;* 108: 11028-11033

9 Tables of Results

Table 7: Summary Statistics by Subsample

	Not connected	Voucher Subsample	Air Quality Subsample Experimental	Air Quality Subsample Non-Exp.
Age of household head	49.02	49.54	48.18	50.79
	(17.55)	(16.82)	(17.49)	(17.32)
Household head is male	0.68	0.69	0.73	0.73
	(0.47)	(0.46)	(0.44)	(0.45)
Average age in hosehold	31.23	30.80	29.84	33.25
	(17.72)	(17.05)	(16.43)	(19.04)
Household size	4.34	4.53	4.64	4.53
	(2.36)	(2.32)	(2.16)	(2.50)
Total dependency ratio	0.44	0.45	0.42	0.44
	(0.28)	(0.27)	(0.26)	(0.28)
Maximum schooling in the household	5.47	6.05	5.90	4.21
	(4.35)	(4.34)	(4.23)	(3.13)
Schooling of the household head	2.23	2.36	2.53	1.45
	(3.55)	(3.47)	(3.71)	(2.18)
household head is literate	0.48	0.54	0.57	0.39
	(0.50)	(0.50)	(0.50)	(0.49)
Income pc, 1000USD per year	0.65	0.77	0.62	0.43
	(1.72)	(4.01)	(1.11)	(0.67)
Monthly expenditure in kerosene	2.49	2.11	2.13	4.83
	(4.17)	(3.90)	(3.65)	(5.12)
Monthly expenditure in propane	1.82	2.09	1.96	0.88
	(2.86)	(2.96)	(3.15)	(2.25)
Monthly expenditure in candles	0.50	0.46	0.36	0.80
	(1.65)	(1.57)	(1.32)	(1.91)
Monhtly expenditure in car battery rchg	0.12	0.08	0.12	0.25
	(0.65)	(0.56)	(0.65)	(0.92)
Cooks with wood	0.72	0.70	0.70	0.90
	(0.45)	(0.46)	(0.46)	(0.30)
Informal electricity	0.46	0.38	0.46	0.04
	(0.50)	(0.49)	(0.50)	(0.20)
Agrees with the following statement				
Electricity illuminates better than kerosene.	0.95	0.96	0.96	0.95
	(0.22)	(0.20)	(0.20)	(0.22)
Powering a TV is cheaper w/elect than battery.	0.76	0.79	0.81	0.77
	(0.43)	(0.41)	(0.39)	(0.43)
Cooking with electricity is not convenient	0.58	0.52	0.53	0.53
	(0.49)	(0.50)	(0.50)	(0.50)
Electricity is very expensive	0.52	0.49	0.54	0.40
	(0.50)	(0.50)	(0.50)	(0.49)
Woodsmoke generates respiratory problems	0.88	0.86	0.86	0.88
	(0.32)	(0.35)	(0.35)	(0.33)
Kerosene is not an expensive source of lighting	0.34	0.35	0.41	0.34
	(0.48)	(0.48)	(0.49)	(0.48)
Kerosene is the best way to illuminate my household	0.22	0.21	0.25	0.28
	(0.41)	(0.41)	(0.43)	(0.45)

Notes: Standard deviation in parenthesis. Samples - Col(1): all *EHEIPCER* households that were off-grid at baseline (N=2014). Col(2): Households in San Miguel and Chalatenango among whom vouchers were randomly allocated (N=500). Col(3): random subset of the households in column 2 (N=150). Col(4): *EHEIPCER* households that had not connected to the grid by the first follow-up (N=207), and from the same departments as those in column 2. See main text for details.

Table 8: Summary Statistics and Balance by Treatment Arm

	Control Group	20% Discount	Diff: C-20%	50% Discount	Diff: C-50%
Age of household head	49.20	50.80	-1.60	48.99	0.21
	(1.47)	(1.25)	(1.92)	(1.29)	(1.96)
Household head is male	0.62	0.72	-0.10*	0.72	-0.10*
	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)
Household size	4.19	4.65	-0.46*	4.82	-0.63**
	(0.18)	(0.19)	(0.27)	(0.18)	(0.27)
Total dependency ratio	0.47	0.44	0.02	0.43	0.03
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Maximum schooling in the household	5.51	5.76	-0.26	5.76	-0.26
	(0.33)	(0.33)	(0.47)	(0.32)	(0.47)
Schooling of the household head	1.90	2.03	-0.14	2.23	-0.33
	(0.25)	(0.25)	(0.36)	(0.26)	(0.37)
Household head is literate	0.49	0.49	-0.00	0.52	-0.03
	(0.04)	(0.04)	(0.06)	(0.04)	(0.06)
Income pc, 1000USD per year	0.55	0.52	0.03	0.57	-0.02
	(0.12)	(0.07)	(0.13)	(0.08)	(0.14)
Monthly expenditure in kerosene	2.96	2.56	0.41	2.20	0.76
	(0.39)	(0.32)	(0.50)	(0.27)	(0.46)
Monthly expenditure in propane	1.69	2.11	-0.42	1.78	-0.09
	(0.25)	(0.22)	(0.33)	(0.22)	(0.33)
Monthly expenditure in candles	0.57	0.55	0.01	0.55	0.01
	(0.14)	(0.13)	(0.19)	(0.13)	(0.19)
Monhtly expenditure in car battery rchg	0.12	0.04	0.08	0.12	0.00
	(0.06)	(0.03)	(0.06)	(0.05)	(0.07)
Cooks with wood	0.76	0.73	0.04	0.73	0.03
	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)
Informal electricity	0.39	0.50	-0.11*	0.48	-0.09*
	(0.04)	(0.04)	(0.06)	(0.04)	(0.06)
Agrees with the following statement					
Electricity illuminates better than kerosene.	0.96	0.95	0.01	0.97	-0.00
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Powering a TV is cheaper w/elect than battery.	0.79	0.74	0.05	0.81	-0.03
	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)
Cooking with electricity is not convenient	0.61	0.46	0.15***	0.50	0.11*
	(0.04)	(0.04)	(0.06)	(0.04)	(0.06)
Electricity is very expensive	0.54	0.43	0.10*	0.47	0.06
	(0.04)	(0.04)	(0.06)	(0.04)	(0.06)
Woodsmoke generates respiratory problems	0.87	0.84	0.04	0.87	-0.00
	(0.03)	(0.03)	(0.04)	(0.02)	(0.04)
Kerosene is not an expensive source of lighting	0.42	0.35	0.07	0.32	0.10*
	(0.04)	(0.04)	(0.06)	(0.03)	(0.05)
Kerosene is the best way to illuminate my household	0.28	0.20	0.08	0.20	0.08
	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)

Notes: Columns 1, 2, and 4 show the mean values for each of the treatment arms at baseline (standard errors in parentheses). Column 3 and 5 report the difference in means between the control group and households that received a 20% or 50% discount voucher, respectively (standard errors in parentheses). Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 9: Validating the Randomization of Voucher Density, OLS estimates, Experimental Sample

	(1) Age Head	(2) Head Male	(3) Head Schooling	(4) Household Size	(5) Adequate Walls	(6) Income	(7) Group Member
Vouchers, 0-100m radius	0.995*	0.004	0.189*	-0.017	-0.007	0.008	0.012
	(0.525)	(0.015)	(0.098)	(0.078)	(0.013)	(0.037)	(0.011)
Vouchers, 100-200m radius	0.589	-0.000	-0.022	-0.072	0.022	0.010	0.005
	(0.565)	(0.017)	(0.106)	(0.083)	(0.014)	(0.039)	(0.012)
Vouchers, 200-300m radius	0.036	0.001	-0.005	-0.043	0.003	0.043	-0.002
	(0.453)	(0.013)	(0.085)	(0.067)	(0.011)	(0.032)	(0.010)
Eligible Neighbors, 0-100m radius	-1.157**	0.002	-0.026	0.015	-0.001	-0.035	-0.005
	(0.496)	(0.015)	(0.093)	(0.073)	(0.012)	(0.035)	(0.011)
Eligible Neighbors, 100-200m radius	-0.452	-0.001	-0.095	-0.026	-0.019**	0.007	0.004
	(0.315)	(0.009)	(0.059)	(0.046)	(0.008)	(0.022)	(0.007)
Eligible Neighbors, 200-300m radius	0.073	0.002	0.027	0.027	-0.028***	-0.025	-0.003
	(0.246)	(0.007)	(0.046)	(0.036)	(0.006)	(0.017)	(0.005)
Mean Dependent Variable	49.76	0.69	2.08	4.58	0.76	0.55	0.13
Number of Observations	486	486	486	486	486	486	486
R squared	0.184	0.054	0.246	0.065	0.199	0.081	0.030

Notes: The dependent variable in each regression is indicated in the column name. Adequate walls indicates adobe, brick, or concrete walls; group member indicates whether any of the household members is a community group member. The controls in each regression include age, sex, and schooling of the household head; number of household members; an indicator for adobe, brick or concrete walls; monthly kerosene expenditure; per capita income; and an indicator for households that have at least one community group member; but when any of these is the dependent variable, it is not included as an explanatory variable. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 10: Discount Vouchers and Connection to the Grid (LPM), Experimental Sample

	(1) Connected	(2) Connected	(3) Connected	(4) Connected
Voucher x Post	0.120***	0.120***		
	(0.041)	(0.041)		
Voucher x Round 2			0.134**	0.131**
			(0.054)	(0.053)
Voucher x Round 3			0.147***	0.148***
			(0.054)	(0.054)
Voucher x Round 4			0.105**	0.109**
			(0.046)	(0.046)
Voucher x Round 5			0.095**	0.096**
			(0.042)	(0.043)
s100 x Post	0.130***	0.130***		
	(0.037)	(0.037)		
s100 x Round 2			0.123**	0.126**
			(0.054)	(0.054)
s100 x Round 3			0.156***	0.156***
			(0.049)	(0.049)
s100 x Round 4			0.109***	0.107***
			(0.041)	(0.041)
s100 x Round 5			0.134***	0.132***
			(0.036)	(0.036)
Household Fixed Effects	No	Yes	No	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	2269	2269	2269	2269
Number of Households	494	494	494	494
Mean Dep.Var.	0.56	0.56	0.56	0.56
R-squared	0.45	0.63	0.45	0.63

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. "s100" is the share of eligible neighbors within 100m that received a voucher. "s200" is the share of eligible neighbors between 100-200m radius that received a voucher. Round is a dummy for each yearly survey round. Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 11: Discount Vouchers and Connection to the Grid with Neighbors <200m (LPM), Experimental Sample

	(1) Connected	(2) Connected	(3) Connected	(4) Connected
Voucher x Post	0.115***	0.116***		
	(0.042)	(0.042)		
Voucher x Round 2			0.141***	0.138**
			(0.054)	(0.054)
Voucher x Round 3			0.149***	0.150***
			(0.055)	(0.055)
Voucher x Round 4			0.092*	0.096**
			(0.047)	(0.047)
Voucher x Round 5			0.080*	0.081*
			(0.042)	(0.043)
s100 x Post	0.126***	0.126***		
	(0.037)	(0.037)		
s100 x Round 2			0.132**	0.136**
			(0.055)	(0.054)
s100 x Round 3			0.158***	0.158***
			(0.050)	(0.049)
s100 x Round 4			0.101**	0.099**
			(0.041)	(0.041)
s100 x Round 5			0.122***	0.121***
			(0.036)	(0.036)
s200 x Post	0.027	0.025		
	(0.035)	(0.035)		
s200 x Round 2			-0.039	-0.043
			(0.054)	(0.054)
s200 x Round 3			-0.010	-0.014
			(0.048)	(0.048)
s200 x Round 4			0.068*	0.066*
			(0.038)	(0.038)
s200 x Round 5			0.086***	0.084**
			(0.033)	(0.033)
Household Fixed Effects	No	Yes	No	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	2269	2269	2269	2269
Number of Households	494	494	494	494
Mean Dep.Var.	0.56	0.56	0.56	0.56
R-squared	0.45	0.63	0.45	0.63

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. "s100" is the share of eligible neighbors within 100m that received a voucher. "s200" is the share of eligible neighbors between 100-200m radius that received a voucher. Round is a dummy for each yearly survey round. Post is a dummy that takes the value of 1 in rounds 2 through 5. Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 12: Connection Fee and Connection to the Grid (LPM), Experimental Sample

	(1) Connected	(2) Connected	(3) Connected	(4) Connected	(5) Connected
Fee x Post	-0.203**	-0.201**			
	(0.085)	(0.085)			
Fee x t					-0.327***
					(0.122)
Fee x t-squared					0.054***
					(0.020)
Fee x Round 2			-0.324***	-0.314***	
			(0.115)	(0.114)	
Fee x Round 3			-0.260**	-0.261**	
			(0.108)	(0.108)	
Fee x Round 4			-0.126	-0.128	
			(0.094)	(0.094)	
Fee x Round 5			-0.108	-0.106	
			(0.088)	(0.088)	
Fee100 x Post	-0.348***	-0.349***			
	(0.099)	(0.099)			
Fee100 x t					-0.400***
					(0.146)
Fee100 x t-squared					0.055**
					(0.024)
Fee100 x Round 2			-0.294**	-0.303**	
			(0.144)	(0.143)	
Fee100 x Round 3			-0.450***	-0.447***	
			(0.127)	(0.126)	
Fee100 x Round 4			-0.297***	-0.296***	
			(0.108)	(0.108)	
Fee100 x Round 5			-0.352***	-0.349***	
			(0.097)	(0.098)	
Household Fixed Effects	No	Yes	No	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	No
Lin + Quad Time Trend	No	No	No	No	Yes
Observations	2269	2269	2269	2269	2269
Number of Households	494	494	494	494	494
Mean Dep.Var.	0.56	0.56	0.56	0.56	0.56
R-squared	0.44	0.62	0.44	0.63	0.62

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. "fee100" is the average fee for households within 100m. "fee200" is the average fee for neighbors between 100-200m radius. Round is a dummy for each yearly survey round. Post is a dummy that takes the value of 1 in rounds 2 through 5. t is a linear time trend (in years). Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 13: Connection to the Grid and Externalities (IV)

	(1) Connected	(2) Connected	(3) Connected	(4) Connected	(5) Connected	(6) Connected
Voucher x Post	0.122***	0.124***				
	(0.040)	(0.040)				
Fee x Post			-0.191**	-0.199**		
			(0.082)	(0.084)		
Fee x t					-0.324***	-0.332***
					(0.116)	(0.117)
Fee x t-squared					0.054***	0.054***
					(0.019)	(0.019)
$\overline{E}$ (percentage)	0.193***		0.222***		0.214***	
	(0.055)		(0.058)		(0.058)	
$\overline{E}$ (number)		0.103***		0.111***		0.101***
		(0.030)		(0.031)		(0.030)
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2256	2256	2256	2256	2256	2256
Number of Households	481	481	481	481	481	481
Mean Dep.Var.	0.56	0.56	0.56	0.56	0.56	0.56
R-squared	0.64	0.64	0.64	0.64	0.64	0.64
F-stat	379.5	190.2	200.9	114.9	88.0	53.1

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. $\overline{E}$ denotes formal electric connections within 100m of the household. Post is a dummy that takes the value of 1 in survey rounds 2 through 5. t is a linear time trend (in years).Excluded instrument in columns (1) and (2) is s100; in columns (3) and (4), fee100. Columns (2) and (4) control for the total number of eligible neighbors within 100m. "F-stat" is the first stage F-statistic of the excluded instruments.Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 14: Discount Vouchers and Switching (LPM)

	None-Formal		Informal-Formal		None-Informal	
	(1)	(2)	(3)	(4)	(5)	(6)
	Switched	Switched	Switched	Switched	Switched	Switched
voucher	0.010	-0.031	0.253***	0.154***	-0.049	0.031
	(0.057)	(0.067)	(0.050)	(0.059)	(0.048)	(0.061)
s100	0.147**	0.103	0.086*	0.111**	-0.009	-0.056
	(0.059)	(0.065)	(0.048)	(0.049)	(0.049)	(0.059)
Baseline Covariates	No	Yes	No	Yes	No	Yes
Subdistrict Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	275	265	264	259	275	265
Mean Dep.Var.	0.75	0.76	0.85	0.85	0.15	0.15
R-squared	0.02	0.29	0.12	0.40	0.00	0.16

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. $\overline{E}$ denotes formal electric connections within 100m of the household. Excluded instrument in columns (1) and (2) is s100; in columns (3) and (4), fee100. Columns (2) and (4) control for the total number of eligible neighbors within 100m. "F-stat" is the first stage F-statistic of the excluded instruments.Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 15: Connection Fee and Switching (LPM)

	None-Formal		Informal-Formal		None-Informal	
	(1)	(2)	(3)	(4)	(5)	(6)
	Switched	Switched	Switched	Switched	Switched	Switched
Fee	0.055	0.076	-0.365***	-0.122	0.085	-0.056
	(0.127)	(0.138)	(0.106)	(0.113)	(0.105)	(0.125)
Fee100	-0.261*	-0.157	-0.392***	-0.360***	0.021	0.088
	(0.155)	(0.167)	(0.131)	(0.132)	(0.128)	(0.152)
Baseline Covariates	No	Yes	No	Yes	No	Yes
Subdistrict Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	275	265	264	259	275	265
Mean Dep.Var.	0.75	0.76	0.85	0.85	0.15	0.15
R-squared	0.01	0.29	0.08	0.39	0.00	0.16

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. $\overline{E}$ denotes formal electric connections within 100m of the household. Excluded instrument in columns (1) and (2) is fee100; in columns (3) and (4), fee100. Columns (2) and (4) control for the total number of eligible neighbors within 100m. "F-stat" is the first stage F-statistic of the excluded instruments.Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 16: Multinomial Choice (Ordered Probit), Connection Type/Choice

	(1) Connection Type	(2) Connection Type	(3) Connection Type	(4) Connection Type
type				
voucher	0.248**	0.272**		
	(0.117)	(0.125)		
s100	0.498***	0.501***		
	(0.130)	(0.135)		
Fee			-0.435	-0.547*
			(0.271)	(0.281)
Fee100			-1.188***	-1.261***
			(0.389)	(0.406)
cut1				
Constant	-0.034	-0.273	-1.766***	-2.153***
	(0.096)	(0.330)	(0.382)	(0.511)
cut2				
Constant	0.321***	0.118	-1.416***	-1.766***
	(0.100)	(0.335)	(0.375)	(0.507)
Baseline Covariates	No	Yes	No	Yes
Observations	1767	1714	1767	1714

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. $\overline{E}$ denotes formal electric connections within 100m of the household. Excluded instrument in columns (1) and (2) is fee100; in columns (3) and (4), fee100. Columns (2) and (4) control for the total number of eligible neighbors within 100m. Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 17: Multinomial Choice Round by Round (Ordered Probit), Connection Type/Choice

	Round 2		Round 3		Round 4		Round 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
main								
voucher	0.211		0.304**		0.299*		0.309*	
	(0.136)		(0.147)		(0.161)		(0.173)	
s100	0.330**		0.538***		0.521***		0.763***	
	(0.142)		(0.161)		(0.183)		(0.204)	
Fee		-0.559*		-0.623*		-0.442		-0.530
		(0.294)		(0.328)		(0.360)		(0.386)
Fee100		-0.750**		-1.477***		-1.341***		-1.865***
		(0.374)		(0.438)		(0.486)		(0.549)
cut1								
Constant	-0.557	-1.907***	-0.729*	-2.887***	-1.176***	-3.050***	-1.062**	-3.553***
	(0.344)	(0.508)	(0.388)	(0.586)	(0.424)	(0.638)	(0.447)	(0.710)
cut2								
Constant	-0.002	-1.354***	-0.411	-2.572***	-0.851**	-2.729***	-0.787*	-3.285***
	(0.343)	(0.506)	(0.387)	(0.584)	(0.422)	(0.636)	(0.446)	(0.707)
Observations	414	414	422	422	446	446	432	432

Notes: The dependent variable in all cases is an indicator of formal connection to the grid. $\overline{E}$ denotes formal electric connections within 100m of the household. Excluded instrument in columns (1) and (2) is fee100; in columns (3) and (4), fee100. Columns (2) and (4) control for the total number of eligible neighbors within 100m. "F-stat" is the first stage F-statistic of the excluded instruments.Standard errors clustered at the household level, reported in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 18: Characteristics of Adopters

	(1) Voucher	(2) No Voucher
Age of household head	-0.000	0.000
	(0.001)	(0.001)
Gender of household head	-0.036	-0.047
	(0.042)	(0.042)
Literacy of household head	0.011	0.031
	(0.042)	(0.041)
Property title	0.082**	0.101***
	(0.039)	(0.039)
Household size	0.019**	0.028***
	(0.008)	(0.008)
Adequate wall materials	0.026	-0.082
	(0.056)	(0.051)
Dirt floor	-0.104**	-0.082*
	(0.046)	(0.046)
Wood for cooking	-0.012	0.003
	(0.051)	(0.050)
Income (thousand USD)	0.014**	0.008
	(0.007)	(0.007)
Informal connection at baseline	0.177***	0.110**
	(0.050)	(0.045)
Constant	0.686***	0.728***
	(0.114)	(0.112)
Subdistrict FE	No	Yes
Observations	349	349

Notes: Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 19: Electrification and Overnight PM2.5 Concentration, Experimental Estimator, OLS.

Panel A: Minute-Level Regressions

	(1) Round 3	(2) Round 3	(3) Round 4	(4) Round 4	(5) Round 5	(6) Round 5
voucher	-1.119***	-1.316***	-0.020	0.054	0.028	-0.056
	(0.286)	(0.294)	(0.161)	(0.161)	(0.236)	(0.253)
s100	-0.340	-0.427	0.096	0.087	-0.163	-0.162
	(0.300)	(0.297)	(0.112)	(0.114)	(0.181)	(0.184)
Baseline covariates	No	Yes	No	Yes	No	Yes
Observations	86284	86284	102398	102398	106869	106869
Households	103	103	122	122	128	128
Mean Control	0.448	0.448	0.074	0.074	0.235	0.235
% Change in PM2.5	-0.674	-0.732	-0.020	0.056	0.029	-0.054
SE Change	0.093	0.079	0.158	0.170	0.243	0.239

Panel B: Household-Level Regressions

	(1) Round 3	(2) Round 3	(3) Round 4	(4) Round 4	(5) Round 5	(6) Round 5
voucher	-1.119***	-1.318***	-0.171	-0.051	0.104	0.013
	(0.408)	(0.438)	(0.152)	(0.174)	(0.349)	(0.366)
s100	-0.344	-0.432	0.073	0.086	-0.135	-0.143
	(0.429)	(0.443)	(0.141)	(0.148)	(0.227)	(0.237)
Baseline covariates	No	Yes	No	Yes	No	Yes
Observations	103	103	122	122	128	128
Mean Control	0.445	0.445	0.073	0.073	0.233	0.233
% Change in PM2.5	-0.674	-0.732	-0.157	-0.050	0.110	0.013
SE Change	0.133	0.117	0.128	0.165	0.387	0.371

Notes: The dependent variable is minute-by-minute log $PM_{2.5}$ concentration from 5pm to 7am. The associated percentage change on $PM_{2.5}$, given by $e^{\hat{\beta}} - 1$, is reported in the lower panel (-1 = 100% reduction). All regressions control for hour-of-the-day, subdistrict, and monitor fixed effects. Even columns also control for baseline characteristics: sex of the household head, literacy status of the household head, use of wood for cooking and type of floor (dirt vs rest). Standard errors in parentheses, clustered at the household level. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 20: Electrification and Overnight PM2.5 Concentration, IV Estimates

Panel A: Minute-Level Regressions

	(1) Round 3	(2) Round 3	(3) Round 4	(4) Round 4	(5) Round 5	(6) Round 5
Connected	-3.235	-3.354	0.183	0.335	-0.215	-0.710
	(2.498)	(2.095)	(0.464)	(0.455)	(2.886)	(3.372)
Baseline covariates	No	Yes	No	Yes	No	Yes
Observations	86284	86284	102398	102398	106869	106869
Households	103	103	122	122	128	128
Mean Control	0.448	0.448	0.074	0.074	0.235	0.235
% Change in PM2.5	-0.961	-0.965	0.201	0.399	-0.194	-0.508
SE Change	0.098	0.073	0.557	0.636	2.327	1.658
F-stat	1.6	2.3	3.7	3.9	0.3	0.3

Panel B: Household-Level Regressions

	(1) Round 3	(2) Round 3	(3) Round 4	(4) Round 4	(5) Round 5	(6) Round 5
(mean) connected	-3.258	-3.372	-0.232	0.085	-0.954	-1.701
	(2.499)	(2.093)	(0.394)	(0.461)	(4.786)	(5.607)
Baseline covariates	No	Yes	No	Yes	No	Yes
Observations	103	103	122	122	128	128
Mean Control	0.445	0.445	0.073	0.073	0.233	0.233
% Change in PM2.5	-0.962	-0.966	-0.207	0.089	-0.615	-0.817
SE Change	0.096	0.072	0.312	0.502	1.843	1.023
F-stat	0.8	1.0	2.3	1.9	0.1	0.1

Notes: The dependent variable is minute-by-minute log $PM_{2.5}$ concentration from 5pm to 7am. The excluded instruments are voucher and s100. The associated percentage change on $PM_{2.5}$, given by $e^{\hat{\beta}}-1$, is reported in the lower panel (-1 = 100% reduction). All regressions control for hour-of-the-day, subdistrict, and monitor fixed effects. Even columns also control for baseline characteristics: sex of the household head, literacy status of the household head, use of wood for cooking and type of floor (dirt vs rest). "F-stat" is the first stage F-statistic of the excluded instruments.Standard errors in parentheses, clustered at the household level. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 21: Electrification and Overnight PM2.5 Concentration, Non-experimental Estimates

	(1) lpm1707	(2) lpm1707	(3) lpm1707	(4) lpm1707	(5) lpm1707	(6) lpm1707	(7) lpm1707
Connected	-0.176*	-0.168	-0.262**	-0.275***		-0.486***	
	(0.101)	(0.144)	(0.103)	(0.104)		(0.147)	
Connected x Round 3					-0.399**		
					(0.163)		
Connected x Round 4					-0.285**		
					(0.132)		
Connected x Round 5					0.131		
					(0.177)		
Connected imputed Round 4[1]							-0.328**
							(0.143)
Time FE	✓	✓	✓	✓	✓	✓	✓
Household FE		✓			✓	✓	✓
Subdistrict FE			✓	✓			
Baseline Covariates				✓			
Observations	791	791	791	783	791	599	786
Number of Households	242	242	242	240	242	240	240
Mean Control	0.21	0.21	0.21	0.21	0.21	0.21	0.21

Notes: The dependent variable is average ln($PM_{2.5}$ concentration) during the respective time window. $PM_{2.5}$ concentration was measured in the room where most household members spent most of their time awake during the evening (typically the living room). For details on the $PM_{2.5}$ measurement protocol, see appendix. Round is a dummy for each yearly survey round. Standard errors in parentheses, clustered at the household level. Col (6) restricts the sample to the first four rounds of data; Col (7) uses all the rounds but replaces connection status at round 5 with connection status at round 4. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 22: Time Allocation Children 6-14 (IV Estimates)

Panel A: Main Activities

	(1) Education	(2) Chores	(3) Work	(4) Leisure
Connected	0.779**	0.963*	-0.230	-0.045
	(0.397)	(0.506)	(0.330)	(0.075)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	747	747	747	747
Individuals	196	196	196	196
Mean Y, t=1	0.221	0.669	0.191	1.000
Mean Y, t>1	0.157	0.577	0.264	0.960
Mean Y $\|Y > 0$	6.123	3.486	6.700	8.682

Panel B: Education

	Education		Studying		Other Educ	
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.779**		0.539*		0.840**	
	(0.397)		(0.292)		(0.398)	
connected x female		0.621		0.566**		0.621
		(0.379)		(0.274)		(0.379)
connected x male		1.209		0.415		1.454
		(1.297)		(0.840)		(1.394)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	747	747	747	747	747	747
Individuals	196	196	196	196	196	196
Mean Y, t=1	0.221	0.221	0.147	0.147	0.213	0.213
Mean Y, t>1	0.157	0.157	0.068	0.068	0.152	0.152
Mean Y $\|Y > 0$	6.123	6.123	2.073	2.073	5.298	5.298

Notes: The dependent variable takes the value of 1 if the respondent participated in the activity indicated in the column and 0 otherwise. The excluded instruments are voucher allocation and s100. Standard errors are clustered at the household level and reported in parenthesis. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 23: Time Allocation Children 6-14 (FE Estimates)

Panel A: Main Activities

	(1) Education	(2) Chores	(3) Work	(4) Leisure
Connected	0.002	0.051	-0.046*	0.013
	(0.031)	(0.035)	(0.027)	(0.012)
Individual and Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	3046	3046	3046	3046
Individuals	922	922	922	922
Mean Y, t=1	0.278	0.629	0.202	0.998
Mean Y, t>1	0.188	0.597	0.228	0.970
Mean Y $\|Y > 0$	6.157	3.363	6.248	8.725

Panel B: Education

	Education		Studying		Other Educ	
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.002		0.022		-0.006	
	(0.031)		(0.022)		(0.031)	
connected==1		0.011		0.025		-0.007
		(0.042)		(0.031)		(0.042)
connected==1 & sex==1		-0.016		-0.010		0.005
		(0.063)		(0.044)		(0.062)
Individual and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3046	3046	3046	3046	3046	3046
Individuals	922	922	922	922	922	922
Mean Y, t=1	0.278	0.278	0.167	0.167	0.265	0.265
Mean Y, t>1	0.188	0.188	0.076	0.076	0.181	0.181
Mean Y $\|Y > 0$	6.157	6.157	0.870	0.870	5.287	5.287

Notes: The dependent variable takes the value of 1 if the respondent participated in the activity indicated in the column and 0 otherwise. Standard errors are clustered at the household level and reported in parenthesis. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 24: Electrification and Income Generating Activities by Gender, Adults 18-65 (IV Estimates)

Panel A: Nonfarm Employment

	All		Females		Males	
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.264*		0.458*		0.111	
	(0.138)		(0.238)		(0.157)	
Time Connected (years)		0.113**		0.187**		0.042
		(0.055)		(0.091)		(0.063)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3728	3728	1968	1968	1760	1760
individuals	952	952	495	495	457	457
Mean Y, t=1	0.000	0.000	0.000	0.000	0.000	0.000
Mean Y, t>1	0.178	0.178	0.211	0.211	0.140	0.140
Mean X, t>1	0.738	1.596	0.758	1.635	0.717	1.553
F-stat	21.6	21.5	19.0	17.4	16.4	17.3

Panel B: Home Business

	All		Females		Males	
Connected	0.122*		0.247**		0.010	
	(0.063)		(0.108)		(0.071)	
Time Connected (years)		0.052**		0.109**		0.004
		(0.025)		(0.046)		(0.027)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5789	5789	3040	3040	2749	2749
individuals	1403	1403	736	736	667	667
Mean Y, t>1	0.070	0.070	0.097	0.097	0.040	0.040
Mean Profits$\|Y=1$, 1000USD/year	1.284	1.284	1.053	1.053	1.519	1.519
F-stat	23.3	22.5	23.1	19.9	18.5	19.5

Notes: **Panel A:** the dependent variable is a dichotomous indicator of participation in non farm employment. The sample is formed by individuals that did not engage in non farm employment in the year leading to the baseline survey. **Panel B:** the dependent variable is a dichotomous indicator of operation of a home business. **Both Panels:** "F-stat" is the first stage F-statistic of the excluded instruments. The excluded instruments are voucher and s100. Standard errors are clustered at the household level and reported in parenthesis. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 25: Electrification and Income Generating Activities by Gender, Adults 18-65 (FE Estimates)

Panel A: Nonfarm Employment

	All		Females		Males	
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.019**		0.022*		0.016	
	(0.009)		(0.012)		(0.013)	
Time Connected (years)		0.013***		0.009		0.017***
		(0.004)		(0.006)		(0.006)
Individual and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14771	14771	7911	7911	6860	6860
individuals					1750	1750
Mean Y, t=1	0.000	0.000	0.000	0.000	0.000	0.000
Mean Y, t>1	0.154	0.154	0.158	0.158	0.149	0.149
Mean X, t>1	0.632	1.331	0.650	1.372	0.611	1.283

Panel B: Home Business

	All		Females		Males	
Connected	0.030***		0.046***		0.012	
	(0.010)		(0.017)		(0.010)	
Time Connected (years)		0.009**		0.006		0.011**
		(0.004)		(0.007)		(0.005)
Individual and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5789	5789	3040	3040	2749	2749
individuals	1403	1403	736	736	667	667
Mean Y, t>1	0.070	0.070	0.097	0.097	0.040	0.040
Mean Profits$\|Y = 1$, 1000USD/year	1.284	1.284	1.053	1.053	1.519	1.519

Notes: **Panel A:** the dependent variable is a dichotomous indicator of participation in non farm employment. The sample is formed by individuals that did not engage in non farm employment in the year leading to the baseline survey. **Panel B:** the dependent variable is a dichotomous indicator of operation of a home business. **Both Panels:** Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 26: Electrification and Household Income (IV Estimates)

Panel A: Gross Income

	Total Income		Labor Income		Non-Labor Income	
Connected	5.449*		5.923**		-0.474	
	(2.854)		(2.883)		(0.510)	
Time Connected (years)		2.252*		2.423**		-0.170
		(1.187)		(1.197)		(0.207)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2147	2147	2147	2147	2147	2147
Households	469	469	469	469	469	469
Mean Y, t=1	1.776	1.776	1.465	1.465	0.312	0.312
Mean Y, t>1	2.567	2.567	2.005	2.005	0.562	0.562
Mean X, t>1	0.730	1.628	0.730	1.628	0.730	1.628
F-stat	23.9	23.1	23.9	23.1	23.9	23.1

Panel B: Net Income

	Total Income		Labor Income		Non-Labor Income	
Connected	1.631		4.792*		-0.474	
	(1.165)		(2.537)		(0.510)	
Time Connected (years)		0.646		1.914*		-0.170
		(0.474)		(1.035)		(0.207)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2147	2147	2147	2147	2147	2147
Households	469	469	469	469	469	469
Mean Y, t=1	1.182	1.182	1.089	1.089	0.312	0.312
Mean Y, t>1	1.883	1.883	1.616	1.616	0.562	0.562
Mean X, t>1	0.730	1.628	0.730	1.628	0.730	1.628
F-stat	23.9	23.1	23.9	23.1	23.9	23.1

Notes: The dependent variable is gross annual income in US$."F-stat" is the first stage F-statistic of the excluded instruments. The excluded instruments are voucher allocation and s100. Standard errors are clustered at the household level and reported in parenthesis. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 27: Electrification and Household Income (FE Estimates)

	Total Income		Labor Income		Non-Labor Income	
	All	> 0	All	> 0	All	> 0
Connected	0.278		0.239		0.038	
	(0.221)		(0.219)		(0.027)	
Time Connected (years)		0.159*		0.104		0.055***
		(0.087)		(0.086)		(0.015)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8582	8582	8582	8582	8582	8582
Households	1911	1911	1911	1911	1911	1911
Mean Y, t=1	1.801	1.801	1.506	1.506	0.296	0.296
Mean Y, t>1	2.666	2.666	2.119	2.119	0.547	0.547
Mean X, t>1	0.635	1.366	0.635	1.366	0.635	1.366

Panel B: Net Income

	Total Income		Labor Income		Non-Labor Income	
Connected	0.165**		0.682***		0.038	
	(0.066)		(0.251)		(0.027)	
Time Connected (years)		0.111***		0.208**		0.055***
		(0.031)		(0.105)		(0.015)
Household and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8582	8582	8582	8582	8582	8582
Households	1911	1911	1911	1911	1911	1911
Mean Y, t=1	1.266	1.266	1.028	1.028	0.296	0.296
Mean Y, t>1	1.927	1.927	1.725	1.725	0.547	0.547
Mean X, t>1	0.635	1.366	0.635	1.366	0.635	1.366

Notes: The dependent variable is gross annual income in US$. Standard errors are clustered at the household level and reported in parenthesis. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey.*

Table 28: Electrification and Changes in Energy Use, IV Estimates

	Kerosene		Candle		Car Battery		Propane		Wood	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	RF	IV	RF	IV	RF	IV	RF	IV	RF	IV
voucher $\times t_2$	-0.241**		0.146*		0.009		0.029		-0.054	
	(0.112)		(0.085)		(0.046)		(0.093)		(0.091)	
voucher$\times t_3$	-0.065		0.121		-0.006		-0.128		0.063	
	(0.099)		(0.091)		(0.035)		(0.134)		(0.133)	
voucher$\times t_4$	-0.091		0.183*		-0.051		-0.068		-0.045	
	(0.118)		(0.097)		(0.067)		(0.117)		(0.105)	
voucher$\times t_5$	0.107		0.115		-0.003		-0.019		-0.047	
	(0.111)		(0.122)		(0.044)		(0.132)		(0.133)	
s100$\times t_2$	0.013		0.024		0.019		-0.077		0.015	
	(0.078)		(0.062)		(0.033)		(0.084)		(0.078)	
s100$\times t_3$	-0.005		0.123*		0.013		-0.087		0.121	
	(0.079)		(0.062)		(0.031)		(0.086)		(0.082)	
s100$\times t_4$	-0.030		0.169***		0.048		-0.039		-0.019	
	(0.077)		(0.058)		(0.035)		(0.084)		(0.065)	
s100$\times t_5$	-0.071		0.148**		0.032		-0.153*		0.106	
	(0.082)		(0.066)		(0.030)		(0.087)		(0.084)	
Connected		-0.332***		-0.099**		-0.017		0.068		-0.009
		(0.051)		(0.043)		(0.016)		(0.046)		(0.040)
Number of Observations	988	988	991	991	987	987	991	991	988	988
F-stat		5.7		6.0		6.5		6.0		5.9

Notes: Each cell reports the IV coefficient of connection on the variable indicated in the left column at the round indicated in the top column. Regressions control for baseline characteristics: literacy, age and sex of the household head, household size, material on the walls, dirtfoor, cooking with wood, household has property title. Connection is instrumented by individual voucher allocation and the number of vouchers within 100m of the household, and controlling for the number of eligible neighbors in that radius (and the baseline characteristics, as in equation **??**). "F-stat" is the first stage F-statistic of the excluded instruments. Huber-White standard errors in parentheses. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 29: Household Appliances, IV Estimates

Panel A: Appliances

	Radio	Stereo	TV	DVD	Fridge	Blender
Connected	-0.026	0.437**	0.578*	0.220*	0.544**	0.246*
	(0.361)	(0.180)	(0.341)	(0.118)	(0.250)	(0.144)
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var., Control	0.335	0.163	0.327	0.089	0.189	0.099
Number of Households	219	386	278	414	362	405
F-stat	2.6	10.7	3.4	10.9	6.5	9.9

Panel B: Appliances (continued)

	Computer	Sewing	Iron	Microwave	Washer	Fan
Connected	0.136***	-0.047	0.241	0.054	0.180***	0.165**
	(0.049)	(0.064)	(0.170)	(0.066)	(0.055)	(0.080)
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var., Control	0.004	0.026	0.199	0.012	0.000	0.035
Number of Households	467	447	367	460	481	481
F-stat	14.8	12.3	9.6	13.3	15.4	15.4

Notes: The dependent variable takes the value of 1 if the household owns the appliance indicated in the column and 0 otherwise. The sample for each regression is households that did not own the appliance at baseline, except for washing machines and fans, because these appliances were not included in the baseline sample. "F-stat" is the first stage F-statistic of the excluded instruments. The excluded instruments are voucher allocation and s100. Standard errors, clustered at the household level, are reported in parenthesis. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 30: Acute Respiratory Infections Among Children 0-6, LPM

	(1) RF	(2) IV	(3) RF	(4) IV
Voucher x Round 3	-0.180*		-0.160*	
	(0.096)		(0.097)	
Voucher x Round 4	0.091		0.106	
	(0.070)		(0.073)	
Voucher x Round 5	-0.032		-0.011	
	(0.089)		(0.091)	
s100 x Round 3	-0.044		-0.048	
	(0.088)		(0.088)	
s100 x Round 4	-0.006		-0.032	
	(0.087)		(0.085)	
s100 x Round 5	-0.055		-0.071	
	(0.076)		(0.080)	
Connected		-0.652*		-0.660*
		(0.389)		(0.359)
Baseline Characteristics	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Subdistrict FE	Yes	Yes	Yes	Yes
Mean Control, Round 3	0.43	0.43	0.43	0.43
Mean Control, Round 4	0.10	0.10	0.10	0.10
Mean Control, Round 5	0.23	0.23	0.23	0.23
Observations	546	546	538	538
Individuals	192	192	189	189
F-stat		1.9		1.7

Notes: The dependent variable in all regressions is an indicator for acute respiratory infection in the four weeks prior to the survey (self-reported). Baseline characteristics: literacy, age and sex of the household head, household size, material on the walls, dirtfoor, cooking with wood, household has property title. Excluded instruments in columns (2) and (4) are voucher and s100 interacted with round. Standard errors, clustered at the household level, are reported in parenthesis. Significantly different than zero at 90(*), 95(**), and 99(***) percent confidence. Source: *Household Electrification Survey*

Table 31: Estimation of PM2.5 Exposure and Health Impacts

	Female Head	Male Head	Female Child	Male child
Time (hours per day)				
Sleeping	8.85	8.83	9.99	9.54
Kitchen	2.54	0.08	0.56	0.09
Outside	2.69	7.77	2.51	4.88
Home, Evening	5.15	5.17	4.01	4.46
Home, Daytime	4.77	2.14	6.93	5.03
Air inhalation (m^3) per hour)				
Sleeping	0.30	0.30	0.30	0.30
Light Activity	0.78	0.78	0.78	0.78
Estimated exposure to $PM_{2.5}$ (mg per day)				
Non-encouraged group	5.68	3.20	4.23	3.72
Encouraged group	3.93	1.43	2.65	2.10
% Change in exposure due to electrification (lower bound)				
Percentage change	-0.328	-0.587	-0.392	-0.457
Bootstrapped standard errors	0.001	0.001	0.003	0.003
Predicted relative risk				
Lung cancer				
Non-encouraged group	3.98	2.68	3.22	2.95
Encouraged group	3.06	1.75	2.39	2.10
Ischemic heart disease				
Non-encouraged group	1.12	1.07	1.09	1.08
Encouraged group	1.08	1.03	1.06	1.05
cardiovascular disease				
Non-encouraged group	1.12	1.07	1.09	1.08
Encouraged group	1.08	1.03	1.06	1.04
Cardiopulmonary disease				
Non-encouraged group	1.15	1.08	1.11	1.10
Encouraged group	1.10	1.04	1.07	1.06

Notes: Exposure $= \sum_{j=1}^{J} time_j \times concentration_j \times inhalation_j$, where $time_j$ is time spent in activity j, $concentration$ is $PM_{2.5}$ concentration in the room while performing activity j, and $inhalation_j$ is the esimated inhalation rate while performing activity j. Time allocation was obtained from the *EHEIPCER* time allocation module. "Home, daytime" is time at home between 0700-1700hrs in any room but the kitchen. "Home, evening" is the time at home between 1700-0700 hrs, while neither sleeping nor in the kitchen. "outside" is time spent outside the home. Air inhalation was taken from the EPA exposure handbook (EPA 2011). "Light activities" include watching TV, desk work, writing, typing, cooking, washing dishes, ironing, walking up to 2.5 mph (2.9 km/h). $PM_{2.5}$ concentration was estimated at 0.41 mg/m^3 for "sleeping" and "home, evening", 0.26 mg/m^3 for "home, daytime", 0.90 mg/m^3 for "kitchen", 0 mg/m^3 for "outside". For the "after electrification" scenario, $PM_{2.5}$ was reduced to 0.15 mg/m^3 (by 63%, as estimated by our main model) for "sleeping" and "home, evening" and held constant for the other instances. Average ages: 43 (female head), 47 (male head), 11 (female child), 13 (male child). Predicted relative risks were calculated with a linearized version of the dose-response function calculated by Pope et al. (2011). R^2 for lung cancer (0.99), ischemic heart disease (0.96), cardiovascular disease (0.86), cardiopulmonary disease (0.80). See main text for details.

10 Appendix 1

Sample Size and Power Calculation

1. Introduction

With the objective of studying the effect of rural electrification program in San Miguel and Chalatenango, 1,533 observations were needed. Due to the high intracluster correlations observed in variables as non-agricultural waged income or time allocated to non-agricultural non-wage labor, the power to detect differences in such variables will be lower, although it may still be possible to detect differences.

The second section deals with survey design issues like the intracluster correlation and assumptions in the variance calculation. Section three covers the main issues regarding power calculation when the treatment is discrete as it is in the rural electrification project. Section four discusses the sample size requirements to assess the rural electrification programs. Section six summarizes the findings and recommends specific sample size.

2. Survey Design

We assumed a clustered, quasi-randomized evaluation design with treatments administered at the cluster level and data collection before and after initiation of the treatments. With this design, impact estimates can be measured using the preferred approach of taking difference-in-differences or “double difference”: the change in the outcome in the treatment group minus the change in the outcome in the quasi-randomized control (or alternate treatment) group. The purpose of the sample size estimates was to determine the minimum impact, Δ, that can be detected for a given number of sampled clusters, g, and households per cluster, m, in each treatment for the evaluation sample.[41] If the impact of the treatment is at least as large as Δ, we will be able to detect it 80 percent of the time in a sample of total size milligram. If the treatment impact is less than Δ, we are less likely to detect it, although it is still possible.

3. Intracluster Correlation

The most controversial issue in sample design is the intracluster correlation, so we made the calculation procedure explicit. DIGESTYC provided detailed geographic information system (GIS) data on the location of all the dwellings electrified in the northern El Salvador. In an ideal scenario, we would have the relevant socio-economic data from the census as well, but at the time of writing this was not available. The intracluster correlation of several variables was calculated from the EHPM Survey 2007. Merging the survey and the GIS data, the cantons electrified were identified and matched to the household survey data. The universe is constituted by the set of cantons that were

[41] In addition to *g* and *m*, the minimum detectable impact, Δ, is a function of the variance of the outcome variable, its intracluster correlation, and the area of influence of the highway being evaluated.

identified. The sub-set of cantons that were also included in the household survey constitutes “level 1.”

For those cantons that were not included in the EHPM survey, the municipality income and time allocation data was imputed. This group plus “level 1” constitutes “level 2.” In turn, for those municipalities that were not included in the survey, the department data was imputed. These sub-set plus “level 2” conforms “level 3.”

Several outcome variables were used in the analysis. We will summarize the results for overall household income, but the analysis also included weekly working hours, and both split by wage agriculture, non-wage agriculture, wage non-agriculture, and non-wage non-agriculture.

4. Scenarios for Variance Calculation

There were three important differences between the proposed sample for evaluation and the EHPM sample, all of which were likely to affect the sample variance in the projected sample relative to that in the EHPM sample: First, we estimated variance of the primary outcomes in the EHPM using only one round of data collection, rather than two. The variance of the difference between the two measures depended upon the variance of each measure as well as the correlation over time between the two measures. We did not know this correlation; so we made assumptions about it, which we vary below. Second, we stratified the sample for the collection of this data, to both balance the sample and reduce the sampling variance. The reduction in sampling variance depended upon the variance between strata means; the larger the difference between the average outcomes across strata, the higher the variance reduction would be. Third, the EHPM measured the variance of outcomes related to different levels of current access to roads. It was likely that the variance of baseline would be smaller given the assumptions of accessibility we imposed.

Since the three differences between the proposed surveys and the EHPM would certainly affect the variance of primary outcomes, we experimented with power calculations using several different variance estimates. First and most conservatively, we simply doubled the variance of the EHPM outcomes. Doing so assumed that the primary outcome would not be correlated across the two surveys, that each strata would have exactly the same mean outcome, and that the treatment would not affect the variance of the treatment. Second, we reduced the doubled variance by 10 percent, to simulate a significant decline in sample variance due to stratification. Third, we simply computed the power calculations using the EHPM variance. Finally, we used the NHS variance less 10 percent, to account for gains from stratification, but also assume between-period correlation of 0.5 and a within-period correlation of -0.5. Since we also ignored the above assertion that the baseline variance in outcomes was likely to be smaller than the EHPM variance, the fourth estimate was likely to be the most realistic and the one we proposed to use.

5. Power Calculations

Discrete Treatment

The impact evaluation was conducted with difference-in-difference estimators. This methodology requires repeat observations on members. Power calculations for this type of survey designs were based on Murray (1998, chapter 9). The main analysis was based on the following three equations (the equation number in Murray's book is in brackets):

$$g = \frac{2(1+(m-1)\hat{ICC})(t_{\alpha/2}+t_{\beta})^2}{m\hat{\Delta}^2}\hat{\sigma}_y^2$$ ...(1) [9.23]

$$\hat{\sigma}_y^2 = \frac{mg}{2(1+(m-1)ICC)}\hat{\sigma}_{\Delta}^2$$ ...(2) [9.17]

$$\hat{\sigma}_{\Delta}^2 = 4\left[\frac{\hat{\sigma}_m^2(1-\hat{r}_{yy(m)})+m\hat{\sigma}_g^2(1-\hat{r}_{yy(g)})}{mg}\right]$$ ...(3) [9.34]

Where:

g: number of clusters in each condition (treatment/control)

m: number of observations per cluster

ICC: Intracluster correlation

α: type I error rate

ß: type II error rate

$\hat{\sigma}_y^2$: estimated variance of the outcome variable

$\hat{\Delta}$: estimated change

$\hat{\sigma}_{\Delta}^2$: estimated variance of the change in the outcome variable

ryy(g): inter-period correlation

ryy(m): intra-period correlation

Replacing (2) in (1), we get:

$$1 = \frac{(t_{\alpha/2}+t_{\beta})^2}{\hat{\Delta}^2}\hat{\sigma}_{\Delta}^2$$ ...(4)

Inserting (3) in (4):

$$1 = \frac{(t_{\alpha/2}+t_{\beta})^2}{\hat{\Delta}^2}\times 4\left[\frac{\sigma_m^2(1-r_{yy(m)})+m\sigma_g^2(1-r_{yy(g)})}{mg}\right]$$ ... (5)

Solving for g:

$$g = \frac{(t_{\alpha/2} + t_{\beta})^2}{\hat{\Delta}^2} \times 4\left[\frac{\sigma_m^2(1 - r_{yy(m)}) + m\sigma_g^2(1 - r_{yy(g)})}{m}\right] \quad \ldots(6)$$

6. Summary of Sample Size Needed to Measure the Impact of the Rural Electrification Program

Conscious of budget limitations, we proposed the study of only two departments, Chalatenango and San Miguel for studying the impact of the rural electrification program. These departments were proposed because, according to the current program plans, they include the largest numbers of cantons that will benefit from the electrification program. In addition, these districts include a number of cantons that will benefit from the road improvement and the electrification programs. Although rather modest, these districts will play a key role in the study of complementarities between road improvement and electrification.

Following the procedure as in section 5, we calculated the minimum sample size for each department. The results are presented in Table 35. We recommend Scenario four given the main assumptions are Type I and II error rates of five percent and 20 percent respectively and a change in incomes of at least 20 percent under a discrete treatment.

Table 32: Number of Clusters per Condition1 and Total Sample Size[2] for Household Income[3] for each Scenario[4]

		Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Intracluster correlation[6]	Clusters per condition	Total sample size	Clusters per condition	Total sample size	Clusters per condition	Total sample size	Clusters per condition	Total sample size
m=25[5]									
Chalatenango	0.030	41	2027	36	1824	20	1014	15	757
San Miguel	0.073	96	4799	86	4319	48	2399	16	816
m=35									
Chalatenango	0.030	31	2147	28	1933	15	1074	11	744
San Miguel	0.073	87	6060	78	5454	43	3030	11	802
m=45									
Chalatenango	0.030	25	2281	23	2053	13	1140	8	737
San Miguel	0.073	81	7334	73	6600	41	3667	9	796

1 The conditions are “treatment” and “control”. The number of clusters in each condition is equal
2 Total sample size (treatment + control)
3 The outcome variable is total monthly household income
4 For the specification of each scenario see section 3.2, and for the formulae, see Appendix 1.
5 Number of observations (households) per cluster
6 Observed in the NHS at the department level.
7 $\alpha = 0.05; \beta = 0.20; \Delta = 0.20$

Table 33: Summary Tables for Other Outcome Variables: Number of Clusters (Discrete Treatment)

	25 observations per cluster				35 observations per cluster				45 observations per cluster			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Chalatenango												
Total income	41	36	20	15	31	28	15	11	25	23	13	8
Agricultural Wage Income	19	17	10	9	13	12	7	6	10	9	5	5
Agricultural Non-wage Income	33	30	17	9	28	25	14	6	25	22	12	5
Non-agricultural Wage Income	10	9	5	5	7	7	4	3	6	5	3	3
Non-agricultural Non-wage Income	36	33	18	8	31	28	16	6	29	26	14	5
Total Labor Hours	27	24	13	10	21	18	10	7	17	15	9	5
Agricultural Wage Hours	108	97	54	51	76	68	38	36	59	53	29	28
Agricultural Non-wage Hours	94	84	47	22	80	72	40	15	73	66	37	12
Non-agricultural Wage Hours	307	276	153	37	288	259	144	26	277	250	139	20
Non-agricultural Non-wage Hours	138	124	69	64	97	87	49	45	75	67	37	35
San Miguel												
Total income	96	86	48	16	87	78	43	11	81	73	41	9
Agricultural Wage Income	25	23	13	9	19	17	10	6	16	15	8	5
Agricultural Non-wage Income	48	43	24	10	41	37	21	7	38	34	19	6
Non-agricultural Wage Income	14	13	7	7	10	9	5	5	8	7	4	4
Non-agricultural Non-wage Income	23	21	12	6	19	17	10	4	17	16	9	3
Total Labor Hours	15	13	7	7	10	9	5	5	8	7	4	4
Agricultural Wage Hours	57	52	29	26	40	36	20	18	31	28	15	14
Agricultural Non-wage Hours	137	123	69	63	96	87	48	44	74	67	37	34
Non-agricultural Wage Hours	107	96	53	18	96	87	48	13	90	81	45	10
Non-agricultural Non-wage Hours	420	378	210	100	359	323	179	70	326	293	163	54