Understanding Patterns and Impacts of Electricity Quality and Access in Northern India

Michael Keppler

Thomas Lutken

Sumin Wang

Advisor: Marc Jeuland

Table of Content

Abstract 4

Introduction 4

Background 5

Electrification and Socioeconomic Development 5

Standard of Living 5

Fuel Consumption 5

Other Impacts 6

Role of Time and Reliability 7

Government Electrification Schemes 7

The Drivers of Electricity Quality 8

The Nature of Electricity Quality 8

Methods 9

Survey 9

Sample Region 9

Survey Design 10

Survey Sampling Methods 10

Index Generation 11

The World Bank Multi-Tier Framework 11

Principal Components Analysis 13

Patterns and Characteristics 14

Patterns of Electrification 14

Patterns of Electricity Quality 15

Relationships between Reliability and Socioeconomic Variables 15

Data Check 16

Results 17

Descriptive statistics of the survey sample 17

Tier scores and PCA scores 18

Comparison to Greater UP and Bihar 20

Comparison of Saubhagya and Unelectrified populations 20

Saubhagya and effects on multi-tier scores 21

Patterns of Electrification 22

Patterns of Electricity Quality 23

Relationships between Reliability and Socioeconomic Variables 27

Discussions 29

Saubhagya 29

Patterns of Electricity Quality 30
Implications for Discoms and policies moving forward 31
Relationships with Socioeconomic Well-being 32
Possible sources of error 33

Conclusions 34

Limitations and Further Study 34

References **Error! Bookmark not defined.**

Appendix 38
Appendix 1: Variables used in models 38
Appendix 2: Robust regression results to identify patterns of electricity quality 40
Appendix 3: Robust socioeconomic impact regression results 44

Abstract

Indians have suffered from rolling outages and blackouts even as millions of citizens have gained access to electricity in recent years, which has led to concerns about electricity quality and its impacts. We conducted an in-person survey of 500 rural households in 40 communities in northern India. We used the World Bank Multi-Tier Framework and Principal Components Analysis to generate indexes of electricity quality, and ran multiple linear regressions to examine the correlations of these measures with socioeconomic indicators. We find that electricity quality varies strongly across zones served by different distribution companies, but is less closely related to the national electrification program Saubhagya. Electricity quality is also positively correlated with legal and older connections that tend to have more safety features, but unexpectedly, not positively correlated with household wealth. We find that access to electricity is strongly correlated with better standard of living, as measured by household consumption, and that better electricity quality is strongly correlated with decreased household fuel consumption.

Introduction

India's population and economy are growing rapidly, and the country is playing an increasingly influential role on the world stage. The demand for electricity to sustain this rapid development is also growing, but even today some Indians still lack connections, and many more have access only to low quality, unreliable service. In rural areas, the electricity deficiency problem is especially significant; 460,000 rural households were without power in 2017, with another 180,000 waiting on their connections [MJ1] (Ministry of Power, 2017). Additionally, electricity development has not been uniform across India; Uttar Pradesh and Bihar, the locations considered in this research, have historically ranked at or near the bottom according to measures of electricity access in the country (Abhishek et al., 2018)

While connection to the grid has been emphasized by the central government (Saubhagya.gov.in, 2018), many studies also point to the importance of quality in realizing the full benefits of electricity (Chakravorty et al., 2014; Rao, 2013) Extending electricity access alone is a capital-intensive investment and is expected to bring significant socioeconomic benefits, but access without reliability could be counter-productive (Khandker et al., 2014). Therefore, understanding exactly how and where electricity is limited, either in supply, reliability, or other quality metrics, is vital to improving social welfare.

While existing literature has examined the drivers and barriers to electricity adoption, less research exists examining how variations in electricity quality impact electrified communities. Our research partner, Prayas Energy Group, was interested in addressing this knowledge gap in India, in particular in the border region between Uttar Pradesh and Bihar since the region struggled with its electricity for years (Abhishek et al., 2018). We conducted a survey of 500 households and 100 enterprises in the area, aiming to understand various aspects of electricity

quality and socioeconomic standing. Using this data, we aim to add to Prayas' understanding of the regions electricity, while also giving some context for future avenues of improving electricity quality.

Background

Electrification and Socioeconomic Development

Standard of Living

A recurring theme in the literature is that there exists a positive relationship between measures of standard of living and electrification. Samad and Zhang (2016, 2017) for example found that household income and expenditure both rise significantly after electrification, and that the household poverty rate decreases with electricity access. In particular, electrification was found to have a statistically significant relationship with per capita non-agricultural and total income, but not with per capita agricultural income, a finding that was echoed by Chakravorty, Pelli, and Marchand (2014).

Similar to the income increase, Samad and Zhang (2017) found that household non-food expenditure increases after electrification, but they found no such relationship between food expenditure and electrification. Furthermore, they found that the returns to electrification (in income and expenditure growth), while positive for all, monotonically increase with households' income and expenditure percentiles; no significant growth was seen below the 60th percentile of income distribution. They believe this difference in income and expenditure growth between wealthier and poorer households can be accounted for by wealthier households' more diversified portfolio of appliances, which allow them to reap more benefits from electrification.

However, not all of the literature echoes these findings. In contrast with much of the existing literature, Burlig and Preonas (2016)[1] implemented a very rigorous causal inference methodology and did not find any economically meaningful relationship between increased energy access and development across multiple outcomes, including income, expenditure, household wealth, and asset ownership.

Fuel Consumption

Another common thread in the literature is a negative relationship between electrification and household fuel consumption, specifically kerosene. Samad and Zhang (2016, 2017) found a negative relationship between household electrification and both time spent collecting biomass fuel, indoor air pollution, and household kerosene consumption. A proposed reason for this among the literature is that a common use for electricity in rural communities is lighting, which is also one of the primary uses of kerosene. This is notable as kerosene is an inferior light source to electricity, a major source of indoor pollution, and contributive to ambient black carbon

emissions (Samad and Zhang, 2016; Samad and Zhang, 2017; Burlig and Preonas, 2016). Conversely, Khandker et al. (2014) did not find that kerosene consumption of electrified households was much less than than that of unelectrified households.

Other Impacts

Apart from standard of living and fuel consumption, electrification is also correlated with a host of other outcomes, including impacts relating to education, employment, and gender. For education, longer study time for children was correlated with electricity access, but higher school attainment was not; conversely, power outages did display a negative relationship with schooling, with respect to both study time and grade completion (Samad and Zhang, 2016). Some literature finds a positive relationship between electrification and reductions in the illiteracy rate, particularly for those with under four years of education, and an even larger positive relationship with years of schooling (Lipscomb et al., 2013).

Analogous to the case of income, some literature reveals a positive relationship between electrification and non-agricultural employment within India's guar belt, but does not find any such relationship in villages in the rest of the country. This is explained as being due, in part, to the rise of firms that are more electricity-intensive and complement agricultural production. This literature does not find that electrification itself is sufficient to confer economic benefits, but that it may be necessary for households to improve their welfare in rapidly shifting economic contexts. Thus, electrification significantly increases non-agricultural employment in villages, but only where the economic opportunity complements the infrastructure; without these conditions there appears to be no discernable effect of electrification (Usmani and Fetter, 2019; Usmani, 2019).

When it came to employment, Samad and Zhang (2016) found that the effects of electrification were gender-sensitive: women's employment-hours increased but not their labor force participation, while men's labor force participation increased, but not their employment-hours. Alternatively, some of the literature finds that electrification has a significant positive relationship with female employment, as well as work hours for both men and women. However, it also appears to have a negative relationship with female wages and a positive relationship with male wages. The literature suggests that electrification increases employment as women who were previously busy with household chores now have more free time to pursue education, employment, social participation, and microenterprise activities. Additionally, electrification was found to have a positive relationship with women's decision-making ability in the home, particularly with respect to children's health care and their own income, and in women's access to increased information, primarily through the TV or radio. Electrification was not found to have any positive effects on boys' study time or men's labor force participation (Dinkleman, 2011; Samad and Zhang, 2017). However, Burlig and Preonas (2016) did not find a positive relationship between electrification and labor, village-level outcomes, or education.

Role of Time and Reliability

Many of the statistical correlations, and possible impacts, of electrification mentioned thus far have also been found to be significantly positively correlated with both the length of time a household has been electrified, and the reliability of its electricity connection. According to some of the literature, length of time is positively correlated with all of the aforementioned potential impacts, but the relationship also appears to diminish over time (Samad and Zhang, 2016; Samad and Zhang, 2017; Chakravorty, Pelli, and Marchand, 2014; Rao, 2013). Reliability appears in the literature to have an impact in one of two ways: higher quality electricity supply has stronger positive relationships with the aforementioned impacts, or power outages reduce the positive relationships. This applies to the relationships pertaining to electrification and standard of living, fuel consumption, and other impacts to such a degree, that some find that small improvements in reliability are correlated with greater positive impacts than electrification itself (Khandker et al., 2014). In line with this, others have found that high quality (i.e. reliable) electricity supply is correlated with increases in income over 200% larger than simple electrification (Chakravorty, Pelli, and Marchand, 2014). One paper estimates that in some developing countries, the potential benefits of improving electricity reliability would confer benefits valued in the hundreds of millions (USD), while improving both reliability and achieving universal electric access would confer benefits in the billions (USD) (Samad and Zhang, 2017).

Government Electrification Schemes

The Indian government has recognized the importance of rural electrification and focused on providing electricity for agricultural irrigation and introducing electricity for households' basic needs (Oda and Tsujita, 2011). A recent scheme is the Rajiv Gandhi Rural Electrification Programme (RGGVY). Launched in 2005, the scheme aimed to electrify 100% of villages by 2012 (RGGVY Evaluation Report, 2014). But the government's definition of electrification is limited - as long as the village has basic electricity infrastructure such as transformers, poles, and distribution lines, and 10% households in the village are connected, that village is considered "electrified." This narrow definition of "electrification" led to critiques of the scheme, and problems with electricity quality remained (Cunha, 2018).

To fill these gaps, Prime Minister Narendra Modi announced the Saubhagya rural household electrification plan in September of 2017. Through Saubhagya, the central government would cover 90% of connection costs for the distribution companies (known as 'Discoms'), aiming to connect unelectrified houses. If a house was too difficult to reach with grid electricity, the household would be provided with an off-grid kit including batteries and LED lights (Saubhagya.gov.in, 2018). Under Saubhagya, houses below the poverty line are given a free connection. For a house that is relatively wealthy, the Discom is instructed to connect the house and add a 500 rupee fee, collected over a 10-month period via the electricity bill (Saubhagya.gov.in, 2018).

The Drivers of Electricity Quality

Many factors contribute to unreliable electricity supply in northern India. From the perspective of the electricity system, two major causes are widespread electricity theft and heavy governmental subsidies to the agricultural sector. Theft and subsidies worsen Discoms' financial ability to make reinvestment in infrastructure and rehabilitation, and to provide sufficient generation, especially as the government tolerates electricity theft during election years (Golden and Min, 2012).

From the perspective of villages, Chakravorty et al. (2014) mentioned that infrastructure investment on electricity depends on the overall economic situation of the village, distance to a big city and population density. Dugoua et al. (2017) observed decreased geographic inequalities of electricity access with respect to distance to the nearest city but remaining inequalities among villages in terms of caste and wealth compositions. Thomas et al. (2018) found that electricity quality is better in villages with early electrification as these electrified villages can reinvest gains from electrification to improve infrastructure.

From the perspective of households, Kemmler (2007) highlighted the spatial disparities in electrification and found that the household connection decision is highly correlated with household educational level and community electrification but weakly correlated with per capita expenditure and income. Households belonging to Scheduled Castes (SC) and Tribes (ST), two historically disadvantaged social groups identified for positive discrimination, also tend to consume less electricity.

The Nature of Electricity Quality

The nature of access to electricity is multifaceted. Different attributes of electricity such as availability (total hours a day with electricity), voltage fluctuations, the frequency of outages, etc. can limit the usefulness of electricity. To evaluate impacts, it is fundamental to understand and measure the quality of power supply. Among relevant studies in the literature, the methods used to create a variable to measure electricity access and quality for economic studies are very different.

Burlig and Preonas (2016) for example used nighttime brightness observed from satellite images as proxies for electrification and energy consumption. Rao (2013) used the density of transmission cables in district level as an instrument for the quality of power supply. Allcott et al. (2016) created an instrument for shortages based on hydroelectric power availability, state-level rainfall and other variables. These latter two methods are applicable in a district-or-state-level study, but not accurate in a village level one. Aklin et al. (2016) measured reliability and voltage stability by the number of days per month with no all-day power outages and with low voltage. None of these studies fully captures all the dimensions associated with electricity supply quality.

Methods

Survey

Sample Region

The region for this study, eastern UP and western Bihar, lies in the Gangetic Plain, an agriculturally rich area south of the Himalayas and home to millions of Indians. From an electricity access perspective, Bihar and UP have historically been among the least electrified in India, though they have seen improvements in the past 3-4 years (Abhishek et al., 2018.)

Figure 1: Communal irrigation system, powered by grid electricity. These systems were found intermittently throughout the survey area. Photo info: Thomas Lutken, Saraya Tiwari, Gorakhpur, Uttar Pradesh (July 17th, 2018.).

All of our villages were rural and primarily agrarian, although many of the small farms were within a day's drive of a relatively large (~10,000s-100,000s population) urban center. This meant that, in addition to farming (the most common occupation,) many people earned income as day laborers, shop employees, and other non-farm jobs. Based on field observations, typical crops were rice, millet, or sugarcane. Every village we visited had irrigated agricultural fields.

The electricity usage in the villages was highly varied. Most houses used some electricity, but many did not, and several houses had decentralized energy solutions, including solar panels. Several villages also had electrical irrigation pumps. Household wealth was highly varied within and across villages as well, as indicated by variation in housing: small one or two room mud houses with thatched roofs to multi-storey brick houses. We conducted surveys in a wide variety of village sizes, some having only a hundred households and others nearly a thousand. Some villages comprised of a single large habitation, while others were spread over several hamlets, sometimes divided along caste lines.

Survey Design

Over the summer of 2018, our team surveyed 500 households and 101 enterprise surveys within 40 village communities.

The household survey consisted of 8 sections: Sections A and B covered household consent and demographic, education, and health information for each inhabitant. Section C covered household electricity quality, appliance usage, and coping costs. Section D asked about household fuel use and expense, section E about economic indicators like income and consumption, and F and G covered daily activities of the household head and their spouse. Section H was for enumerator observations, including home construction material and a rating of the appearance of legal or illegal connections. Lastly, a Raven's childhood aptitude test was given to all children in the house between the ages of 5 and 17. This household survey was administered in 10 to 12 households within each village we visited. The average time for conducting the survey was under 2 hours.

An enterprise survey was also administered to 2 or 3 businesses in each village. These were randomly selected but were located in the same area as the village where our household surveys took place. Finally, a community survey was given to the head of each village or to community leaders. All components of the survey were given IRB approval by Duke University, and the survey team completed IRB training as required.

To implement the survey, the research team worked with a local firm named Morsel, which had experience translating and collecting electricity access survey data in Uttar Pradesh and Bihar. The team relied on this firm's expertise for translating the survey from English to Hindi, and implementing the survey in the field, including interviewing and sampling of respondents.

Survey Sampling Methods

We used the village as the sampling unit. The National Rural Drinking Water Program (NRDWP)'s database of village was used to identify potential survey locations. We limited our survey to four Discoms, two of which operate in Uttar Pradesh (Purvanchal and Madhyanchal Discoms) and two of which operate in Bihar (North and South Bihar Discoms). Both Purvanchal and Madhyanchal Discoms are subsidiaries of the Uttar Pradesh Power Corporation Limited (UPPCL). See **Figure 2** which places our sample area in geographical context of Northern India. Villages were chosen based on region of interest.

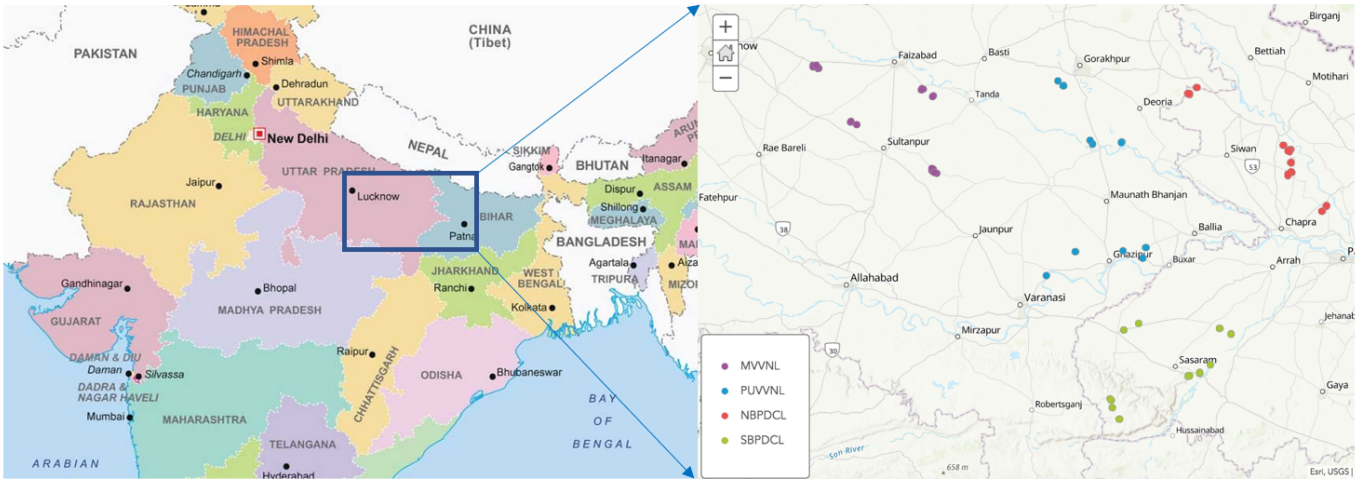

Figure 2: Left shows our survey area in Northern India, Right shows our villages and which distribution company provides their electricity.

Each Indian state is comprised of administrative units called districts, which themselves contain blocks. To ensure that our sample contained villages representing a range of electricity quality levels, we systematically chose blocks within each Discom and district that were characterized as "good" or "bad" based on the hours of reported outages as obtained from the government's Urban electrification database (URJA), as well as advice provided by Morsel. Blocks are made up of Gram Panchayats (GP), analogous to a city or county. We randomly selected these GPs using the NRDWP's list, and then we randomly sampled villages from the list in each GP.

Of the 40 villages, 10 selected from each of the Discoms and 20 in each State. Each village had 12 or 13 household surveys, to reach our *a priori* goal of 500 total household surveys. To avoid geographic biases, if a village had multiple hamlets, we would assign surveys based on the proportion of the village's population living in each hamlet. This also helped ensure representation of a range of socioeconomic standings in our sample, because these hamlets often corresponded to caste or wealth. Within this geographic distribution, we randomly selected houses to be surveyed.

Index Generation

The World Bank Multi-Tier Framework

In 2015, the World Bank Group devised the Multi-Tier Framework to define access to electricity. The framework ranks households from tier 0 to 5 based on seven key attributes: capacity, availability, reliability, quality, affordability, legality, and health & safety. The framework is still relatively new and has seldom been applied to generate independent variables for socioeconomic research. Our study is therefore pioneering in adopting this methodology.

We made several modifications when adopting the framework. The first modification was to add a new attribute called "Predictability" to take it into consideration since the issue of unpredictable electricity supply was ubiquitous in the context of India. Second, we tailored some categorizing criteria to fit our survey questions because we designed the survey independently, causing discrepancies between our survey and the standard World Bank survey for measuring several attributes (**Table 1**). Third, in the World Bank Multi-Tier Framework, the aggregated tier score that represents a household's overall access to electricity is the minimum score obtained in any of the attributes (2015). However, in our study, an aggregated tier score is calculated by taking the average of tier scores for all eight attributes to weight them equally.

$$Aggregated\ Tier = \frac{Sum\ \ (Capacity, Availability, Reliability, Quality, Affordability, Legality, Safety\ \&\ Health, Predictability)}{8}$$

We decided not to use the minimum tier score because it could only generate a variable in a range of 0,1,2,3,4,5 to represent quality, whereas taking an average could allow us to differentiate households with somewhat more granularity (for example, households with scores of 2.2 and 2.7). Also, the minimum tier score method tends to underestimate the electricity quality by overweighting lower quality dimensions. In our data, households would never exceed Tier 3 and the majority of households would be classified as Tier 2 or lower, levels that usually represents off-grid communities (the World Bank Group, 2015). In reality, we observed households with reasonably good electricity supply that used medium to high-power appliances that more typically qualify Tier 4 conditions. Thus, the method of taking an average was deemed to be better aligned with the reality.

Table 1. Modified Multi-Tier Framework

Attribute	**Definition**	**Tier 0**	**Tier 1**	**Tier 2**	**Tier 3**	**Tier 4**	**Tier 5**
Capacity	Peak capacity of appliances operating at the same time in a household.	Unelectrified	3W	50W	200W	800W	2000W
Availability	Minimum hours per day households get electricity	Unelectrified	-	4	8	16	23
Reliability	Frequency and total duration of electricity interruption happened in this community in the past week	Unelectrified	-	-	Frequency: At least 3-5 times a day AND Duration: >4 hours/day	Frequency: no more than once or twice per day AND Duration: =<4 hours/day	Frequency: only once. AND Duration: <2 hours/week
Quality (voltage)	The frequency of home appliances	Unelectrified	A lot (>5 times)	Often (4-5 times)	Occasionally (2-3 times)	Very rarely (1 time)	Never

	damage due to voltage surge in the past year						
Affordability	Household's feeling for the affordability of electricity	Unelectrified	-	Electricity is expensive	Not sure if electricity is expensive	-	Electricity is not expensive
Legality	Whether a household received an electricity bill and whether the surveyor think the household has a legal connection	Unelectrified	Household: never received a bill AND Surveyor: illegal connection	-	Household: never received a bill Or Surveyor: illegal connection	-	Household: received a bill AND Surveyor: legal connection
Safety & Health	Whether a household feel the electricity hookups in house are safe and whether a household have MCBs (miniature circuit breaker), earth grounding wire and fuse installed	Unelectrified	Unsafe hookups or Unsure safety AND No MCBs, earth grounding wire and fuse installed	Unsafe hookups OR Have MCBs, or earth grounding wire or fuse installed	Unsure safety OR Have MCBs, or earth grounding wire or fuse installed	Safe hookups OR Have MCBs, or earth grounding wire or fuse installed	Safe hookups AND Have MCBs, earth grounding wires and fuse installed
Predictability	Whether a household can predict the time when outages will occur	Unelectrified	Completely random	-	Somewhat but not very predictable	Mostly predictable	Always Predictable

Principal Components Analysis

In addition to the Multi-Tier Framework, we also explored to use of Principal Components Analysis (PCA) to measure electricity quality. PCA is a statistical tool that can reduce a correlated multivariate dataset into uncorrelated components. We aim to apply it to reduce the large amount of information we collected around electricity quality via our survey into a single summary index that captures relevant variation in quality. The first principal component is a linear combination of variables that has maximum variance, which captures most of the variation in a dataset (Jackson, 1991). PCA has been widely applied to generate indexes measuring socioeconomic status for households, often based on asset ownership information (Vyas and Kumaranayake, 2006; Gunnsteinsoon et al., 2010).

In our study, the survey responses related to electricity quality were aggregated into an index to measuring power quality, by using data on availability, outage frequency, and predictability. Compared to other studies that use density of transmission cables to signal electricity quality, PCA directly utilizes the most relevant data in a particular situation, thus

creating a more relevant representation of the real situation in a given micro-level study. Compared to the World Bank Multi-Tier framework, PCA will weigh each variable with different loadings and better capture the inherent the relationships between variables incorporated into the analysis.

Unfortunately, in our case, the household level data was not suitable for PCA for two reasons. First, the household data was nested. Interactions of both community and household variables would contradict PCA's assumption of an identical linear relationship of variables. Second, the household survey arguably captured less useful information on electricity quality than the community survey did. This observation was largely made on the basis of what we observed in application of a household-level PCA estimation, which offered only mixed evidence that such an index provided a useful electricity quality concept. Therefore, PCA was only conducted with community survey data. After verifying that the loadings represent an intuitive relationships between these variables and electricity quality, we generated an index out of the first principal component to represent community-level electricity quality. Since PCA was not implemented on our household-level data, we did not use this index in our household regression analyses.

Patterns and Characteristics

Patterns of Electrification

Before looking at electricity quality, we firstly analyzed how discoms, geographic and socioeconomic characteristics of households were related to electrification decisions. We used probit regression, with binary electrification status as the dependent variable, and created different specifications to control for different combinations of household characteristics. Hindu households belonging to Open/General caste and Discom Madhyanchal is the omitted discom in our regression. The specifications are

Base (1): $\text{Electrification}_1 = \sum \beta_{k1}\text{discom}_{k1} + \beta_1 + \varepsilon_1$
Geographic (2): $\text{Electrification}_2 = \sum \beta_{k2}\text{discom}_{k2} + \sum \beta_{k2}\text{geo}_{k2} + \beta_2 + \varepsilon_2$
Demographic (3): $\text{Electrification}_3 = \sum \beta_{k3}\text{discom}_{k3} + \sum \beta_{k3}\text{dem}_{k3} + \beta_3 + \varepsilon_3$
Wealth (4): $\text{Electrification}_4 = \sum \beta_{k4}\text{discom}_{k4} + \sum \beta_{k4}\text{wealth}_{k4} + \beta_4 + \varepsilon_4$
Full (5): $\text{Electrification}_5 = \sum \beta_{k5}\text{discom}_{k5} + \sum \beta_{k5}\text{geo}_{k5} + \sum \beta_{k5}\text{dem}_{k5} + \sum \beta_{k5}\text{wealth}_{k5} + \beta_5 + \varepsilon_5$

j: 'Base' (j=1), 'Geographic' (j=2), 'Demographic' (j=3), 'Wealth' (j=4), and 'Full' (j=5) specifications
Electrification_j = a binary dependent variable for j^{th}specification (0 for unelectrified and 1 for electrified)
β_{kj}= coefficient term for the k^{th}variable in j^{th}specification
ε_j= error term for the j^{th}specification

The full list of variables appears in **Appendix 1**. We selected these variables based on literature review. For variables with less than 10% values missing, we replaced them with village means.

For variables with more than 10% values missing, we left those values as missing and omitted those samples in regressions to avoid biasing the result.

Patterns of Electricity Quality

Next, to understand the patterns of electricity quality, we first compare the performance of Discoms with respect to the eight electricity quality attributes by running the following regression:

$$Y_i = \sum \beta_{ki} discom_{ki} + \beta_i + \varepsilon_i$$

Where,

Y_i = dependent variable for the i^{th}model's j^{th}specification. They are tier scores for Capacity (i=1), Availability (i=2), Reliability (i=3), Quality (i=4), Legality (i=5), Affordability (i=6), Safety (i=7), Predictability (i=8).

Second, on top of the regressions above, we picked regressions for several key aspects of electricity quality (availability, reliability, predictability, and the aggregated quality) and then controlled for geographic, demographic, wealth, electricity infrastructure, and the combined effects of these variables in the five other specifications. The six specifications are:

Base (1): $Y_{i1} = \sum \beta_{ki1} \text{discom}_{ki1} + \beta_{i1} + \varepsilon_{i1}$
Geographic (2): $Y_{i2} = \sum \beta_{ki2} \text{discom}_{ki2} + \sum \beta_{ki2} \text{geo}_{ki2} + \beta_{i2} + \varepsilon_{i2}$
Demographic (3): $Y_{i3} = \sum \beta_{ki3} \text{discom}_{ki3} + \sum \beta_{ki3} \text{dem}_{ki3} + \beta_{i3} + \varepsilon_{i3}$
Wealth (4): $Y_{i4} = \sum \beta_{ki4} \text{discom}_{ki4} + \sum \beta_{ki4} \text{wealth}_{ki4} + \beta_{i4} + \varepsilon_{i4}$
Electricity (5): $Y_{i5} = \sum \beta_{ki5} \text{discom}_{ki5} + \sum \beta_{ki5} \text{electricity}_{ki5} + \beta_{i5} + \varepsilon_{i5}$
Full (6): $Y_{i6} = \sum \beta_{ki6} \text{discom}_{ki6} + \sum \beta_{ki6} \text{geo}_{ki6} + \sum \beta_{ki6} \text{dem}_{ki6} + \sum \beta_{ki6} \text{wealth}_{ki6} + \sum \beta_{ki6} \text{electricity}_{ki6} + \beta_6 + \varepsilon_6$

i: hours of electricity availability (i=1), reliability tier (i=2), predictability tier (i=4), and the aggregated tier (i=5).
j stands for 'Base' (j=1), 'Geographic' (j=2), 'Demographic'(j=3), 'Wealth'(j=4), 'Electricity' (j=5), and 'Full' (j=6) specifications.

Y_{ij} = dependent variable for the i^{th}model's j^{th}specification
β_{kij}= coefficient term for the k^{th}variable in i^{th}model's j^{th}specification
ε_{ij}= error term for the i^{th}model's j^{th}specification

Relationships between Reliability and Socioeconomic Variables

We were interested in two socioeconomic variables that are potentially related to energy access and quality: standard of living and consumption of alternative fuels. We represented standard of living with household consumption (minus fuel and electricity consumption) and alternative fuel consumption with household kerosene consumption. We specified two linear

regression models to determine the importance of relationships between electrification, the eight attributes of electricity quality, and other potential confounding variables, on the one hand, and these potential outcomes, on the other. Both of these models have the same two specifications, each with four sub-specifications. The purpose of the different specifications for the models is to study to the effect of electrification overall, as well as the effects of its various characteristics. We acknowledge that there may be endogeneity that renders a causal interpretation of these relationships difficult.

The two specifications of the models are: binary and tiered. The 'binary' specification includes all households and aims to access relationships with electrification status overall, whilst the 'tiered' specification includes only electrified households and the various characteristics of their electricity supply to determine those characteristics' relationships with the variables of interest.

The four sub-specifications of the model exist to account for different sets of potentially confounding variables when studying the effects of electrification and its characteristics. The models were specified as follows,

Base: $Y_{i1} = \sum \beta_{k1} elec_{k1} + \beta_{i1} + \varepsilon_{i1}$
Demographic: $Y_{i2} = \sum \beta_{k2} elec_{k2} + \sum \beta_{k2} dem_{k2} + \beta_{i2} + \varepsilon_{i2}$
Wealth: $Y_{i3} = \sum \beta_{k3} elec_{k3} + \sum \beta_{k3} wealth_{k3} + \beta_{i3} + \varepsilon_{i3}$
Full: $Y_{i4} = \sum \beta_{k4} elec_{k4} + \sum \beta_{k4} dem_{k4} + \sum \beta_{k4} wealth_{k4} + \beta_{i4} + \varepsilon_{i4}$

Y_{ij} = Dependent variable, either standard of living or household fuel consumption, for the i^{th}specification's j^{th}sub-specification
β_{ij}= constant term for the i^{th}specification's j^{th}sub-specification
ε_{ij}= error term for the i^{th}specification's j^{th}sub-specification
elec = key electricity independent variables (binary, fully-tiered, or semi-tiered)
dem = relevant household demographic variables
wealth = relevant household wealth variables

The full list of variables again appears in **Appendix 1**.

Data Check

All the models were also subjected to heteroskedasticity and multicollinearity checks. Heteroskedasticity is when there is differing variance in the error term of a model; it can have multiple potential causes such as measurement errors or omitted variables. For linear regressions, constant variance in the error term (homoscedasticity) is an important assumption. When this assumption is violated, standard errors will be biased. We tested for heteroskedasticity using the Breusch-Pagan and Cook-Weisberg test for heteroskedasticity, and found it to be present in all specifications of all models. We corrected for this using robust standard errors.

Total multicollinearity arises when one independent variable can be represented by a linear combination of other independent variables. If a large amount of multicollinearity exists, the standard errors of coefficients might be inflated. We employ the vif (variance inflation factor) command in Stata and all the models have average vif values below 10, which means that no strong multicollinearity exists among variables considered.

Results

Descriptive statistics of the survey sample

Around seventy percent of the households we surveyed were below the poverty line. Around 90% of them belonged to historically disadvantaged caste groups such as Scheduled Castes, Scheduled Tribes and OBC.

Many households were still suffering from a lack of access to electricity and unreliable electricity supply. As of August 2018, around 23% of the households, 67% of the enterprises, and 50% of community farms still had no access to electricity. Only 65% of the electrified households and 55% of the electrified enterprises were satisfied with electricity supply. Ninety-seven percent of electrified households reported that electricity supply was unpredictable, 39% did not receive electricity bills, 70% had no miniature circuit breakers, 54% had no fuses, and 92% had no grounding wire.

Households received an average of 12.8 hours of electricity per day, though there was significant variation (SD of about 4 hours) within the sample set (**Figure 3**). Around 70% of electrified households had been connected in the previous 5 years. Fourteen percent of electrified houses in Bihar received their connection from Saubhagya, and 9.6% of electrified houses in UP got electricity through Saubhagya.

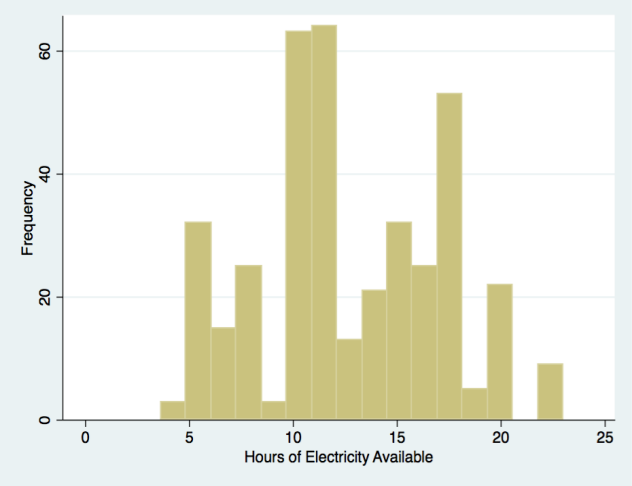

Figure 3. Hours of electricity available to surveyed households

The enterprise survey revealed that only 33 of 101 firms surveyed used electricity at all. Of those firms that did not use electricity, “far and away from lines” or “connection costs are too high” were the most common reasons for a lack of electrification. By a significant margin, firms also asserted that lighting was their most important use of electricity. Most connected firms reported issues with their electricity supply. All electrified firms had at least one outage a day. However, 25 of the firms reported that these issues had no effect on their business; only 2 reported a severe effect that jeopardizes their business operations.

Tier scores and PCA scores

The aggregated tier index for electrified households averages close to Tier 3, which means that households living in this region suffer short duration, unpredictable and unreliable electricity supply (The Saubhagya section has a detailed comparison of electrified and unelectrified households). None of the households had Tier 5 access, which means that there is a large room for improvement. No strong spatial pattern is observed (**Figure 6**) in plots of the village-averaged aggregated tier score. However, large variations are observed across and within different communities.

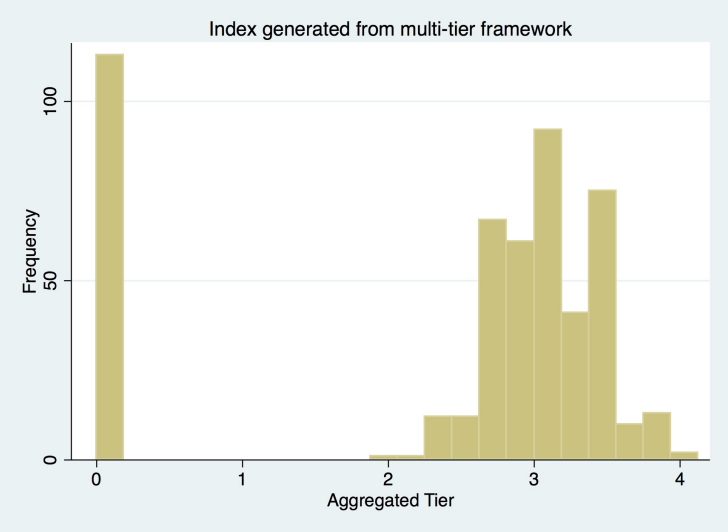

Figure 4: Histogram of number of households in given Aggregated Tier scores.

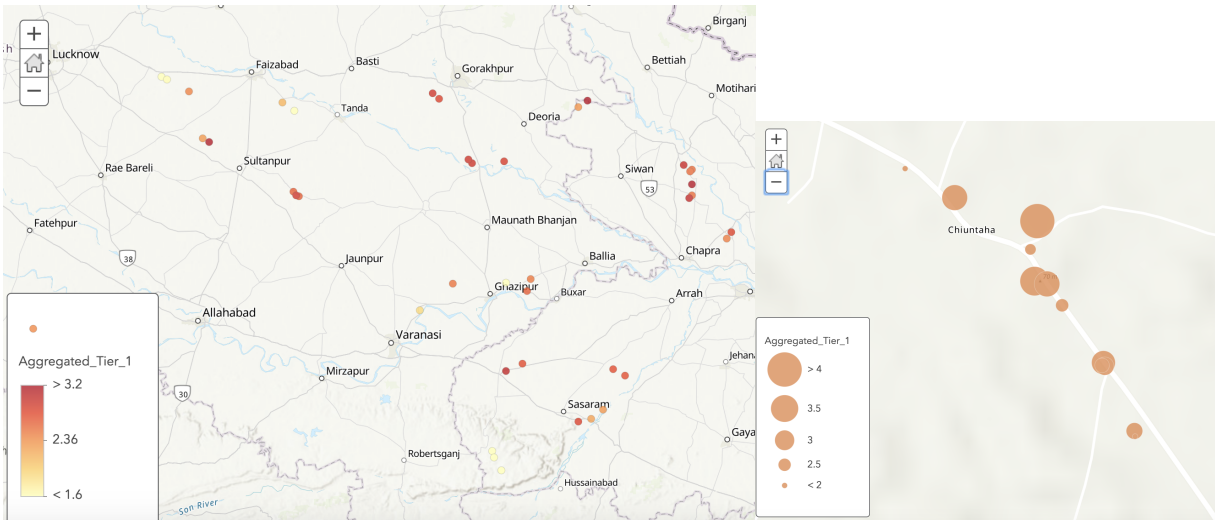

Figure 5: Left: community-averaged aggregated tier scores (the darker color represents better electricity quality). Right: aggregated tier score for several households in proximity (larger circle represents better electricity quality).

Since the PCA procedure only used community-level variables on availability, reliability and predictability, the PCA score was only comparable to the combination of tier scores in availability, reliability and predictability. To compare tier scores with the community-level PCA indexes, we took an average of the three relevant tiers and found that results were relatively consistent with each other with a correlation of 0.48. As seen, tier scores capture variation within a community, however, and were therefore used in the later regressions.

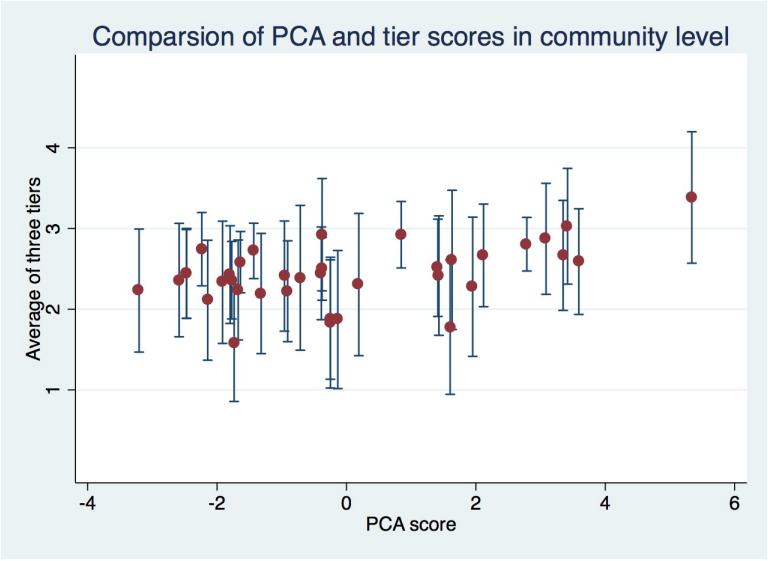

Figure 6: Comparison of PCA and tier scores (average of Availability, Reliability and Predictability tier scores) at the community level. Red dots present one community. The error bar represents one standard deviation of tier scores for each community.

Comparison to Greater UP and Bihar

After our survey was completed, the Council on Energy, Environment, and Water published "Access to Clean Cooking Energy and Electricity: Survey of States 2018." This survey was given to over 9,000 households in 756 villages across India (Abhishek et al., 2018). Compared to this larger survey, our data look similar to the state-wide averages. **Table 2** compares our survey with the Access survey's UP and Bihar averages. The main difference is in the percent of houses listing "high connection costs" as a barrier to electrification.

Table 2: Comparison of Access Survey and our survey. Adhaura's 32 unelectrified homes make Bihar's "high connection cost" value 23%. Removing this subset brings that value to 50%.

State	Access Survey	Our Survey
Median hrs/day electricity		
Bihar	15 hr	12 hr
UP	15 hr	14 hr
% using grid electricity		
Bihar	75%	74%
UP	59%	69%
% citing cost as barrier to electrification		
Bihar	64%	23% (50%)
UP	66%	62%

Comparison of Saubhagya and Unelectrified populations

We compared households which received Saubhagya connections to those which remained unelectrified, as well as to those which should have received Saubhagya, to see if there were patterns[MJ1]. Those households which benefited from Saubhagya were roughly as poor as the unelectrified homes, with 91% of Saubhagya households and 83% of unelectrified households falling below the poverty line. Only 64% of electrified households are below poverty line. There were some geographic patterns of Saubhagya adoption, with several blocks having no households receiving the program. However, it is pertinent to mention that our analysis was done in the middle of Saubhagya's implementation, so it is very possible that many of these houses have since received the scheme.

Two-way t-tests were run between Saubhagya households, other electrified households, and other households which ought to have received Saubhagya. Saubhagya houses were found to have statistically significantly higher incomes than other BPL houses, though total property

values were not found to be different. Comparing the Saubhagya households to those which ought to have received Saubhagya yielded a similar result, with the Saubhagya houses having higher incomes but similar property values. It is also worth noting that households which did not receive Saubhagya had also not received any other government subsidy; this could imply that the problem is more one of access to government assistance in general than in Saubhagya's specific implementation.

See **Table 3** below for a breakdown of caste by populations. Of note are the differences between the Scheduled Tribes and Open/General makeup of the unelectrified households and the Saubhagya households. Relating to Saubhagya, 47 unelectrified households said they were not connected because doing so would be too expensive, though 40 of these said they did not know how much such a connection would cost.

Table 3: Caste breakdown of survey sample, based on electrification status as well as those who should receive Saubhagya and those who have received Saubhagya.

Caste Makeup of survey sample			
Caste	**Should be Saubhagya (N=47)**	**Saubhagya (N=61)**	**Other Electrified Houses (N=329)**
SC	40.40%	27.90%	29.50%
ST	2.10%	1.60%	2.40%
OBC	51.10%	50.00%	54.40%
General	6.40%	20.00%	12.50%

Saubhagya and effects on multi-tier scores

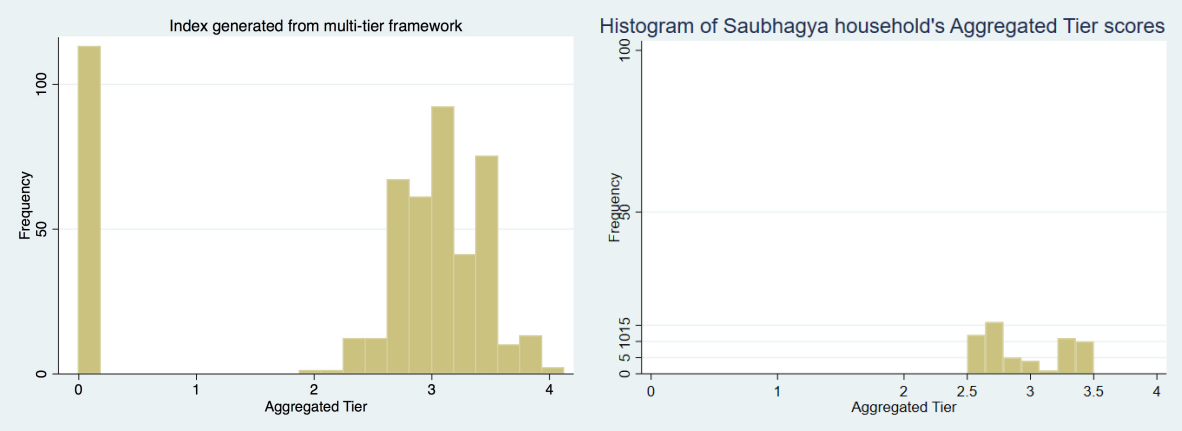

Figure 7: Comparison of tier scores for all households(left) to Saubhagya households(right.) Scale is preserved to show relative proportions of Saubhagya connected homes.

Eighty five percent of Saubhagya households had meters, while only 59% of non-Saubhagya electrified households had a meter installed. Still, correlation between the two variables was not high, at only 0.19, because a sizeable number of houses (193) had meters installed but were not part of Saubhagya.

We ran regression models on a Saubhagya-specific series of variables. Overall, 59 households in our survey sample had their electricity installed as part of the Saubhagya program, representing approximately 15% of the electrified households (Bihar 18%, UP 12%). To ensure a more uniform sample, only houses below the poverty line were included in further analysis.

Dependent variables were each individual tier score and the aggregated tier score. Independent variables were presence of a Saubhagya connection, the presence of a meter, the total years electrified, State ID, property value, number of wage earners, and caste. Saubhagya was only significantly correlated with the reliability tier score (= -0.572 $p<0.001$), and affordability tier score (=0.375, $p=0.051$). The presence or absence of a meter had no significant relationships with any dependent variable. VIF tests for these variables produced values between 1 and 1.2. Models were run with robust standard errors to correct for heteroskedasticity.

We also ran basic statistical tests to see how much of our sampled population received any kind of government subsidy, as a way of approximating overall efficiency in distributing such programs to our region. Overall, we found similar rates of government subsidized households in both states, 36% for Bihar and 32% for UP. Among electrified homes, those receiving any kind of government subsidy had a mean Aggregated Tier score of 2.95, while those without subsidy scored 3.04. A t-test on these populations showed significant differences between them, $t=2.607$ ($p=0.01$).

Patterns of Electrification

As shown in the Base Model (**Table 4)**, households in South Bihar are less likely to be electrified. But after controlling the distance of houses from electric pole, we find that distance from the electric pole is a more significant impediment to electrification and the real reason behind fewer households getting electrified in South Bihar is that more households live farther from the electric poles. Compared to our baseline, the probit model shows that households in Discom Purvanchal, North Bihar and South Bihar are more likely to be electrified than households in Madhyanchal, at the 10% significance level.

We also find that households with more adults, more elders, higher education level for the head of household, higher annual income and higher property value are more likely to be electrified. For households that belong to Scheduled Tribes, a historically disadvantaged caste group in India, and that live further from the electric poles, electrification rates are lower.

Table 4: Probit regression results associated with electrification.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Base Model	Geographic Model	Demographic Model	Wealth Model	Full Model
discom_purvanchal	0.256	0.337*	0.389*	0.173	0.413*
	(0.180)	(0.186)	(0.199)	(0.185)	(0.211)
discom_northbihar	0.630***	0.621***	0.719***	0.483**	0.638***
	(0.197)	(0.198)	(0.217)	(0.208)	(0.228)
discom_southbihar	-0.379**	0.533**	-0.0541	-0.402**	0.687***
	(0.167)	(0.219)	(0.190)	(0.175)	(0.248)
dem_num_adults			0.245***		0.254***
			(0.0535)		(0.0603)
dem_num_youngchild			-0.0965		-0.107
			(0.0675)		(0.0839)
dem_num_teenager			0.0556		0.0312
			(0.0532)		(0.0594)
dem_num_elders			0.346***		0.293***
			(0.103)		(0.110)
dem_num_earners			-0.0743		-0.0332
			(0.0743)		(0.0829)
dem_head_edu_lvl			0.200***		0.213***
			(0.0566)		(0.0632)
dem_muslim_hh			-0.0444		0.363
			(0.299)		(0.372)
dem_caste_SC			-0.340		0.125
			(0.267)		(0.289)
dem_caste_ST			-1.562***		-0.343
			(0.362)		(0.536)
dem_caste_OBC			-0.144		0.0278
			(0.257)		(0.269)
geo_dist_headquarter		0.00183			0.00223
		(0.00430)			(0.00495)
geo_dist_electricpoles		-0.00276**			-0.00317**
		(0.00126)			(0.00145)
wealth_annual_inc				8.31e-06***	5.03e-06*
				(2.39e-06)	(2.62e-06)
wealth_prop_val				3.56e-07*	-2.45e-08
				(1.84e-07)	(1.92e-07)
Constant	0.674***	0.666***	-0.348	0.106	-0.940**
	(0.122)	(0.217)	(0.362)	(0.182)	(0.448)
Observations	499	461	499	499	461

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Patterns of Electricity Quality

After we identified that the Discoms variables were related to electrification, we compared the quality of electricity across them as well.

Table 5 shows that, compared to Madhyanchal, Discom Purvanchal has significantly better performance with respect to capacity, availability, reliability and affordability. However, more households in Purvanchal have illegal connections. North Bihar has significantly higher scores in reliability and affordability, but it offers significantly fewer hours of electricity than Madhyanchal. And South Bihar offers significantly longer electricity supply.

When combined with the results for electrification, we can observe a phenomenon that in Uttar Pradesh, while more households have illegal connections, Discom Purvanchal does a good job in terms of electrification and providing electricity. Although North Bihar has more households electrified, fewer hours of electricity are available for these households. In contrast, although South Bihar has a lower percentage of electrified households, households enjoy longer hours of electricity available if electrified.

Table 5. Comparison of Discoms' electricity quality for different attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Capacity	Availability	Reliability	Quality	Legality	Affordability	Safety	Predictability
Purvanchal	0.198*	0.257***	0.317***	0.0919	-0.600***	0.236**	0.0598	0.139
	(0.112)	(0.0964)	(0.0645)	(0.106)	(0.211)	(0.113)	(0.163)	(0.149)
North Bihar	0.0510	-0.251***	0.0973***	-0.0344	0.198	0.344***	0.0718	-0.0480
	(0.100)	(0.0865)	(0.0280)	(0.110)	(0.205)	(0.128)	(0.154)	(0.147)
South Bihar	-0.0651	0.271***	0	0.129	-0.0612	0.124	-0.00642	-0.248
	(0.103)	(0.0956)	(3.21e-09)	(0.121)	(0.225)	(0.120)	(0.187)	(0.152)
Constant	2.312***	3.118***	3***	4.495***	3.581***	2.409***	3.344***	1.871***
	(0.0717)	(0.0731)	(3.21e-09)	(0.0803)	(0.142)	(0.0783)	(0.119)	(0.108)
Observations	387	387	387	387	387	387	387	387
R-squared	0.016	0.120	0.107	0.007	0.040	0.021	0.001	0.016

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

We next controlled for more factors related to geographic, demographic, socioeconomic and electricity infrastructure characteristics of these households to better understand the drivers of electricity quality. In addition to running a regression on the aggregated tier score, we also ran regression on specific attributes of electricity. The robust results for all regressions can be found in the **Appendix 2**, examples of which are shown below. There are some confounding factors influencing the results.

Table 6. Robust regressions of aggregated tiers on variables

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Base Model	Geographic Model	Demographic Model	Wealth Model	Electricity Model	Full Model
discom_purvanchal	0.0870*	0.0847*	0.108**	0.114**	0.0979**	0.152***
	(0.0501)	(0.0508)	(0.0489)	(0.0534)	(0.0448)	(0.0470)

discom_northbihar	0.0536	0.0519	0.0959**	0.0620	0.0386	0.109**
	(0.0491)	(0.0491)	(0.0465)	(0.0513)	(0.0442)	(0.0504)
discom_southbihar	0.0180	0.0305	-0.00601	0.0285	0.0681	0.0565
	(0.0540)	(0.0547)	(0.0521)	(0.0557)	(0.0465)	(0.0542)
dem_num_adults			-0.0255**			-0.0254***
			(0.0117)			(0.00932)
dem_num_youngchild			-0.0243*			-0.0230
			(0.0145)			(0.0147)
dem_num_teenager			-0.0207*			-0.0130
			(0.0120)			(0.0112)
dem_num_elders			-0.0260			-0.0156
			(0.0210)			(0.0184)
dem_num_earners			0.136***			0.0959***
			(0.0193)			(0.0205)
dem_head_edu_lvl			0.0290**			0.0182*
			(0.0117)			(0.0106)
dem_muslim_hh			0.0561			0.0107
			(0.0702)			(0.0671)
dem_caste_SC			-0.0128			0.0167
			(0.0544)			(0.0477)
dem_caste_ST			-0.213*			-0.124
			(0.122)			(0.0870)
dem_caste_OBC			-0.00872			0.0396
			(0.0468)			(0.0439)
geo_dist_headquarter		0.00157				0.00153
		(0.00109)				(0.000980)
geo_dist_pole		-0.000202				-0.000174
		(0.000411)				(0.000271)
wealth_annual_inc				1.40e-07		-3.12e-07
				(2.49e-07)		(1.92e-07)
wealth_prop_val				1.99e-08		-5.77e-08*
				(3.56e-08)		(3.09e-08)
elec_peak_usage					0.000191***	0.000288***
					(4.55e-05)	(4.43e-05)
elec_circuit_breaker					0.128***	0.0971***
					(0.0309)	(0.0355)
elec_fuse					0.0647***	0.0299
					(0.0204)	(0.0206)
elec_grounding					-0.0324	0.0131
					(0.0565)	(0.0552)
elec_meter					-0.00879	0.00113
					(0.0347)	(0.0342)
elec_bill_effi					0.106***	0.120***
					(0.0347)	(0.0386)
elec_hhonly_interrupt					0.0274	0.0103
					(0.0433)	(0.0480)
elec_legal_obs					0.241***	0.205***
					(0.0360)	(0.0348)

elec_connection_yr					0.00443*	0.00430*
					(0.00228)	(0.00246)
Constant	3.016***	2.956***	2.956***	2.988***	2.647***	2.609***
	(0.0363)	(0.0582)	(0.0763)	(0.0420)	(0.0513)	(0.0891)
Observations	386	386	386	365	386	365
R-squared	0.009	0.015	0.197	0.018	0.325	0.435

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*A. Robust Regression on overall electricity quality (represented by the aggregated tier score) (**Table 6**)*

At a 10% or lower significance level, results show that Discom Purvanchal and North Bihar offers better overall electricity supply than Madhyanchal. Demographically, households with more adults tend to have worse electricity supply, whereas households with more earners and higher education level have better electricity supply. Protective devices (e.g. miniature circuit breaker and fuse) have a significantly positive correlation with overall electricity quality. Households that pay bills regularly and have legal connection also enjoy better electricity supply and have higher peak electricity consumption. Older connections are also better. Counter-intuitively, property value is negatively correlated with electricity quality at the 10% significance level.

*B. Robust Regression on electricity availability (represented by hours of electricity available) (**Appendix.2.1**)*

The statistical significance (at 5% or lower) of Discoms is consistent across models for electricity availability. Hours of electricity available are the greatest in South Bihar and lowest in North Bihar. Households with longer electricity availability also tend to use more electricity during peak hours. Households with legal connections have longer electricity availability. Counterintuitive findings are that metered households with more earners tend to have fewer hours of electricity available, and that Scheduled Castes, another socially disadvantaged castes, tend to have more.

*C. Robust Regression on reliability (represented by reliability tier score) (**Appendix.2.2**)*

Discom Purvanchal and North Bihar offer more reliable supply than Madhyanchal at the 5% level. Different from Scheduled Castes who have significantly longer electricity availability, households belonging to Scheduled Tribes suffer from less reliable supply at a 5% significance level. Counterintuitively, households with higher property value, higher billing efficiency and earth grounding wires installed get less reliable electricity. Villages living farther away from district headquarters enjoy more reliable electricity supply.

*D. Robust Regression on predictability (represented by predictability tier score) (**Appendix.2.3**)*

At 10% significance level, electricity outage is less predictable in South Bihar and more predictable in Purvanchal. Households with more adults and young children tend to have less predictable electricity, but households with more earners and higher education level tend to have more predictable supply. What is counter-intuitive is that households with earth grounding wires installed, higher annual income, and legal connections report less predictable electricity. Households with older connections have more predictable supply.

Relationships between Reliability and Socioeconomic Variables

The robust results for all regressions can be found in the **Appendix 3**; key results are summarized below.

Table 7. Robust results for the tiered consumption model

	(1)	(2)	(3)	(4)
VARIABLES	Base Model	Demographic Model	Wealth Model	Full Model
capacity_tier	1,857***	1,421**	1,180*	871.7
	(682.4)	(661.8)	(606.0)	(616.8)
quality_tier	-2,861***	-2,689***	-2,960***	-2,770***
	(949.2)	(933.5)	(982.2)	(954.3)
num_male_adults		1,540**		1,405**
		(681.1)		(704.7)
num_male_elders		1,646*		1,294
		(984.4)		(945.9)
hh_head_edu_lvl		827.3*		875.3*
		(459.5)		(455.1)
prop_val			0.00469***	0.00448***
			(0.000764)	(0.000662)
num_chicken			-2,736	-2,986***
			(1,792)	(927.0)
Constant	14,960**	7,886	11,531*	5,309
	(6,201)	(6,425)	(5,948)	(6,200)
Observations	387	387	387	387
R-squared	0.109	0.187	0.211	0.270

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

Table 8. Robust results for the tiered fuel model

	(1)	(2)	(3)	(4)

VARIABLES	Base Model	Demographic Model	Wealth Model	Full Model
availability_tier	-15.56***	-15.61***	-13.66***	-13.73***
	(4.361)	(4.693)	(4.209)	(4.350)
quality_tier	-6.806	-11.49**	-7.311	-12.16**
	(5.187)	(5.116)	(4.998)	(5.109)
legality_tier	-7.940***	-6.259***	-8.504***	-6.375***
	(1.809)	(2.117)	(1.913)	(2.174)
predictability_tier	-14.06***	-9.839**	-12.42***	-7.628**
	(4.637)	(4.372)	(3.732)	(3.580)
num_male_earners		-13.66***		-14.67***
		(4.444)		(4.827)
num_female_earners		-9.318		-12.23**
		(5.843)		(6.112)
num_chicken			-28.04***	-16.77**
			(7.556)	(6.589)
Constant	238.1***	246.4***	217.2***	226.0***
	(42.79)	(56.82)	(38.75)	(49.64)
Observations	387	387	387	387
R-squared	0.060	0.103	0.083	0.130

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

A. Results for the Consumption Model

Electrification is found to have a statistically significant positive correlation with standard of living (at the 1% level) across all sub-specifications. However, when looking at the characteristics of electrification for electrified households, only capacity (i.e. peak usage) and quality of electrification are found to have statistically significant correlations with this outcome.

Capacity is found to have a statistically significant positive correlation with standard of living (at the 1, 5, or 10% level, depending on the sub-specification), which is intuitive as higher capacity allows households to reap greater benefits from electrification. Quality tier score, however, is found to have a statistically significant negative relationship with standard of living (at the 1% level) across specifications and sub-specifications.

With respect to the demographic and wealth variables, household property values were found to have a statistically significant positive relationship (at the 1% level) across specifications and sub-specifications. The number of male adults was found to have a positive correlation (at the 5% level) for all demographic sub-specifications. The head of household's education level is found to have a statistically significant positive correlation (at the 5 or 10% level, depending on (sub-)specification) with standard of living.

B. Results for the Fuel Model

For fuel consumption, electrification is not found to have any statistically significant relationship. However, amongst electrified households (in the full-tier and semi-tier specifications), the availability, legality, and predictability of electricity are all found to have a statistically significant negative correlation with fuel consumption (at the 1% level), additionally quality of electricity is found to have a statistically significant negative correlation (at the 5% level) in the demographic and full models of full-tier and semi-tier specifications.

Amongst the demographic and wealth variables, the number of male earners in the household was found to have a statistically significant negative relationship with fuel consumption (at the 1% level) across (sub-)specifications. The number of male earners in the household was found to have a statistically significant negative correlation with fuel consumption (at the 1, 5, or 10% level) among multiple (sub-)specifications. Household head education level was found to have a statistically significant negative relationship with fuel consumption (at the 10% level) among all sub-specifications of the binary specification, but not among either of the other two specifications. Alternatively, household income was found to have a statistically significant negative correlation with fuel consumption (at the 10% level), but only in the full models for the full- and semi-tier specifications. The fuel model likely suffers from aforementioned heteroskedasticity.

Discussions

Saubhagya

On the one hand, we see that household consumption does appear higher when a house gets electrified, which supports Saubhagya's goal of improving socioeconomic status through electrification. Saubhagya households also appear to have a positive correlation with a better affordability score, which is logical for a subsidized electricity connection. However, the apparent negative relationship with reliability could imply reverse causality and require further refining of the analysis. Assuming reverse causality (namely, that a Saubhagya household is not impaired by the free connection, but rather the free connection targeted an impaired household), Saubhagya could be being limited by the local power grid in its positive impacts. Looking through the lens of fuel consumption, predictability, availability, and quality appear as the three most important characteristics we found to reduce fuel use, none of which are significantly correlated with a house being connected via Saubhagya. To realize the full potential of connecting these homes to the grid, it could be beneficial to focus new government and discom policy on improving these specific metrics.

Additionally, while many households have benefitted from Saubhagya, the number of unconnected households citing connection costs as a barrier to electrification is unsettling. This corresponds to the results of the Access Survey, which found an even greater percentage of households citing costs as a barrier. Given that Saubhagya's specific method of implementation

relies on free connections, this suggests that there is an information gap in rural households. If our sample is representative of rural households in our region, then over 10% of total houses should be receiving electricity connections via Saubhagya but are not. Increasing the awareness of the program at the village level could be an avenue for reaching these households. If we assumed we could connect every household in an electrified village which cited cost as its barrier to electrification, we would add 47 houses to the 61 Saubhagya recipients, roughly an 80% increase. Direct comparisons of the populations also show some differences in income between the households, though causality is difficult to prescribe, since poorer houses might have more difficulty applying for this government subsidy.

It is worth mentioning the timing of our survey, however; it was administered in the midst of the Saubhagya program's implementation. The Access survey came to a similar conclusion and occurred several months later, though a true evaluation of the program should probably take place some time after the March 2019 goal of full household electrification.

Patterns of Electricity Quality

Implications for Discoms are discussed in the next section. Looking at the household level, first, head of household education level has a significantly positive correlation with electricity quality. This is consistent with Kemmler (2007)'s finding that households with higher education can make better electricity decisions. However, that electricity availability is negatively correlated with the number of earners and reliability is negatively correlated with the number of elders might be because earners usually are not at home and observe less hours of electricity available; and elders usually stay at home and can observe more outages and inform the respondent. These might be observation biases rather than the actual drivers of electricity quality, which requires further investigation.

Second, we observed evidence of potential discrimination with respect to castes but with opposite directions. Scheduled Castes and Scheduled Tribes are both historically disadvantaged caste groups who represents a fair proportion of our sample. What is interesting is that Scheduled Tribes have less reliable supply, but Scheduled Castes have more hours of supply. One potential reason might be drawn from Golden and Min (2012)'s finding that the victories of Bahujan Samaj Party (BSP), a political party representing SC constituents, in recent elections helped improve the electricity quality received by these households. They observed that line losses are lower in districts with more Scheduled Castes (SC). However, these political victories did not positively influence Scheduled Tribes.

Third, households with older connections tend to have better overall electricity quality and more predictable supply. Similar to the explanation under the Saubhagya section, this might be because electrification efforts intentionally seek out problematic households. Or households with early access tend to have better electricity quality as they can reinvest gains to improve

infrastructure (Thomas et al., 2018). Another explanation can be that recent electrification efforts focus on getting more households electrified quickly, at the expense of the quality of electrification. We need further study to fully disentangle the relationship.

Fourth, protective devices have mixed relationship with electricity quality. Miniature circuit breakers and fuses have significant positive correlations with overall electricity quality, but earth grounding wires have negative correlations with reliability and predictability. There might be confounding factors influencing the relationship here. A potential policy subsidizing the installation of protective electrical devices in households can, in a limited degree, improve electricity quality.

Fifth, legal connections and on-time bill payments do not necessarily lead to better electricity quality. Though households with legal connections reported longer electricity availability, they also reported less predictable supply. And households who paid bills on time reported less reliable supply.

Sixth, the relationship of income and property value with electricity quality is small and negative. This is different from our original expectation of wealthy households receiving better electricity. Throughout the models, electricity quality is negatively correlated with property value or annual income and the magnitude of the correlation is very small. For example, the regression results for the aggregated tier score shows that 1 million (in Rupees) increase in property value can only reduce the tier score by 0.057. While Kemmler (2007)'s found that electrification has a weak correlation with income, no relevant literature can be found to explain this negative relationship between electricity quality and household wealth status, which requires further study. One conclusion we can derive from this study is that we did not observe discrimination of electricity quality based on wealth.

Implications for Discoms and policies moving forward

Based on our data, Purvanchal appears to be the highest performing Discom when it comes to overall electricity quality. The significant variation within each Discom leads us towards the conclusion that targeting Districts or even Blocks, rather than Discoms, could be the best policy area to focus on when trying to implement improvements to electricity quality. Though Discoms would still need to be involved, it could mean that a tailored program for Adhaura, of expanding the infrastructure or installing a microgrid could be the best solution, while improving hours of electricity per day could be a better fit for the North Bihar Discom.

Discoms put an emphasis on billing efficiency. While it is true that improved billing efficiency does have a positive and statistically significant correlation with the aggregated tier score, houses which have better billing efficiency tend to have more damaged appliances and trouble predicting outages. This could be due to omitted variable biases or reverse causality, so

we would need further analysis to fully endorse or refute the Discom's current policy of improving bill collection as a means of improving electricity overall.

Relationships with Socioeconomic Well-being

As seen in the results for the binary models, electrification has a clear statistically significant, positive relationship with standard of living. This could be due to decreased fuel costs, thereby allowing increased household consumption. As expected we see that variables that could be proxies for household socioeconomic status, such as education level or property value, also display a statistically significant, positive relationship with standard of living. This, along a lack of significance for income and savings, could be due to property being the primary store of value for households in the communities surveyed. Thus, households with more valuable property are the wealthier households, who intuitively will have a higher standard of living.

In the tiered consumption specification, the only attributes we see that have a statistically significant relationship with standard of living are the capacity and quality. The capacity tier score displays a positive relationship, which is intuitive as peak capacity (i.e. usage) means households may be reaping greater benefits from electrification, but that relationship decreases in statistical significance as more variables are added; this would imply that while capacity may have a positive relationship with standard of living, the relationship between standard of living and other variables is stronger. Quality tier score, on the other hand, consistently displays a highly statistically significant negative relationship with standard of life, defying our initial expectations. However, when we account for the fact that quality is measured in the number of damaged appliances a household has due to electricity surges, we realize that this relationship may be due to the quality tier acting as a proxy variable for the number of appliances a household has and uses; presumably the more appliances a household has in use, the more are likely to be damaged by electricity surges. Furthermore, wealthier households are likely to have a higher standard of living (measured through household consumption) and own more appliances (e.g. electric fans, televisions) that are sensitive to electrical surges. Once more, education level and property value display a statistically significant, positive relationship with standard of living.

The binary fuel model reveals no statistically significant relationship with standard of living, however it displays highly statistically significant, negative relationships between the number of both male and female earners and fuel consumption. Further, it shows a statistically significant negative relationship between education level and fuel consumption. Of note, however, is that the constant term displays a highly statistically significant relationship with the dependent variable in all sub-specifications.

This story is quite different in the case of the Tiered Fuel Model. In it we see highly statistically significant negative relationships between fuel consumption and the availability and legality of electricity supply, a somewhat less significant negative relationship between fuel

consumption and the predictability of electricity supply, and a less (but still fairly) statistically significant negative relationship between fuel consumption and electricity supply quality. The explanation for this could lie in that the fuel consumption we looked at was for kerosene, which is largely used for lighting among households. More available and predictable electricity makes light bulbs a more viable alternative to kerosene lamps. More legal electricity could possibly result in better electricity supply to the same effect. The effect of electricity quality is likely less significant since light bulbs are less sensitive to electrical surges, though better quality of electricity supply should still result in fewer light bulbs being damaged, further increasing their viability as a light source.

Household wealth and demographic variables display extant, but decreased, statistical significance, while the significance of the constant term remains unaffected. This suggests that either these variables are correlated, or that demographic or household wealth variables served as rough proxies for the different characteristics of electricity supply.

Possible sources of error

Our study area was selected based on the fact that it was a region with relatively poor electricity quality. This probably led to some bias relative to a representative survey. This intentional selection will limit the study's ability to apply to greater India.

One possible source of systematic error stems from administrative differences between the states. Because Uttar Pradesh tended to have fewer villages in each Gram Panchayat than Bihar did, we often "randomly" selected the 1 village in UP which was the entire Gram Panchayat. This could mean that our UP households tended to be located in larger, more diverse villages than Bihar, only because of administrative designation. We hope to have accounted for some of this bias by including variables like distance to district headquarters.

Many of our answers are based on self-reported data, which will most likely be imprecise. To try and compensate, we iterated the survey by testing it on residents in Lucknow, and examining questions they struggled to answer. For some difficult questions estimating expenses, enumerators were instructed to walk the respondent through a daily example and then use multiplication to extrapolate. We also assumed uniform imprecision; or that residents of a poor household in UP would be just as accurate at recalling information about electricity outages or expenditures as a wealthy household in Bihar.

At the outset, we decided to randomly select enterprises in the same villages where we did the household surveys. This was done to ensure representation of the enterprises in the area, instead of systematically seeking out those enterprises which used electricity. As a result, the majority of enterprises sampled were small shops, which had little to no electricity use. While this limits the viability of running regressions analyzing how this electricity use impacts

businesses, it does provide us with a representative picture of the kinds of enterprises in rural villages in northern India.

Conclusions

Our study looks at the electricity quality in northern India from three perspectives. First, Saubhagya was successful at connecting many homes, though some otherwise qualified houses were not electrified. Additionally, the benefits of Saubhagya appear limited by issues with electricity quality, which suggests that future improvements will need to focus on this quality problem as well as connectivity.

Second, there is large variation in electricity quality among and within communities and Discoms have significant correlations with electricity quality. In Uttar Pradesh, Discom Purvanchal is significantly better than Discom Madhyanchal. Better electricity quality is also associated with household composition of more working members and higher education level, better safety features and early electrification, although legal connections and on-time bill payments do not necessarily lead to better electricity quality. We observed discrimination of electricity quality with respect to castes (though in the opposite directions) but no discrimination with respect to wealth.

Third, electrification has a statistically significant positive relationship with household standard of living. Electrification of a household does not have a statistically significant relationships with kerosene consumption. However, among electrified households and across multiple specifications, electricity availability, legality, predictability, and (to a lesser extent) quality all have a highly statistically significant negative relationship with household kerosene consumption.

Limitations and Further Study

Our study examined Saubhagya during the program's implementation; it is possible that some of the non-Saubhagya households we surveyed have been helped by the program in the intervening 10 months. Similarly, we came across several unelectrified communities through the course of our survey, and those communities could have been connected since we visited them because of the current government's push for electrification.

Given the nature of our analysis, we could not prove that our variables of interest were *drivers* of electrification or electricity quality, or that electrification or electricity quality were *drivers* of household standard of living or fuel consumption. We would recommend further causal study and survey work to address this.

While we find a number of interesting relationships with respect households' socioeconomic well-being, many of which are also found in the literature, we stop short of purely causal claims. For our consumption model there are clear endogeneity and reverse causality concerns, particularly with respect to variables that characterize household wealth. While these concerns are less present in the fuel model, due to one of the most common uses for electricity amongst households being lighting (thereby arguably replacing kerosene sources of lighting), they are nonetheless extant. As with the drivers of electrification and electricity quality, we would recommend further causal study.

References

Abhishek Jain, Saurabh Tripathi, Sunil Mani, Sasmita Patnaik, Tauseef Shahidi, and Karthik Ganesan (2018) 'Access to Clean Cooking Energy and Electricity: Survey of States 2018', November. https://www.ceew.in/sites/default/files/CEEW-Access-to-Clean-Cooking-Energy-and-Electricity-11Jan19_0.pdf

Aklin, Michaël, Chao-yo Cheng, Johannes Urpelainen, Karthik Ganesan, and Abhishek Jain. 2016. "Factors Affecting Household Satisfaction with Electricity Supply in Rural India." *Nature Energy* 1 (11): 16170. https://doi.org/10.1038/nenergy.2016.170.

Allcott, Hunt, Allan Collard-Wexler, and Stephen D O`Connell. 2016. "How Do Electricity Shortages Affect Productivity ? Evidence from India." *American Economic Review* 106 (3): 587–624. https://doi.org/10.1257/aer.20140389.

Banerjee, Sudeshna Ghosh, Douglas Barnes, Bipul Singh, Kristy Mayer, and Hussain Samad. 2015. Power for All: Electricity Access Challenge in India. World Bank Studies. Washington, DC: World Bank. doi: 10.1596/978-1-4648-0341-3. License: Creative Commons Attribution CC BY 3.0 IGO http://documents.worldbank.org/curated/en/562191468041399641/pdf/922230PUB0978100Box385358B00PUBLIC0.pdf

Burgig, Fiona, Louis Preonas. 2016. "Out of the Darkness and Into the Light? Development Effects of Rural Electrification". *Energy Institute at Haas.* https://ei.haas.berkeley.edu/research/papers/WP268.pdf

Chakravorty, Ujjayant, Martino Pelli, and Beyza Ural Marchand. 2014. "Does the quality of electricity matter? Evidence from rural India." *Journal of Economic Behavior and Organization* 107 (A), 228-247. https://www.sciencedirect.com/science/article/pii/S0167268114001164?via%3Dihub

Cunha, Suparna Dutt D'. 2018. "Modi Announces '100% Village Electrification', But 31 Million Indian Homes Are Still In The Dark." Forbes. 2018. https://www.forbes.com/sites/suparnadutt/2018/05/07/modi-announces-100-village-electrification-but-31-million-homes-are-still-in-the-dark/#40699f9163ba.

Dinkelman, Taryn. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." American Economic Review. December 2011. Accessed April 16, 2019. https://www.aeaweb.org/articles?id=10.1257/aer.101.7.3078.

Dugoua, Eugenie, Ruinan Liu, and Johannes Urpelainen. 2017. "Geographic and Socio-Economic Barriers to Rural Electrification: New Evidence from Indian Villages." *Energy Policy* 106 (October 2016): 278–87. https://doi.org/10.1016/j.enpol.2017.03.048.

Golden, Miriam, and Brian Min. 2012. "Theft and Loss of Electricity in an Indian State." Working Paper Version 20. https://www.theigc.org/wp-content/uploads/2014/09/Golden-Min-2012-Working-Paper.pdf.

Gunnsteinsson, Snaebjorn, Alain B Labrique, Keith P West, Parul Christian, Sucheta Mehra, Abu Ahmed Shamim, Mahbubur Rashid, Joanne Katz, Rolf D W Klemm, and Rolf D.W. Klemm. 2010. "Constructing Indices of Rural Living Standards in Northwestern Bangladesh." *Journal of Health, Population, and Nutrition* 28 (5): 509–19. http://www.ncbi.nlm.nih.gov/pubmed/20941903.

Hussain, Samad, and Fan Zhang. 2016. “Benefits of Electrification and the Role of Reliability: Evidence from India.” *World Bank Group*, Policy Research Working Paper 7889. https://openknowledge.worldbank.org/bitstream/handle/10986/25691/WPS7889.pdf?sequence=1

Hussain, Samad, and Fan Zhang. 2017 “Heterogeneous Effects of Rural Electrification: Evidence from Bangladesh.” *World Bank Group,* Policy Research Working Paper 8102. http://documents.worldbank.org/curated/en/904491497275742546/pdf/WPS8102.pdf

Jackson, J. Edward. 1991. *A User’s Guide to Principal Components*. John Wiley & Sons, Inc.

Kemmler, Andreas. 2007. “Factors Influencing Household Access to Electricity in India.” *Energy for Sustainable Development* 11 (4): 13–20. https://doi.org/10.1016/S0973-0826(08)60405-6.

Khandker, Shahidur R., Hussain A. Samad, Rubaba Ali, and Douglas F. Barnes. 2014. “Who Benefits Most from Rural Electrification? Evidence in India.” *Energy Journal* 35 (2): 75–96. https://doi.org/10.5547/01956574.35.2.4.

Lipscomb, Molly, A. Mushfiq Mobarak, and Tania Barham. "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil." American Economic Journal: Applied Economics. April 2013. Accessed April 16, 2019. https://www.aeaweb.org/articles?id=10.1257/app.5.2.200.

Ministry of Power, 2017. “Guidelines for Pradhan Mantri Sahaj Bijli Har Ghar Yojana (Saubhagya).” *Government of India* Guidance letter, October 20th 2017. https://powermin.nic.in/sites/default/files/webform/notices/Guidelines_of_SAUBHAGYA.pdf

Rao, Narasimha D. 2013. “Does (Better) Electricity Supply Increase Household Enterprise Income in India?” *Energy Policy*, 2013. https://doi.org/10.1016/j.enpol.2013.02.025.

Saubhagya dashboard, 2018. “About Saubhagya”, *Government of India*, https://saubhagya.gov.in/

The World Bank Group. 2015. “Beyond Connections: Energy Access Redefined.” *The World Bank*, 1–244. http://www.worldbank.org/content/dam/Worldbank/Topics/Energy and Extract/Beyond_Connections_Energy_Access_Redefined_Exec_ESMAP_2015.pdf.

Thomas, Daniel Robert, and Johannes Urpelainen. 2018. “Early Electrification and the Quality of Service: Evidence from Rural India.” *Energy for Sustainable Development* 44 (June): 11–20. https://doi.org/10.1016/J.ESD.2018.02.004.

United Nations Energy, 2005. “The Energy Challenge for Achieving the Millennium Development Goals”, *United Nations*, https://www.undp.org/content/undp/en/home/librarypage/environment-energy/sustainable_energy/the_energy_challengeforachievingthemillenniumdevelopmentgoals.html

Vyas, Seema, and Lilani Kumaranayake. 2006. “Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis.” *Health Policy and Planning* 21 (6): 459–68. https://doi.org/10.1093/heapol/czl029.

Appendix

Appendix 1: Variables used in models

Electricity Variables		Geographic Variables		Demographic Variables		Wealth Variables	
Variable Name	Descriptor	Variable Name	Descriptor	Variable Name	Descriptor	Variable Name	Descriptor
elec_circuit_breaker	Number of miniature circuit breaker the household has	discom_purvanchal	Whether a household is in PUVVNL	dem_num_adults	Number of adults in a household	Wealth_annual_inc	Household annual income (in Rupees) from all sources (including farming, remittance, and government support)
elec_fuse	Whether the household has fuse installed	discom_northbihar	Whether a household is in NBPDCL	dem_num_youngchild	Number of young children (<5 years old) in a household	wealth_savings	Household savings (in Rupees)
elec_grounding	Whether the household has earth grounding wires installed	discom_southbihar	Whether a household is in SBPDCL	dem_num_adolescents	Number of adolescents (5-17 years old) in a household	wealth_prop_val	The market value expected from sale of properties (land and house)
elec_meter	Whether the household has meters installed	Geo_dist_headquarter	Distance of the village from district headquarter	dem_num_elders	Number of elders (>60 years) in a household	wealth_hh_debt	Household debt (in Rupees)
elec_bill_effi	Whether a household pays bill on time (once or twice a month)	geo_dist_pole	Distance of the household from district headquarter	dem_num_earners	Number of members of a household who earn money on a regular basis	wealth_num_cows	Number of cows the household owns
elec_hhonly_interrupt	Whether a household encounter interruption only occur in this household but not in neighboring households			dem_head_edu_lvl	Head of household education level	wealth_num_buffalo	Number of buffalo the household owns
elec_legal_obs	Based on surveyor's observation, whether the household has legal connection			dem_muslim_hh	Whether the household is a Muslim household	wealth_num_goats	Number of goats the household owns
elec_connection_yr	Number of years since first connection			dem_caste_SC	Whether the household belongs to Scheduled Caste	wealth_num_chicken	Number of chickens the household owns
elec_peak_usage	Total capacity of appliances used at the same time (in Watts)			dem_caste_ST	Whether the household belongs to Scheduled Tribes	wealth_num_pigs	Number of pigs the household owns
elec_trified	Whether the household is electrified			dem_caste_OBC	Whether the household belongs to OBC (Other Backward Class)	wealth_other_ls	Number of other livestock the household owns
elec_time_electrified	Time period the household has been electrified (in months)			dem_num_male_adults	Number of male adults in a household		

Electricity Variables		Geographic Variables		Demographic Variables		Wealth Variables	
Variable Name	Descriptor	Variable Name	Descriptor	Variable Name	Descriptor	Variable Name	Descriptor
capacity_tier	Capacity of electricity supply (in tiers)			dem_num_ female_adults	Number of female adults in a household		
availability _tier	Availability of electricity supply (in tiers)			dem_num_ male_YC	Number of male young children (<5 years old) in a household		
reliability _tier	Reliability of electricity supply (in tiers)			dem_num_ female_YC	Number of female young children (<5 years old) in a household		
quality_tier	Quality of electricity supply (in tiers)			dem_num_ male_OC	Number of male adolescents (5-17 years old) in a household		
affordability_tier	Affordability of electricity supply (in tiers)			dem_num_ female_OC	Number of female adolescents (5-17 years old) in a household		
legality_tier	Legality of electricity supply (in tiers)			dem_num_ male_elders	Number of male elders (>60 years) in a household		
safety_tier	Safety of electricity supply (in tiers)			dem_num_ female_elders	Number of female elders (>60 years) in a household		
predictability_tier	Predictability of electricity supply (in tiers)			dem_num_ male_earners	Number of male members of a household who earn money on a regular basis		
elec_peak _usage	Capacity of electricity supply (in Watts)			dem_num_ female_earners	Number of female members of a household who earn money on a regular basis		
elec_hours _day	Availability of electricity supply (in hours)						

discom = distribution companies. We use three binary variables for Discoms, Purvanchal, South Bihar and North Bihar, to avoid multicollinearity.
geo = relevant geographic variables
dem = relevant household demographic variables
wealth = relevant household wealth variables
elec = relevant electricity variables

Appendix 2: Robust regression results to identify patterns of electricity quality

Appendix 2.1 Robust regression results of household-level electricity availability

VARIABLES	(1) Base Model	(2) Geographic Model	(3) Demographic Model	(4) Wealth Model	(5) Electricity Model	(6) Full Model
discom_purvanchal	1.690***	1.822***	1.879***	1.952***	1.480**	2.081***
	(0.637)	(0.647)	(0.648)	(0.661)	(0.695)	(0.719)
discom_northbihar	-1.900***	-1.928***	-1.664***	-1.797***	-1.970***	-1.637**
	(0.561)	(0.558)	(0.580)	(0.581)	(0.594)	(0.634)
discom_southbihar	2.164***	2.106***	2.368***	2.294***	2.315***	2.742***
	(0.669)	(0.686)	(0.695)	(0.683)	(0.677)	(0.739)
dem_num_adults			0.109			0.162
			(0.133)			(0.136)
dem_num_youngchild			-0.317			-0.229
			(0.218)			(0.228)
dem_num_teenager			0.0998			0.120
			(0.161)			(0.167)
dem_num_elders			0.119			0.0470
			(0.293)			(0.292)
dem_num_earners			-0.129			-0.474**
			(0.202)			(0.235)
dem_head_edu_lvl			0.0662			-0.0516
			(0.145)			(0.155)
dem_muslim_hh			-0.270			0.239
			(0.945)			(0.938)
dem_caste_SC			1.023			1.434*
			(0.746)			(0.762)
dem_caste_ST			-2.425			-1.995
			(1.473)			(1.429)
dem_caste_OBC			0.00639			0.415
			(0.685)			(0.757)
geo_dist_headquarter		-0.00511				-0.00520
		(0.0137)				(0.0154)
geo_dist_pole		-0.00615				-0.00265
		(0.00486)				(0.00514)
wealth_annual_inc				1.71e-06		-2.35e-07
				(3.30e-06)		(3.17e-06)
wealth_prop_val				-1.11e-07		-5.99e-07
				(4.54e-07)		(4.98e-07)
elec_peak_usage					0.00135**	0.00194***
					(0.000583)	(0.000663)
elec_circuit_breaker					0.254	0.252
					(0.415)	(0.455)
elec_fuse					0.278	0.299

					(0.271)	(0.305)
elec_grounding					-0.500	-0.836
					(0.748)	(0.811)
elec_meter					-0.380	-0.870*
					(0.504)	(0.515)
elec_bill_effi					0.412	0.665
					(0.457)	(0.482)
elec_hhonly_interrupt					0.0710	0.132
					(0.572)	(0.566)
elec_legal_obs					0.774	1.408**
					(0.512)	(0.555)
elec_connection_yr					0.0253	0.0246
					(0.0341)	(0.0372)
Constant	12.48***	12.83***	11.71***	12.30***	11.36***	10.68***
	(0.477)	(0.747)	(1.049)	(0.532)	(0.755)	(1.386)
Observations	386	386	386	365	386	365
R-squared	0.141	0.145	0.166	0.152	0.175	0.234

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

Appendix 2.2. Robust regression results of household-level reliability tier score

VARIABLES	(1) Base Model	(2) Geographic Model	(3) Demographic Model	(4) Wealth Model	(5) Electricity Model	(6) Full Model
discom_purvanchal	0.320***	0.308***	0.318***	0.346***	0.294***	0.296***
	(0.0651)	(0.0606)	(0.0643)	(0.0720)	(0.0654)	(0.0642)
discom_northbihar	0.0973***	0.0942***	0.0878***	0.110***	0.0637*	0.0855**
	(0.0280)	(0.0293)	(0.0308)	(0.0297)	(0.0358)	(0.0418)
discom_southbihar	0	0.0351*	-0.0180	-0.00184	0.00304	0.0250
	(2.25e-09)	(0.0199)	(0.0189)	(0.00379)	(0.0171)	(0.0349)
dem_num_adults			-0.00918			-0.00819
			(0.0121)			(0.0126)
dem_num_youngchild			0.00714			0.00450
			(0.0224)			(0.0224)
dem_num_teenager			-0.0153			-0.0207
			(0.0140)			(0.0153)
dem_num_elders			-0.0585**			-0.0599***
			(0.0233)			(0.0229)
dem_num_earners			-0.0103			-0.0169
			(0.0179)			(0.0207)
dem_head_edu_lvl			-0.0115			-0.00417
			(0.0157)			(0.0159)
dem_muslim_hh			0.0637			-0.0396
			(0.111)			(0.110)
dem_caste_SC			-0.0879			-0.0842

			(0.0698)			(0.0743)
dem_caste_ST			-0.200**			-0.136
			(0.0870)			(0.0909)
dem_caste_OBC			-0.00791			0.0293
			(0.0641)			(0.0694)
geo_dist_headquarter		0.00430**				0.00479**
		(0.00199)				(0.00200)
geo_dist_pole		-0.000244				7.23e-05
		(0.000430)				(0.000503)
wealth_annual_inc				-8.60e-08		-2.01e-07
				(2.48e-07)		(2.63e-07)
wealth_prop_val				-6.25e-08*		-8.18e-08*
				(3.56e-08)		(4.55e-08)
elec_peak_usage					4.26e-05	8.46e-05
					(7.82e-05)	(8.36e-05)
elec_circuit_breaker					0.0440	0.0406
					(0.0456)	(0.0431)
elec_fuse					0.0381	0.0389
					(0.0270)	(0.0266)
elec_grounding					-0.117***	-0.0895**
					(0.0378)	(0.0430)
elec_meter					0.0458	0.00823
					(0.0424)	(0.0423)
elec_bill_effi					-0.0701**	-0.0729**
					(0.0328)	(0.0365)
elec_hhonly_interrupt					0.0253	0.0293
					(0.0528)	(0.0571)
elec_legal_obs					-0.0721	-0.0312
					(0.0492)	(0.0518)
elec_connection_yr					-0.00209	0.000220
					(0.00341)	(0.00340)
Constant	3***	2.828***	3.155***	3.035***	3.025***	2.981***
	(2.22e-09)	(0.0801)	(0.0798)	(0.0184)	(0.0446)	(0.123)
Observations	386	386	386	365	386	365
R-squared	0.109	0.140	0.142	0.124	0.143	0.231

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

Appendix 2.3. Robust regression results of household-level predictability tier score

VARIABLES	(1) Base Model	(2) Geographic Model	(3) Demographic Model	(4) Wealth Model	(5) Electricity Model	(6) Full Model
discom_purvanchal	0.148	0.125	0.177	0.300*	0.141	0.319*
	(0.147)	(0.152)	(0.152)	(0.153)	(0.153)	(0.163)
discom_northbihar	-0.0480	-0.0391	0.0546	0.0158	-0.0828	0.164

	(0.144)	(0.148)	(0.136)	(0.149)	(0.157)	(0.150)
discom_southbihar	-0.248	-0.256*	-0.328**	-0.238	-0.192	-0.276*
	(0.158)	(0.153)	(0.138)	(0.153)	(0.154)	(0.148)
dem_num_adults			-0.0885***			-0.0813**
			(0.0312)			(0.0331)
dem_num_youngchild			-0.0831			-0.103*
			(0.0527)			(0.0550)
dem_num_teenager			0.0163			0.0125
			(0.0354)			(0.0374)
dem_num_elders			-0.0593			-0.0584
			(0.0605)			(0.0624)
dem_num_earners			0.389***			0.427***
			(0.0490)			(0.0593)
dem_head_edu_lvl			0.0720**			0.0734**
			(0.0365)			(0.0365)
dem_muslim_hh			0.329			0.167
			(0.245)			(0.256)
dem_caste_SC			0.0589			0.0930
			(0.161)			(0.157)
dem_caste_ST			-0.113			-0.164
			(0.284)			(0.285)
dem_caste_OBC			0.0921			0.144
			(0.144)			(0.139)
geo_dist_headquarter		-0.00156				0.000579
		(0.00313)				(0.00316)
geo_dist_pole		0.00163*				0.000494
		(0.000895)				(0.00079)
wealth_annual_inc				-8.93e-07		-1.20e-06*
				(7.05e-07)		(6.93e-07)
wealth_prop_val				-1.66e-07		-1.41e-07
				(1.01e-07)		(1.11e-07)
elec_peak_usage					-0.000197	8.02e-05
					(0.000151)	(0.00015)
elec_circuit_breaker					0.203	0.0872
					(0.124)	(0.122)
elec_fuse					0.0987	-0.0423
					(0.0728)	(0.0716)
elec_grounding					-0.621***	-0.388***
					(0.133)	(0.142)
elec_meter					0.0254	0.152
					(0.115)	(0.113)
elec_bill_effi					-0.0176	0.0233
					(0.122)	(0.111)
elec_hhonly_interrupt					-0.102	-0.115
					(0.151)	(0.160)
elec_legal_obs					0.0108	-0.202*
					(0.119)	(0.118)
elec_connection_yr					0.0252***	0.0263***

					(0.00736)	(0.00671)
Constant	1.871***	1.898***	1.576***	1.991***	1.694***	1.490***
	(0.107)	(0.169)	(0.229)	(0.121)	(0.174)	(0.298)
Observations	386	386	386	365	386	365
R-squared	0.017	0.022	0.200	0.048	0.100	0.307

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 3: Robust socioeconomic impact regression results

Appendix 3.1. Robust Binary Consumption Model

VARIABLES	(1) Base Model	(2) Demographic Model	(3) Wealth Model	(4) Full Model
electrified	3,845***	2,010***	3,084***	1,812***
	(545.1)	(526.1)	(537.9)	(551.1)
num_male_adults		1,298**		1,157**
		(573.9)		(587.9)
num_female_adults		845.5		410.1
		(739.9)		(671.9)
num_male_YC		-556.6		-343.0
		(409.8)		(384.6)
num_female_YC		502.0		403.4
		(624.5)		(593.6)
num_male_OC		804.1*		736.6
		(457.1)		(455.6)
num_female_OC		410.7		360.1
		(322.8)		(346.1)
num_male_elders		1,218		817.7
		(780.9)		(737.9)
num_female_elders		-190.3		-409.0
		(681.7)		(651.1)
num_male_earners		889.7		864.6
		(610.2)		(606.9)
num_female_earners		959.9		981.1
		(645.9)		(670.3)
hh_head_edu_lvl		899.7**		892.6**
		(385.2)		(385.2)
muslim_hh		1,555		2,099
		(2,058)		(2,085)
SC_hh		-1,853		-761.3
		(1,375)		(1,105)
ST_hh		-2,130		-551.2
		(1,433)		(1,190)
OBC_hh		-1,710		-1,086

		(1,411)		(1,113)
prop_val			0.00456***	0.00420***
			(0.000737)	(0.000621)
num_cows			747.3**	468.0
			(337.8)	(330.5)
num_buffalo			608.7	299.4
			(513.2)	(588.1)
num_goats			91.89	20.75
			(216.8)	(233.2)
num_chicken			-383.0	-293.2
			(360.3)	(384.2)
num_pigs = o,			-	-
other_ls = o,			-	-
Constant	5,965***	1,522	3,776***	-298.4
	(271.8)	(1,607)	(420.3)	(1,443)
Observations	500	500	500	500
R-squared	0.036	0.142	0.148	0.225

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

Appendix 3.2. Robust Tiered Consumption Model

VARIABLES	(1) Base Model	(2) Demographic Model	(3) Wealth Model	(4) Full Model
time_electrified	1.208	-2.510	-3.094	-6.294
	(6.867)	(7.083)	(6.948)	(7.124)
capacity_tier	1,857***	1,421**	1,180*	871.7
	(682.4)	(661.8)	(606.0)	(616.8)
availability_tier	6.571	209.3	324.4	397.9
	(761.9)	(724.1)	(738.0)	(712.7)
reliability_tier	-412.6	4.162	244.3	556.6
	(1,224)	(1,278)	(1,136)	(1,170)
quality_tier	-2,861***	-2,689***	-2,960***	-2,770***
	(949.2)	(933.5)	(982.2)	(954.3)
affordability_tier	777.6	732.2	354.2	368.2
	(663.2)	(610.5)	(585.1)	(560.5)
legality_tier	227.4	197.9	34.55	48.11
	(286.5)	(313.6)	(266.2)	(303.1)
safety_tier	358.0	416.9	575.6	549.9
	(399.5)	(418.2)	(364.9)	(392.8)
predictability_tier	321.6	178.9	707.1	604.4
	(626.9)	(659.8)	(609.0)	(647.0)
num_male_adults		1,540**		1,405**
		(681.1)		(704.7)

num_female_adults		679.3		345.1
		(904.7)		(847.5)
num_male_YC		-1,102		-900.7
		(716.2)		(685.7)
num_female_YC		844.4		768.6
		(894.5)		(862.6)
num_male_OC		771.6		691.6
		(533.6)		(550.7)
num_female_OC		562.8		561.3
		(388.5)		(402.5)
num_male_elders		1,646*		1,294
		(984.4)		(945.9)
num_female_elders		-227.0		-384.0
		(815.7)		(780.7)
num_male_earners		215.0		143.9
		(636.0)		(608.1)
num_female_earners		592.7		547.1
		(1,031)		(1,030)
hh_head_edu_lvl		827.3*		875.3*
		(459.5)		(455.1)
muslim_hh		1,125		1,641
		(2,398)		(2,535)
SC_hh		-1,357		-254.3
		(1,381)		(1,224)
ST_hh		-1,018		874.1
		(2,100)		(1,931)
OBC_hh		-1,714		-1,062
		(1,448)		(1,264)
prop_val			0.00469***	0.00448***
			(0.000764)	(0.000662)
num_cows			743.7	359.8
			(495.4)	(483.8)
num_buffalo			851.5	488.2
			(629.9)	(713.5)
num_goats			-107.6	-259.7
			(361.5)	(370.4)
num_chicken			-2,736	-2,986***
			(1,792)	(927.0)
num_pigs = o,			-	-
other_ls = o,			-	-
Constant	14,960**	7,886	11,531*	5,309
	(6,201)	(6,425)	(5,948)	(6,200)
Observations	387	387	387	387
R-squared	0.109	0.187	0.211	0.270

Robust standard errors in parentheses

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

Appendix 3.3. Robust Binary Fuel Model

VARIABLES	(1) Base Model	(2) Demographic Model	(3) Wealth Model	(4) Full Model
electrified	-6.244	-10.69	-10.54	-11.10
	(7.593)	(11.88)	(6.899)	(11.61)
num_male_adults		3.984		3.056
		(6.754)		(6.972)
num_female_adults		3.539		1.159
		(4.215)		(4.365)
num_male_YC		-5.454		-4.328
		(4.507)		(3.971)
num_female_YC		1.874		1.821
		(5.027)		(4.959)
num_male_OC		-5.329		-6.059
		(4.479)		(4.779)
num_female_OC		3.049		2.519
		(4.882)		(4.915)
num_male_elders		24.72		24.06
		(16.06)		(15.54)
num_female_elders		-4.899		-7.040
		(13.40)		(14.30)
num_male_earners		-20.21***		-19.53***
		(4.220)		(3.981)
num_female_earners		-14.95***		-15.82***
		(4.331)		(4.482)
hh_head_edu_lvl		-5.589*		-5.834*
		(3.280)		(3.358)
muslim_hh		-0.211		6.041
		(9.363)		(10.13)
SC_hh		7.010		11.80
		(14.51)		(17.36)
ST_hh		-21.88		-12.73
		(13.73)		(17.37)
OBC_hh		-0.135		-0.262
		(8.556)		(9.010)
prop_val			1.57e-05	1.55e-05
			(1.53e-05)	(1.55e-05)
num_cows			-3.003	-2.687
			(2.772)	(3.101)
num_buffalo			10.85	13.52
			(7.221)	(8.435)
num_goats			-2.348	-1.877
			(2.076)	(2.188)
num_chicken			-10.49**	-4.920

			(4.135)	(3.287)
num_pigs = o,			-	-
other_ls = o,			-	-
Constant	87.01***	104.6***	80.96***	98.57***
	(5.848)	(17.27)	(8.450)	(15.75)
Observations	500	500	500	500
R-squared	0.001	0.072	0.023	0.093

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

Appendix 3.4. Robust Tiered Fuel Model

VARIABLES	(1) Base Model	(2) Demographic Model	(3) Wealth Model	(4) Full Model
time_electrified	-0.00627	-0.00721	-0.0222	-0.0174
	(0.0591)	(0.0633)	(0.0576)	(0.0576)
capacity_tier	1.965	1.161	-1.082	-2.038
	(4.472)	(4.672)	(6.028)	(6.325)
availability_tier	-15.56***	-15.61***	-13.66***	-13.73***
	(4.361)	(4.693)	(4.209)	(4.350)
reliability_tier	-5.547	-5.517	-2.576	-2.409
	(5.841)	(5.915)	(5.678)	(6.527)
quality_tier	-6.806	-11.49**	-7.311	-12.16**
	(5.187)	(5.116)	(4.998)	(5.109)
affordability_tier	-4.509	-1.603	-5.657	-2.206
	(3.913)	(3.964)	(4.590)	(4.376)
legality_tier	-7.940***	-6.259***	-8.504***	-6.375***
	(1.809)	(2.117)	(1.913)	(2.174)
safety_tier	0.304	1.703	1.523	2.790
	(4.690)	(3.834)	(4.606)	(4.045)
predictability_tier	-14.06***	-9.839**	-12.42***	-7.628**
	(4.637)	(4.372)	(3.732)	(3.580)
num_male_adults		1.010		0.138
		(8.298)		(8.433)
num_female_adults		3.588		1.276
		(4.686)		(4.675)
num_male_YC		-8.557		-6.879
		(7.099)		(6.327)
num_female_YC		0.267		-0.455
		(6.427)		(6.483)
num_male_OC		-7.565		-8.425
		(5.143)		(5.539)
num_female_OC		2.579		2.077
		(6.198)		(6.261)

num_male_elders		28.04		27.18
		(19.11)		(19.08)
num_female_elders		-8.963		-10.59
		(16.33)		(17.10)
num_male_earners		-13.66***		-14.67***
		(4.444)		(4.827)
num_female_earners		-9.318		-12.23**
		(5.843)		(6.112)
hh_head_edu_lvl		-6.220		-5.943
		(4.135)		(3.933)
muslim_hh		-1.408		5.679
		(11.38)		(13.03)
SC_hh		13.27		19.03
		(16.81)		(20.13)
ST_hh		-17.79		-10.22
		(19.16)		(22.07)
OBC_hh		6.381		7.003
		(7.941)		(8.635)
prop_val			1.91e-05	1.98e-05
			(1.88e-05)	(1.88e-05)
num_cows			0.620	-1.054
			(3.337)	(4.232)
num_buffalo			12.04	16.12
			(9.393)	(11.07)
num_goats			-1.259	-1.481
			(2.005)	(2.278)
num_chicken			-28.04***	-16.77**
			(7.556)	(6.589)
num_pigs = o,			-	-
other_ls = o,			-	-
Constant	238.1***	246.4***	217.2***	226.0***
	(42.79)	(56.82)	(38.75)	(49.64)
Observations	387	387	387	387
R-squared	0.060	0.103	0.083	0.130

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