WP. No.: SAUFE-WP-2020-010

Electrification and Cooking Fuel Choice in Rural India

Ridhima Gupta
Assistant Professor
Faculty of Economics,
South Asian University
Akbar Bhawan, Chanakyapuri
New Delhi 110021, INDIA
Email: ridhima@sau.int

and

Martino Pelli
Department of Economics,
Université de Sherbrooke and Québec
Canada CIREQ,
CIRANO, and GREDI
Email: martino.pelli@usherbrooke.ca

Working Paper Number: SAUFE-WP-2020-010

http://www.sau.int/fe-wp/wp010.pdf

FACULTY OF ECONOMICS
SOUTH ASIAN UNIVERSITY
NEW DELHI
September, 2020

Electrification and Cooking Fuel Choice in Rural India*

Ridhima Gupta† Martino Pelli‡

April 2020

Abstract

This study investigates the causal link between electrification and the adoption of modern (and cleaner) cooking fuels, more specifically Liquefied Petroleum Gas (LPG). In order to correct for the potential endogeneity in the placement of electrical infrastructure, we exploit an instrumental variables approach. Our instrument interacts state-level supply shifts in hydro-electric power availability with the initial level of electrification of each district. The results are consistent with an expansion of households' choice set under a fixed budget constraint. We find that electrification leads to an increase in the probability of adoption of (free) biomass fuels and a decrease in the probability of adoption of (costly) modern cooking fuels. These results are statistically significant only for the poorest households in our sample, while they become statistically insignificant when we move to richer households. The same is true for the share of expenditure in a specific fuel. These results seem to indicate that electrification by creating an additional strain on households' finances pushes them back on the energy ladder.

JEL classification: O12, O13, Q56

Keywords: rural electrification, cooking fuel, energy ladder, fuel stacking.

*We thank, without implicating them, Eswaran Somanathan, Daniel Kaffine, Sarah Jacobson, Jörg Peters, Soham Sahoo, Subhrendu Pattanayak, and participants at the Canadian Economic Association Meeting (2019), the Canadian Resource and Environmental Economics Association Meeting (2019), the 13^{th} Annual Meeting of Environment For Development (2019), the 4^{th} Meeting of the Sustainable Energy Transition Initiative (2019), and the Annual Economic Growth and Development Conference at ISI Delhi (2018), as well as seminar participants at Kansas University and Lulea University of Technology for helpful comments and suggestions. Rakesh Sinha provided outstanding research assistance. Financial support from the Sustainable Energy Transition Initiative (SETI) is gratefully acknowledged. Pelli also received financial support from the *Fonds de Recherche Société et Culture Québec*.

†Faculty of Economics, South Asian University, India; Email: ridhima@sau.int.

‡Corresponding author, Department of Economics, Université de Sherbrooke, Sherbrooke, Quebec, Canada CIREQ, CIRANO, and GREDI. Email: martino.pelli@usherbrooke.ca.

1 Introduction

A large literature on households' energy transition has focused on the energy ladder hypothesis. This hypothesis explains the substitution of traditional fuels with modern ones through increases in income and in the socio-economic status of the household (see for instance Bruce et al., 2000; Hosier and Dowd, 1987; Leach, 1992; Van der Kroon et al., 2013). The traditional view of the energy ladder sees energy transitions as a series of disjointed steps. Yet, more recently, a growing body of literature (e.g. Heltberg, 2004; Masera et al., 2000; Ruiz-Mercado, 2015) has shown that instead of switching from one fuel to the other, households simultaneously use multiple fuels. This is known as fuel stacking.

The fuel choice of households for everyday activities, such as cooking and heating, impacts on several factors influencing general wellbeing, from health to time use and exposure to financial risks (see for instance Bruce et al., 2000; Dherani et al., 2008; Khandker et al., 2012; Kishore and Spears, 2014; Peters and Sievert, 2016; Po et al., 2011; Van de Walle et al., 2017). In India, electricity is rarely used for cooking. The benefits of electricity consist primarily of improved lighting and of providing power for consumer appliances or agriculture. Only 19% of the rural households use Liquified Petroleum Gas (LPG) as a primary fuel for cooking, while the majority of the rural population still cooks with biofuels such as fuelwood and crop residues.[1]

For these reasons, it is not surprising that domestic fires are the largest contributor to emissions of black carbon in South Asia (Bond et al., 2013). Black carbon is the second most important greenhouse agent after carbon dioxide. About 400,000-550,000 premature deaths occur annually in India from indoor air pollution exposure of children under five and adult women (Smith, 2000). The daily average particle level in these households ranges from between 1000 and 2000 $\mu g/m^3$ of PM_{10} (particles of mass smaller than 10 micron diameter, Smith, 2000). This is 10-20 times higher than the national ambient air quality standard of PM_{10} set by the government of India (Pant et al., 2019). Exposure to ambient and household $PM_{2.5}$ (particles of mass smaller than 2.5 micron diameter) is among the largest single causes of premature mortality in India (Cohen et al., 2017). It is therefore crucial to understand the determinants of fuel choice and stacking and the role played by the electrification status of a household in spurring adoption of modern cooking fuels.

We investigate the existence of a causal link between electrification and the adoption of modern (and cleaner) cooking fuels, like LPG. Other papers, such as Barron and Torero

[1]The primary cooking fuel is the fuel used in the majority of cases by the household.

(2017), focus on the impact of electrification on indoor pollution. We focus more specifically on the channel through which indoor air pollution may be affected.

The arrival of an electric connection may impact households in two different ways. Electricity is a fundamental input for the utilisation of a large variety of appliances and, therefore, gaining access to a connection opens up a lot of new consumption possibilities. The size of the choice set of a recently electrified household increases significantly. With a larger choice set, households face a new trade-off. Households may decide to: *i)* invest in new appliances that are going to increase their productivity level (such as an irrigation pump or a small mill) or help them optimize their labor supply (for instance a fridge that allows less frequent visits to the market), or *ii)* invest in entertainment technology (such as a TV or a Radio). Clearly, the way in which a household decides to use the recently acquired electricity is going to determine the impact that electricity has on its budget and, through this bias, its impact on other consumption choices, like the ones related to energy.

If households decide to use electricity for productive activities, following the traditional energy ladder theory and results from a large literature (see for instance International Evaluation Group, 2008; Khandker et al., 2012; Lipscomb et al., 2013; Chakravorty et al., 2014, 2016; Van de Walle et al., 2017), we would conclude that the increase in income generated by electrification would push households to increase their use of LPG, a normal good, and decrease their use of fuelwood, dung and crop residues, inferior goods. Instead, if households are going to favour entertainment goods, or any other use of electricity which is not productivity enhancing, we would not observe an increase in the household income, but we would observe an increase in their expenditures.

In order to disentangle these two effects, we use a nationally representative repeated cross-section of rural households constructed from two waves of the National Sample Survey (NSS). The waves used are 2004 and 2009. Unravelling this new and important effect of electrification may play a role in the design of new policies aimed at increasing the uptake of modern cooking fuels such as LPG.

Identifying the impact of electrification on modern fuel use presents a number of empirical challenges. These challenges are common in the literature that identifies causal effects of big infrastructure projects (see for instance Allcott et al., 2016; Duflo and Pande, 2007; Röller and Waverman, 2001; Aschauer, 1989; Garcia-Mila and McGuire, 1992; Holtz-Eakin, 1993). Electrification, like other infrastructure investments, is not randomly assigned. Governments may aim infrastructure investment to areas that are already growing faster and, therefore, electrification may always depend on a range of unobservables. In the specific case of this paper, some omitted variable could contemporaneously impact a higher rate of electrification and a higher rate of adoption of modern cooking fuels. We tackle this endogeneity problem by using an instrumental variable approach.

The use of an instrumental variable allows us to capture the part of the variation in electrification which is not related to factors that are also likely to affect a household's choice of cooking fuel. Our instrument is constructed starting from the work of Allcott et al. (2016). In their paper, they instrument electricity shortages with state-level supply shifts in hydro-electric power availability. Following Bartik (1994) and Chakravorty et al. (2014), we multiply these state level shifts in hydro-electric power availability by the initial level of electrification of each district. In other words, we weight the supply shifts by the initial level of electrification of each district. The initial level of electrification is going to be defined as the district's electrification level 11 years before the beginning of our sample. We measure a district's initial electrification rate using the mean light intensity emanating from it, measured using satellite data.[2] We analyse both the intensive and the extensive margin of decision regarding the choice of the primary cooking fuel. The extensive margin focuses on the adoption decision to use a given cooking fuel. Hence our outcome of interest is a binary indicator variable indicating whether a household uses a particular cooking fuel or not. The intensive margin, instead, denotes the intensity of use of a given fuel. We measure the intensity of use by calculating the share of expenditure on a given fuel type over total monthly expenditure by a household. We separately examine the choice of Liquified Petroleum Gas (LPG) and fuelwood.

The results from the main specification point in the direction of the second mechanism highlighted above: households receiving a connection invest in new goods which are not going to enhance their productivity. For this reason, we find that electrification leads to an increase in the probability of adoption of fuelwood and a decrease in the probability of adoption of LPG. One way to test whether the results found are related to a change in the choice set is to look at the heterogeneity of the results across households' wealth levels. We find that these effects are statistically significant for the poorer part of the sample – households whose expenditures per month are below average – while they are statistically insignificant for the richest part. The same is true for the intensive margin. We also find that these effects are more accentuated in male-lead households compared to female-lead ones. This result is consistent with Duflo (2003) and Qian (2008), which show that, on average, women take less impulsive decisions, aimed at improving the welfare of the household. We then run a series

[2]The 11 years gap is dictated by the availability of the night lights data.

of robustness tests using alternative definitions of a household's wealth level.

Basically, electricity allows people to include goods such as TVs, Radios or mobile phones in their consumption basket. Goods which (with the exception of mobile phones) they were not able to use before. TV sets do not have any close substitute, while LPG does have a substitute, which is completely free. This paper does not claim that the only adjustment that households will make after receiving an electrical connection is through energy consumption, yet, energy is going to be one of the dimensions along which households adjust their spendings.

Precise estimations of these effects are crucially important for policymakers when taking decisions related to rural electrification. These electrification initiatives are very expensive and, therefore, have historically occurred mainly in places near the main axes of communication or the existing power grid. The cost to extend the power grid to a village which is more than 15km away from the existing infrastructure is estimated at around 150,000 dollars (Greenstone et al., 2014). For this reason, in order to justify these important expenditures, it is important to provide accurate figures on the benefits of a new connection. As noted above, this paper focuses on a potential additional benefit of a connection to the grid that has been previously neglected. Yet, in light of our results, it seems that, for the poor, access to electricity makes the switch to cleaner cooking fuels even harder. As a consequence, one of the main take away of the paper for policymakers is that the introduction of electricity should not be done without a careful consideration of its impact on the poorest households and the provision of subsidies (for instance on LPG cylinders) in order to help them avoid the negative consequences of an additional regular expenditure on an already tight budget.

Our results are in line with recent findings by Lee et al. (2020). They also conclude that only certain subgroups of the populations – those willing to pay for it – benefit from electrification. As a consequence, electrification should never come alone, but always be combined with complementary programs.

The remainder of the paper is organized as follows. Section 2 presents the data used. Section 3 focuses on the empirical strategy adopted and Section 4 discusses the results. Section 5 describes robustness tests and, finally, Section 6 draws some conclusions.

2 Data

Data on households' expenditures are taken from surveys conducted by the National Sample Survey Organisation (NSSO). These data are representative of the rural and urban population of India. For our analysis, we use data for the rural sample for the years 2004-2005 (61^{st} Round), and 2009-2010 (66^{th} Round). These two datasets give us a repeated cross-section sample containing 136,221 households.[3] Table 1 reports summary statistics for the variables used in the paper. Panel A reports statistics for the variables of interest and the dependent variables, Panel B for the instruments used and, finally, in Panel C we present all the control variables used in the different specifications.

68% of the households in our sample have a connection to the power grid.[4] Only 16.2% of the households in the sample use LPG as the primary fuel for cooking, while 73% of rural households still rely on fuelwood.[5] NSS asks households to recall expenditures incurred and quantities purchased for almost all items of domestic consumption over the last 30 days. Total expenditures on any item include money spent on purchases and value of consumption out of home production. The latter is valued at the average retail prices prevailing in the household's village of residence. We use this data to calculate expenditure shares for all fuel items for each household. Panel A of Table 1 shows that while LPG represents roughly 1% of a household's monthly expenditures, fuelwood constitute 5.3%. These expenditure shares are computed conditional on the household using the specific fuel. The difference between the two shares is partly explained by the fact that households using LPG are on average richer than households using fuelwood as the primary cooking fuel.

Panel B presents statistics for the instrument used. As discussed in the introduction, the instrument we use in this paper is composed by the interaction of two separate variables. The first variable comes from Allcott et al. (2016) and is predicted hydro generation as a share of predicted electricity demand at the state level.[6] These supply shifts, related to hydro generation, represent on average 13% of predicted electricity demand, yet there is a lot of variation across states.[7] The shocks are then interacted with the initial electrification status of each district. As mentioned above, the initial electrification status of a district is measured by the level of light intensity emanating from the district 11 years before the

[3] 136,211 is reached after dropping outliers. The sample is composed of 78,017 households interviewed in 2004, and 58,204 in 2009.

[4] For most households in India, having a power connection does not mean receiving regular power, as the quality of supply is fairly low. Unfortunately, the NSS survey has no information about the quality of the power supply.

[5] All these statistics are based on what each household declares as the primary cooking fuel and, therefore, it is possible that for a household reporting fuelwood as the primary cooking fuel we still observe expenditures in LPG. The remaining 11% is divided in the following way: 7.34% use biomass; 2.09% are in *other category*; 0.92% use kerosene; 0.39% has no cooking arrangement; and 0.06% use electricity.

[6] More details about the construction of this variable are presented in Section 3.

[7] In 2004/05 hydro generation represented 14.38% of total generation in India, Central Electricity Authority (2006)

beginning of our sample, i.e. in 1993. 11 years is as far back as we can go due to data availability.

The database on night lights for India has been constructed by the University of Michigan in collaboration with the World Bank, using images taken by the Defence Meteorological Satellite Program (DMSP), run by the U.S. Department of Defence. This satellite program took pictures of the earth at night for 20 years, from 1993 to 2013. These images have a resolution of 30 arc-seconds (i.e. roughly 1 square km at the equator). Each pixel is assigned a value between 0 and 63, where 0 indicates no light output and 63 is the highest level of light output. The India Lights API data is based on these images and is freely available for each district and month from 1993 to 2013.[8] Figure 1 shows district averages for the reference year (1993) used in the construction of our instrument. Darker shades represent a higher light output. It is easy to identify the biggest cities as the poles of highest light intensity, Delhi, Mumbai, Kolkata, Bangalore and Chennai to name a few. No district reaches the maximum light output of 63. The maximum observed over the sample is of 55.44. The average light intensity is low, at 2.53. The interaction between this variable and the predicted supply shift variable takes into account the fact that districts that already had more electricity connections (and, therefore, a higher light output in 1993) benefit more from the positive hydro supply shifts than districts with lower degrees of electrification.

The NSS survey also collects information on several household-level demographic and economic characteristics such as household size, religion, caste and occupation, which we use as household level controls. We also obtain prices paid by households for the various items purchased by dividing expenditures by the number of units bought. Since many households do not consume some of the fuel items in which we are interested – such as LPG – we use the average price of the item in the state of residence of the household as a proxy for the price. In order to make expenditure and prices comparable across rounds we use industrial and agricultural consumer price indices reported by the Labour Bureau of India to deflate to 2004 equivalent values.

Panel C of Table 1 shows the large difference in price between fuelwood and LPG. The rest of the control variables concerns the socio-economic status of each household, measured through its size, land holdings, religion, the age, gender and education level of its head and whether the household belongs to a schedule caste/tribe or another backward caste.

[8] `http://india.nightlights.io/#/nation/2006/12`

3 Empirical Strategy

In order to study the effect of a connection to the grid on the choice of cooking fuel we split the analysis in two parts. First, we analyze the extensive margin response, i.e. the adoption of a specific cooking fuel as a primary cooking fuel and, second, we look at the intensive margin response, i.e. how is the share of expenditure in a given fuel going to be affected. The specification we use in both cases takes the following form:

$$y_{hdst} = \alpha_0 + \alpha_1 E_{hdst} + \alpha_2 X_{hdst} + \alpha_3 V_{dst} + \delta_d + \delta_t + \varepsilon_{hdst} \quad (1)$$

where y represents different outcome variables. Subscripts h, d, s and t denote household, district, state and year, respectively. First, we will investigate the impact of an electric connection on adoption. In this case, the outcome variable is binary and takes the value 0 if a household does not use LPG (fuelwood) and 1 otherwise. Second, we will investigate the intensive margin, i.e. does an electric connection increase the amount a household spends on LPG (fuelwood). In this case, we use the relative share of expenditure on LPG (fuelwood) in total monthly expenditure of the household. E is the electrification status of a household and X includes time-varying household specific controls, i.e. household size, religion, total land owned, information on the household head age, sex and education level, and finally whether the household belongs to a scheduled caste, tribe or another backward caste. V is a matrix of village level controls, and it contains the price of fuelwood and the price of LPG. δ_d and δ_t denote district and year fixed effects, respectively. Finally, ε is the error term, clustered at the district level across all specifications.

Fuel choice and the presence of an electricity connection could depend on a variety of unobservable factors. If this were the case, equation (1) cannot be interpreted causally. As mentioned above, establishing causality in the case of important infrastructure investments presents important econometric challenges. These challenges are mainly due to the endogeneity in the placement of infrastructure. In order to tackle this endogeneity and to establish causality going from electrification to the adoption of modern cooking fuels, we use a standard instrumental variable approach.

A series of factors may hide behind the decision to electrify some areas rather than others. For instance, governments could reserve new investments in infrastructure to areas already experiencing more growth. Other unobservable economic trends may also influence investment decisions. For example, a richer village may have a higher probability of being connected than a poor village. The probability of connection may also depend on the proximity to a major city or population density in certain areas. For all these reasons, it is difficult to isolate the effect of infrastructure investments on development outcomes. This issue has been widely discussed in the literature, see for example, Allcott et al. (2016), Duflo and Pande (2007), Röller and Waverman (2001), Aschauer (1989), Garcia-Mila and McGuire (1992) and Holtz-Eakin (1993), just to mention a few.

We are going to instrument electrification using electricity supply-shifts due to the coming online of new hydroelectric power plants. Hydroelectric power in India is much more reliable than coal generated power (see Chan et al. (2014) which shows that between 1994 and 2009 Indian coal power plants were offline 28% of the time). An increase in the available supply is going to be correlated with an increase in the probability of receiving a new connection. Another advantage of hydroelectric generation is that the location for new power plants is closely related to the topology of the terrain and not so much to economic considerations, which would make these new plants endogenous. Our instrument is constructed starting from the work done by Allcott et al. (2016). In their paper, they instrument electricity shortages with state-level supply shifts in hydro-electric power availability.

Allcott et al. (2016)'s instrument is constructed at the state level. Following Bartik (1994) and Chakravorty et al. (2014), we multiply state level shifts in hydro-electric power availability by the initial level of electrification of each district in order to obtain a district-level instrument. This is equivalent to weighting the supply shifts by the initial level of electrification of each district. The importance that the supply shifts measured by Allcott et al. (2016) will play in each district is likely to be contingent on the district's initial level of electrification. A higher initial level of electrification allows us to use more effectively the positive supply shifts, while no initial electrification would render them virtually useless. Following the same narrative, negative supply shifts should have a bigger impact on districts with a lower initial level of electrification and a smaller one on districts with a higher initial level of electrification. The initial level of electrification is going to be defined as the district's electrification level 11 years before the beginning of our sample. We measure a district's initial electrification rate using the mean light intensity emanating from it. The mean light intensity is measured as the average of the light intensity of each pixel composing a district.

The instrument constructed by Allcott et al. (2016) consists of predicted state-level supply shifts from hydroelectric generation.[9] More precisely, it is predicted hydro generation as a

[9] Allcott et al. (2016) is interested in quantifying the impact of power outages on industrial production in Indian states. Yet, power outages and industrial production may be correlated. For instance, if a state is growing faster and, consequently, its energy demand is higher, it may experience more power outages.

share of predicted electricity demand. Allcott et al. (2016) divides predicted hydro generation by predicted demand in order to obtain the relative share of hydro generation across the different states. Actual demand could be affected by shortages and, for this reason, the instrument is not based on actual demand but on predicted demand. Demand is predicted using the average share demanded by a given state between 1992 and 2010 multiplied by electricity demand in all other states.[10] Computing predicted electricity demand in this way minimizes fluctuations due to important shortages.

Allcott et al. (2016) also need to use predicted hydro generation, because – since water can be kept in reservoirs – in years of low industrial demand less hydro energy will be produced and, therefore, industrial production and hydro generation may be correlated. Predicted generation capacity is computed in a similar way as predicted demand, and is based on reservoirs inflow and run-of-river plants. The first element needed is the "state predicted annual hydro generation capacity" (C), which is computed in the same way as predicted demand. This term is then multiplied by a "state average capacity factor". The capacity factor depends on two main elements. First, on the "share of output of a given reservoir which is contractually allocated to the state" and, second, on the "demeaned inflow predicted capacity factor for a given reservoir in a given year". This last element is computed individually for each reservoir-year couple. Allcott et al. (2016) first regress generation on inflow (for each reservoir) in order to produce inflow-predicted generation, which is then divided by the generation capacity of the reservoir (capacity x 8,760 hours in a year).[11]

In order to eliminate this endogeneity problem, Allcott et al. (2016) look for an instrumental variable (the use of an instrumental variable also solves for the possible measurement error coming from the way in which shortages are measured). For this reason, the authors need an instrument affecting shortages but not affecting industrial production, or better, that affects industrial production only through its impact on power shortages.

[10]Denoting by s and t the state and year at hand, and by r and y other states and years, predicted electricity demand $\tilde{D}_{st}$ is defined as

$$\tilde{D}_{st} = \sum_{r \neq s} D_{rt} \cdot \sum_{y=1992}^{2010} \frac{D_{sy}}{\sum_{r \neq s} D_{ry}} \tag{2}$$

The second multiplicative term computes the electricity demand/consumption in a given state s as a share of total electricity demand/consumption in all other states and takes the average of this share over the period 1992 to 2010. Once obtained this average share, it is multiplied by total consumption in all other states in the year of interest, obtaining in this way a prediction of consumption for that particular year.

[11]As shown in Table 1 the average of the value obtained with this methodology is 13%, which compares well with average hydro generation in India, which was 14.38% in 2004/05, Central Electricity Authority (2006).

Our first stage specification takes the following form

$$E_{hdst} = \beta_0 + \beta_1(H_{st-1} * L_{sd0}) + \beta_2 H_{st-1} + \beta_3 X_{hdst} + \delta_d + \delta_t + u_{hdst} \tag{3}$$

where L represents a district's light intensity and H is Allcott et al. (2016)'s state-level instrument. We use the lag of the hydro instrument to allow enough time for the positive supply shifts to take effect and impact the electrification level. δ_d and δ_t denote district and year fixed effects, respectively. $H_{st} * L_{sd0}$ is our instrument, the interaction between the Allcott et al. (2016)'s instrument and a district's initial light intensity.[12] Finally X is a matrix of time-varying household specific controls and u is the error term, clustered at the district level.

Table 2 reports results for first stage estimations. The table is organized as follows, columns (1) and (2) contain first stage estimations for the whole sample, first with only the instrument and the relevant fixed effects and then with the full set of household controls. Column (3) shows the first stage estimation for individuals whose household's per adult equivalent monthly expenditure is below the mean level (of about 724 Rs, equivalent to roughly 10 USD), while column (4) shows it for households whose income is above the mean level. Let us focus on the effect of the interaction, i.e. our instrument in Column (2), the full specification. Its impact is positive and statistically significant at the 1% level across most specifications. This means that a hydro shock combined with a higher level of initial electrification in the district increases the probability of electrification by roughly 11 percentage points. The direction of this coefficient is the one expected. The negative value on *Lag Hydro* cannot be interpreted on its own, it has to be interpreted together with the coefficient on the interaction term. The best way to look at this result is to compute the marginal effect of a supply shift across the full spectrum of initial levels of electrification. We plot the marginal effect in Figure 2. The gray histogram shows that most of the districts had low levels of initial electrification. We can also notice that for districts with no initial electrification, the impact of a supply shift is statistically insignificant. Yet, as the initial

[12] The condition which needs to be satisfied in order for our identification to be consistent is that $E[L_{ds0} * H_{st} * \varepsilon_{idst}] = 0$. We can easily verify this by taking the limit over districts and states $\lim_{D,S\to\infty} \frac{1}{D*S} \sum_d \sum_s (L_{ds0} * H_{st} * \varepsilon_{idst}) = 0$. Since H_{st} does not depend on district, we can extract it from the sum and re-write it as $\lim_{D,S\to\infty} \frac{1}{D*S} \sum_s H_{st} \sum_d (L_{ds0} * \varepsilon_{idst}) = 0$. In order for this to be verified we only need to argue that $\lim_{D\to\infty} \frac{1}{D} \sum_d (L_{ds0} * \varepsilon_{idst}) = 0$, or equivalently that $E[L_{ds0} * \varepsilon_{idst}] = 0$. This means that shocks in cooking fuels decisions in 2004 and 2009 need not to be correlated with the district light intensity in 1993. While the trend in cooking fuel choices may be related to the initial level of electrification of a district, shocks to the path are highly unlikely to be related to initial electrification. For this reason, we are confident in claiming that this condition is satisfied.

level of electrification increases this insignificant effect becomes positive and statistically significant.

When running the first stage on the whole sample, we obtain an F statistic of 12.1, above the critical level of 10. When we split the sample the value of the F statistic decreases to 6.76 and 8.03, respectively. This means that our instrument becomes weaker when applied to the two subsamples. In order to insure the robustness of our inference, we compute the Anderson-Rubin F statistics for weak instruments, which in both cases rejects the hypothesis that the coefficients of interest may be equal to zero. Therefore, although the instrument becomes weaker, the inference is still robust.

4 Results

First, we present the results for the extensive margin, focusing on the adoption of LPG and fuelwood as primary cooking fuels. Second, we look at the intensive margin, measuring the impact of electrification on spending in LPG and fuelwood.

Extensive margin

Table 3 reports results for the extensive margin estimations, i.e. adoption. For each estimation we report OLS and IV coefficients for a linear probability model.[13] Columns (1) and (2) show results for LPG, while (3) and (4) for fuelwood.

Let us focus on the IV results for the linear probability model.[14] The results on the variable of interest, *Electricity*, are all statistically significant at least at the 10% level. These coefficients are negative for LPG and positive for fuelwood. This implies that, when a household gets connected to the power grid, the probability that it will use LPG for cooking decreases by 56.6 percentage points, while the probability that it will use fuelwood increases

[13] We also run the model using a probit specification instead of a linear probability model, the results are very similar in terms of magnitude, sign and statistical significance.

[14] From Table 3, and all the following tables, one sees that the direction of the bias in the OLS specification is positive in the case of LPG and negative in the case of fuelwood. This is easily explained by the sign of the covariances: *i)* between the electricity connection variable and the instrument and, *ii)* between the electricity connection variable and the dependent variable. While the covariance between electricity connection and the instrument is positive in both cases (see results of the first stage specification), the covariance between electricity and LPG use is expected to be positive and the one between electricity and fuelwood use is expected to be negative. For instance, we expect areas experiencing a higher rate of economic growth to have a larger number of electricity connections and at the same time more people using LPG and less people using fuelwood. Since $cov(E, y)$ and $cov(E, (H * L))$ move in the same direction in the case of LPG, we expect the bias to be positive. In the case of fuelwood, instead, $cov(E, y)$ and $cov(E, (H * L))$ move in opposite directions and, therefore, we expect the bias to be negative.

by 52 percentage points. Between 2004 and 2009, the electrification rate in India went up from 0.64 to 0.74, implying a change of ten percentage points. These numbers, together with our coefficients, imply that, *ceteris paribus*, fuelwood adoption increased by roughly 5.53 percentage points and LPG adoption decreased by 6.02 percentage points over this 5-years period as a result of the increase in the electrification rate.[15] According to the energy ladder theory, one would expect electricity to constitute a push towards modernization and, therefore, towards the adoption of more modern and less polluting cooking fuels. Yet, as we discussed in the introduction, electrification also expands the choice set of households. A new connection allows households to invest in new appliances for productivity or for entertainment, putting households, especially the poorer ones, in front of difficult choices when deciding how to allocate their limited resources.

The signs of the coefficients on the control variables are all as expected. The price of fuelwood has a positive impact on the adoption of modern fuels and a negative one on the adoption of biomass, while the price of LPG has the opposite effect. Richer households, with larger landholdings and a higher level of education, use less biomass and more modern (and usually more expensive) fuels. A more educated and older head of household is also usually associated with a higher economic status, and this is exactly what we find in our data. This typology of household will use modern fuels more often than biomass. Belonging to a lower caste also means a higher probability of cooking using biomass. A male head of household is more likely to use fuelwood, while a female has a higher probability to use LPG. Women being usually the sole responsible of cooking, it make sense that when they have the choice they opt in larger numbers for the cleaner fuel.

As we mentioned above, our results may be explained by an expansion of the choice set available to households. A natural way to test for this is to look at the behaviour of households as a function of the stringency of their budget constraints. Cooking with LPG is more expensive than with biomass. Biomass, consisting mainly of fuelwood, dung and crop residue, can easily be collected (for free), while LPG has to be bought (a five-member household cooking almost exclusively with LPG spends roughly 530 Rs. per month, based on a per capita consumption of about 5.2 kg/month at a price of 20.3 Rs/kg. The price comes from our summary statistics).[16] Households connected to the grid have to pay

[15]This numbers are obtained by multiplying the coefficients 0.52 and -0.566 by the improvement in the electrification rate, 0.7432-0.6369=0.1063, so 10.63 percentage points.

[16]The price of 20.3 Rs/kg is an average of the price between 2004 and 2009. In 2020, the price of LPG is of around 56 Rs/kg and, therefore, the same five-member household would spend roughly 1,456 Rs. per month.

for the electricity they consume, irrespective of whether they use it for irrigation, light or entertainment (for appliances such as TVs or Radios), and this without accounting for the costs of investments such as a TV, a radio or a pump. These extra expenditures leave a smaller disposable budget for cooking fuels. In order to verify whether this is what is happening, we split our sample in two. The split is based on the average monthly per capita expenditure. In the average household, the monthly per capita expenditure is about 724 Rs (the exact value is 724.09 Rs, about 10 USD). Therefore, in the first sub-sample we keep the 92,773 households characterized by a monthly per capita expenditure lower than 724 Rs, while in the second sub-sample we keep the 43,381 richer households, with expenditures per capita equal or higher than 724 Rs.[17]

Panel A of table 4 presents results for this test, which seems to confirm our hypothesis. Columns (1) through (4) present results for LPG and fuelwood adoption for households whose monthly per capita expenditures are below 724 Rs., while columns (5) through (8) present them for households situated above the 724 Rs. threshold. The IV coefficients obtained are statistically significant at the 5% level for the poorest part of the sample, while they are statistically insignificant for the richer part. For households below the average expenditure level, the coefficient on LPG adoption is -0.526, so slightly smaller in magnitude than the one obtained for the baseline estimation. The coefficient for fuelwood adoption instead increases in magnitude to a value of 0.758. The electrification rate for this group of households went from 0.55 to 0.66, i.e. an increase of 11 percentage points. Taken together with the coefficients, this change implies that over this 5-year period, *ceteris paribus*, we would have observed a decrease in the adoption of LPG by 5.8 percentage points and an increase in the adoption of fuelwood by 8.34 percentage points.[18]

These results seem consistent with the expansion of the choice set hypothesis. Electricity allows households to adopt new consumption habits, for instance, 35.2% of the households with expenditures below the mean own a TV. These new choices impose an additional constraint on their tight budget. An easy way to free up some extra budget is to revert to fuelwood for cooking, instead of buying LPG cylinders, household members collect fuelwood for free. Conversely, richer households do not suffer from the additional financial strain, and, for them, obtaining a power connection does not have an impact on the choice of cooking fuel.

[17] If we use the median instead of the mean in order to divide the sample, the results are similar in terms of sign and magnitude but less statistically significant. These results are available upon request.

[18] The exact numbers for the electrification rate are 0.6621 for 2009 and 0.5526 for 2004, giving a difference of 10.95 percentage points, which we then multiply by the two coefficients.

A considerable body of literature (see for instance Duflo, 2003; Qian, 2008) shows that women tend to take decisions that are less impulsive and aimed more at increasing the general welfare of the household, while men tend to take more impulsive decisions. It is therefore interesting to investigate whether we observe a difference in behaviour between male-headed and female-headed households when faced with a change in the available choice set. In order to do this, we interact the electricity dummy with a gender dummy for the head of the household. This leaves us with two endogenous variables. In order to take care of this additional endogenous variable we also interact the instrument with the gender dummy. The results of this specification are presented in Table 5 and are not surprising. In columns (1) to (3), we present results for LPG, first for the whole sample, followed by households below and above the average level of expenditures. Columns (4) to (6) have the same structure and show results for fuelwood. We observe that male led households have higher positive effect on the probability of reverting to fuelwood and a higher negative effect on the probability of adopting LPG compared to female led household, both effects are statistically significant. Moreover, once we interact with gender, the effect appears also for households whose expenditures level is above the average, and behaves similarly.

Intensive margin

Table 6 presents intensive margin results, i.e. how does spending on LPG and fuelwood evolve as a consequence of electrification. Columns (1) and (2) show results for LPG, while columns (3) and (4) report fuelwood results. The picture presented here is similar to the one presented above for the extensive margin.

A new electrical connection decreases the probability of adoption of LPG as a cooking fuel, but seems to have no impact on the share of spending households dedicate to it. The coefficient obtained is 0.094 and it is statistically insignificant. When focusing on fuelwood we observe the same pattern as for the extensive margin, a new electrical connection increases spending in fuelwood, the coefficient here is 6.835, but it is only statistically significant at the 10 percent level. This coefficient implies an increase in the share of expenditures in fuelwood over the 5 years of 72.7 percentage points. At first, this number may seem large. Yet, the largest share of a household expenditure on fuelwood in the NSS sample is imputed using consumption multiplied by the prevailing regional price. This means that a large part of the increase does not come from an actual increase in spending but from an increase in the collection of fuelwood. Poorer households seem to decide to spend more on consumption goods becoming available because of the new form of energy, while cutting spending wherever possible, for instance by collecting fuelwood instead of buying LPG cylinders.

The pattern of the intensive margin results is similar to the one observed for the extensive margin. This similarity, reinforces our believe in an explanation related to an increase of the consumption choice set linked with a tight budget constraint. As before, in order to investigate this claim we split the sample according to the average monthly per capita expenditure. Results for this can be found in Panel B of Table 4. Columns (1) through (4) present results for households with monthly per capita spending below the 724 Rs. threshold, while columns (5) through (8) present results for households with spending above it. As in the extensive margin analysis, results are statistically significant only for the poorer part of the households. These households, when faced with new expenditures derived from electric power have to cut back on expenditures. We observe a decrease on expenditures on expensive fuels and, an increase in the share of income dedicated to fuelwood, the cheaper alternative.

Panel B of Table 5 shows results for the interaction with the gender of the head of the household for the intensive margin. Also in this case, results imply that women tend to spend more on better fuels compared to men.

5 Robustness

In order to further verify the claim that households revert to the use of biomass because electrification opens up a new array of possible consumption choices, we perform several robustness checks. First, instead of using expenditures to measure the wealth of each household, we use caste and education level. Second, we verify our claim that electricity puts further stress on the poorest households' budget by looking at the impact of electrification on their energy budget.

Alternative measures of wealth

To further confirm our previous results, we use two additional measures of wealth in order to split the sample. First, we divide households according to the caste to which they belong, and, second, we interact the impact of electrification with the education level of the head of the household.

Caste The caste system used to be linked to individuals' wealth level. It has been shown that a correlation between caste and wealth still exists (see for instance Zacharias and Vakulabharanam, 2011). We split our sample in five different ways. First, we identify the three backward castes: scheduled tribes, scheduled castes and other backward castes, we then take all the backward castes together and, finally, the other (upper) castes. All the backward castes taken together represent roughly 72.3% of our sample. Table 7 reports results for the extensive margin, while Table 8 for the intensive margin.

The extensive margin results, reported in Table 7, confirm what we observed before. When we take all the backward castes together, in column (4), an electric connection decreases the probability of adopting LPG, while it increases the probability of adopting fuelwood. The size of the coefficients obtained is similar to the one obtained in Table 4. Instead, when we look at the other castes, the result become statistically insignificant. We obtain the strongest results for other backward castes, in column (3) and schedule tribes, in column (1), while the results for scheduled castes are statistically insignificant (column 2).

Table 8 shows results for the intensive margin, which are also similar to the one obtained in Table 4. When pulling all the backward castes together we find the same results we had for households below mean expenditures, no impact on LPG expenditures and a statistically significant, albeit only at 10%, result on fuelwood expenditures. Interestingly, when looking at the other castes we find a positive and statistically significant result on LPG expenditures following electrification. This result fits the standard energy ladder hypothesis.

Education Education also serves as a proxy for wealth. On average, richer households have a higher level of education. We focus on primary education and create a dummy variable that takes value 1 if the head of the household as an education level higher than primary school and 0 if the level is at most primary school. In India primary school goes from grade 1 to grade 4. 49.8% of households have a head with at most a primary school education level.

We then proceed in the same way used when dealing with the gender of the head of household and interact the education dummy with the household's electrification status. We also interact it with our instrument in order to obtain a second instrument. Results are shown in Table 9. As expected, a lower education level further increases adoption of fuelwood and decreases the likelihood of picking up LPG. A higher level of education is likely to be linked with a better understanding of the dangers of indoor air pollution, while households with a lower level of education are more likely to be more oblivious to the dangers of it and use electricity for other purposes. The effect becomes noticeable even for households above the mean level of monthly expenditures.

Budget reaction

Finally, if a new connection to the power grid poses an extra constraint on households' budget, we should be able to pick this up. In Table 10, we regress the share of monthly fuel expenditures out of a household's total monthly expenditures on access to a connection to the power grid. Columns (1) and (2) show the OLS and IV results for the whole sample and seem to indicate, as expected, an increase in the share of income that is dedicated to energy. When we split the sample between households below and above the mean monthly per capita expenditures, we find an additional argument strengthening our claim. Poorer households already spend a larger share of their income on energy (on average 17% of their expenditures are dedicated to energy), and the new power connection increases this share in a significant way, by 16 percentage points on average, and this result is statistically significant at the 5% level. If instead we move to the richer half of the sample, that on average dedicates 14% of its expenditures to energy, the coefficient becomes statistically insignificant. Therefore, it seems that the new connection does not put any additional financial stress on richer households.

6 Conclusion

The goal of this paper is to identify a causal relationship between access to electricity and the adoption of modern cooking fuels, such as LPG. Access to a new electricity connection may impact households in two different ways. First, electricity opens up a whole new set of consumption opportunities, both productive – such as irrigation pumps – and non-productive – such as entertainment. Faced with these new consumption opportunities, a household may decide to divert to them some of the money that previously was devoted to energy, given the availability of free choices for energy (fuelwood, dung, crop residues). Second, according to the existing literature (e.g. International Evaluation Group, 2008; Khandker et al., 2012; Lipscomb et al., 2013; Chakravorty et al., 2014, 2016; Van de Walle et al., 2017) a connection to the grid contributes to increase rural households' incomes. This increase in socio-economic status, according to the energy ladder theory should push households towards the adoption of more efficient and less polluting fuels, such as kerosene or LPG. According to the fuel stacking literature, this may not be a clear jump, but we should observe the appearance of modern fuels together with the persistence of biomass. This causal relationship would constitute an additional channel through which rural electrification contributes to an improvement in households' welfare.

The results show that the first possibility is the one at play. We observe that, while electrification has no impact on the choice of a cooking fuel for richer households, its impact is consistent with what one would expect for poorer households when offering them a wider choice set. Poorer households (below mean expenditure), when faced with the possibility of having access to electric light, will take it, since electricity allows them to have good quality lightning. Yet, these households run on a tight budget constraint. As a consequence, in order to be able to afford electricity, they cut back on other energy related expenses, which leads them to revert from modern cooking fuels, such as LPG, to freely available biomass.

In order to improve our understanding of the results obtained, we should ideally investigate how these coefficients are affected by the quality of power supply. Presently, in rural India, electricity is available on average for only about 12 and a half to 13 hours in a day (Chakravorty et al., 2014; Aklin et al., 2016). Chakravorty et al. (2014) found that while a good quality connection may increase a household income by up to 27%, a low quality connection will not do better then 9%. The bad quality of power supply could have two consequences. First, incomes may not increase sufficiently enough to compensate for the new bill that needs to be paid and enable the poorest household to afford LPG. Second, because of the low quality of the power supply, households may not invest in many electrical appliances, i.e. their choice set would actually not be affected. This lack of appliances together with the fact that they only receive power for less than 13 hours per day may result in small expenditures for electricity and, therefore, we may not observe the change in cooking fuel choice for poor households receiving a bad quality of power supply. Unfortunately, NSS does not contain information about the quality of power supply and, therefore, we are unable to further investigate this channel.

Indoor pollution generated by the burning of biomass is a major health concern in India, between 400,000 and 500,000 people die of household air pollution every year. Pollution from dirty biomass fuels used for cooking contributes to 12% of still birth. Moreover, this is also a gender issue, since women are the primary cooks in many Indian households and household air pollution leads to a range of diseases, every year we observe: over 2.4 million cases of chronic bronchitis, over 300,000 cases of tuberculosis and over 5 million cases of cataract.[19] For these reasons the government of India is trying to push LPG adoption through the Ujjwala scheme.[20] Introduced in 2016, this scheme provides a subsidy to government-owned oil manufacturing companies for each LPG gas connection that they install in poor rural households without one. Yet, this program only focuses on the sunk costs related to the

[19] "How withdrawal of LPG subsidy can raise India's healthcare costs," *Business Standard*, August 3rd 2017.

[20] `http://www.pmujjwalayojana.com`

uptake of LPG, what may be inferred from our results is that the running costs, e.g. the costs of refilling the LPG cylinder, may still be a hurdle for the poorer households. Subsidies on LPG cylinders are currently low and decreasing in India.[21]

Our findings align with Lee et al. (2020) along two dimensions. First, they imply that, for the poor, access to electricity makes the switch to cleaner cooking fuels even harder. Second, one of the main take away of the paper, is that the introduction of electricity should always be accompanied by complementary policies. In this specific case, they could consist of subsidies for the adoption of cleaner cooking fuels, in order to avoid a worsening of indoor pollution for the poorest households.

[21] "Govt orders LPG prices to be hiked by Rs 4 per month", *The Times of India*, July 31st 2017; Singh, S., "How withdrawal of LPG subsidy can raise India's healthcare costs," *Business Standard*, August 3rd 2017.

References

Aklin, M., Cheng, C.-y., Urpelainen, J., Ganesan, K., and Jain, A. (2016). Factors affecting household satisfaction with electricity supply in rural india. *Nature Energy*, 1(11):16170.

Allcott, H., Collard-Wexler, A., and O'Connell, S. D. (2016). How do electricity shortages affect industry? evidence from india. *The American Economic Review*, 106(3):587–624.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of monetary economics*, 23(2):177–200.

Barron, M. and Torero, M. (2017). Household Electrification and Indoor Air Pollution. *Journal of Environmental Economics and Management*, 86:81–92.

Bartik, T. J. (1994). The effects of metropolitan job growth on the size distribution of family income. *Journal of Regional Science*, 34(4):483–501.

Bond, T. C., Doherty, S. J., Fahey, D., Forster, P., Berntsen, T., DeAngelo, B., Flanner, M., Ghan, S., Kärcher, B., Koch, D., et al. (2013). Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research: Atmospheres*, 118(11):5380–5552.

Bruce, N., Perez-Padilla, R., and Albalak, R. (2000). Indoor Air Pollution in Developing Countries: A Major Environmental and Public Health Challenge. *Bulletin of the World Health Organization*, 78:1078–1092.

Central Electricity Authority (2006). *Review of Performance of Hydro Power Stations 2005-2006*. Government of India, Ministry of Power.

Chakravorty, U., Emerick, K., and Ravago, M.-L. (2016). Lighting up the last mile: The benefits and costs of extending electricity to the rural poor. *Resources for the Future Discussion Paper*, pages 16–22.

Chakravorty, U., Pelli, M., and Marchand, B. U. (2014). Does the quality of electricity matter? evidence from rural india. *Journal of Economic Behavior & Organization*, 107:228–247.

Chan, H., Cropper, M., and Malik, K. (2014). Why Are Power Plants in India Less Efficient Than Power Plants in the United States? *American Economic Review*, 104(5):586–590.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082):1907–1918.

Dherani, M., Pope, D., Mascarenhas, M., Smith, K., Weber, M., and Bruce, N. (2008). Indoor Air Pollution from Unprocessed Solid Fuel Use and Pneumonia Risk in Children Aged under Five Years: A Systematic Review and Meta-Analysis. *Bulletin of the World Health Organization*, 86:390–398C.

Duflo, E. (2003). Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa. *World Bank Economic Review*, 17(1):1–25.

Duflo, E. and Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 122(2):601–646.

Garcia-Mila, T. and McGuire, T. J. (1992). The contribution of publicly provided inputs to states' economies. *Regional science and urban economics*, 22(2):229–241.

Greenstone, M. et al. (2014). Energy, growth and development. *International Growth Center Evidence Paper*.

Heltberg, R. (2004). Fuel switching: evidence from eight developing countries. *Energy economics*, 26(5):869–887.

Holtz-Eakin, D. (1993). State-specific estimates of state and local government capital. *Regional science and urban economics*, 23(2):185–209.

Hosier, R. and Dowd, J. (1987). Household Fuel Choice in Zimbabwe. *Resources and Energy*, 9:347–361.

International Evaluation Group (2008). *The welfare impact of rural electrification: A reassessment of the costs and benefits*. The World Bank.

Khandker, S. R., Barnes, D. F., and Samad, H. A. (2012). The welfare impacts of rural electrification in bangladesh. *The Energy Journal*, 33(1).

Kishore, A. and Spears, D. (2014). Having a Son Promotes Clean Cooking Fuel Use in Urban India: Women's Status and Son Preference. *Economic Development and Cultural Change*, 62(4):673–699.

Leach, G. (1992). The Energy Transition. *Energy Policy*, 20(2):116–123.

Lee, K., Miguel, E., and Wolfram, C. (2020). Does Household Electrification Supercharge Economic Development? *Journal of Economic Perspectives*, 34(1):122–44.

Lipscomb, M., Mobarak, A., and Barham, T. (2013). Development Effects of Electrification: Evidence from the Geologic Placement of Hydropower Plants in Brazil. *American Economic Journal: Applied Economics*, 5(2):200–231.

Masera, O., Saatkamp, B., and Kammen, D. (2000). From Linear Fuel Switching to Multiple Cooking Strategies: A Critique and Alternative to the Energy Ladder Model. *World Development*, 28(12):2083 – 2103.

Pant, P., Lal, R. M., Guttikunda, S. K., Russell, A. G., Nagpure, A. S., Ramaswami, A., and Peltier, R. E. (2019). Monitoring particulate matter in india: Recent trends and future outlook. *Air Quality, Atmosphere & Health*, 12(1):45–58.

Peters, J. and Sievert, M. (2016). Impacts of rural electrification revisited – the African context. *Journal of Development Effectiveness*, 8(3):327–345.

Po, J., Fitzgerald, J., and Carlsten, C. (2011). Respiratory Disease Associated with Solid Biomass Fuel Exposure in Rural Women and Children: Systematic Review and Meta-Analysis. *Thorax*, 66:232–239.

Qian, N. (2008). Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance. *The Quarterly Journal of Economics*, 123(3):1251–1285.

Röller, L.-H. and Waverman, L. (2001). Telecommunications infrastructure and economic development: A simultaneous approach. *American economic review*, pages 909–923.

Ruiz-Mercado, I., M. O. (2015). Patterns of Stove Use in the Context of Fuel-Device Stacking: Rationale and Implications. *EcoHealth*, 12:45–56.

Smith, K. R. (2000). National burden of disease in india from indoor air pollution. *Proceedings of the National Academy of Sciences*, 97(24):13286–13293.

Van de Walle, D., Ravallion, M., Mendiratta, V., and Koolwal, G. (2017). Long-term gains from electrification in rural india. *The World Bank Economic Review*, 31(2):385–411.

Van der Kroon, B., Brouwer, R., and van Beukering, P. (2013). The Energy Ladder: Theoretical Myth or Empirical Truth? Results from a Meta-Analysis. *Renewable and Sustainable Energy Reviews*, 20:504–513.

Zacharias, A. and Vakulabharanam, V. (2011). Caste stratification and wealth inequality in india. *World Development*, 39(10):1820–1833.

7 Tables

Table 1: Descriptive statistics

Variable	*Mean*	*SE*	*Min*	*Max*
<u>*Panel A:*</u>				
Share of electricity	0.68	0.47	0	1
Share of LPG‡	0.16	0.37	0	1
Share of fuelwood‡	0.73	0.44	0	1
LPG share of expenditures (%)	0.94	2.16	0	27.82
Fuelwood share of expenditures (%)	5.29	4.59	0	64.88
<u>*Panel B:*</u>				
Nightlights 1993	2.53	2.85	0	55.44
Hydro Instrument	0.13	0.19	0	0.93
Interaction Hydro Instrument	0.32	0.60	0	4.91
<u>*Panel C:*</u>				
Price Fuelwood (Rs./Kg)	1.44	1.76	0.05	112.97
Price LPG (Rs./ Kg)	20.27	9.99	8.36	215.83
Household Size	5.00	2.46	1	30.00
Total land owned (Hectares)	1.01	2.42	0	310.31
Education head of household (Years)	4.39	3.10	1	13.00
Age head of household	46.44	13.29	14	100
Sex head of household‡	0.89	0.31	0	1
Hindu‡	0.77	0.42	0	1
Scheduled tribe‡	0.16	0.37	0	1
Scheduled caste‡	0.18	0.38	0	1
Other backward caste‡	0.38	0.49	0	1
Below Primary‡	0.50	0.50	0	1

Notes: Our sample contains 136,221 households, 78,017 interviewed in 2004 and 58,204 interviewed in 2009. *Panel A* reports statistics for the variables of interest and the dependent variables, *Panel B* for the instruments used and, *Panel C* for all the control variables used in the different specifications. The share of expenditures is computed over the last 30 days. *Sex head of household* equals 1 if the head of the household is a male and 0 otherwise. ‡ denotes indicator variables.

Table 2: First stage

	Dep. variable: electricity connection			
	All		**Below**	**Above**
	(1)	(2)	(3)	(4)
Hydro	-0.301^{**}	-0.274^{**}	-0.323^{**}	0.022
	(0.133)	(0.123)	(0.147)	(0.137)
*Hydro*Initial elec*	0.140^{***}	0.118^{***}	0.123^{***}	0.045^{**}
	(0.026)	(0.025)	(0.034)	(0.023)
Price Fuelwood		0.002	0.002	−0.001
		(0.001)	(0.002)	(0.002)
Price LPG		−0.000	−0.000	−0.001
		(0.001)	(0.001)	(0.001)
HH size		0.010^{***}	0.016^{***}	0.011^{***}
		(0.001)	(0.001)	(0.001)
Total land owned		0.004^{***}	0.004^{***}	0.001
		(0.001)	(0.001)	(0.001)
Education head of HH		0.028^{***}	0.028^{***}	0.015^{***}
		(0.001)	(0.001)	(0.001)
Sex of head of HH		-0.019^{***}	-0.012^{**}	-0.021^{***}
		(0.005)	(0.006)	(0.005)
Age head of HH		0.002^{***}	0.002^{***}	0.001^{***}
		(0.0001)	(0.0001)	(0.0001)
Hindu		0.010	0.010	0.002
		(0.007)	(0.007)	(0.007)
Scheduled tribe		-0.143^{***}	-0.144^{***}	-0.080^{***}
		(0.013)	(0.013)	(0.014)
Scheduled caste		-0.074^{***}	-0.071^{***}	-0.046^{***}
		(0.006)	(0.006)	(0.008)
Other backward caste		-0.027^{***}	-0.028^{***}	-0.020^{***}
		(0.005)	(0.006)	(0.005)
Year F.E.	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes
Observations	136,221	136,154	92,773	43,381
F-stat	16.68	12.08	6.76	8.03
AR F-stat	7.40	13.86	22.24	5.81

Notes: *All* includes the whole sample, *Below* is the sub-sample of households below the mean of monthly household consumption expenditure of 724 Rs. (roughly 10 USD) and *Above* is the sub-sample of households below the mean of monthly household consumption expenditure of 724 Rs. AR F-stat stands for the Anderson Rubin F statistic for weak instruments. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Baseline – extensive margin

	Dep. variable: adoption			
	LPG		**Fuelwood**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Electricity	0.081***	−0.566**	−0.074***	0.520*
	(0.005)	(0.231)	(0.007)	(0.308)
Price Fuelwood	0.003**	0.004**	−0.005***	−0.006***
	(0.001)	(0.002)	(0.002)	(0.002)
Price LPG	−0.003***	−0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
HH size	−0.003***	0.004	0.006***	−0.001
	(0.001)	(0.002)	(0.001)	(0.003)
Total land owned	0.001	0.004***	−0.000	−0.003*
	(0.001)	(0.001)	(0.001)	(0.002)
Education head of HH	0.037***	0.056***	−0.036***	−0.053***
	(0.001)	(0.007)	(0.001)	(0.009)
Sex of head of HH	−0.047***	−0.060***	0.044***	0.055***
	(0.004)	(0.007)	(0.004)	(0.008)
Age head of HH	0.002***	0.004***	−0.002***	−0.003***
	(0.000)	(0.001)	(0.000)	(0.001)
Hindu	0.005	0.011	−0.004	−0.010
	(0.006)	(0.008)	(0.008)	(0.009)
Scheduled tribe	−0.075***	−0.168***	0.098***	0.183***
	(0.008)	(0.035)	(0.009)	(0.046)
Scheduled caste	−0.078***	−0.126***	0.075***	0.119***
	(0.005)	(0.012)	(0.006)	(0.022)
Other backward caste	−0.056***	−0.074***	0.045***	0.062***
	(0.005)	(0.009)	(0.006)	(0.011)
Year F.E.	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes
Observations	136,154	136,154	136,154	136,154

Notes: The dependent variable is a dummy taking value 1 if LPG (fuelwood) is adopted as the primary cooking fuel by a household and 0 otherwise. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Baseline – below and above mean expenditures

	Below				Above			
	LPG		**Fuelwood**		**LPG**		**Fuelwood**	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: extensive								
Electricity	0.046***	−0.526**	−0.040***	0.758**	0.163***	−0.091	−0.148***	−0.148
	(0.003)	(0.229)	(0.006)	(0.384)	(0.009)	(0.600)	(0.011)	(0.592)
Observations	92,773	92,772	92,773	92,772	43,381	43,381	43,381	43,381
Panel B: intensive								
Electricity	0.409***	−1.512	−1.080***	12.250**	1.027***	3.017	−1.260***	−1.122
	(0.028)	(1.167)	(0.068)	(6.032)	(0.052)	(3.088)	(0.117)	(4.006)
Observations	92,591	92,590	92,469	92,468	43,200	43,200	43,259	43,259
Year F.E.	yes	yes	yes	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes	yes	yes	yes	yes

Notes: *Below* denotes the sub-sample of households below the mean of monthly household consumption expenditures of 724 Rs. (roughly 10USD) while *Above* denotes the sub-sample of households above the mean of monthly household consumption expenditures of 724 Rs. In *Panel A*, the dependent variable is a dummy taking value 1 if LPG (fuelwood) is adopted as the primary cooking fuel by a household and 0 otherwise. In *Panel B*, the dependent variable is the share of expenditure in LPG (fuelwood) over total monthly expenditures. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5: Gender

	LPG			Fuelwood		
	All	Below	Above	All	Below	Above
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Extensive						
Electricity	-0.493^{**}	-0.491^{**}	0.009	0.469	0.724^{*}	-0.207
	(0.225)	(0.227)	(0.583)	(0.299)	(0.380)	(0.577)
*Gender*Elec*	-0.086^{***}	-0.041^{***}	-0.109^{***}	0.062^{***}	0.041^{**}	0.059^{***}
	(0.012)	(0.016)	(0.026)	(0.014)	(0.019)	(0.022)
Observations	136,154	92,772	43,381	136,154	92,772	43,381
Panel B: Intensive						
Electricity	0.508	-1.183	3.451	6.520^{*}	12.348^{**}	-1.308
	(1.195)	(1.154)	(2.995)	(3.550)	(5.988)	(3.892)
*Gender*Elec*	-0.487^{***}	-0.389^{***}	-0.467^{***}	0.354^{**}	-0.121	0.201
	(0.076)	(0.107)	(0.138)	(0.169)	(0.244)	(0.161)
Observations	135,791	92,590	43,200	135,728	92,468	43,259
Controls	yes	yes	yes	yes	yes	yes
Year F.E.	yes	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes	yes	yes

Notes: This table reports only instrumental variable estimations. The *Gender* variable is a dummy for the gender of the head of the household. It takes value 1 for male and 0 otherwise. *All* includes the whole sample, *Below* is the sub-sample of households below the mean of monthly household consumption expenditure of 724 Rs. (roughly 10 USD) and *Above* is the sub-sample of households above the mean of monthly household consumption expenditure of 724 Rs. In *Panel A*, the dependent variable is a dummy taking value 1 if LPG (fuelwood) is adopted as the primary cooking fuel by a household and 0 otherwise. In *Panel B*, the dependent variable is the share of expenditure in LPG (fuelwood) over total monthly expenditures. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Baseline – intensive margin

	Dep. variable: income share			
	LPG		**Fuelwood**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Electricity	0.614***	0.094	−1.451***	6.835*
	(0.030)	(1.230)	(0.072)	(3.652)
Price Fuelwood	0.017**	0.018**	0.100**	0.085**
	(0.007)	(0.008)	(0.042)	(0.042)
Price LPG	−0.018***	−0.018***	0.003	0.006
	(0.004)	(0.004)	(0.007)	(0.009)
HH size	−0.038***	−0.033**	−0.187***	−0.274***
	(0.003)	(0.013)	(0.009)	(0.041)
Total land owned	−0.008**	−0.006	−0.060***	−0.095***
	(0.003)	(0.006)	(0.013)	(0.024)
Education head of HH	0.179***	0.193***	−0.338***	−0.572***
	(0.005)	(0.035)	(0.010)	(0.105)
Sex of head of HH	−0.304***	−0.314***	−0.230***	−0.066
	(0.027)	(0.038)	(0.061)	(0.111)
Age head of HH	0.012***	0.013***	−0.012***	−0.030***
	(0.001)	(0.003)	(0.001)	(0.008)
Hindu	0.105***	0.110***	−0.070	−0.156
	(0.033)	(0.037)	(0.073)	(0.105)
Scheduled tribe	−0.458***	−0.533***	1.115***	2.303***
	(0.049)	(0.184)	(0.117)	(0.557)
Scheduled caste	−0.457***	−0.496***	0.877***	1.494***
	(0.028)	(0.095)	(0.071)	(0.275)
Other backward caste	−0.303***	−0.317***	0.345***	0.575***
	(0.025)	(0.042)	(0.052)	(0.114)
Year F.E.	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes
Observations	135,791	135,791	135,728	135,728
F-stat		12.30		11.87

Notes: The dependent variable is the share of expenditure in LPG (fuelwood) over total monthly expenditures. The sample is smaller than for the extensive margin because 420 households report using LPG, yet, do not report the amount of their expenditures. Moreover, some households (24) report their LPG expenditures but do not use LPG as the primary cooking fuel. As a consequence, the F-stat of the first stage are slightly different from those reported in Table 2. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Caste – extensive margin

	Dep. variable: adoption				
	ST	SC	OBC	All Backward	Other
	(1)	(2)	(3)	(4)	(5)
Panel A: LPG					
Electricity	-0.855^{*}	0.199	-0.597^{*}	-0.495^{**}	-0.235
	(0.470)	(0.274)	(0.315)	(0.227)	(0.395)
Observations	21,811	24,253	52,270	98,419	37,720
Panel B: Fuelwood					
Electricity	0.724	-0.568	1.572^{***}	0.581^{*}	-0.339
	(0.492)	(0.445)	(0.610)	(0.321)	(0.409)
Observations	21,811	24,253	52,270	98,419	37,720
Controls	yes	yes	yes	yes	yes
Year F.E.	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes	yes

Notes: This table reports only instrumental variable estimations. *Panel A (B)* reports results for LPG (fuelwood) adoption and, therefore, the dependent variable is a dummy taking value 1 if LPG (fuelwood) is adopted as the primary cooking fuel by a household and 0 otherwise. *ST*, *SC* and *OBC* represent scheduled tribe, scheduled caste and other backward caste, respectively. *All Backward* is all the aforementioned castes combined into one category and *Other* includes all upper castes. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8: Caste – intensive margin

	Dep. variable: income share				
	ST	SC	OBC	All Backward	Other
	(1)	(2)	(3)	(4)	(5)
<u>*Panel A: LPG*</u>					
Electricity	−3.373	2.314	−1.646	−1.528	4.359*
	(2.422)	(1.455)	(1.582)	(1.123)	(2.510)
Observations	21,707	24,216	52,152	98,160	37,618
<u>*Panel B: Fuelwood*</u>					
Electricity	5.086	−0.796	16.875**	7.975*	−3.589
	(6.422)	(3.449)	(8.067)	(4.669)	(3.611)
Observations	21,727	24,184	52,117	98,113	37,600
Controls	yes	yes	yes	yes	yes
Year F.E.	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes	yes

Notes: This table reports only instrumental variable estimations. *Panel A (B)* reports results for the intensive margin of LPG (fuelwood) and, therefore, the dependent variable is the share of expenditure in LPG (fuelwood) over total monthly expenditures. *ST*, *SC* and *OBC* represent scheduled tribe, scheduled caste and other backward caste, respectively. *All Backward* is all the aforementioned castes combined into one category and *Other* includes all upper castes. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Schooling

	Dep. variable: adoption and income share					
	LPG			**Fuelwood**		
	All	Below	Above	All	Below	Above
	(1)	(2)	(3)	(4)	(5)	(6)
<u>*Panel A: Extensive*</u>						
Electricity	−0.632**	−0.559**	0.079	0.533*	0.753*	−0.301
	(0.253)	(0.249)	(0.557)	(0.317)	(0.388)	(0.560)
Below Primary	−0.285***	−0.163***	−0.254***	0.244***	0.179***	0.182***
	(0.024)	(0.026)	(0.027)	(0.031)	(0.040)	(0.028)
Observations	136,183	92,792	43,390	136,183	92,792	43,390
<u>*Panel B: Intensive*</u>						
Electricity	−0.135	−1.813	3.548	6.852*	12.652**	−2.111
	(1.303)	(1.334)	(2.910)	(3.594)	(6.099)	(3.711)
Below Primary	−1.084***	−1.022***	−0.861***	2.304***	2.361***	0.958***
	(0.124)	(0.152)	(0.137)	(0.356)	(0.637)	(0.187)
Observations	135,820	92,610	43,209	135,755	92,486	43,268
Controls	yes	yes	yes	yes	yes	yes
Year F.E.	yes	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes	yes	yes

Notes: This table reports only instrumental variable estimations. *All* includes the whole sample, *Below* is the sub-sample of households below the mean of monthly household consumption expenditure of 724 Rs. (roughly 10USD) and *Above* is the sub-sample of households above the mean of monthly household consumption expenditure of 724 Rs. The variable *Below Primary* is a dummy coded 1 if the head of the household complete at most primary education and 0 otherwise. In *Panel A*, the dependent variable is a dummy taking value 1 if LPG (fuelwood) is adopted as the primary cooking fuel by a household and 0 otherwise. In *Panel B*, the dependent variable is the share of expenditure in LPG (fuelwood) over total monthly expenditures. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

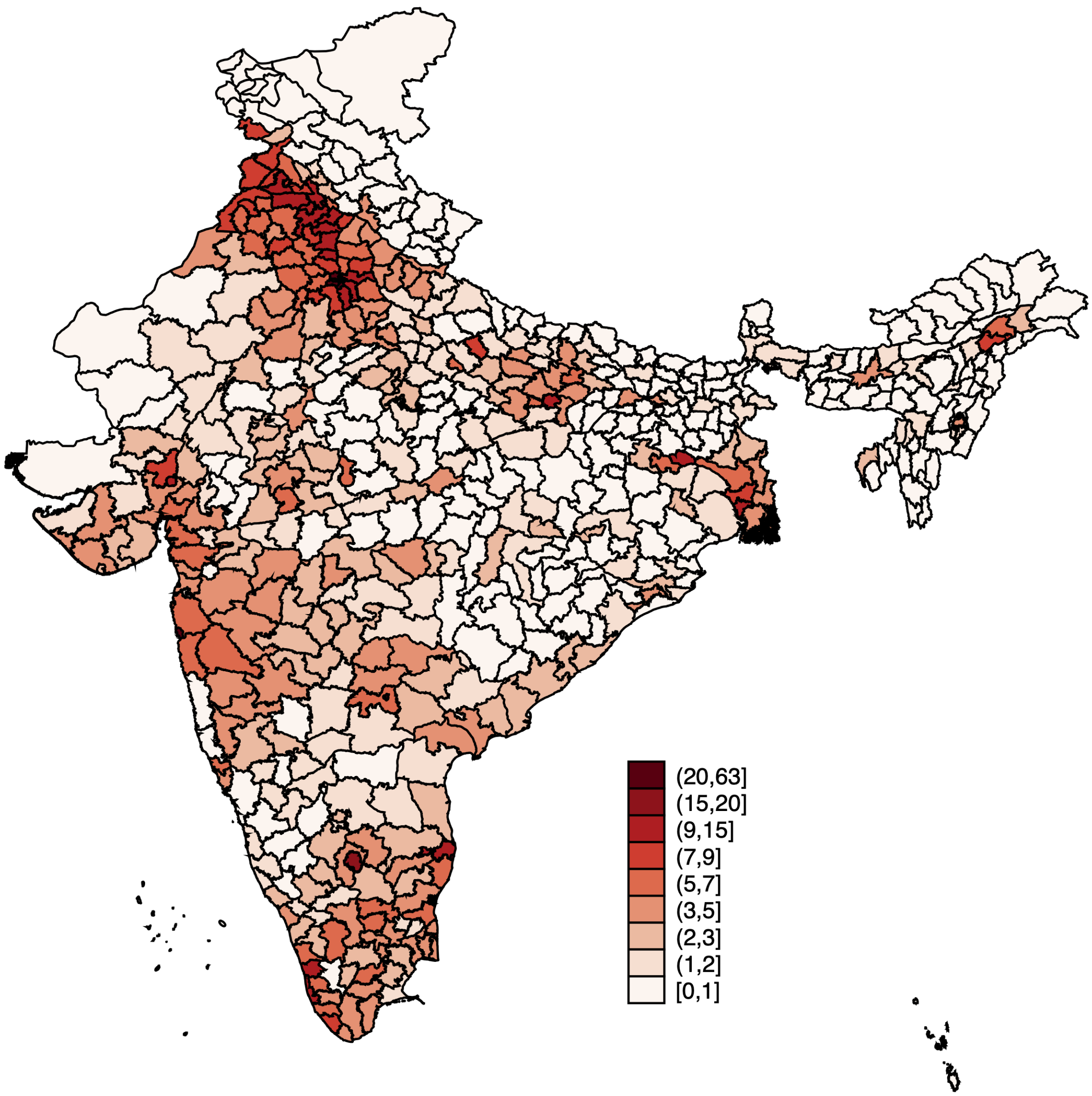
Table 10: Share of fuel expenditures

	Dep. variable: share of fuel expenditures					
	All		**Below**		**Above**	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Electricity	1.436***	17.631***	2.053***	22.676***	1.713***	13.048
	(0.093)	(5.389)	(0.096)	(8.263)	(0.165)	(9.103)
Price Fuelwood	0.126**	0.096	0.104	0.058	0.194***	0.179**
	(0.058)	(0.061)	(0.068)	(0.078)	(0.066)	(0.076)
Price LPG	−0.015*	−0.010	−0.019**	−0.015	0.001	−0.013
	(0.008)	(0.013)	(0.009)	(0.017)	(0.008)	(0.015)
HH size	−0.661***	−0.831***	−0.860***	−1.193***	−0.804***	−0.641***
	(0.014)	(0.062)	(0.017)	(0.138)	(0.020)	(0.104)
Total land owned	−0.118***	−0.186***	−0.065**	−0.138**	0.008	0.019
	(0.024)	(0.040)	(0.027)	(0.057)	(0.014)	(0.016)
Education head of HH	−0.346***	−0.803***	−0.161***	−0.733***	−0.160***	0.067
	(0.011)	(0.156)	(0.012)	(0.234)	(0.013)	(0.142)
Sex of head of HH	−1.215***	−0.903***	−1.617***	−1.367***	−0.842***	−1.154***
	(0.086)	(0.174)	(0.097)	(0.201)	(0.102)	(0.242)
Age head of HH	0.000	−0.035***	0.015***	−0.025	0.024***	0.037***
	(0.002)	(0.012)	(0.002)	(0.016)	(0.003)	(0.009)
Hindu	0.048	−0.115	0.024	−0.182	0.075	0.104
	(0.083)	(0.152)	(0.086)	(0.194)	(0.106)	(0.145)
Schedued tribe	0.785***	3.099***	0.352***	3.323***	0.341*	−0.826
	(0.135)	(0.833)	(0.133)	(1.246)	(0.195)	(0.798)
Scheduled caste	0.982***	2.185***	0.533***	2.011***	0.122	−0.566
	(0.091)	(0.424)	(0.090)	(0.605)	(0.120)	(0.454)
Other backward caste	0.358***	0.804***	0.101	0.693***	0.118	−0.167
	(0.068)	(0.179)	(0.078)	(0.263)	(0.079)	(0.210)
Year F.E.	yes	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	yes	yes	yes
Observations	136,070	136,070	92,725	92,724	43,345	43,345

Notes: *All* includes the whole sample, *Below* is the sub-sample of households below the mean of monthly household consumption expenditure of 724 Rs. (roughly 10 USD) and *Above* is the sub-sample of households below the mean of monthly household consumption expenditure of 724 Rs. The dependent variable is the share of expenditure in energy over a household's total monthly expenditures. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

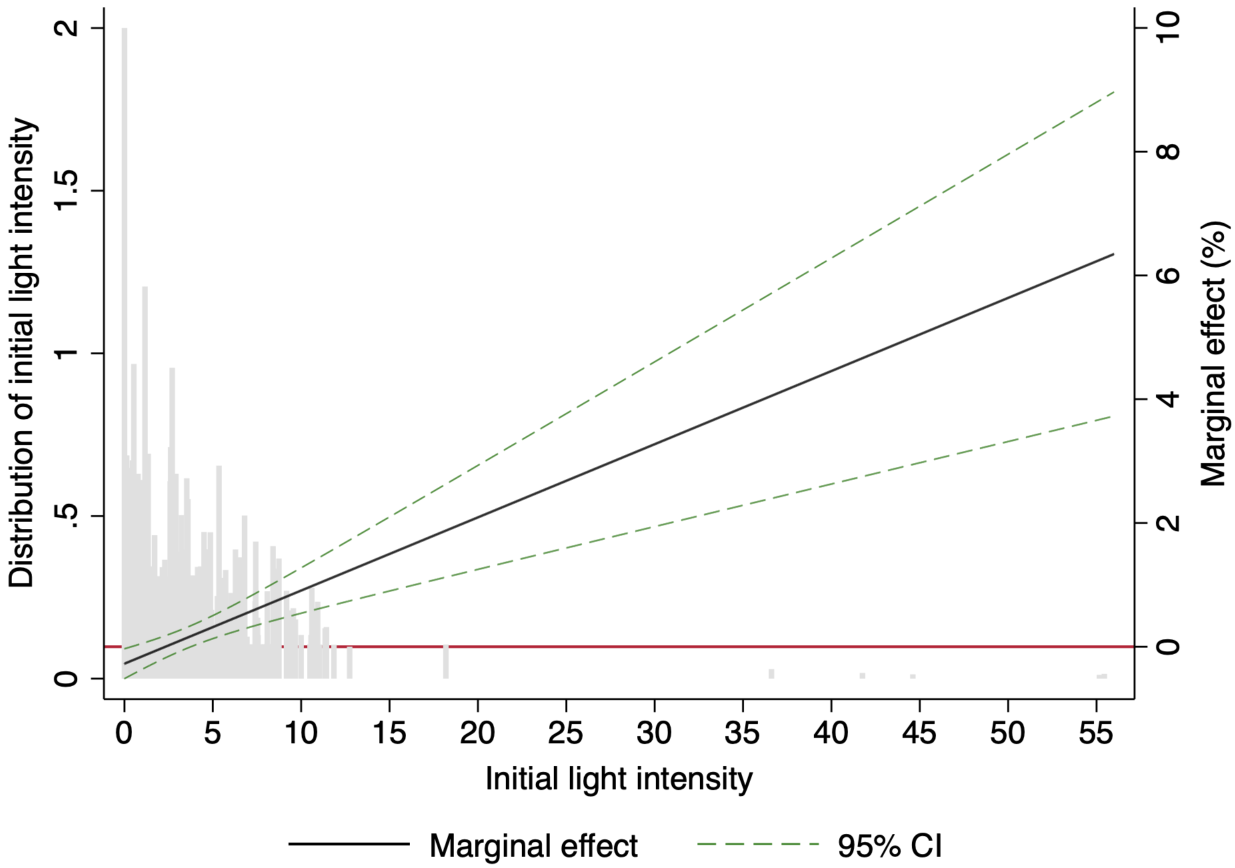
8 Figures

Figure 1: Night light intensity by district, 1993



Notes: Night light intensity varies between a minimum of 0 (no light output) and a maximum of 63 (maximum light output). Darker shades of red indicate a high light output, while lighter shades indicate a smaller light output. The figure represents the situation of 1993. Source: `http://india.nightlights.io/#/nation/2006/12`

Figure 2: Marginal effect of a supply shift



Notes: The graph represents the marginal effect of the instrument in the first stage estimation. The marginal effect is computed using the coefficients from equation (3) in the following way: $\frac{\partial E_{hdst}}{\partial H_{st-1}} = \beta_1 L_{sd0} + \beta_2$. The gray histogram represents the distribution of districts across the initial level of electrification (1993).