Navarra Center for International Development

Working Paper nº 04/2022

Quality Matters: Power Reliability and Grid Connection in Rural Guatemala

Federico M. Accursi

University of Navarra

Quality Matters: Power Reliability and Grid Connection in Rural Guatemala.

Federico M. Accursi*

November 10, 2022

Electrification rates have been increasing within low and middle-income countries. However, the prevalence of outages is still a relevant issue for rural households when considering whether to connect to the grid or not. We test this by exploiting a shock in quality that unequally affected different municipalities in Guatemala during 2012-2014. Our main estimates, which are robust to the use of an instrumental variable strategy, suggest that households affected by severe outages are about 18-27p.p. less likely to get a connection to the grid. We further check this result by combining household-level data from the 2018 Census with a complete register of electricity quality service. Although 2018 was a good year in terms of quality, 1p.p. increase in the number of outage hours affected the probability of connection by 2.5p.p. Efforts to expand the electricity grid to rural areas should thus be analyzed in parallel with actual power grid quality levels.

Keywords: power reliability; energy access; rural households; Instrumental Variables; Latin America

JEL classification codes: Q49;D10;O10

*Address: Dept. Economics, Edificio Amigos, Campus Universitario s/n, 31009, Pamplona, Spain. E-mail: faccursi@unav.es.

1 Introduction

Even though the worldwide access to electricity has been rising from 82% in 2008 to 89% in 2018, almost 800 million people still do not have access to it (World Bank, 2021). Lack of electricity access is a particular pronounced problem in rural settings whose access rate is 15 p.p. less than urban.[1] Although many rural households have benefited from off-grid energy devices like solar panels, more research is needed to study the barriers that impede them from fully exploiting all the advantages electricity grid provides.[2]

As the Sustainable Development Goal #7 (SDG-7) stresses, access to electricity supply goes beyond the classic dichotomous variable of grid connection, and entails affordability, reliability and sustainability. These characteristics are not independent from each other, and access (i.e. grid expansion) does not necessarily mean truly and reliable connections. In particular, Lee et al. (2014) distinguishes between households that are "off grid" and those "under grid". The former are households too far away from the grid —and therefore too expensive— to be connected, and the latter are close enough to get connected at a reasonable low cost. In their study, they emphasize the affordability issue as a barrier to electrification: 50% of their unconnected households were "under grid".

In this paper we empirically study the role of the lack of reliability as a barrier for rural households to get a connection to the grid or also discourage households already connected to continue with the service. Poor quality service could also provoke conflicts in the form of theft and illegal connections or unpaid bills, triggering a "vicious circle" for an utility company: decreasing firm revenues and, therefore, increasing outages Dzansi et al. (2018). This issue concerns from a public policy perspective since the investment done to spread the low voltage grid needs social returns. Our results show that these efforts could end wasted if quality decreases because of, for example, insufficient complementary investments (e.g. transmission lines).

[1] As an example of regional inequality, Smith and Wills (2018) emphasize that oil discoveries have benefited the urban areas leaving rural areas behind in terms of electricity access.

[2] See for example Bayer et al. (2020) for a literature review. Grimm et al. (2020) suggest in their field experiment in rural Rwanda that willingness to pay for off-grid solar sometimes do not cover the costs of off-grid electrification.

To study this policy-relevant question, we use data from Guatemala. This country offers an appropriate context for several reasons. Considered an upper middle-income country by the World Bank, ended a civil war in 1996 and started since then a reform process, enhancing the rural electricity access rate from 48% to 74% in a decade (World Bank, 2021). However, in 2016, total energy consumption from residential sector came 90% from firewood and only 5% from electricity (Ministerio de Energía y Minas, 2019). Firewood, which is still the main primary energy source in the country specially for cooking and heating in rural areas, is typically associated with indoor pollution, and, hence, with well-documented negative health consequences (Parikh, 2011; Smith-Sivertsen et al., 2009).

Notwithstanding the growth in rural electricity access, the urbanization process remains steadily and differences between urban and rural areas persist. According to the Instituto Nacional de Estadística (INE), in 2017 the average Metropolitan urban labor income more than doubled the rural one; meanwhile the poverty index was 32% in the Metropolitan Area in 2014, and in the North Region of the country —mostly rural—, reached 77%. This gap also exists in electricity supply. In the last decade, on average, rural areas suffered 35% of more service interruption in duration, and 14% in frequency. Importantly, after 2011, the number of outages suddenly increased in rural Guatemala. In this study, we take advantage of this plausible exogenous shock (mainly attributable to managerial reasons) to analyze the causal relationship between power reliability on rural households' disposal to connect to thegrid.

To address our research question, we use data from the National Commission of Electricity Energy of Guatemala (CNEE), which we combine with two household level datasets; namely, the National Survey of Living Conditions (ENCOVI, hereafter) of 2011 and 2014 and the recent 2018 National Population Census. On the one hand, the ENCOVI dataset allows us to exploit spatial and time variation at more aggregated level. On the other hand, the Census data allows us to exploit spatial variation at a more granular level.

Regarding ENCOVI, the particular variation of quality observed in time will help us for the identification strategy, given the aforementioned unexpected shift in quality after 2011 that unequally affected different departments of Guatemala. Contrary to previous studies, our empirical analysis uses official (objective) data on outages from CNEE, avoiding two classical empirical problems: self-selection bias and measurement error. Using this data, we find strong evidence that there is a positive effect of quality on rural household connections to the grid. This evidence is robust to a battery of robustness check, including an instrumental variable regression, and it is further supported when using cross-sectional data from the Census. These empirical results have important policy implications, as they directly speak on the grid reliability vs grid expansion trade-off.

To the best of our knowledge, most of the previous literature that studies the impact of power reliability has mainly focused on the industrial sector. Special attention has received the effect on productivity (Allcott et al., 2016; Grainger and Zhang, 2019), on average unit costs (Fisher-Vanden et al., 2015), firm sales (Cole et al., 2018), or strategic behavior such as investment on back up generation (Oseni and Pollitt, 2015).

At the household level, most of the previous papers have concentrated on the effect of electrification on household outcomes, taking reliability for granted.[3] According to the literature review by Bonan et al. (2017), there are few studies that focus on barriers to electricity connections —the majority of which focus on liquidity constraints—, while the role of reliability as a barrier is not analyzed. However, since 2017 there has been an increasing interest on reliability itself. For example, Dang and La (2019) stresses the positive effect of power quality on rural income in Vietnam by the extensive margin (new connections because of better quality) and the intensive margin (more electricity usage), Bajo-Buenestado (2021) states that blackouts discourage electricity connections in Kenya, and Sedai et al. (2021) examines the effect of reliability on reducing gender differences in labor market in India.

It is important to remark that the way in which previous literature deals with the issue of measuring "quality" is varied. The problem of its definition depends essentially on the information and data structure available. For example, Chakravorty et al. (2014) uses a dummy variable to define either good or bad quality with a threshold value, based on self-reported hours of effective supply and frequency of outages. Alternatively, other authors

[3]See for example the effect of electrification on time distribution in Guatemala (Grogan, 2018), or on education (Arraiz and Calero, 2015)

use a continuous measure. For instance, Dang and La (2019) use data from a three-round household dataset in Vietnam and *counts* for the number of days without power outages. Using data from an *opinion survey*, Millien (2017) builds a weighted severity index of reliability uncertainty based on perception data in Kenya. Our paper overcomes the usual problems associated to self-reported data on quality since, as discussed above, we use detailed (objective) data on outages from official sources.

Finally, there is another related strand of the literature interested in measuring willingness to pay for ensuring a reliable power supply. For example, Hashemi (2021) points out the heterogeneity valuation of reliable supply across and within customer categories in India specially for industrial consumers. Also, Kennedy et al. (2019) construct village reliability supply variables from an average *self-reported information*, using daily hours of supply, availability of electricity at night, frequencies of outages and damages of electric equipment due to voltage fluctuations. They state the importance of high-quality service for rural households, as they are willing to pay more for better service resulting thus in additional connections if quality is improved.

This paper contributes to the nascent literature on the impact of reliability on the uptake of electricity connections. Our findings are aligned with Millien (2017) and Kennedy et al. (2019). In terms of policy implications, keeping a good quality service would be as important as grid extension as Chakravorty et al. (2014) document for the Indian case, with their positive impact on household incomes. In addition, Chaurey and Le (2022) argues that infrastructure maintenance (i.e. electrification and roads connectivity) play an important role in enhancing rural economic activity.

The rest of the paper is organized as follows. Section 2 provides a brief background on the Guatemala electricity sector. Section 2.3 describes the data —further explained in the Appendix. Then, Section 2.4 explains the empirical strategy and discusses how we tackle some potential issues as threats to identification. Section 3 explains the main results with additional robustness checks. Finally, Section 4 concludes.

2 Materials And Methods

2.1 Background

2.1.1 Recent History of Guatemalan Power Sector

In this section we provide some figures to put Guatemala (and its power sector) in context. According to the World Bank, its GDP per capita (PPP) grew 3.1% annually on average in 2000s and 3% in 2010s. Rural population access to electricity has also experienced a steady growth path, growing from 55% in 2000 to 93% in 2018. However, according to Census data, 77.7% of rural households are connected to the energy grid, 6% have solar panels, 12% use candle and 3.7% other sources of lightning. Almost half of population (46%) live in rural areas, and one third of labor force are farmers.

After the 1960-1996 civil war, there was an evident lack of infrastructure and services, particularly in the energy sector. In 1995, the electricity access was 88.9% in urban areas, while 39% for rural population. The outstanding growth in grid expansion begun in 1996 with the General Electric Law (LGE), which established a new scheme for electricity market based on liberalization and competition. In order to increase electrification rates, the LGE established the obligation for the utilities to connect households which were closer than 200 meters from any of their installation. Also, it allowed the Government to gather the necessary resources to expand the grid beyond that area (Iorio and Sanin, 2019). Therefore, 80% of the most important public utility firm (EEGSA) was sold in 1998, as well as the 80% of the other two large utility firms (DEOCSA and DEORSA). Part of the money obtained from these privatizations financed the grid expansion, especially in rural areas (Benavides and Dussan, 2004), where the main investments were done up to 2005. According to Iorio and Sanin (2019), 76% of new connections made between 1999-2014, were done during the first five years. In addition, according to Paz Antolín (2009), this grid network expansion was not accompanied by the necessary investments on the transportation line, resulting in a lower quality service.

The LGE also created new institutions that regulate and supervise the energy sector such as the National Commission of Electricity Energy (CNEE) and the Wholesale Market

Operator (AMM). CNEE is the Government agency in charge of ensuring the compliance of the LGE and its regulations, monitoring quality of the energy supplied, penalizing utility companies, and approving retail prices each term following the conditions approved in the Tariff Agreement (which is renewed every five years).

2.2 Generation and distribution of electricity

Guatemala generation relies mainly on renewable sources which follows a seasonal pattern. Hydro reaches more than 50% in rainy season (May to October), and biomass generation (mainly from sugarcane) is concentrated in first trimester. Figure 1 displays the evolution of installed capacity according to data from the AMM.

Although the load factor has been raising from 57% in 2001 to 70% in 2018, the installed capacity has been enough to fulfill its national consumption and even to export. In fact, Guatemala has been a net exporter over this period. That is, in contrast to some low-income countries, the lack of reliability is unlikely to arise due to generation constraints, but it is usually linked to issues in the distribution stage.

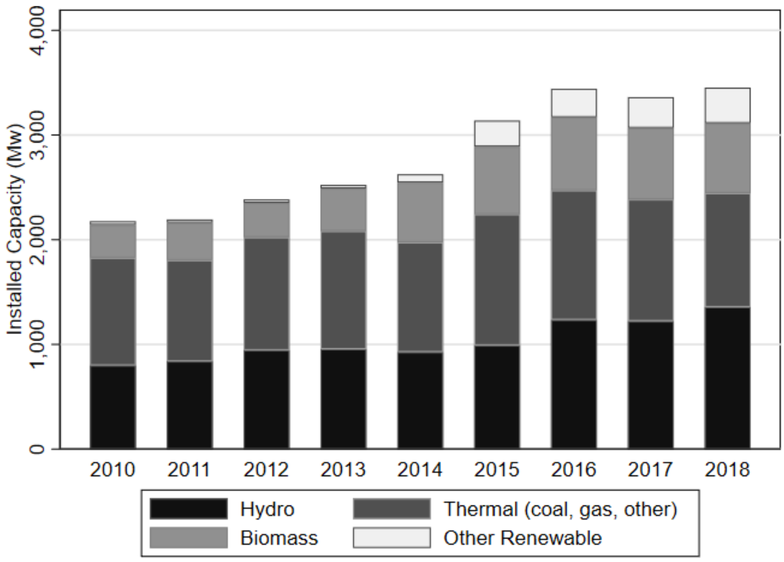

Figure 1: Evolution of installed capacity.

The distribution of electricity is mainly done by three large utility companies which have almost a zonal distribution. EEGSA (38% of total customers) provides to the densest urban area —Guatemala department—; DEOCSA (33%) and DEORSA (22%) provide to the West and East areas of the country, respectively. These last two firms supply an area where 92% of the Guatemalan rural population live.[4]

Figure 2 shows the average number of hours with outages in a semester. There is an upsurge of outages in DEOCSA and DEORSA area in 2012-2014, returning to its mean afterwards. This increment in outage hours coincides contemporaneously with a large number of service cut offs in Table 1. We exploit this quality variability over time and space —as different departments in rural Guatemala were unevenly affected— to estimate the impact of reliability on electrification, using the National Survey of Living Conditions (EN- COVI) waves 2011 and 2014.

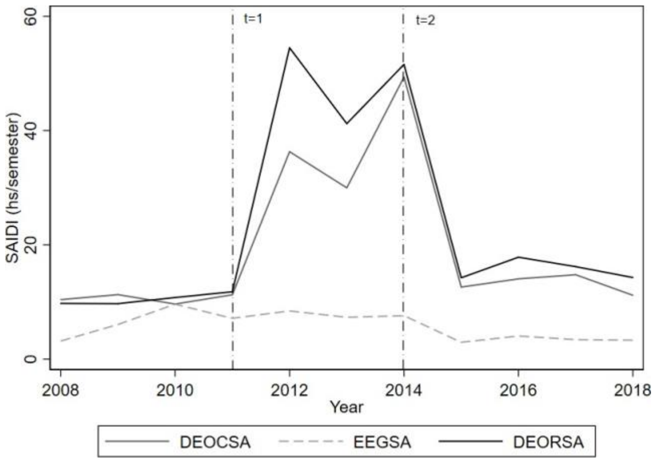

Figure 2: Evolution of the number of hours with power outages (SAIDI).

In addition to regulating minimum quality standards, CNEE also sets the electricity rates. They are based on an agreement between the government agency and each individual firm every five years. In this agreement, among other aspects, it is decided the way to update prices quarterly accordingly to input costs. CNEE announces publicly the new prices with a resolution. Thus, the final consumer price comprises four elements: a fixed and a variable

[4]In addition, there are sixteen small firms in some cities that provide energy only to its urban area. In 2018, they represented only 7% of low voltage consumers. See Appendix

component approved by CNEE, plus a Public Light Tariff (PLT) established by each of the 340 municipalities in Guatemala, and finally the value added tax.

In 2019, 94% of the Guatemalan households belonged to the residential category named Social Tariff, which includes households that consume less than 300kw/h per month. Additionally, this group could also be benefited by the so called "INDE contribution", which is a tiered subsidy up to 100kw/h.[5] In practice, the subsidy entails a maximum price *only for the variable component of the rate*. In other words, the difference between the Social Tariff and that price cap is paid by INDE. [6] Figure 6 shows the parallel evolution of Social Tariff and subsidies, entailing a price cap quite stable along time. According to official data, in December 2018 almost 70% of DEOCSA and DEORSA Social Tariff consumers received the INDE contribution. Since 70% of all DEOCSA customers live in rural zones, and 63% in DEORSA, we can infer that INDE contribution benefits nearly all rural households.[7]

Although electricity rates are homogeneous for rural consumers due to INDE contribution, the final bill is not. Dispersion across the country comes mainly from the way municipality collects the PLT. In EEGSA zone is charged an ad valorem tax, while in DEOCSA and DEORSA area there is a lump sum. Simulating a monthly 50kw/h consumption, the maximum bill (16.5 US$) is almost three times more than the minimum (5.6 US$). However, since we drop EEGSA municipalities, final bill is quite homogeneous between DEOCSA and DEORSA municipalities. More importantly, the PLT did not suffer important modifications across time.[8]

[5]The Instituto Nacional de Electrificación (INDE) is a public, autonomous and decentralized organization, which owns the main hydro plant of the country Chixoy (300Mw). Hydro plants profits let finance subsidies.

[6]For example, if a household consumes 150kw/h in a month, it pays the fixed cost plus the variable cost —in this case the Social Tariff—, and then it receives the INDE contribution for the first 100kw consumed. For instance, in November 2019, the scheme was: Q0.50 from 1-60kw/h; Q0.81 from 61-88kw/h. Simulating the bill for a hypothetical consumer in DEOCSA zone, assuming a consumption of 150kw/h, the final cost will be:

$$Q14.8 + 150 * (1.86) - [60 * (1.86 - 0.5) + 28 * (1.86 - 0.81)] = Q183$$

So far, it represents a 60% discount. Then, the VAT of 12% is added, and finally, the public lighting fee to final bill. The way to calculate the bill is available at: http://www.cnee.gob.gt/Calculadora/index.php

[7]The price cap has been Q 0.50 for customers that consumes between 1-50 kw/h per month. Although, the subsidy scheme has occasionally been changed (specially the scales), the lowest category has almost always had the same price. The poorest household are represented in that range. According to Centro de Investigaciones Económicas Nacionales (2015), 40% of Guatemalan families were in this category, and 30% between 51-100kw/h.

[8]CNEE gave us data on monthly PLT by municipality from 2015. Average PLT in 2015 was Q36.9 and in

Table 1: **Utilities main characteristics (2014)**

	EEGSA	**DEOCSA**	**DEORSA**
Total Consumers	1,108,352	975,717	598,550
% Rural	45%	70%	63%
Social Tariff Consumers	997,668	952,152	576,215
Per capita consumption (kw/h per month)	104.66	68.61	79.06
Social Tariff (Quetzal/kw/h)	1.63	2.02	1.92
Large Consumers	769	9	49
Compensation (Quetzals)	Q767,967	Q46,211,187	Q54,525,905
Services cut off (%)	6%	17%	22%
KvA installed per user	2.58	0.99	1.44

Notes: Services cut off is a proportion of total consumers. Compensation data is from 2013 as well as the proportion of rural consumers. Source: CNEE.

2.3 Data Description

Reliability is defined as "the attribute of energy supply that implies ability to draw energy when needed for use of energy services" (Bhatia and Angelou, 2015). One of the commonly metrics used for measuring quality is System Average Interruption Duration Index (SAIDI). In particular, the CNEE uses the following formula:

$$SAIDI_m = (\Sigma_{cm}\Sigma_j(Duration_{cm\,j}))/\Sigma_{cm},$$

which essentially captures the average duration of the lack of electricity supply for a customer c in municipality m. The duration of the outage is indexed by j. We have semi-annual data on rural areas from almost all municipalities (338 out of 340) in the time span of 2006 to 2018. To obtain quality service at municipality level, we get the mean of the two semesters observations over the year, and finally we get an unweighted mean for each department.[9] Figure 3 displays three departments with the best quality, and all of them belongs to EEGSA, which serves the capital city of Guatemala.

In addition to SAIDI evolution over time portrayed in Figure 2, Table 2 adds some

2018 Q35.2. Reading CNEE reports we found no evidence of an important change in PLT scheme between years 2011 and 2014. For simulation bills in different municipalities see Figure 7.

[9]Due to regulation requirements (CNEE Resolution 9/1999) quality of service supply must be measured twice a year. The normative settles some limits per consumer, establishing the right to be compensated if that limit is exceeded. The edge for duration is 6 for urban area, and 8 for rural. For more information of raw data see Appendix.

details regarding reliability performance in some districts. Over 2008/2018 period, DEOCSA failed to reach the legal maximum of SAIDI hours almost 5 years out of 11 in each municipality, and DEORSA over 5. This directly speaks on the heterogeneity of SAIDI performance in some districts in DEOCSA and DEORSA region.

Table 2: SAIDI descriptive statistics

	DEOCSA					DEORSA				
	Obs	Mean	Std	Min	Max	Obs	Mean	Std	Min	Max
SAIDI 2011 (hs/semester)	172	10.3	8.4	0.2	56.0	110	10.4	6.9	2.0	54.0
SAIDI 2014 (hs/semester)	176	41.4	29.2	5.6	164.9	113	45.3	23.7	12.8	137.7
SAIDI 2018 (hs/semester)	173	10.3	5.8	0.4	39.4	113	13.8	6.6	3.2	36.8
Frequency SAIDI >14hs (2008/18)	162	4.7	2.1	0	11.0	110	5.3	2.0	1.0	11
Frequency SAIDI >14hs (2015/18)	166	1.4	1.2	0	4.0	113	1.8	1.7	0	4.0

Note: Obs is the number of municipalities. SAIDI is an unweighted mean by firm. It is expressed in total hours by semester. Frequency is the number of times that SAIDI exceeds the quality regulation of 14hs in a municipality. We only consider municipalities that have no missing data over that period. Source: CNEE

Household information come from two sources, namely, the National Survey of Living Conditions (ENCOVI) —waves 2011 and 2014—, and 2018 Census database (INE, 2018). The geographic level of aggregation is at the department level in the former, and at the municipality level in the latter. The Census was conducted between July and August 2018, and collects detailed information about the universe of households in the country (3,275,931), such as dwelling characteristics, level of education, labor and migrant condition.

The output dummy variable of interest, *grid connection*, equals 1 if the head of the household answers "yes" to the closed question whether the house is connected to an electrical distribution network. In Census questionnaire is stated in another way, asking what type of lightning is mainly used at home. Any answer that is not "electricity" (e.g. gas, candle) is recorded as zero. Figure 3 shows that the East part of the country has lower values of electrification rate.

2.4 Empirical Strategy

2.4.1 Regression Model

We explore the impact of reliability on grid connection by exploiting variation in SAIDI over 2011-2014 period —coinciding with ENCOVI database— and then compare those results with those obtained using the 2018 Census. Therefore, the empirical strategy will

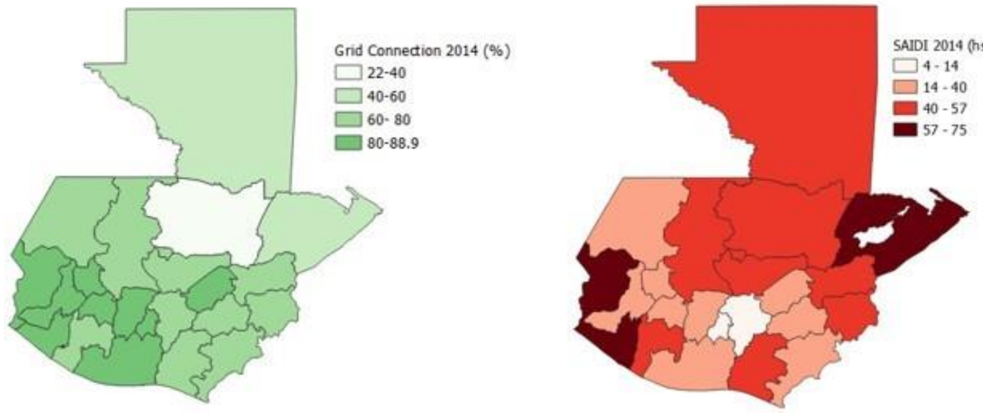

Figure 3: Rural electrification rate and SAIDI at department level (2014)

be twofold. First, we use the repeated cross section dataset (ENCOVI) at department level with time variation in a lineal probability model, which is as follows

$$Y_{hdrt} = \alpha_0 + \alpha_1 * LnSAIDI_{drt} + H_{hdrt} + D_{drt} + \eta_r + \theta_t + \varepsilon_{hdrt} \tag{1}$$

where subscripts *h, d, r, t* mean household, department, region and time respectively.[10] Y_{hdrt} is a dummy variable whether the household *h* is connected to the grid in department *d* in region *r* at time *t*. Our quality measure is the natural logarithm of SAIDI in department *d*, calculated as described above. H_{hdrt} is a set of control variables at the household level, D_{drt} are some characteristics at department level, η_r are region-specific dummies and θ_t denote year fixed effects. The interaction of both is included in some specifications, and finally ε_{hdrt} is the error term.

In equation (1), our parameter of interest is α_1 which measures the average effect of 1% increase in the lack of reliability on the probability of a rural household to be connected to the grid. The identification of this effect relies on the assumption that the 2012 change in SAIDI is orthogonal to households' electrification decision. Thus, we hypothesize thata reduction in quality —an increase in SAIDI— at department level, reduces the

[10] Guatemala is divided into 8 regions, 22 departments and 340 municipalities. Regions and departments are geographically divisions and administrative areas but with low political power. Because of geographical, economical or social reasons, departments are grouped into regions. Only Petén department is a region on itself.

expected benefits of electrification for households, resulting in a lower number of connections.

Next, in order to confer robustness to our results, we present a second regression model estimated using cross-sectional and more granular data. More precisely, we combine the Census database with CNEE records at municipality level to estimate the following regression model:

$$Y_{hmd} = \beta_0 + \beta_1 * LnSAIDI_{md} + H_{hmd} + M_{md} + \eta_d + \varepsilon_{hmd} \quad (2)$$

where M_{md} are some municipality-level characteristics in department d, η_d are department fixed effects, ε_{hmd} is the error term, and the rest of the variables are as defined above. In equation (2) β_1 captures the average effect of 1% increase in the lack of reliability on the probability of a rural household to be connected to the grid, at municipality level in 2018.

Different sets of control variables at *dwelling and household level* are considered. Regarding the dwelling unit, the following dummy variables are created: ownership, housing materials (i.e. thatched roof, metal sheet wall, or having no floor), shared house with another family, having some facilities (e.g. kitchen, own toilet access). At household level we control for remittances reception, whether any household member has ever emigrated; and whether the household is overcrowded.[11] Besides, we have income data from ENCOVI as well as some household´s assets (e.g. radio, motorbike).[12] Finally, we include dummy variables that give information of the head of household, sex, age, level of education, marital status, spoken language, ethnicity and type of job (e.g.: self-employer, farmer, retailer).

Furthermore, to capture socioeconomic conditions at *municipality or depart- ment level* we add some mean characteristics in their rural area: employment, literacy and schooling rate.[13] We also control for the proportion of households with a child between

[11]The usual standard to define overcrowding is when more than three people sleep in a room. This measure is the ratio between all the members of the household and the number of rooms, not including the kitchen. However, ECLAC warns that for some indigenous culture, many people are used to sleep in large rooms, so overcrowd could be overestimated.

[12]We deflate total familiar income, and then we transform it by the inverse hyperbolic sine. We make this transformation –instead of logarithm– in order to avoid losing zero income observations.The coefficient interpretation is not income elasticity, but the sign provides information if it is a normal or an inferior good. See Bellemare and Wichman (2020)

[13]Literacy is defined as people older than 15 who knows reading and writing. Schooling rate is defined as: kids from 4 to 14 years who attends the school, divided total population of kids 4-14.

7 and 12 years old working, has its own toilet, has water access, owns a motorbike, and has poor dwelling conditions as already defined.

In addition to these socioeconomic variables, we include Cooling Degree Days (CDD) as a demand driver and Public Lightning Tariff.[14] Due to the fact that PLT data is available since 2015 at municipality level, we incorporate it in Census regression model.

Full descriptive summary statistics are available in the Appendix, Tables 11 and 12. Although similar in SAIDI levels and in socioeconomic variables, regions supplied by DEOCSA and DEORSA are quite different in two aspects: more indigenous ethnic composition in the West, and better access to some other facilities besides grid connection (e.g. water pipe access). Finally, although average total familiar income is larger in the East, more households are considered in "extreme poverty" reflecting more inequality in that region.

2.4.2 Main Identification Assumptions

Our objective reliability measure (SAIDI) is in principle an endogenous variable. However, Figure 2 gives evidence of a potential exogenous shock *only* to DEOCSA and DEORSA area. This variation in time and space (across municipalities), if exogenous induced, could let us find an unbiased estimation of α_1 in equation (1).

In terms of what reasons could be behind this sudden change, we are inclined to think it was due to insufficient transmission infrastructure and managerial problems. DEOCSA and DEORSA were sold from Union Fenosa to Actis Group in May 2011 and later resold in 2016 to ENERGUATE (the current owner of both firms). In addition, the proximity of Tariff Agreement expiration with CNEE in 2014, would have probably discouraged important investments. This hypothesis is in line with anecdotal evidence from the local press, which stated there had not been investments neither from Union Fenosa in its last years nor from Actis.[15] Therefore, it seems reasonable to think that the sudden change in quality in 2012 it is not linked in principle, to demand factors.

[14] Some part of the literature includes weather conditions that push demand, such as cooling or heating degree days (e.g.Allcott et al. (2016)). In the Guatemalan context, the demand of electricity for heating is useless in many parts of the country due not only for tropical weather, but also for the widespread use offirewood.

[15] See Estrategia y Negocios (2016). The highest amount of compensation (US$14 million) imposed by CNEE for different kind of infractions was in 2013, and almost 90% corresponded to these two firms.

Nevertheless, there are other plausible explanations. First, an unforeseen increase in grid-connected consumers entailing an endogenous shift in quality. Second, an insufficient energy supply in long-term contracts that had forced firms to buy in the spot market at higher prices. Third, firms were getting into a "vicious circle" of bad quality and not payment behavior conducting to financial constraints, or in some cases electricity theft.

The first potential cause seems not to have anchor in data. The total number of customers (rural and urban) increased between 2007-2011, at a different path along firms: EEGSA experienced an annually growth rate of 3.1%, meanwhile DEOCSA and DEORSA were both 2.1%. These growth rates are lower in comparison with the later period 2015-2018: DEOCSA (2.3%) and DEORSA (3.4%).

Secondly, although the need of buying energy at a more expensive spot market would have affected all firms similarly, EEGSA did improve during those years, while DEOCSAand DEORSA did not. As tariffs could be updated and, more importantly, they did not implya larger rural household spending because of the subsidy scheme, it should not affectfirms' financial health. Therefore, and in relation with the third cause, if families did notpay their bills, the main reason of nonpayment would have not probably come from the priceside, but from the quality one. In addition to the lack of payment, which can be deducedfrom the high number of disconnections in Table 1, there is also evidence of an increase in illegal connections MEM (2014). Both are potential outcomes of an unsatisfactory service.

We do not include in the main regression setup households in EEGSA service zone for many reasons. First and foremost, because the shock was in DEOCSA and DEORSA area. Second, in ENCOVI dataset we do not have information on household's municipality, and there are two municipalities in a mainly EEGSA department that are supplied by DEOCSA. Therefore, it is impossible to distinguish. Finally, 92% or rural population live in the areasupplied by DEOCSA and DEORSA. However, as robustness check, we include them in

the 2018 Census.

2.4.3 Potential Threats to Identification

So far, we provide several arguments that lead us to believe that the sudden variation in SAIDI in 2012-2014 is not driven by demand-related factors. Still, one might be concern of a potential endogeneity problem. In order to mitigate this concern, we provide additional evidence using IV regressions with rainfall as instrument.[16] Previous studies (e.g. Cole et al. (2018)) rely on the assumption that more rainfall means more water disposal to hydrogeneration and, therefore, more electricity availability.

However, we argue the opposite in the rural Guatemalan context. One reason is that the country can substitute different fuels to generate electricity in case there is not enough hydro generation (e.g.in the dry season). Also, when heavy rains come (frequently together with strong winds, like in tropical storms), muddy or flooded roads makes maintenance tasks complicated, and power outages sometimes appear even before the storm starts a precaution measure in rural areas.

Regarding the exclusion restriction, it could be argued that in wet areas where there are less possibilities to do outdoor activities, and therefore, watching TV becomes a potential leisure activity. Thus, rainfall could also be affecting households' decision to connect to the grid, driving more demand in those areas. However, Guatemala data shows that 35% of electrified rural houses do not have a TV, so this concern seems to be a second order issue.

To build the rainfall variables for the IV strategy, we gathered weather data from the national weather agency of Guatemala (INSIVUMEH) and the Statistics Institute (INE), which provides information of 49 weather stations that belongs to the studied area. We assign each municipality to the closest Weather Station, and then we aggregate data into department level for ENCOVI regression.

As regards equation (2), the empirical strategy is different. Census dataset is cross-sectional, without time varying information, and we can not exploit any shift in quality nor in rainfall. Besides, our former measure of rainfall is a weak instrument at municipality

[16] The way literature exploits exogenous variation on quality is varied. For the sake of simplicity: lightning density (Andersen and Dalgaard, 2013), a river-flow modelling and its impact on hydro-power generation (Cole et al., 2018), temperature (Fisher-Vanden et al., 2015), lightning activity and distance to the closest generator (Millien, 2017)

level because is difficult that only 49 weather stations could represent the variability needed for 340 municipalities.[17]

Therefore, we follow Dang and La (2019) and Sedai et al. (2021) procedure, which use as instrumental variable the average quality at district level obtained from the rest of the districts that belongs to the same department. In our context, the instrument z would be:

$$z_{md} = \frac{1}{S}\sum_{j=1}^{S} SAIDI_j, \quad \text{being} \quad j \neq m, m \notin S, m \in D \quad \text{and D=1,2,.....S,m} \tag{3}$$

where z_{md} is the average SAIDI in municipality m in department d, using SAIDI information from all districts that belongs to the same department d, except municipality m. The exogeneity condition is expected to hold because quality in other municipalities does not directly affect the household decision to grid connect in its own district.[18]

Finally, one could also be concerned of the aforementioned "off-grid" households. There are rural settlements so far away from grid lines, that having a grid connection is not feasible. Although the probability of having such isolated households is very low in ENCOVI dataset, it is not the case for Census where the entire population is surveyed. Thus, we estimate our main regression in Census model after dropping rural households that declare having solar panels (6% of rural households), because they are most likely too far from the power grid. Then, as an additional robustness check we re-include them in Section 3.2 and main results hold.

3 Results

3.1 Main Results

The main results obtained using the ENCOVI regression —equation (1)— are presented in Table 3. The four columns reflect different estimated models, considering different subsets of controls. Robust standard errors are clustered at Primary Sample Unit level.[19]

[17]For an in-depth discussion for the instrumental variable, see Appendix.

[18]Although it could be argued that families could move to municipalities that have better facilities, the richness of our database allow us to control for migration condition.

We find a negative relationship between SAIDI and the probability of a rural household to get connected in columns (3) and (4). The magnitude varies from 14p.p. to 19p.p. in our preferred and most complete specification. In column (4) we have additional controls at household and department level, year and region fixed effects and the interaction between them. If quality decreases 1%, the probability of connection goes down by 18.6 percentage points. In the context of 2011-2014 period, this means that reducing SAIDI at department level by almost 30 minutes per year, will increase an expected number of 54,325 new customers, representing 1.3 US$ million dollars in annual revenues.[20]

The rest of the control variables show the expected sign. Those correlated with wealth (e.g. dwelling materials, income) are positively correlated with a higher probability to have grid connection. In addition, the positive sign of CDD implies that, conditional on the rest of control variables, those departments where the climate is hotter, there is a relatively higher demand in power grid connections.

Despite the fact power grid connection raises in rural areas between 2011-2014 —see Table 11—, these results provide evidence that this growth would have been larger if quality had been better. In other words, rural grid connection increased in that period as well as quality declined. However, other variables positive correlated with grid connection also improved in this period, such as education level and income (e.g. real GDP growth was on average 3.6% in 2011-2014).

Having exploited spatial and time variation in the ENCOVI regression setup, we now complement them with those obtained using more granular data from the Census. Results of this second model specifications are resumed in Table 4, where robust standard errors are clustered at the municipality level.

[19]Primary Sample Units (PSU) are random selected areas. Each one has its own weight to expand results to the whole population. In ENCOVI we can either cluster at PSU level or department level. Since having only 19 clusters is too few, we estimate Wild cluster bootstrap standard errors (see Cameron et al. (2008) and Cameron and Miller (2015)). Results are robust to both methods. Although clustering at department-year level would have doubled the number of clusters, we would be assuming that quality -or rainfall in the IV case- are not autocorrelated, which could be a strong assumption.

[20]In 2011-2014, on average SAIDI was 26.9 hours by semester or 54 hours per year. Also, on average, 34 out of 100 did not have grid connection, so improving 18.6pp means: 0.34*0.186= 0.063, six out of 100 unconnected households will be expected to connect, representing 54,325 new households. At the end of 2014, with a monthly fix fee of 15 quetzales per customer represented, it gives approximately 23.7 dollars annually per customer in revenues.

Table 3: The impact of SAIDI on Grid Connection (ENCOVI 2011-2014)

	Outcome: Grid connection			
	(1)	(2)	(3)	(4)
Ln SAIDI	-0.150***	-0.071*	-0.135**	-0.184**
	(0.050)	(0.043)	(0.062)	(0.087)
Cooling Degree Days	-0.017***	-0.011***	0.009	0.013**
	(0.004)	(0.004)	(0.006)	(0.006)
Poor housing materials	-0.346***	-0.272***	-0.201***	-0.201***
	(0.016)	(0.015)	(0.014)	(0.014)
Real income per capita (asinh)	0.017***	0.013***	0.007***	0.007***
	(0.003)	(0.003)	(0.002)	(0.003)
Year FE	Yes	Yes	Yes	Yes
Dwelling characteristics	No	Yes	Yes	Yes
Household controls	No	No	Yes	Yes
Department controls	No	No	Yes	Yes
Region FE	No	No	Yes	Yes
Region FE* Year FE	No	No	No	Yes
Observations	12,940	12,914	12,825	12,825
Adjusted R^2	0.174	0.235	0.320	0.322

Dwelling specific controls include: ownership, shared house, connected to water grid, own toilet access, own kitchen, overcrowd. Households' controls comprise if household receives remittances and Head of Household variables: gender, recent migrant (5 years), age, labor, marital status, spoken language, ethnic and education. Robust clustered standard errors at PSU level (1,138)

* p<0.10, ** p<0.05, *** p<0.01

Table 4: The impact of SAIDI on Grid Connection (Census, 2018)

	Outcome: Grid connection			
	(1)	(2)	(3)	(4)
Ln SAIDI	-0.0508**	-0.0536***	-0.0228*	-0.0245*
	(-3.03)	(-3.84)	(-2.11)	(-2.34)
Cooling Degree Days	0.00733*	0.00285	0.00475	0.00576
	(2.45)	(1.07)	(1.79)	(1.71)
Public Light Tariff	-0.00153	-0.00196**	-0.000902	-0.00105
	(-1.88)	(-2.71)	(-1.61)	(-1.91)
Asset: Motorbike	0.0640***	0.0467***	0.0457***	0.0416***
	(9.53)	(8.85)	(11.30)	(10.93)
Asset: Radio	0.135***	0.126***	0.113***	0.111***
	(21.07)	(22.28)	(23.24)	(23.11)
Dwelling characteristics	Yes	Yes	Yes	Yes
Head of household controls	No	Yes	Yes	Yes
Municipality controls	No	No	No	Yes
Department FE	No	No	Yes	Yes
Adjusted R^2	0.223	0.249	0.283	0.288
Observations	1,178,160	1,171,243	1,171,243	1,171,243

Dwelling specific controls include: ownership, shared house, connected to water grid, own toilet access, housing materials, own kitchen, overcrowd. Head of household controls comprise: gender, recent migrant, age, labor, marital status, spoken language, ethnic, education, and if there is any kid working in the house. Robust clustered standard errors at municipality level (266)

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

First, as it was also the case in the ENCOVI regression, all variables correlated with income —e.g. assets— do also have a positive impact on the probability to have a grid connection. Nevertheless, two variables at municipality level show no effect: Cooling Degree Days and Public Light Tariff. The effect of SAIDI remains being statistically significant and negative. Now, 1% of reduction in SAIDI (15 minutes annually on average at municipality level), raises the probability of getting connected in 2.5 p.p., representing 20,489 new rural household connections. Although at first glance one could think that 15 minutes per year and its effect is meaningless, we should have in mind that quality level has improved significantly in comparison with ENCOVI period, which could lead us to think that *the level* of quality matters. Although results are robust to both regression setups and database, some additional estimates should be performed.

3.2 Robustness Checks

We now estimate equations (1) and (2) by 2SLS. Each column of Table (5) and Table (6) replicates the same model as in Table (3) and Table (4) respectively. We test the strength of the instrument using Kleinbergen-Paap F-Statistic.

In the first three columns of Table (5) we can not reject the null hypothesis of being a weak instrument. Once we include the interaction term between Region and Year, we have evidence rainfall is an adequate instrument. Column (4) estimates that 1% reduction in SAIDI increases 27.6 percentage points the probability of a rural household to connect to the grid. This estimation is 9 p.p. larger than OLS. In terms of new connections, this would have represented 80,741 new rural connections and 1.9 US$ million dollars in annual revenues.

Meanwhile, F-test in Table 6 shows there is no evidence to support rainfall as an instrument. The less variation across municipalities and, on top of everything, the loss of variation on time, make rainfall an inadequate instrument. However, using z_{md} from equation 3 as instrument in Columns 3 and 4, the 2SLS estimation gives account of a negative effect of SAIDI —3% on average— on households' willingness to connect to the grid. Our preferred model specification (model 4) presents an estimation 0.5 p.p.

larger than OLS in Table 4.

Table 5: The impact of SAIDI on Grid Connection (2SLS estimation-ENCOVI)

	Outcome: Grid connection			
	(1)	(2)	(3)	(4)
Ln SAIDI	-0.670	-0.320	-0.234**	-0.275***
	(0.506)	(0.463)	(0.096)	(0.049)
Cooling Degree Days	0.001	-0.003	0.009**	0.014***
	(0.020)	(0.016)	(0.004)	(0.003)
Poor housing materials	-0.328***	-0.270***	-0.201***	-0.201***
	(0.030)	(0.025)	(0.016)	(0.016)
Real income per capita (asinh)	0.015**	0.013**	0.007*	0.007*
	(0.006)	(0.005)	(0.004)	(0.004)
Year FE	Yes	Yes	Yes	Yes
Dwelling characteristics	No	Yes	Yes	Yes
Household controls	No	No	Yes	Yes
Department controls	No	No	Yes	Yes
Region FE	No	No	Yes	Yes
Region FE* Year FE	No	No	No	Yes
Adjusted R^2	0.119	0.223	0.320	0.321
F_test	3.872	3.898	8.587	20.492
Observations	12,940	12,914	12,825	12,825

In parenthesis Robust Standard Errors clustered at department level. Results remain with Wild Bootstrapped Robust standard errors with 400 replications clustered at department level. F-statistic is the heteroskedasticity and cluster robust Kleibergen-Paap weak instrument test. With a 5% level of confidence and a potential bias of 10%, the instrument is not considered weak if F>16, see Stock and Yogo (2005). M1-M4 means models 1 to 4 of of Table 3. Instrumental variable: rainfall from Weather Stations.
* p<0.10, ** p<0.05, *** p<0.001

In addition to 2SLS estimation, we perform some robustness checks regarding ENCOVI and Census regression setups. Firstly, from ENCOVI we should be aware of the method of aggregation used to calculate SAIDI at department level. So far, SAIDI has been calculated as a simple average without considering the population in each municipality. Therefore, we construct an alternative variable: $SAIDI_w$. It is a weighted average by the number of households a municipality has, according to 2018 Census. With this new measure, on the one hand, we could have a more representative SAIDI measure at department level. On the other hand, we could be giving a double —and possible wrong— weight to each household observation, since the only geographical information provided by ENCOVI is from which department household belongs to, but not the municipality. The ENCOVI, as any survey, has its weights to extrapolate results to whole population, in this case to rural. Results are in Table 7, and SAIDI remains being significant. In comparison with Table 3, point estimations are different when we do not control for year FE.

Table 6: The impact of SAIDI on Grid Connection (2SLS Estimation-Census)

	Outcome: Grid connection			
	(1)	(2)	(3)	(4)
Ln SAIDI	-0.010	0.043	-0.033*	-0.030**
	(0.199)	(0.379)	(0.017)	(0.014)
Cooling Degree Days	0.005	0.004	0.005*	0.006*
	(0.003)	(0.007)	(0.003)	(0.003)
Public Light Tariff	-0.001	-0.001	-0.001	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)
Asset: Motorbike	0.046***	0.045***	0.046***	0.042***
	(0.004)	(0.006)	(0.004)	(0.004)
Asset: Radio	0.113***	0.112***	0.113***	0.111***
	(0.005)	(0.006)	(0.005)	(0.005)
Household controls	Yes	Yes	Yes	Yes
Department Fixed effects	Yes	Yes	Yes	Yes
Municipality controls	No	No	No	Yes
Adjusted R^2	0.283	0.277	0.283	0.288
F_test	0.526	0.214	143.929	163.892
Observations	1,171,243	1,171,243	1,171,243	1,171,243

In parenthesis Robust Standard Errors clustered at municipality level. Results remain with Wild Bootstrapped Robust standard errors with 400 replications clustered at department level. F-statistic is the heteroskedasticity and cluster robust Kleibergen-Paap weak instrument test. With a 5% level of confidence and a potential bias of 10%, the instrument is not considered weak if F>16, see Stock and Yogo (2005). Columns one to four are the same models as Table 4. Instrumental variable: rainfall for columns 1 and 2, and quality as defined in equation (3). Column 1 is Model 3 of Table 4 assigning the 45 Weather Stations rainfall registers to closest municipality. Column 2 is the same model, but the way of estimating rainfall for each municipality is by interpolating values from the WS. Columns 3 and 4 are models 3 and 4 of Table 4 using the instrument variable as in equation (3)

* p<0.10, ** p<0.05, *** p<0.001

Table 7: The impact of SAIDI on Grid Connection (using $SAIDI_{weighted}$). ENCOVI (2011-2014)

	(1)	(2)	(3)	(4)
Ln $SAIDI_{weighted}$	-0.125***	-0.074***	-0.071***	-0.154***
	(0.017)	(0.016)	(0.025)	(0.038)
Cooling Degree Days	-0.018***	-0.011***	0.009***	0.014***
	(0.001)	(0.001)	(0.003)	(0.003)
Poor housing materials	-0.348***	-0.272***	-0.201***	-0.201***
	(0.008)	(0.008)	(0.008)	(0.008)
Real income per capita (asinh)	0.016***	0.013***	0.007***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Year FE	Yes	Yes	Yes	Yes
Dwelling characteristics	No	Yes	Yes	Yes
Household controls	No	No	Yes	Yes
Department Controls	No	No	Yes	Yes
Region FE	No	No	Yes	Yes
Region FE* Year FE	No	No	No	Yes
Adjusted R^2	0.173	0.235	0.320	0.322
Observations	12,940	12,914	12,825	12,825

Robust clustered standard errors at PSU level (1,138). In this table all model specifications are the same as in Table 3. Results remain if clustering is at department level and Wild bootstrap is performed with 400 replications.
* p<0.10, ** p<0.05, *** p<0.01

Secondly, Table 8 shows some robustness check with Census dataset. In the first column we check if the aforementioned exclusion of households with solar panels could be affecting the results in Census, and main result remains. Then, because all districts performed very similarly in 2018 (see Figure (2)), we include EEGSA zone. Including thosedistricts do not change the negative impact. Finally, in the last two columns we perform twoplacebo tests, checking if power reliability affects some variables that beforehand should not. Results confirm that quality supply would be affecting only grid connection.

Thirdly, one concern may be the presence of one department so different from the rest that could be misleading the results. For example, Alta Verapaz has almost the lowest quality level and is by far the department with lowest electrification rate. Table 9 shows the estimation results excluding this department. As expected, all regressions have lower point estimations with the only exception of column 2, which in fact is 7 p.p. larger. This could be due to interaction term of region and year, that can be capturing a relative better performance of this department between 2011 and 2014.

Table 8: Robustness Check: Census

	(1) Grid connection	(2) Grid connection	(3) Water grid	(4) Garbage
Ln SAIDI	-0.030**	-0.018**	-0.011	-0.006
	(0.013)	(0.009)	(0.019)	(0.011)
Public Light Tariff	-0.001*	-0.001	0.005***	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Head of household variables	Yes	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes	Yes
Department Fixed effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.313	0.286	0.198	0.183
Observations	1,253,405	1,286,060	1,171,243	1,171,243

Robust clustered standard errors at Municipality level. Model (1) includes households with solar panel. Model (2) includes EEGSA districts. PLT for these districts were estimated using a monthly consumptionof 50Kw/h. Model (3) and (4) are placebo tests, using connection to water grid and garbage collection asoutcome dummy variables.

* p<0.10, ** p<0.05, *** p<0.01

Table 9: Regressions without Alta Verapaz

	Outcome: Grid connection			
	ENCOVI		Census	
Ln SAIDI	-0.115*	-0.263**	-0.018**	-0.020**
	(0.065)	(0.103)	(0.009)	(0.009)
Year Fixed effects	Yes	Yes		
Dwelling characteristics	Yes	Yes		
Household controls	Yes	Yes		
Department Controls	Yes	Yes		
Region Fixed effects	Yes	Yes		
Region FE * Year FE	No	Yes		
Department Fixed effects			Yes	Yes
Municipality Controls			No	Yes
Adjusted R^2	0.251	0.252	0.210	0.215
Observations	12,153	12,153	1,046,385	1,046,385

Columns 1 and 2 corresponds to model 3 and 4 of Table 3. Columns 3 and 4 corresponds to models 3 and 4 of Table 4. Robust clustered standard errors at PSU level (1,079) in columns 1 and 2. Robust clustered standard errors at municipality level (249) in columns 3 and 4

* p<0.10, ** p<0.05, *** p<0.01

4 Conclusion and discussion

In this paper we study the relationship between reliability in power supply and rural households grid connection in Guatemala. Taking advantage of the main attributes of two different household dataset, combined with a unique and objective quality data at municipality level for a time span of ten years, we find supported evidence that there is a positive effect between quality and grid connection.

Considering quality and reliability as synonyms, we use the System Average Interruption Duration Index (SAIDI) as the reliability measure. In the first regression setup for 2011-2014 period, we find that a 1% reduction in outages duration at *department level*, increases probability of grid connection between 18-27 percentage points. In the second regression setup, with a better and more stable quality level in 2018, a 1% reduction in outages duration at *municipality level* increases probability in 3 percentage points. Results are robust to different model specifications and robustness checks.

The interpretation of these results could be that not only good quality is a way of getting new connections, but also not losing existing ones. From a policy perspective, these means that grid network extension should not be done without the guarantee of achieving a minimum level of quality. As Chaurey and Le (2022) points out, rural infrastructure and maintenance are critical for rural electrification and connectivity, which are potential drivers for microenterprise growth. Besides, considering Grogan´s (2018) results, where rural Guatemalan women increased their work paid hours by 2 hours per day thanks to rural electrification, it is not difficult to realize the importance of reliability for reducing gender disparities and providing more opportunities for all.

Further potential research questions arise at the end of this paper. Does historical performance play a role in actual levels of electrification rates? Does reliability affect households in a different way according to their historical connection status? And specially from a public policy perspective: is it the lump sum Public Tariff a barrier to electrification? Can we estimate the regressive effect that it has on income distribution, given that the vast majority of rural households consume less than 50kw/h per month? Although some of our model specifications could be giving some evidence of its counteracting effect, we do not have enough data to assure that. Some of the poorest municipalities have the highest tariffs, perhaps because there are not enough connections for maintaining the system or, inversely, they are few because it is expensive. What is clear, however, is that the actual system of PLT combined with INDE contribution that tries to help poorest household, loses efficiency with this kind of lump sum scheme.

Not only are they potential barriers to electrification, but they are also a spoke in the wheel of development for many people and regions. While some areas continue to grow, others lag with all the potential problems that entail.

References

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

Andersen, T. B. and C.-J. Dalgaard (2013). Power outages and economic growth in africa. *Energy Economics 38*, 19 – 23.

Arraiz, I. and C. Calero (2015). From candles to light: the impact of rural electrification. Technical report, IDB Working Paper Series.

Bajo-Buenestado, R. (2021). The effect of blackouts on household electrification status: Evidence from kenya. *Energy Economics 94*, 105067.

Bayer, P., R. Kennedy, J. Yang, and J. Urpelainen (2020). The need for impact evaluation in electricity access research. *Energy Policy 137*, 111099.

Bellemare, M. and C. Wichman (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics 82*(1), 50–61. cited By 37.

Benavides, J. and M. I. Dussan (2004). Economía política de las finanzas y subsidios del sector eléctrico de Guatemala. Technical report, Banco Interamericano de Desarrollo.

Bhatia, M. and N. Angelou (2015). Beyond connections: Energy access redefined. Technical report, ESMAP Technical Report;008/15. World Bank.

Bonan, J., S. Pareglio, and M. Tavoni (2017). Access to modern energy: a review of barriers, drivers and impacts. *Environment and Development Economics 22*(5), 491–516.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *The Journal of Human Resources 50*(2), 317–372.

Centro de Investigaciones Económicas Nacionales (2015). Diagnóstico y propuestas para infraestructura sector eléctrico, sector portuario y aeroportuario. https://cien.org.gt/wp-content/uploads/2018/09/Sector_Eectrico_Portuario_y_Comercio.pdf. Accessed: 2021-26-3.

Chakravorty, U., M. Pelli, and B. Ural Marchand (2014). Does the quality of electricity matter? Evidence from rural India. *Journal of Economic Behavior and Organization 107*(PA), 228–247.

Chaurey, R. and D. T. Le (2022). Infrastructure maintenance and rural economic activity: Evidence from India. *Journal of Public Economics 214*, 104725.

Cole, M. A., R. J. Elliott, G. Occhiali, and E. Strobl (2018). Power outages and firm performance in sub-saharan africa. *Journal of Development Economics 134*, 150 – 159.

Dang, D. A. and H. A. La (2019). Does electricity reliability matter? Evidence from rural Viet Nam. *Energy Policy 131*, 399–409.

Dzansi, J., S. L. Puller, B. Street, and B. Yebuah-Dwamena (2018, October). The Vicious Circle of Blackouts and Revenue Collection in Developing Economies: Evidence from Ghana . *Working Paper. International Growth Centre. E-89457-GHA-1*.

Estrategia y Negocios (2016, April). Actis vende energuate por us$299,5 millones. Retrieved from https://www.estrategiaynegocios.net/inicio/916595- 330/actis-vende-energuate-por-us2995-millones, on 2020-7-24.

Fisher-Vanden, K., E. T. Mansur, and Q. J. Wang (2015). Electricity shortages and firm productivity: Evidence from China's industrial firms. *Journal of Development Economics 114*, 172–188.

Grainger, C. A. and F. Zhang (2019). Electricity shortages and manufacturing productivity in Pakistan. *Energy Policy 132*(May), 1000–1008.

Grimm, M., L. Lenz, J. Peters, and M. Sievert (2020). Demand for off-grid solar electricity:

Experimental evidence from rwanda. *Journal of the Association of Environmental and Resource Economists 7*(3), 417–454.

Grogan, L. (2018). Time use impacts of rural electrification: Longitudinal evidence from guatemala. *Journal of Development Economics 135*, 304–317.

Hashemi, M. (2021). The economic value of unsupplied electricity: Evidence from Nepal. *Energy Economics 95*, 105124.

Huffman, J. et al. (2019). Gpm imerg final precipitation l3 1 month 0.1× 0.1 degree v06, greenbelt, md, goddard earth sciences data and information services center (ges disc). *NASA*. Accessed: March 1, 2022. url= 10.5067/GPM/IMERG/3B-MONTH/06.

INDE (2019). Aporte social del inde al mes de enero de 2019. Technical report, Instituto Nacional de la Energía.

INE (2018). Glosario XII Censo Nacional de Población y VII de Vivienda. Technical report, "Instituto Nacional de Estadística (INE)".

Iorio, P. and M. E. Sanin (2019). Acceso y asequibilidad a la energía eléctrica en América Latina y El Caribe. Technical report.

Kennedy, R., A. Mahajan, and J. Urpelainen (2019). Quality of service predicts willingness to pay for household electricity connections in rural India. *Energy Policy 129*, 319–326.

Lee, K., E. Brewer, C. Christiano, F. Meyo, E. Miguel, M. Podolsky, J. Rosa, and C. Wolfram (2014). Barriers to electrification for "under grid" households in rural Kenya. Technical report, National Bureau of Economic Research.

McNally, A. et al. (2018). Fldas noah land surface model l4 global monthly 0.1× 0.1 degree (merra-2 and chirps). *Atmos. Compos. Water Energy Cycles Clim. Var*.

MEM (2014). Robo de energía rebasa los Q250 millones al año. Accessed October 1, 2022. https://mem.gob.gt/blog/robo-de-energia-rebasa-los-q250-millones-al-ano/ .

Millien, A. (2017, June). Electricity supply reliability and households decision to connect to the grid. Documents de travail du Centre d'Economie de la Sorbonne 2017.31 - ISSN : 1955-611X.

Ministerio de Energía y Minas (2019). Estadísticas del subsector eléctrico 2018. http://www.mem.gob.gt/wp-content/uploads/2019/01/EstadÃŋsticas-Subsector-ElÃľctrico-1.pdf.

Oseni, M. O. and M. G. Pollitt (2015). A firm-level analysis of outage loss differentials and self-generation: Evidence from African business enterprises. *Energy Economics 52*, 277–286.

Parikh, J. (2011). Hardships and health impacts on women due to traditional cooking fuels: A case study of Himachal Pradesh, India. *Energy Policy 39*(12), 7587–7594. Clean Cooking Fuels and Technologies in Developing Economies.

Paz Antolín, M. J. (2009). Efectos de la expansión de empresas transnacionales en el sector eléctrico en guatemala. *Problemas del Desarrollo. Revista Latinoamericana de Economía 35*(137).

Sedai, A. K., R. Vasudevan, A. A. Pena, and R. Miller (2021). Does reliable electrification reduce gender differences? Evidence from India. *Journal of Economic Behavior Organization 185*, 580–601.

Smith, B. and S. Wills (2018). Left in the dark? oil and rural poverty. *Journal of the Association of Environmental and Resource Economists 5*(4), 865–904.

Smith-Sivertsen, T., E. Díaz, D. Pope, R. T. Lie, A. Díaz, J. McCracken, P. Bakke, B. Arana, K. R. Smith, and N. Bruce (2009, 05). Effect of Reducing Indoor Air Pollution on Women's Respiratory Symptoms and Lung Function: The RESPIRE Randomized Trial, Guatemala. *American Journal of Epidemiology 170*(2), 211–220.

Stock, J. H. and M. Yogo (2005). *Testing for Weak Instruments in Linear IV Regression*, pp. 80–108. Cambridge University Press.

World Bank (2021). World Bank Data. Accessed on 4-3-2021.

World Meteorological Organization (2017). *WMO Guidelines on the Calculation of Climate Normals*. WMO-No. 1203. Accessed: 2021-29-3.

5 Appendix

5.1 Quality data and descriptive statistics.

Approximately 180 rural districts are supplied mainly by DEOCSA and 115 by DEORSA. Since 2011, seven districts have been created separating from a larger one. For ENCOVI regression setup, it is not an issue since they all belong to the same department. There is only one of the recent created municipalities that CNEE has no data: Petatán. Nevertheless, it represents less than 0.01% of observations since there are only 752 rural households in Census.

When merging CNEE and Census data, there are 21 municipalities –17 departmental capitals– that Census considers only urban. Table 10 resumes the original CNEE data.

Table 10: SAIDI raw statistics from CNEE

	DEOCSA				**DEORSA**			
	N° Different districts			**Mean SAIDI**	**N° Different districts**			**Mean SAIDI**
Year	Sem 1	Sem 2	Year	(Hs/semester)	Sem 1	Sem 2	Total	(Hs/semester)
2006	132	140	164	4.03	80	86	97	5.16
2007	119	123	152	6.11	89	1	91	7.67
2008	174	174	176	9.05	109	106	111	8.80
2009	169	161	176	9.75	109	97	111	9.10
2010	173	166	177	8.50	109	100	111	9.65
2011	164	158	172	10.31	109	100	111	10.57
2012	177	177	176	32.30	110	111	113	44.97
2013	177	177	176	26.33	111	111	113	36.30
2014	177	176	176	41.38	111	112	114	45.23
2015	169	165	175	11.47	112	111	114	12.30
2016	173	173	176	13.48	112	110	114	17.87
2017	168	168	175	14.23	112	109	114	15.39
2018	171	167	173	10.28	111	112	114	13.73

Note: Number of districts with no missing data by semester and year. Rural SAIDI unweighted mean by firm. It is expressed in total hours by semester. Source: CNEE

5.2 Weather DATA: Cooling Degree Days (CDD) and Rainfall.

We use three different data sources. First, National Institute of Climatology (INSIVUMEH) provided us daily observations from 48 weather stations (e.g. temperature, rainfall, wind), mainly with a time span of 2000-2018. Second, Institute of National Statistics (INE) has an

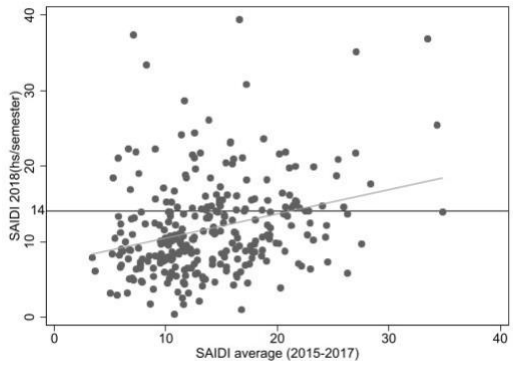

Figure 4: Actual and past SAIDI performance by municipality.

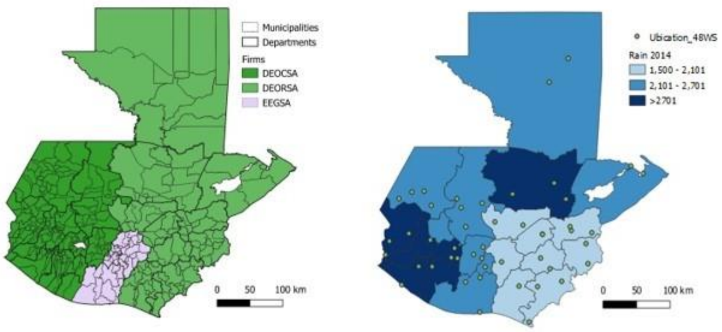

Figure 5: Zonal distribution of Utility Firms and rainfall in 2014

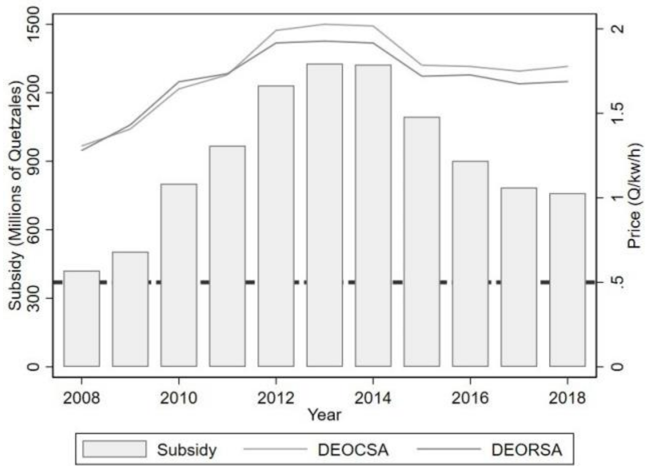

Figure 6: Social Tariff and total subsidies (INDE, 2019).

Table 11: Descriptive Statistics: ENCOVI

	DEOCSA				DEORSA			
	2011		2014		2011		2014	
	mean	sd	mean	sd	mean	sd	mean	sd
Rural Households (%)	57.51	49.4	54.55	49.8	66.0	47.3	64.17	47.9
Rural SAIDI (hs/semester)	10.31	8.3	41.38	29.22	10.36	6.9	45.31	23.73
Rural households connected to grid (%)	71.91	45.0	79.71	40.2	52.32	50.0	54.79	49.8
Dwelling variables								
Shares the dwelling (%)	2.09	14.3	0.44	6.6	1.60	12.6	0.18	4.3
Owner (%)	85.73	35.0	87.04	33.6	85.86	34.9	87.60	33.0
Poor housing materials (%)	50.15	50.0	47.40	49.9	57.93	49.4	56.93	49.5
Kitchen (%)	51.80	50.0	71.12	45.3	51.15	50.0	51.25	50.0
Own Toilet access (%)	7.78	26.8	9.57	29.4	4.67	21.1	5.78	23.3
Connected to water grid (%)	62.33	48.5	70.19	45.8	54.33	49.8	57.11	49.5
Overcrowd (%)	51.89	50.0	45.98	49.8	50.83	50.0	47.33	49.9
Head of Household variables								
Age	44.78	15.5	45.98	15.7	43.84	15.6	45.63	15.3
Female (%)	17.66	38.1	17.47	38.0	15.51	36.2	13.91	34.6
Recent migrant (< 5 years) (%)	1.90	13.6	0.77	8.7	2.29	15.0	2.29	15.0
Indigenous (%)	61.77	48.6	58.31	49.3	37.96	48.5	38.29	48.6
Does not speak Spanish (%)	10.64	30.8	10.61	30.8	17.94	38.4	16.39	37.0
Married (%)	82.23	38.2	82.39	38.1	82.63	37.9	84.33	36.4
Works (%)	83.92	36.7	84.21	36.5	83.99	36.7	87.37	33.2
Works as a farmer (%)	60.54	48.9	58.52	49.3	62.10	48.5	62.87	48.3
Works as a family worker (%)	0.27	5.2	0.29	5.4	0.69	8.3	0.39	6.2
A kid (7-14) is working in the house	2.91	16.8	2.85	16.6	1.10	10.4	3.34	18.0
The house has received remittances	0.11	0.3	0.12	0.3	0.08	0.3	0.09	0.3
Level of education: primary (%)	56.45	49.6	61.30	48.7	58.22	49.3	58.30	49.3
Income per capita (Quetzales)	224.61	283.8	449.52	537.8	264.45	591.5	537.00	1047.3
Observations	3,319		2,695		3,958		2,970	

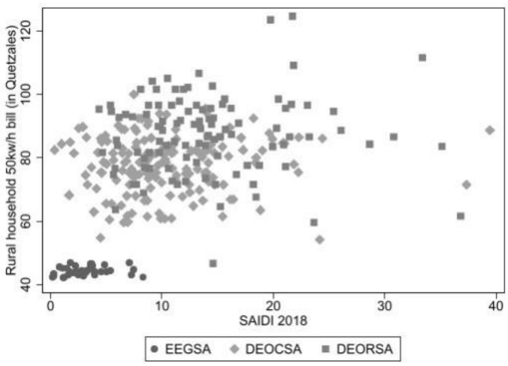

Figure 7: Bill simulation of 50kw/h consumption by municipality.

Table 12: Descriptive Statistics: Census

	DEOCSA 2018		DEORSA 2018	
	mean	sd	mean	sd
Rural Households (%)	60.44	48.9	61.57	48.6
Rural SAIDI(hs/semester)	10.28	5.8	13.80	6.6
Public Light Tariff (Quetzales)	29.97	9.3	38.45	13.0
Rural households connected to grid (%)	84.65	36.0	65.64	47.5
Rural households with solar panel (%)	3.77	19.05	9.91	29.9
Dwelling variables				
Shares the house (%)	1.80	13.3	1.95	13.8
Owner (%)	88.22	32.2	86.61	34.1
Poor housing materials (%)	40.38	49.1	50.95	50.0
Kitchen (%)	68.31	46.5	68.66	46.4
Own toilet (%)	24.19	42.8	22.77	41.9
Water grid access (%)	62.62	48.4	58.07	49.3
Overcrowd (%)	32.24	46.7	30.21	45.9
Receives remittances (%)	12.17	32.7	7.91	27.0
Asset: motorbike (%)	16.22	36.9	19.36	39.5
Asset: Radio (%)	60.39	48.9	48.97	50.0
Head of House variables				
Age (years)	45.94	15.8	45.55	16.4
Female (%)	21.68	41.2	19.46	39.6
Married (%)	82.38	38.1	80.89	39.3
Indigenous (%)	60.66	48.9	46.09	49.8
Migrant (%)	10.95	31.2	21.47	41.1
Recent migrant (< 5 yrs) (%)	1.33	11.4	2.09	14.3
Does not speak Spanish (%)	0.47	6.8	1.85	13.5
Works (%)	64.20	47.9	72.13	44.8
Works as a farmer (%)	38.84	48.7	48.02	50.0
Works as a retailer (%)	6.99	25.5	5.05	21.9
Works as Self Employer (%)	27.11	44.5	40.86	49.2
HH works as family worker (%)	2.68	16.1	1.60	12.5
HH completed at least Primary	63.42	48.2	62.90	48.3
A kid (7-14) is working in the house (%)	1.20	10.9	1.18	10.8
Observations	671,572		516,747	

Note: All summary statistics, except the first four variables, are calculated without taking into account households that have a solar panel.

almost complete monthly register of rainfall from 49 weather stations. Finally, we complement this information with satellite images from NASA.

We follow the Guide of Climatological Practices (World Meteorological Organization, 2017) that recommends not calculating a monthly mean if either of this criteria is not satisfied: missing observations for 11 or more days; or for a period of 5 or more consecutive days during the month. In case this condition is not satisfied, we assume a missing value for that specific month.

Cooling Degree Days (CDD) is calculated as the difference between mean temperature and 18 Celsius degrees. If negative, the value of CDD for that day is zero. Due to a large number of missing values in 2018, we use a Time Averaged Map of Surface air temperature monthly 0.1 degrees from NASA website (FLDAS model), and we calculate the zonal statistics for each municipality with Qgis software (McNally et al., 2018).

With weather station data we follow these steps: 1) consolidate and get an homogeneousdataset at weather station level; 2) calculate distance between the capital city of each munic-ipality to its closest weather station; 3) assign each weather station record to that district[21] and 4) collapse data at department level for ENCOVI.

Once obtained a daily measure of CDD we average first by month and then by year. We follow this procedure because as there are some missing values, we estimate monthly missing values with historical data, and then we average. After checking data consistency, the final dataset consists of 42 weather stations for 2011 and 2014.

As regards rainfall, our main dataset provides from INE. If a monthly record is missing, we check if that data is in INSIVUMEH dataset. If we can not complete the register, we estimate the missing value. Unfortunately, there are some weather stations with almost all 2018 monthly values were unrecorded, so we must dismiss them in our 2018 estimation. The final rainfall dataset is 45 weather stations.

[21]Mean distance from the city to closest weather station is 15km

5.2.1 Instrumental variable discussion

Our main argument for considering rainfall as an instrumental variable is that heavy rainfalls where roads are mainly unpaved, could be a potentially exogenous source of variation in reliability. This idea entails an ideal granular dataset: rainfall and quality data should be even more disaggregated.

Although we do not have that ideal dataset, we can test our main hypothesis. First, if the main drivers of low quality are unpaved roads and isolated communities, we should expect no effect in the urban area or at least, rainfall should result a weaker instrument. Second, as there is seasonality in rainfall, we should also expect —following our hypothesis of quality affected by rain in rural area— a better performance of the IV strategy in the second semester. Table 14 confirms the weakness of the instrument for urban area, and the robustness of the instrument in second semester, more related to the hurricane season.

Up to this point, it seems reasonable that our instrument at department level could be reflecting the underlying phenomena. Nevertheless, the way we measure rainfall in Census regression —at municipality level— fails to be a good instrument. A first plausible interpretation is that 45 weather stations are not enough to represent the variability in rainfall over 340 municipalities across the country. So, we test two other alternatives: interpolating weather station rainfall data or using NASA satellite images.[22] Using the Giovani APP from NASA website, we obtain the annual map accumulated rainfall for 2018. Each image represents a spatial resolution of around 100km^2. Then, with QGis software we estimate the average annual rainfall for each municipality. Although satellite information provides variation across municipalities (there are not two municipalities with the same rain record), we also acknowledge that rainfall variance is lower reducing the possibility of registering extreme events, useful for our identification strategy. And the last but not the least possible reason, we lack variation in time, which is present in ENCOVI regressions.

Once the 45 stations are collapsed by department, we can compare these estimations with NASA database in Table 13. Reasonably, summary statistics that come from point estimates –weather stations– have more standard deviation that satellite images.

[22]See Huffman et al. (2019)

Table 13: Weather data

	Rainfall				Cooling Degree Days			
Weather Station Data	Mean	Std	Min	Max	Mean	Std	Min	Max
Year 2011 (WS=49)	2,019.07	1,001.12	846.2	4,843.6	4.97	3.81	0	10.61
Year 2014 (WS=49)	1,717.66	965.54	440.1	5,026.7	5.11	3.71	0	10.99
Year 2018 (WS=45)	1,483.11	949.68	409.6	4,392.2	2.60	3.14	0	9.74
Municipality level (340)								
Using 45 Weather Stations	1,520.37	1,006.35	409.6	4,392.2				
Zonal statistics to interpolation image	1,446.72	244.34	874.63	2,636.77				
Zonal statistics to NASA satellite GPM image	1,742.78	456.25	1,032.73	2,733.13				

Note: Rainfall is in millimetres and Cooling Degrees Days in Celsius. 2018 CDD corresponds to FLDAS annual average surface temperature.
Source: INE, INSIVUMEH and NASA

Table 14: Robustness check: Instrument

	(1) IV_urban	(2) IV_1 Sem	(3) IV_2Sem
Ln $SAIDI_{urban}$	-0.383		
	(0.207)*		
	[0.303]		
Ln SAIDI_1 semester		-0.823	
		(0.326)**	
		[0.108]	
Ln SAIDI_2 semester			-0.205
			(0.041)***
			[0.005]***
Adjusted R^2	0.210	0.300	0.325
F_test	4.289	3.483	**42.949**
Observations	6,856	12,825	12,825

Robust Clustered standard errors at department level in parenthesis. Model 1 estimates the IV model for urban area. Model 2 uses SAIDI and rain for first semester in rural area; and model 3 does the same but with second semester. In square brackets the p-value of Wild Bootstrapped Clustering with 400 replications. All models have household and state variable controls, year and region fixed effects, and interaction between region and year.
* p<0.10, ** p<0.05, *** p<0.001

Acknowledgements

I would like to thank Raúl Bajo for all his useful comments, as well as participants in Best Student Paper competition at the 1st IAEE virtual conference in 2021, and in X International Academic Symposium organized by Institute of Economics of Barcelona in 2022. Also, to the Guatemalan offices CNEE and INSIVUMEH, who have been very patient with me giving me all data I requested. Finally, I am grateful to the Scholarship Program of Universidad de Navarra.