Power reliability and grid connection: evidence from rural Guatemala

Federico M. Accursi*

Dept. of Economics, University of Navarra

March 31, 2021

Abstract

Electrification rate have been increasing, but the prevalence of outages is still relevant for rural households when considering whether to connect to the grid or not. We combine two different households dataset with a complete register of electricity quality service in rural Guatemala at municipality level. Exploiting the particular evolution of power reliability and the precision of Census database, we find evidence that power reliability affects rural household willingness to connect to power grid. Our estimates are robust to different model specifications, including an IV strategy using rainfall. Moreover, results suggests an heterogeneous effect depending also on past performance. Efforts to expand the grid line to rural areas should be analyzed in concordance with actual power grid quality levels.

Keywords: power reliability - energy access - rural households- IV - Latin America

JEL classification codes: Q49-D10-O10

*e-mail: faccursi@unav.es. I would like to thank Raul Bajo for all his usefull comements. Also, to CNEE and INSIVUMEH who have been very patient with me giving me all data I asked for. All errors are of my own.

1 Introduction

Despite the worldwide access to electricity rate has been rising from 82% in 2008 to 89% in 2018 (World Bank Data), almost 800 million people still do not have access to electricity. 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. Besides, these characteristics are not independent from each other, and access does not necessarily mean truly and reliable connections. In other words, poor quality service could discourage new connections, although electricity access is available. Moreover, a faulty service could entail conflicts in the form of unpaid bills, theft and illegal connections. As an example, Dzansi et al. (2018) analyze the vicious circle for an utility company in Ghana, where households exposed to rolling blackouts are more prone to have unpaid bills, decreasing firm revenues and struggling its financial state.

The lack of access to electricity is a particular pronounced problem in rural settings. In 2018, the urban population had 97% of electricity access, while the rural, 15 p.p. less. Although many rural households have been benefited from renewable energy devices like solar panels, it has not been investigated enough if they can really fully exploit all the advantages electricity supply provides.[1] In places where the grid was actually extended, access to electricity from that the grid provides a least-cost solution, but it is well-documented that there are reliability issues.

In this paper we focus on the role of lack of reliability as a barrier to rural electricity connection. Reliability, or better said availability, is defined as "the attribute of energy supply that implies ability to draw energy when needed for use of energy services" and it is measured as the time and duration of supply (Bhatia and Angelou, 2015) . The importance of reliability is straightforward not only for firms productivity, but also for the residential sector that relies on a wide range of services electricity provides. Also, it concerns from a public policy perspective since many effort could be done to spread the low voltage grid, but this effort could end wasted if quality decreases because of insufficient complementary investments (e.g. transmission lines). To study this issue, we use detailed data from rural Guatemala.

Guatemala offers an appropriate context to study this question for several reasons. Although it is considered by the World Bank an upper middle income country, the contrast between the urban area of Guatemala city and the rest of the country is marked. According to Instituto Nacional de Estadística (INE), in 2017 the average of 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

[1] See Bayer et al. (2020) literature review for the impact of electricity access. They state that from 31 studies, one third are related to "off-grid" solutions.

country, reached the outstanding value of 77%.

Ending a civil war in 1996 and starting a reform process, it enhanced the rural electrification rate from 48% to 74% in a decade (World Bank). However, according to 2018 Census, 77.7% of rural households uses the grid as the first way of lightning, while 12.5% still remains using candle or 2.4% even gas. On the contrary, this proportion reaches 95.8% in urban households. Additionally, firewood is still the main primary energy source in the country –specially for cooking and heating in rural areas– entailing indoor pollution. The Ministry of Energy reported that in 2016, total energy consumption from residential sector came 90% from firewood and only 5% from electricity (Ministerio de Energía y Minas (2019)). Likewise, the quality gap in electricity supplied is notorious. In the last decade, on average, rural area suffered 35% of more service interruption in duration, and 14% in frequency. After 2011, following a series of management changes, rural Guatemala suddenly increased the number of outages in comparison to the capital areas. We take advantage of that to study the causal relationship between power reliability on rural households disposal to connect to the grid.

To study this question, we use data from National Commission of Electricity Energy of Guatemala (CNEE), which we combine with two set of data at the household level: the recent 2018 National Population Census, and the National Survey of Living Conditions (ENCOVI, hereafter) of 2011 and 2014. 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 the ENCOVI, the particular variation of quality observed in time will help us for the identification strategy, since quality experienced an unexpected shift after 2011. In addition, the use of official, objective and accurate data on outages from CNEE avoids two classical empirical problems: self selection bias and measurement error. Thus, we find plausible evidence that there is a positive effect between quality and rural household connections to the grid. This evidence is robust to an IV strategy –with rainfall as an instrument– and further supported using cross section data from Census. As a policy implication, the expansion of the grid should not be at the expense of quality, or at least the hidden costs of a not reliable service should be taken into account.

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

Moreover, the vast majority of literature that relies on data at household level, concentrates on the effect of electrification on household outcomes, taking for granted reliability.[2] According to Bonan et al. (2017) literature review, there are very few experimental studies focused on barriers to electricity, which are specifically focused on liquidity constraints, meanwhile reliability is not included, or is at best a secondary goal. For instance, Chakravorty et al. (2014) studies the effect of both grid connection and quality on household´s rural income in India, and finds that less frequency in outages increases household income. Nevertheless, in recent years there has been an increasing interest on reliability itself: Dang and La (2019) stresses the positive effect of power quality on rural income in Vietnam, and Bajo-Buenestado (2021) states that blackouts in Kenya discourage electricity connections.

It is important to remark that the way literature deals with the issue of measuring "quality" is varied. The problem of its definition depends essentially on the information and data structure available. Those who gather data from Surveys, where people or firms are asked frequency of outages, the usual approach is to build a dummy variable, or a categorical one if questions are based on a Likert scale. For example, the before mentioned paper of Chakravorty et al. (2014) uses a dummy approach for defining good and bad quality with a threshold, using for that self reported hours of effective supply and frequency of outages. Alternatively, others uses a continuous measure. Dang and La (2019) use data from a three-round household dataset in Vietnam and *counts* for the number of days without power outages. Then, Millien (2017) from an *opinion survey* builds a weighted severity index of reliability uncertainty based on perception data in Kenya. Using an IV approach, he finds out that higher reliability entails higher probability of connections, being highest for middle-rich households. Also, households are more sensitive in areas where outages are less frequent. Our paper overcomes the usual problems associated to self-reported data on quality since, as we discussed above, we use detailed (objective) data on outages from official sources.

Finally, there is a strand in 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. Using Heckman model, they state the importance of high-quality service for rural households, being willing to pay more for better service. They stress that more households will connect, if quality is improved.

The main contribution of this paper would be one of the few that focus on reliability as a barrier to

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

rural households electrification, as well as using a unique database that does not rely on memory nor in people perceptions. We add new empirical evidence –specifically in a rural Latin American context– to these least developed area of the literature. 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. Furthermore, bearing in mind the similarities that Chakravorty et al. (2014) document for the Indian case, we also advocate their suggestion that bringing new households to the grid is as important as providing high quality service because they had at least an equally significant and positive effect on household incomes.

The rest of the paper is organized as follows. Section 2 attempts to provide a brief and insightful background of Guatemala electricity sector. Section 3, structures the basement of the empirical strategy and explains how we tackle some potential issues as threats to identification. Then, Section 4 provides data description –further explained in appendix–, Section 5 explains the main results fulfilled with and IV strategy, additional robustness checks, and heterogeneous effects. Finally, Section 6 concludes.

2 Background

2.1 Recent history of Guatemalan Power Sector

After a long civil war, Guatemala was in a very fragile situation economically speaking, with a GDP per capita comparable to Eswatini. There was an evident lack of infrastructure and services, including electricity –the rural population electricity access was 39% in 1995–.

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 Law established the obligation for the utilities to connect households which were closer than 200 metres 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). Consequently, in 1998 the 80% of the most important public utility firm (EEGSA) was sold, 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). The main investments were done up to 2005. The Rural Electrification Plan remains, but with less resources. According to Iorio and Sanin (2019), 76% of new connections made between 1999-2014, were done during the first five years. Also, in Paz Antolín (2009) opinion, these grid network expansion was not accompanied by the necessary investments on the transportation line, resulting in a lower quality service.

The main institutions that the Law created in the energy sector are: Ministry of Mining and Energy (MEM), the National Commission of Electricity Energy (CNEE) and the Wholesale Market Manager (AMM). CNEE is the Government agency in charge of ensuring the compliance of the General Law of Electricity and its Regulations, monitoring quality of the energy supplied, penalizing utility companies, approving retail prices each term following the conditions approved in the Tariff Agreement which is renewed every five years. Lastly, the AMM allows the negotiation between generators, transporters, distributors, big customers and marketers. The energy policy has been basically fostering private investment –specially on renewable–, subsiding consumers (by a Social Tariff created in 2000) and promoting some rural electrification -done by The National Institute of Electrification (INDE).

According to World Bank Data, its GDP per capita (PPP) has growth 3.1% annually on average in 2000s and 3% last decade. Rural population electricity rate access has experienced also a steady growth path, growing from 55% in 2000, and reaching 93% in 2018. However, these figures contrasts with Census data, where 77.7% rural households are actually connected to the energy grid, 6% has solar panels, 12% uses 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.

2.2 Generation and distribution of electricity

In the last decade, Guatemala generation relied mainly on renewable sources (64%), 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 AMM. There has been a significant growth in the so called Renewable Generating Distributors (GDR), which are small power plants –with less than 5Mw capacity– that can sell directly to wholesale market.

Although the load factor has been raising from 57% in 2001 to 70% in 2018, own generation has been enough to fulfill national consumption and even export. In fact, Guatemala has been a net exporter over this period, with the only exception of 2010-2012. In contrast to low income countries, the lack of reliability rarely comes from generation constraints, but it is usually linked to issues in the distribution stage.

In fact, distribution of electricity is mainly done by three large utility companies, which have mostly a zonal distribution.[3] EEGSA provides mainly to the most dense area -Guatemala department-. DEOCSA and DEORSA provides to areas of the west and east of the country, respectively, that includes the bulk of

[3] See Appendix A

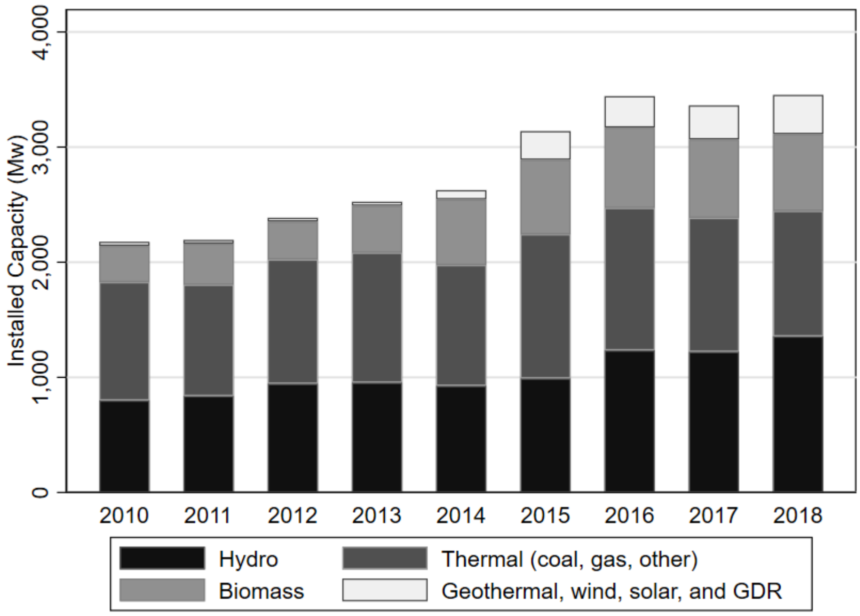


Figure 1: Evolution of installed capacity.

rural customers nation-wide. Also, 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. Meanwhile, EEGSA provides to 38%, DEOCSA (33%) and DEORSA (22%).

As we want to measure the impact of electricity supply reliability on rural households disposal to grid connect, we restrict our analyses to DEOCSA and DEORSA geographical regions since 92% of rural households live in the area they supply. Also, EEGSA (and its area) is fairly different from the rest. Table 1 resumes very well those differences as well as Figure 2 does displaying quality evolution.

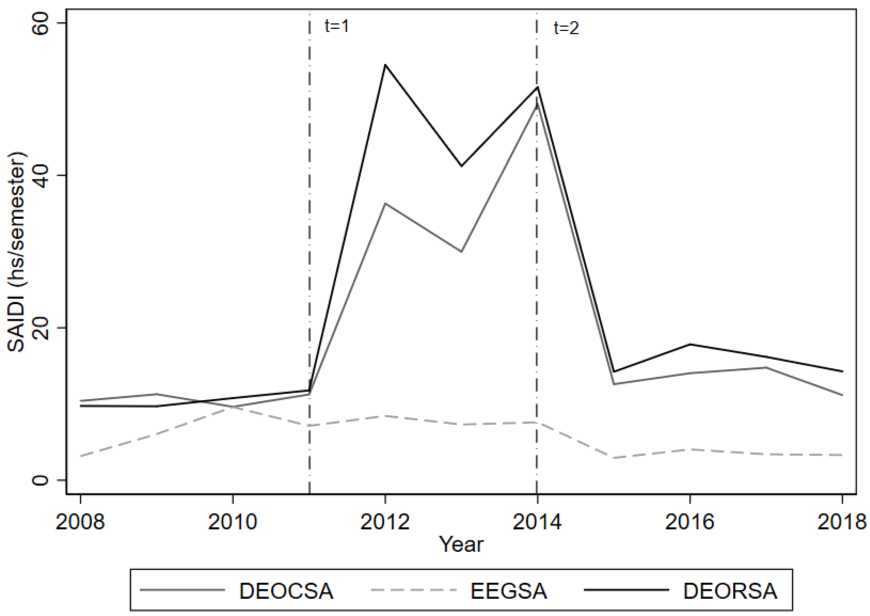
Table 1: **Utilities main characteristics (2014)**

	EEGSA	**DEOCSA**	**DEORSA**
Total Consumers	1,108,352	975,717	598,550
Social Tariff Consumers	997,668	952,152	576,215
Per capita consumption (kw/month)	104.66	68.61	79.06
Social Tariff (Quetzal/kwh)	1.63	2.02	1.92
Large Consumers	769	9	49
Compensation (Quetzales)	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. Source: CNEE.

Figure 2 reveals the evolution of quality along time, measured by the widely used System Average Interruption Duration Index (SAIDI), obtained from CNEE. EEGSA has almost always had a better quality service, and its evolution in time is uptrend. On the contrary, although quality for DEORSA and

DEOCSA are quite similar and stable, they experienced three consecutive years -from 2012 to 2014- of a deficient service provision. We find at least three possible reasons why this happens, which are mainly related to managerial issues.[4]



Note: SAIDI is a weighted average -by rural population- according to zones served by each utility. Data was obtained from CNEE

Figure 2: Evolution of rural quality (SAIDI).

One possible reason is the lack of negotiating capacity -or purchasing behavior in the wholesale market- what had differentiated firms. Even if there would have been some generation or transportation constraints (e.g the growth in generation capacity could have been insufficient to an outgrowing demand), this situation should had affected all firms equally. However, EEGSA did improve along those years, and DEOCSA and DEORSA did not.

Secondly, some managerial problems could had arisen with DEOCSA and DEORSA sale from Union Fenosa to Actis Group in May 2011 and its later resold in 2016 to ENERGUATE, actual owner of both firms. Additionally, the proximity of Tariff Agreement expiration with CNEE in 2014, would have probably discouraged important investments. This hypothesis is in concordance with anecdotal evidence from the local press, who stated there had not been investments neither from Union Fenosa in its last years nor from Actis.[5]

[4]It is important to remark that this is unlikely to a sudden increase in consumers. In fact, the total number of consumers (rural and urban) barely increased between 2007-2011. In this period EEGSA experienced an annual growth rate of 3.9%, meanwhile DEOCSA and DEORSA were 2.6% and 2.7% respectively. Note that population growth rate was 1.9% annually.

[5]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% were for these two firms.

Thirdly, it could be argued that households decided not paying the bill or stealing it because electricity was being expensive in a context of low quality. This situation would have more probably happened in urban areas than in rural. Looking at Figure 3, tariffs were increasing since 2008 reaching its maximum in the worst quality period 2012-2014. At this point, it is important to figure out how tariffs are settled.

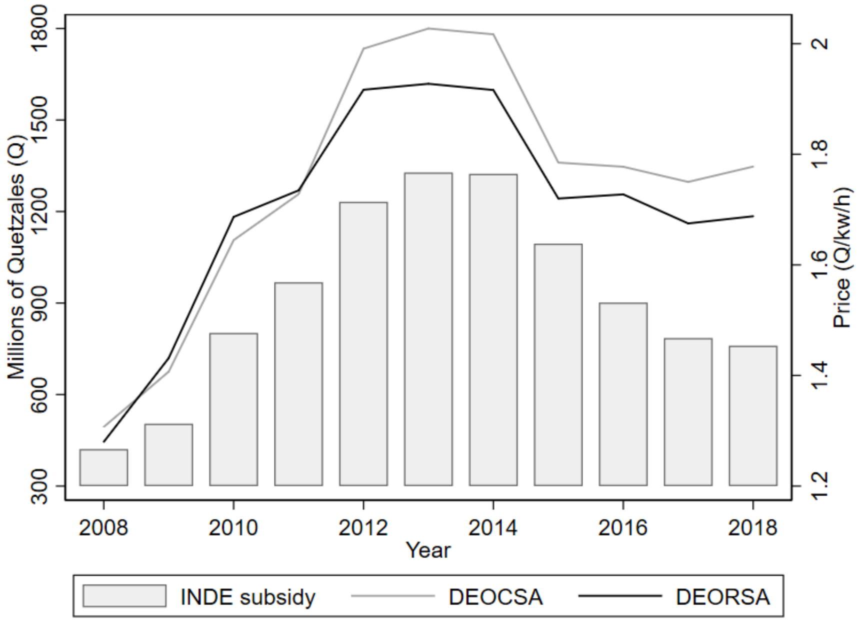


Figure 3: Social Tariff evolution.

According to the General Law of Electricity, final consumer prices are regulated based on an agreement between the CNEE and the firm, which changes every five years. In this agreement, it is decided the way to update tariffs every three months according to the cost of buying energy -mainly settled in a bidding process- and the distribution added value (i.e. the mean cost of capital), which is decided for a longer period of time. The tariff scheme for households is based on two prices: one for those who consume more than 300kw/h per month (i.e. based tariff called BTS), and a Social Tariff for those who consume less. In 2019, 94% of Guatemalan belong to this category. Additionally, this group could also be benefited by the so called "INDE contribution", which is a tiered subsidy up to 100kw/h. In practice, the subsidy entails a maximum price for the customer, because the difference of the Social Tariff and that price cap is paid by INDE. 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.[6]

[6]For example, in November 2019, the scheme was: Q0.50 from 1-60kw; 0.81 from 61-88. Simulating the bill for an 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

Since the price cap has been quite stable along time, the total amount of subsidies increases as well as the Social Tariff . It reached its maximum in 2013 (almost 180 US$ million dollars). 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 almost all rural households.[7]

Therefore, in next section we formulate our empirical analysis in order to exploit these variation of quality in time.

3 Empirical strategy

3.1 Regression model

We exploit quality variation over the period 2011-2014 that coincides with ENCOVI database, and then compare those results with Census along a more stable year. Therefore, the empirical strategy will be twofold. Firstly, we use a repeated cross section at department level with time variation in a lineal probability model:

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

Where subscripts h,d,r,t mean household, department, region and time respectively.[8] 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, *lnSAIDI*, is the natural logarithm of SAIDI in department "d". H is a set of control variables at the household level, D_{drt} are some characteristics at department level, η_r are region dummy variables and θ_t year fixed effects. Then, an interaction of region and year fixed effects, and finally ε_{hdt} 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 level shift in SAIDI is exogenous to rural households decisions. We hypothesize that a reduction in quality at department level, reduces the expected benefits

to calculate the bill is available at: http://www.cnee.gob.gt/Calculadora/index.php

[7]Basically, 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 intervals), 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]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. Each department has a governor who is a President´s delegate. Only municipalities do have representatives and its own institutions, as well as dictate their own laws. Because of geographical, economical or social reasons, departments are grouped into regions. Only Petén department is a region on itself.

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

In order to confer robustness to our results, we present a second regression setup using a cross section with more granular data (at municipality level). Merging Census database with CNEE records at municipality level the regression model is:

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

Now, M_{md} are some characteristics at municipality level in department d, η_d are department fixed effects and ε_{hmd} is the error term. Now β_1 is capturing 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.

3.2 Potential threats to identification

Although we have exposed the reasons why we believe SAIDI variation in 2012-2014 is not driven by demand-related factors, still we might be concern of a potential endogeneity, due to reverse causality. In order to mitigate this concern we provide additional evidence based on an IV regression.[9]

In our empirical setting, we believe rainfall is an appropriate instrument. As an exogenous variable, although it could increase electricity generation by hydro and biomass power plants, it could also affect negatively electricity distribution, specially on rural zones. Heavy rains produce electricity cuts as a precaution measure, and muddy or flooded roads makes maintenance tasks really hard. In Guatemala, seasonality in rainfall is present. Although many papers use rainfall as an IV and the sign of the first stage is negative -more rains affects negatively power outages-, we expect the contrary in the Guatemala context.[10]

In pursuance of building the instrumental variable, weather data come from the national weather agency of Guatemala (INSIVUMEH) providing information of 43 weather stations that belongs to the studied area. We aggregate rainfall data into department -for ENCOVI regression- or municipality level -for Census-. At this level, each municipality was assigned to only one weather station.[11]

[9]The way literature exploits exogenous variation on quality is varied. For the sake of simplicity we can cite: quality of the other villages of the same province as an instrument (Dang and La, 2019), 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)

[10]We give two more arguments supporting the adequacy of the instrument. First, if we are reasoning rainfalls is positively correlated with SAIDI, we should expect no effect in the urban area or at least, be 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. Looking in the Appendix at Table A.2 we confirm the weakness of the instrument for urban area, and the robustness of the instrument in second semester, more related to the wet station.

[11]See Appendix B for details

Still, one might be concerned that many municipalities would have the same rainfall value, or that two weather stations can not represent a whole department. Therefore, we use another dataset: NASA satellite images. Each image is a monthly rainfall estimation with a spatial resolution of around 100km^2. With QGis we get the annual rainfall for each municipality or department. On the one hand, satellite information provides variation along municipalities (there are not two municipalities with the same rain record). On the other hand, we lose variance in the variable "rainfall" reducing the possibility of registering extreme events, useful for our identification strategy, specially at department level.

Implementing the same IV strategy with Census regression is not exempt from some caveats ex-ante. On the one hand, we have more spatial variation (given by the number of municipalities and by NASA estimations). On the other hand, we lack of time variation. Also, some municipalities are so small that rainfall in municipality *m* could also be affecting the quality of neighbourhood municipalities, specially when they share the same subsection of the grid line.

Finally, one could also be concern that rural settlements could be so far away from grid lines, that lacking of electricity is a fact, and not a choice. Although the probability of having such isolated households is very low in ENCOVI regression , it is not the case for Census where the entire population is surveyed. Thus, in order to tackle this matter, we drop rural households that declare having solar panels as main source of lightning in Census regression because we consider them as isolated. Despite they represent 6% of observations, we show in Section 5 that our results are robust to its inclusion.

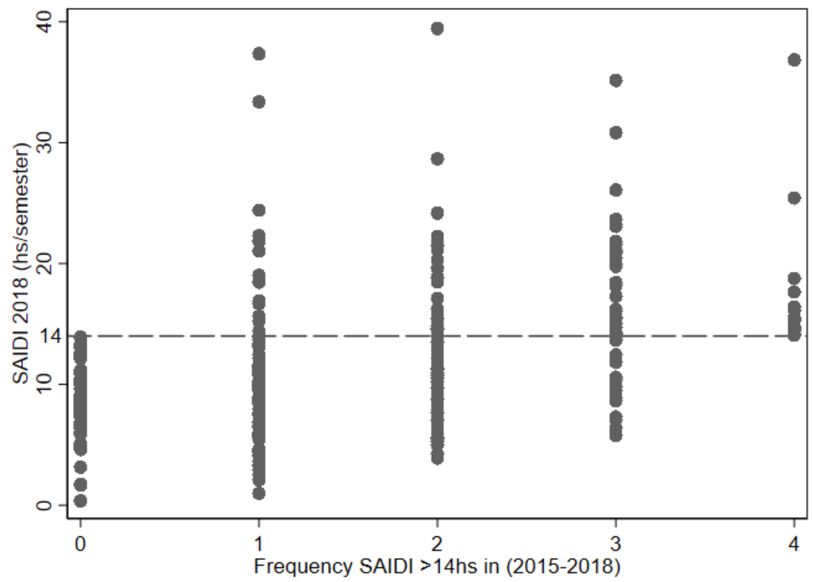
4 Data description

Our *reliability* variable is System Average Interruption Duration Index, provided by CNEE. We have data on rural areas from almost all districts (339 out of 340) in the time span of 2006 to 2018. Fortunately, in our study period, information is quite balanced. To recover quality service at municipality level we get the mean of the two semesters observations, and finally we get an unweighted mean by department.[12]

Regarding SAIDI evolution, Figure 2 shows a quite stable performance with exception of period 2012-2014. Although quality has improved since 2015, it is worth differentiating the quality performance in some districts over time. In that sense, Figure 4 reflects a sort of "reputation", counting the number of times that the firm exceeded the minimum of 14 hours -required by law- in that municipality. For instance, more than a half of non-compliers districts in 2018 are "regular defaulters" as they have exceeded the standards three or even all the times since 2015, while there are 13 that defaulted for first

[12]CNEE Resolution 9/1999 establishes that quality of service supply will 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 B

time in 2018. Hence, this reflects an heterogeneity in quality supply over municipalities.



Note: Rural municipalities quality in 2018 and its frequency of infringement SAIDI standard in 2015-2018. The horizontal line shows the law requirement of 14 hours. Data was obtained from CNEE

Figure 4: Frequency of infringement of SAIDI standards (2015-2018)

Then, we count with two household dataset. The first is the National Survey of Living Conditions (ENCOVI) of 2011 and 2014, containing above all income information. The second is the Census, conducted between July and August in 2018. It gathered detailed information on 3,275,931 households, such as dwelling characteristics, level of education, labor and migrant condition.

The output dummy variable *grid connection* comes from question 8 of the Census, which asks the kind of lightning mainly used at home. The answer is unique. From ENCOVI survey, the questionnaire asks whether the house is connected to an electrical distribution network. From both dataset a full set of controls is created.

Firstly, dummy variables at *household level*: if they are the owners of the dwelling; if it is shared with another family or not; if the family usually receives remittances and if relatives of that household has ever emigrated from the household, and finally if the house is overcrowded.[13] Then, some *assets* are taken into account: motorbike and car; poor dwelling materials -having earth as floor, or a thatched roof, or a metal sheet wall-. Finally, if the house has some *basic utilities*: water pipe access (inside or outside the dwelling); public or private garbage collection; access to an own toilet; sewerage; having own kitchen to cook.

[13]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, CEPAL warns that for some indigenous culture, many people are used to sleep in large rooms, so overcrowd could be overestimated.

From ENCOVI, income differences are captured by the inverse hyperbolic sine of total familiar real income. We make this transformation –instead of logarithm– to avoid losing zero income observations.[14] Although Census lacks of income data, it gives some labor condition information that could give an idea of purchasing power of the household. More control variables are introduced trying to capture these underlying income differences. For example, from labour category some dummy variables are created (e.g. family worker, self-employer) and from type of job (e.g. farmer, retailer) and the dependence ratio at household level –calculated as the division between those who work at home divided the inactive inhabitants–.

Also, we build additional *head of household variables*: age, square age, gender, marital status, ethnic group, if is able to speak in Spanish, if he/she works, four dummies for different levels of education; recent migrant (not more than five years), a "whole life migrant" (he migrated more than 5 years ago), and labor status as before mentioned.

Table 2: Rural Area Descriptive statistics

	ENCOVI				CENSUS	
	2011		2014		2018	
Variable	DEOCSA	DEORSA	DEOCSA	DEORSA	DEOCSA	DEORSA
Rural Household (%)	57.5%	66.0%	54.6%	64.2%	60.4%	61.6%
SAIDI (hours/semester)	10.3	10.3	41.4	45.3	10.3	13.8
Grid Connection (%)	71.9%	52.3%	79.7%	54.8%	84.7%	65.6%
Dwelling characteristics						
Owns the house	85.7%	85.9%	87.0 %	87.6%	88.2%	86.6%
Poor materials	50.1%	57.9%	47.4%	56.9 %	40.3 %	51.0%
Head of Household Variables						
Primary completed	57.1%	58.7%	61.3%	58.3%	63.4 %	62.9%
Farmer (% from workers)	66.8%	71.0%	67.7%	70.8%	60.5%	66.6%
Recent Migrant	1.9%	2.3%	0.8%	0.2%	1.3%	2.1%
Indigenous	61.0%	37.7%	58.3%	38.2%	60.7%	46.1 %
Average monthly TFI	Q 1,433.3	Q 1,368.9	Q 1,682.5	Q 1,844.0	nd	nd
Observations	3,319	3,958	2,695	2,970	671,572	516,747

Notes: 2011 and 2014 data is from ENCOVI, and 2018 from Census. All summary data is from the area supplied by DEOCSA and DEORSA. SAIDI data come from CNEE and is a weighted average by each municipality population. TFI means Total familiar income. From Census data, rural households who have panel solar are not taken into account for dwelling and HH characteristics

Then, at *municipality and department level* some mean characteristics have been calculated for rural area : employment, literacy and schooling rates[15], proportion of households at district level who: has a child under 10 years working, has toilet, water access, motorbike, and poor dwelling conditions as already defined. Also, we include Cooling Degree Days (CDD) [16] and Public Lightning Tariff. Each

[14]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)

[15]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.

[16]Some part of the literature includes weather conditions that push demand, such as cooling or heating degree days

local government charges to the final bill a lump sum or percentage. Theoretically, this earnings finance the cost of provision of public light in streets, but sometimes they turn into a hidden tax. Unfortunately we only have reliable data from 2015, so we include it only in Census regression.

Finally, for both regression setups we include a dummy variable if household belongs to DEOCSA supplied area. Also, in ENCOVI setup, we include an interaction term between DEOCSA and year dummy for precaution because Social Tariff (ST) was the same in 2011 for DEOCSA and DEORSA, but in 2014 was 5% more expensive in the former.

5 Empirical results

5.1 Main results

Results in the ENCOVI regression expresses the negative relationship between SAIDI and probability of a rural household to get connected. If quality improves 1%, the probability of connection raises 23 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 69,413 new customers, representing 1.6 US$ million dollars in annual revenues.[17] The rest of the control variables show a negative correlation between poor housing materials and grid connection, and positive with being the owner and income. Results are robust to different ways of clustering.[18]

These conditional analysis contrasts with the unconditional relationship that appears in Table 2, where grid connection grows in parallel with lower quality supply in 2011-2014 period. The identification strategy followed has let us disentangle the negative effect of poor quality on rural grid connection.

However, these preliminary results should be taken with caution. In addition to the potential issues before mentioned, the main assumption made so far is that households are making their own decisions based on average quality at rural department level, which could be too aggregated data. Therefore, Census regression setup could bring a closer look, since quality information is at municipality level.

Results of this second model specifications are resumed in Table 4. Firstly, as well as in ENCOVI

(v.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 aforementioned- widespread use of firewood.

[17]On average in 2011-2014 SAIDI was 26.9 hours by semester, 54 hours per year. On average, 34 out of 100 did not have grid connection, so improving 23pp means: 0.34*0.23= 0.078, eight out of 100 unconnected households will be expected to connect, representing 69,413 new households. At the end of 2014, with a monthly fix fee of 15 quetzales for each customer represented approximately 23.7 dollars annually per customer.

[18]Primary Sample Units (PSU) are the random areas selected for the Survey. Each one has its own weight in order to expand results to the whole population. In ENCOVI we can cluster at PSU level or department level. Clustering at department-year level although would double the number of clusters, would be assuming that quality -or rainfall in the IV case- are not autocorrelated, which could be a strong assumption. 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)).

Table 3: ENCOVI results 2011-2014

	(1) Grid	(2) Grid	(3) Grid	(4) Grid
Ln SAIDI	-0.111**	-0.108**	-0.141**	-0.230**
	(0.041)	(0.039)	(0.061)	(0.084)
Owns the house	0.038**	0.035**	0.042**	0.043**
	(0.019)	(0.017)	(0.017)	(0.017)
Poor Housing materials	-0.269***	-0.199***	-0.196***	-0.197***
	(0.016)	(0.014)	(0.014)	(0.014)
CDD		-0.002	0.000	0.006
		(0.003)	(0.005)	(0.005)
Real Income		0.009***	0.010***	0.010***
		(0.003)	(0.003)	(0.003)
Year FE	Yes	Yes	Yes	Yes
HH controls	No	Yes	Yes	Yes
Department controls	No	No	Yes	Yes
Region FE	No	No	Yes	Yes
Region*Year	No	No	No	Yes
Adjusted R^2	0.226	0.306	0.323	0.324
Observations	12,914	12,825	12,825	12,825

Robust cluster errors at Primary Sample Unit (PSU) level, N=1,138. Department controls are mean characteristics at that level such as: dwelling, water and toilet access; literacy, schooling and employment rate, and proportion of indigenous living in the department, and proportion oh households that have at least one child working. Results remains in model 4 if clustering is at department level, and Wild bootstrap is performed with 400 replications

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

model, all variables correlated with income -e.g. assets- are also positive correlated with grid connection. Then, Cooling Degree Days is still being not significant in rural Guatemala. It is possible that by insufficient wealth conditions, CDD would not be a demand driver of electricity. For instance, in rural ENERGUATE region, Census data reveals that those who have grid connection, 63% has a TV, 35% a fridge and 7% a washing machine. Finally, Public Light Tariff is not significant neither.[19] All in all, the

Table 4: 2018 Census estimation

	(1)	(2)	(3)	(4)
	grid	grid	grid	grid
Ln SAIDI	-0.067***	-0.025**	-0.029***	-0.021*
	(0.023)	(0.012)	(0.011)	(0.011)
CDD		-0.003	-0.003	-0.003
		(0.003)	(0.003)	(0.003)
Public Lightning Tariff		-0.000	-0.001	-0.001
		(0.001)	(0.001)	(0.001)
Poor housing materials		-0.166***	-0.145***	-0.147***
		(0.008)	(0.007)	(0.007)
Owns the house		0.017***	0.026***	0.025***
		(0.005)	(0.005)	(0.004)
Usually receives remittances		0.027***	0.023***	0.022***
		(0.003)	(0.003)	(0.003)
Asset: motorbike		0.061***	0.050***	0.050***
		(0.005)	(0.004)	(0.004)
Household variables	No	Yes	Yes	Yes
Municipality controls	No	No	Yes	Yes
Department Fixed effects	No	No	No	Yes
Adjusted R^2	0.008	0.242	0.255	0.271
Observations	1,178,160	1,161,698	1,161,698	1,161,698

Robust clustered standard errors at municipality level(266)
* p<0.10, ** p<0.05, *** p<0.01

effect of SAIDI remains being statistically significant and negative. Nevertheless, the effect is ten times smaller. Now, 1% of reduction in SAIDI (15 minutes annually on average at municipality level), raises the probability of getting connected in 2 p.p., representing 16,391 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 to think that *the level* of quality matters. Although results are robust to both regression setups and database, some additional estimates should be performed.

[19]The introduction of this covariate as a control is important because it is what really differentiates the cost of electricity that household faces.

5.2 Robustness checks: IV regressions

Recalling our concerns in the empirical section, we now present additional evidence using the IV regressions. In particular, we perform these regressions using rainfall as instrument, based on two possible data sources; namely, INSIVUMEH and NASA.

Table 5 includes the second stage –and some basic information of the first– results for the two more robust specifications of our ENCOVI model. First, columns 1 and 2 use the weather station data for rainfall as an instrument of SAIDI. Both models give a similar point estimation. However, following the Kleinbergen-Paap F-Statistic, we reject the null hypothesis of weak instrument only in model 2. Assuming the conditional exogeneity of the instrument, model 2 gives account that one 1% reduction in SAIDI increases 39.6 percentage points the probability of a rural household to connect to the grid. This estimation is 16 p.p. larger than OLS. In terms of new connections, this would have represented 116,267 new rural connections and 2.8 US$ million dollars in annual revenues.

On the other hand, columns 3 and 4 use NASA rainfall estimation as an instrument of SAIDI. Again, both models accounts for a negative relationship between SAIDI and grid connection. However, results should be read with care, since both are showing evidence of being weak according to F-Test. The difference in the F-Test for both sources of rainfall estimations could be due to the characteristics of both series: Weather Stations (WS) doubles the variation of NASA satellite image rainfall estimations, making the former a better instrument. [20]

5.3 Additional robustness checks

In this section some robustness checks are performed regarding ENCOVI and Census regression setups. Firstly, from ENCOVI we should be aware of the method of aggregation employed to get the variable SAIDI at department level. Therefore, we construct a second variable: $SAIDI_w$, a weighted average by the number of households a municipality has according to 2018 Census. On the one hand, we could have the "true" 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 information we have in 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 6, and SAIDI remains being significant with a lower point estimation.

Finally, for Census regression some caveats could come from the way CDD has been built and the

[20] See First Stage estimations in Appendix A for both regression setups. As discussed in Section 3.2, rainfall in Census context is a weak instrument.

Table 5: IV Regression models-ENCOVI

	(1)	(2)	(3)	(4)
	M3 $Rain_{WS}$	M4 $Rain_{WS}$	M3 $Rain_{NASA}$	M4 $Rain_{NASA}$
LnSAIDI	-0.388**	-0.396***	-0.854*	-0.771***
	(0.138)	(0.075)	(0.450)	(0.170)
	[0.038]	[0.005]	[0.118]	[0.008]
Household controls	Yes	Yes	Yes	Yes
Department controls	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Region Dummies	Yes	Yes	Yes	Yes
Region*Year	No	Yes	No	Yes
Adjusted R^2	0.320	0.324	0.295	0.316
Clusters	19	19	19	19
First stage F-statistic	9.988	23.197	2.750	9.146
Observations	12,825	12,825	12,825	12,825

Robust Standard errors clustered at department level in parenthesis. Wild Bootstrapped Robust Standard errors clustered at department level (400 replications). Between square brackets it is the p-value. F-statistic is for the heteroskedasticity and cluster robust Kleibergen-Paap weak instrument test. M3 and M4 means models 3 and 4 of Table 3

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

Table 6: Robustness check: weighted measure

	(1)	(2)	(3)	(4)
	Grid	Grid	Grid	Grid
Ln $SAIDI_w$	-0.126***	-0.149***	-0.086	-0.135*
	(0.037)	(0.038)	(0.054)	(0.074)
Year FE	Yes	Yes	Yes	Yes
Household controls	No	Yes	Yes	Yes
Department controls	No	No	Yes	Yes
Region FE	No	No	Yes	Yes
Region*Year	No	No	No	Yes
Adjusted R^2	0.227	0.308	0.323	0.324
Observations	12,914	12,825	12,825	12,825

Robust Standard errors clustered at PMU level. These are the same models as Table 4. Results remains in model 4 if clustering is at department level, and Wild bootstrap is performed with 400 replications

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

before mentioned exclusion of households with solar panels. We check this with different model setups. Results are in Table 7. Whether the inclusion or not of CDD does not affect results, nor the inclusion of households with solar panels, which elevates slightly the effect. To conclude, we perform some placebo tests, testing if power reliability affects some variables that beforehand should be not, and results confirm that quality supply would be affecting only to grid connection.

Table 7: Robustness Check Census.

	(1)	(2)	(3)	(4)	(5)
	grid	grid	grid	water	garbage
Ln SAIDI	-0.022**	-0.028**	-0.027**	-0.006	-0.007
	(0.011)	(0.013)	(0.013)	(0.020)	(0.011)
Household controls	Yes	Yes	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes	Yes	Yes
Department Fixed effects	Yes	Yes	Yes	Yes	Yes
adj. R^2	0.270	0.302	0.303	0.182	0.197
Observations	1,161,698	1,243,221	1,243,221	1,161,698	1,161,698

Robust clustered standard errors at municipality level. Model (1) does not include CDD as a control; Model (2) does not includes CDD but includes households with solar panel. Model (3) includes both. Model (4) and (5) are placebo tests. In these models, CDD, DEOCSA dummy variable, and the mean access to water and garbage in each municipality are not included as control variables

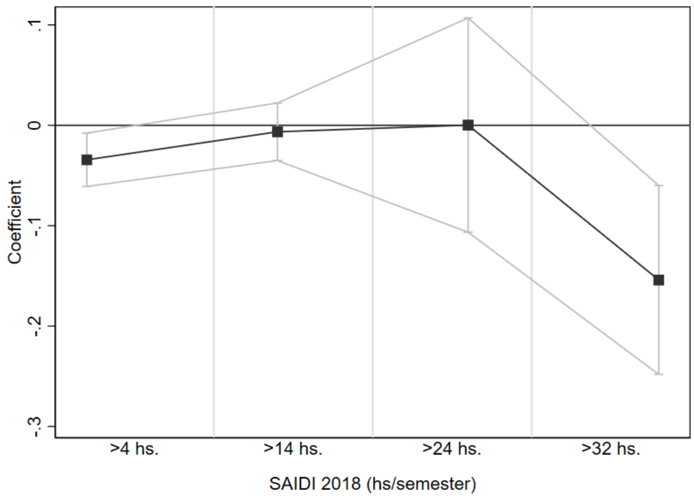
* p<0.10, ** p<0.05, *** p<0.01

5.4 Heterogeneous effects

Taking into consideration the relative stability achieved in quality standards by DEOCSA and DEORSA since 2015, we can explore potential heterogenous effects. We can take advantage of the granularity of Census and CNEE data. Figure 4 outlines different situations along municipalities regarding actual and past performance. We hypothesize an heterogeneous effect of SAIDI depending on current levels (let us call it "intensity") and on past quality ("reputation").

In fact, scaling SAIDI as an indicator dummy variable is depicted in Figure 5. The explanatory variable *Ln SAIDI* has been replaced by a dummy variable that equals 1 if SAIDI is more than a x number. For instance, districts that suffered in 2018 more than 32 hours of service interruption, has a lower conditional probability of being connected -15%- in comparison with the other group of districts that had SAIDI less than 32 hours. The no effect and larger confidence intervals in dummy variables of 14 and 24 hours, could be hiding an extra heterogeneity not explicit yet (e.g. reputation).

As quality standards oblige to have a maximum SAIDI of 14 hours, we choose it as threshold and we split the sample: municipalities where ENERGUATE complied and those where it did not in 2018. Results of estimating our baseline model in the splitted sample are in Table 8. The absence of a significant



Note: Model 4 of Table 4, but replacing SAIDI with one dummy variable indicator.

Figure 5: Estimations with different indicators of SAIDI intensity

effect in Column 2 is probably hiding an heterogeneity. It does not seem reasonable that quality would not have had any influence on grid connection in communities that did not satisfy SAIDI maximum levels in 2018. In fact, in that group there are not only serial defaulters, but also districts that failed to satisfy law requirements for their first time. Consequently, in order to model such disparity, we add in model

Table 8: OLS Census full model specification with different subsamples

	(1) SAIDI<14hs	(2) SAIDI>14hs	(3) SAIDI>14hs
Ln SAIDI	-0.031***	-0.050	-0.186***
	(0.009)	(0.058)	(0.061)
Ln SAIDI* Bad Reputation			0.233**
			(0.115)
Bad Reputation=1			-0.663*
			(0.335)
Household controls	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes
Department Fixed effects	Yes	Yes	Yes
Clusters	184	82	82
Adjusted R^2	0.267	0.298	0.300
Observations	767,308	394,390	394,390

Robust clustered standard errors at municipality level. Model 1 includes the municipalities that had SAIDI levels under 14 hours. Models 2 and 3, upper 14 hours. All models have the same controls as model 4 in Table 4

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

3 a dummy variable "reputation" and an interaction term.[21] Now, quality matters along all defaulters

[21] Reputation is a dummy variable that contains municipalities where ENERGUATE defaulted the law three or four times in 2015-2018 period.

districts in 2018, but districts´ historical performance also does. In municipalities that almost always suffers from larger SAIDI levels, the effect is larger. For instance, a 1% reduction in SAIDI levels would mean on average a positive effect of 61.6% in the probability to grid connect, while in the other districts would be 18.6%. However, results should be read with caution as reputation is a recursive process, so this should be interpreted as a very short term effect.[22]

Finally, focusing into the number of times did ENERGUATE exceed the 14 hours´ threshold in each municipality, we now split the sample into "frequency groups". In contrast to models in Table 8, subsamples are now defined by frequency, and not by SAIDI levels of 2018. We can interpret the frequency group of 0 as the group of compliers, and –on the opposite side– serial defaulters those who have frequency of 3 and 4.

As Table 9 shows, reliability is still a predictor for the group that always complies. Although it is half lower, it does not necessarily mean that quality is not important. In fact, having constantly levels of SAIDI below the threshold has probably contributed to these districts having reached higher levels of grid connection (84.6%). The more frequent the law is unaccomplished, the larger the effect is perceived. However, SAIDI is no significant in column 3. A likely conjecture is that group is in a sort of transition from one reputation group to the other, and there is much more heterogeneity inside it.

Table 9: OLS Census full model specification and frequency of law compliance

	(1) Frequency=0	(2) Frequency=1	(3) Frequency=2	(4) Frequency>3
Ln SAIDI	-0.010*	-0.036**	0.025	-0.081*
	(0.006)	(0.015)	(0.018)	(0.046)
Household variables	Yes	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes	Yes
Department Fixed effects	Yes	Yes	Yes	Yes
Clusters	56	78	68	64
Adjusted R^2	0.245	0.305	0.260	0.287
Observations	230,265	305,907	294,171	331,355
$SAIDI_{18}$ mean	8.52	10.32	12.54	16.48
$SAIDI_{18}$ min	0.37	0.99	3.89	5.79
$SAIDI_{18}$ max	13.91	37.35	39.44	36.84
Grid connected (%)	84.5	72.0	75.6	74.6
Solar Panels (%)	3.4	7.6	6.8	7.4

Robust clustered standard errors at municipality level in parenthesis. All models have the same controls as model 4 in Table 4. The information in the last five lines are summary statistics for each frequency group. Data of solar panels is just informative, and it does not mean it was included in the estimations.

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

In conclusion, there seems to be an heterogeneous effect of SAIDI depending, on the one hand on

[22] If model 1 includes Bad Reputation dummy, it results no significant and SAIDI point estimation remains significant.

current levels (being larger at larger levels of SAIDI), and on the other, on past performance: the worse the reputation is, the worse is the conditional probability to grid connect.

6 Conclusion

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 as the reliability measure. In first regression setup for the period 2011-2014 we find that a 1% reduction in outages duration at *department level*, increases probability of grid connection between 23-39 percentage points. In the second regression setup, with a more stable quality level in 2018, a 1% reduction in outages duration at *municipality level* increases probability in 2 percentage points. In addition, this effect would not be homogeneous between municipalities, since there is plausible evidence that past performance also matters. Restricting 2018 sample to the group of non-compliers districts in that year, the effect is larger: 18 percentage points for those who rarely are non-compliers, and 61 p.p. in municipalities where ENERGUATE is a serial defaulter.

Results are robust to different model specifications and robustness checks, and are in concordance with Kennedy et al. (2019) who state that improving quality is critical to improving energy access in rural India.

To sum up, we hope this paper could contribute to highlight the importance of power reliability and its direct effect on contributing to grid access.

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, Irani Calero, C. (2015). From Candles to Light: The impact of rural electrification.

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.

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, August). 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.

Grogan, L. (2018). Time use impacts of rural electrification: Longitudinal evidence from Guatemala. *Journal of Development Economics 135*(September 2017), 304–317.

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

Iorio, P. and M. E. Sanin (2019). Acceso y asequibilidad a la energía eléctrica en américa latina y el caribe.

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.

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/Estadísticas-Subsector-Eléctrico-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.

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).

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

Appendix A

A.1 Additional Figures and Tables

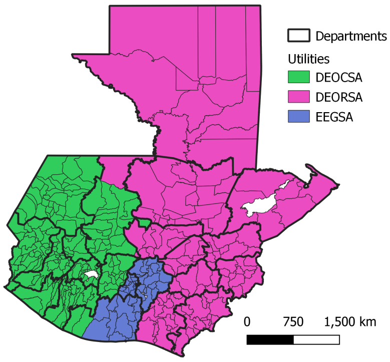
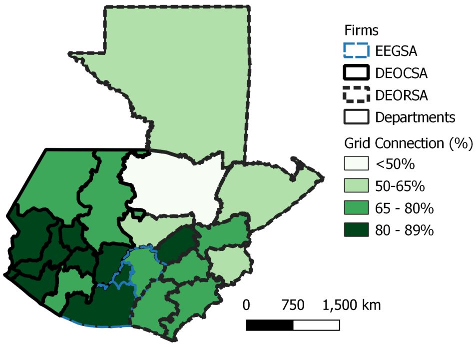
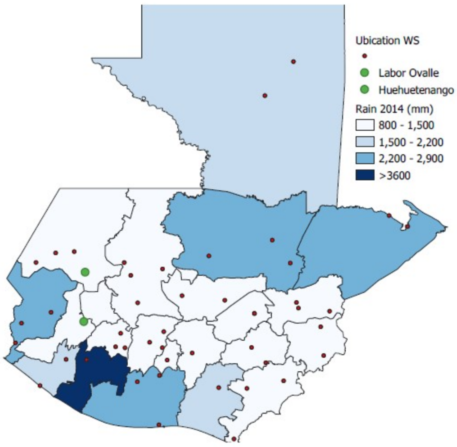


Figure A.1: Zonal distribution of Utility Firms



(a) Rural Grid connection in 2014



(b) Weather stations and Rainfall in 2014

Table A.1: Department rainfall estimations in mm

	Data source	Mean	SD	Min	Max
2011	Weather Stations	2,051	874	1,139	4,537
	NASA	2,321	469	1,617	3,042
2014	Weather Stations	1,770	1014	884	5,026
	NASA	2,251	477	1,544	2,934

Table A.2: 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

Table A.3: First Stage ENCOVI

	(1) LnSAIDI (1)	(2) LnSAIDI (2)	(3) LnSAIDI (3)	(4) LnSAIDI (4)
Ln $Rainfall_{ws}$	0.234**	0.269***		
	(0.074)	(0.056)		
	[0.04]	[0.03]		
Ln $Rainfall_{NASA}$			0.382	0.485**
			(0.231)	(0.160)
			[0.26]	[0.13]
Adjusted R^2	0.981	0.993	0.978	0.990
First stage F-stat	9.988	23.197	2.750	9.146
Observations	12,825	12,825	12,825	12,825

Robust Standard errors clustered at department level (19) in parenthesis. Wild Bootstrapped robust clustered at same level, with 400 replications, p-value in square brackets. F-statistic is Kleibergen-Paap heteroskedastic cluster robust weak instrument test.The IV estimation corresponds to models 3 and 4 of Table 3

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

Table A.4: First Stage-Census

	(1) LnSAIDI	(2) LnSAIDI	(3) LnSAIDI	(4) LnSAIDI
Ln $Rainfall_{ws}$	0.016	-0.063		
	(0.078)	(0.094)		
Ln $Rainfall_{NASA}$			0.247	0.027
			(0.156)	(0.333)
Household variables	Yes	Yes	Yes	Yes
Municipality controls	Yes	Yes	Yes	Yes
Department Fixed effects	No	Yes	No	Yes
Adjusted R^2	0.151	0.213	0.159	0.212
First stage F-stat	0.965	0.002	3.393	0.009
Observations	1,161,698	1,161,698	1,161,698	1,161,698

Robust clustered standard errors at municipality level (266). The IV estimation corresponds to models 3 and 4 of Table 4. F-statistic is Kleibergen-Paap heteroskedastic cluster robust weak instru ment test

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

Table A.5: SAIDI raw statistics from CNEE

	DEOCSA				**DEORSA**			
	N° Different districts				**N° Different districts**			
Total	**Sem 1**	**Sem 2**	**Year**	**Mean SAIDI**	**Sem 1**	**Sem 2**	**Total**	**Mean SAIDI**
2006	132	140	164	4.03	80	86	97	5.16
2007	119	123	152	6.09	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	8.52	109	100	111	9.65
2011	164	158	172	10.31	109	100	111	10.57
2012	177	177	176	32.28	110	111	113	44.97
2013	177	177	176	26.34	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	14.22	112	109	114	15.39
2018	171	167	173	10.30	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

Appendix B.

A.2 Weather DATA: CDD and rainfall.

Daily temperature and rainfall data comes from the National Institute of Climatology (INSIVUMEH), having data since 2001 in most cases. Also, we combine it with monthly rainfall register from Institute of National Statistics (INE). Although combining both sources we get 51 weather stations, we have to discard some because of long missing data. For estimation years 2011-2014 we use 43 stations in DEOCSA and DEORSA supplying areas, and for Census in 2018, 39. Camantulul and Todos Santos weather stations did not have observations along that year, and Cuilco and Tikal lack off many observations, too.

In order to collapse data at monthly level, 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: observations are missing for 11 or more days during the month; or observations are missing for a period of 5 or more consecutive days during the month. In case this condition is fulfilled, we assume a missing value and then we estimate it using historical data. Cooling Degree Days (CDD) is calculated as the difference between mean temperature and 18 degrees. If negative, the value of CDD for that day is zero. Once obtained a daily measure of CDD we average by month and year. Similarly process is done with rainfall, but we get monthly rainfall and we add it up to have total rainfall by year.

Considering that almost all departments have weather stations -except Totonicapán-, we get department CDD and rainfall values by averaging weather station data that belongs to the same department. For Totonicapán, we estimate it by combining data from two weather stations that belongs to the same climate region: Labor Ovalle and Huehuetenango. Those stations are highlighted in figure A.2b and we give them the same weight.

The criteria for assigning weather stations to municipalities has been twofold: similarities in weather conditions and distance. For the first criteria, a subset of stations is chosen if they are in the same state or in its border. Then, from that subset, the closest to the district is assigned. Closeness is defined as the distance from the capital city of each municipality to each weather station.[23]

Once the 43 stations are collapsed by department, we can compare these estimations with NASA database.[24] Reasonably, summary statistics that come from point estimates –weather stations– has more variability.See Table A.1

A.3 Quality data

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 A.5 resumes the original CNEE data. Notice that these summary statistics (unweighted) are very similar of those in Table 2 (weighted).

[23]The centroid would have been another option, but many municipalities have convex shape, meaning that the centroid could be outside of the district itself

[24]See https://disc.gsfc.nasa.gov/