The Welfare Effects from Electricity Theft: Evidence from Brazil

Gabriel Richter Marcelo Sant'Anna André Trindade *

January 25, 2020

PRELIMINARY & INCOMPLETE—PLEASE DO NOT CITE

Abstract

Electricity theft is a serious issue in much of the developing world. It can have an impact on generation decisions, grid reliability, global emissions, among others. However, few papers have been written on this topic, mainly due to the lack of data. Having access to detailed micro data from a major electric utility in Brazil, we estimate a structural model of electricity demand. Consumers make a discrete and a continuous decision: to steal or not, and how much to consume. We use the structural model to simulate counterfactual scenarios where: (i) electricity theft is not possible, and (ii) the utility can price discriminate across regions with high vs low theft rates. We find that eliminating theft increases consumer welfare.

Keywords: Electricity Theft, Discrete-Continuous Demand, Structural Estimation, Brazil, Welfare.

JEL codes: D12, L94, Q41

*Richter: FGV EPGE, gabriel.richter@outlook.com. Sant'Anna: FGV EPGE, marcelo.santanna@fgv.br. Trindade: FGV EPGE, andre.trindade@fgv.br. We thank Eduardo Coelho, Angela Gomes, Lavinia Hollanda, Rudi Rocha, Joisa Saraiva for useful comments and suggestions. The usual disclaimer applies.

1 Introduction

In most of the developing world, Non-Technical Losses (NTL) - electricity that is consumed but not billed (i.e. stolen from the grid[1]) - is pervasive and can be a serious problem. For example, the percent of electricity losses out of all the energy injected into the system is 14% in Africa and 17% in Latin America and the Caribbean (Jiménez et al. (2014))[2], but it can be much larger than that in some countries. There are several potential negative impacts on the energy sector as a consequence of NTL. Among them: 1) *a less reliable grid*, with more power outages, as demand becomes more unstable and difficult to predict; 2) *energy waste*, because consumers that steal energy pay a price of zero per MWh and do not internalize the generation and distribution costs; 3) *excessive prices* as the electric utilities typically pass on the cost of the stolen electricity to formal consumers; 4) *personal injuries* due to illegal connections that cause electric shocks; and so on. Moreover, there can be environmental costs due to the waste of energy. This is important to consider as it is forecasted that by 2035 the energy demand in the developing world will be twice that of the developed world (Wolfram et al. (2012)). Not addressing the issue of NTL can contribute to an increase in CO_2 emissions from electricity generation worldwide.

Despite the relevance of the question, there exists almost no literature studying NTL. That is noted in a recent survey (Lee et al. (2017)) where the authors list the issue of NTL among the key areas for future research. In particular, the authors call for a better understanding of how utilities and policymakers should respond to NTL. One of the reasons for the lack of past work is the difficulty in obtaining detailed micro data on electricity theft. See for example Jacobi and Sovinsky (2016) and Galenianos and Gavazza (2017) for other studies that discuss the difficulties of empirical work in markets with limited access to consumer data.

In this paper we try to address this hole in the literature and understand better the economics behind electricity theft. In particular we try to answer the following questions: (i) How much (if any) welfare is lost due to NTL? (ii) Can the amount of NTL be reduced with the use of differentiated tariffs? (iii) Which policy is more efficient to reduce the amount of NTL?

We obtained access to detailed data from a large electric utility in Brazil. Brazil is one of the countries in the world where the energy theft problem is the most severe (ANEEL,

[1]The formal definition of NTL is wider than electricity theft. It may include, for example, consumption mismeasurement due to faulty meters. Nevertheless it is understood that most NTL is composed of power theft, particularly in developing countries. Therefore, it is common to treat the two concepts as quasi-synonyms.

[2]These percentages include both Technical and Non-Technical Losses. Technical Losses are small amounts of energy that are lost naturally in the system due to transmission.

the sector regulator, reports over 33 TWh of stolen energy in 2018). The firm that we study provides electricity to over 10 million people, and is located in an area where power theft is particularly severe. We obtained and combined a number of different datasets: a long time series with aggregated data on billed consumption and NTL over time, a cross-section of all the "formal" consumers and, importantly, a panel with NTL information at the month and feeder level[3]. Since, by definition, there is no direct information available on consumers that engage in electricity theft, the disaggregated information at the feeder level is as good as one can get. At the feeder level, the utility knows how much electricity was transmitted, and the amount of technical losses. Therefore, NTL is just the difference between the two. This is the traditional and best available method to compute NTL (Lewis (2015)).

We start by documenting patterns in our data. We show that both the amount of billed consumption and NTL are highly seasonal, and that the socio-demographic characteristics of an individual affect its propensity to steal energy. Moreover, we present evidence that consumers respond to permanent price increases by migrating to informality.

In order to evaluate the welfare impact of different policies, we then set up and estimate a structural demand model. In the model, consumers make a discrete and a continuous decisions. First they decide if they want to be formal consumers of the firm (paying the full price) or steal the product (and pay zero). Then, conditional on that decision, they decide how much electricity to consume. The trade-off that consumers face is clear: by moving to NTL they face a price of zero for each unit of electricity and hence are able to increase their utility from consumption. On the other hand they incur in a non-pecuniary fixed cost (which represents the costs of the illegal connection, lost benefits from not being a formal consumer, etc).

We use the primitives from the model to simulate different counterfactual scenarios. We start by removing the possibility of theft from the choice set of consumers and, at the same time, reducing the electricity price in a way that the revenues from the utility firm do not change. We find that welfare for the average consumer goes up by 15.9 R$ per month. We are currently working on evaluating the impact of introducing differentiated tariffs for the different municipalities.

Relevant Literature

There has been a strong interest recently in studying the electricity sector in developing countries. Examples of questions being asked are: the economic effects of electrification, the relation between the income distribution and demand for electricity, among others. For example, Lipscomb et al. (2013) and Costa and Gerard (2018) look at the case of Brazil,

[3] Feeder are power lines that connect electricity from a substation to the final consumer.

McRae (2015) at Colombia, Gertler et al. (2016) study Mexico, Allcott et al. (2016) and Burlig and Preonas (2016) focus on India, and Auffhammer and Wolfram (2014) on China. See Lee et al. (2017) for a recent survey on the literature on electrification in developing countries.

There is no work that we are aware, however, that looks at the welfare costs from electricity theft or that evaluates the relative efficiency of different policies to mitigate that problem. The only work in economics that looks at NTL is Smith (2004b) that does a cross-country comparison, and Min and Golden (2014) that look at the relation between the political cycles and energy theft.

In this paper we estimate a discrete-continuous demand for electricity model. Several other papers have tried to empirically understand how consumers make decisions in this sector. For example, Ito (2014) use spatial discontinuities to provide evidence that consumers respond to electricity average price and not marginal, McRae and Meeks (2016) use a survey to ilicit consumer information about price schedules, and Deryugina et al. (2017) use a difference-in-differences matching estimator to measure quantity responses to changes in prices. However, the closest papers to ours are those that estimate a structural econometric model of electricity demand, namely Dubin and McFadden (1984), Reiss and White (2005), and McRae (2015). In particular, we also estimate a discrete-continuous model like Dubin and McFadden (1984), although in their case the discrete decision is which appliances to purchase while in our case is whether to steal energy or be a formal customer.[4]

There is also a small literature on nonpaying consumers of public utilities, although with a focus in the water sector. For example, Szabo (2015) analyzes the residential water sector in South Africa, estimates a structural model, finds that the policy of giving a free water allowance is suboptimal and derives the optimal nonlinear water schedule. Szabó and Ujhelyi (2015) use an experimental design in the same setting to evaluate the impact of water education campaigns.

In the next section we describe the relevant instutional details. In section 3 we detail the different datasets that we have available, and present descriptive statistics and figures. Then, in section 4 we introduce and estimate our empirical model. The recovered primitives are then used to simulate different counterfactual scenarios, which we do in section 5. Finally, in section 6, we conclude.

[4]Other examples of discrete-continuous demand models in sectors other than electricity are Smith (2004a) and Magnolfi and Roncoroni (2016).

2 Institutional Details

In 2017 the total electricity consumed in Brazil was 467 TWh, making the country one of the 10 largest in the world. The total installed capacity in the same year was over 157 GW, roughly 60% of which was hydropower and the remaining mostly a combination of natural gas, biomass and nuclear (EPE (2018)).

There are around 50 different local monopolies that distribute electricity in Brazil. Most of them are privately owned but several are public (state owned). The largest 5 distributors, in terms of the number of customers served are, in order: Cemig, Eletropaulo, Coelba, Copel, and Light (EPE (2018)).

The sector is regulated by *Agência Nacional de Energia Elétrica* - ANEEL, which is supervised by the Ministry of Mines and Energy. The consumer price of electricity is regulated. Up to 1993 there was a single electricity price for all of Brazil. From that point onwards, the regulated price was allowed to vary across utilities - but not within. The idea is that the different tariffs reflect the heterogeneity across utilities in terms of productive efficiency, demand conditions, and so on. The residential price varies with the quantity consumed[5]. Some low-income consumers qualify for a lower "social rate". The discount in that case will be a negative function of the quantity consumed, but it can go up to 65% (for low income, low consumption households). In 2015, ANEEL introduced a system of "tariff flags" that change each month and introduce some variation in the final price that consumer pays, depending on the color of the flag (red, yellow or green). The color of the flag represents the general conditions of the electric generation system and the goal is for consumers to internalize part of the differences in generation costs over time and adjust consumption accordingly.

There are two types of losses in the distribution of electricity: technical (TL) and non technical (NTL). The former are just natural losses inherent to the activity of transporting electricity from one place to another, and are a function of the quality of the infrastructure. The latter mostly consists of electricity theft or measurement error. In 2018 the total electricity lost in Brazil, as a percent of the electricity injected in the system, was 14%, roughly equally divided across TL (7.5%) and NTL (6.6%). The total amount of NTL in that year was above 33 TWh. Those percentages are a little misleading because most of the NTL take place in the residential sector. Therefore, while it is natural for the denominator of TL to be the amount of electricity injected, the usual approach is to compare NTL with the total amount of electricity in that sector. In that case, the percentage of NTL goes up to 14.3%). Again, this hides some heterogeneity: at least 7 utilities have NTL higher than

[5] The different intervals currently are: up to 50 kWh, from 51 to 300 kWh, from 301 to 450 kWh, and above 450 kWh.

30% of residential consumption. That is the case for the firm that we will study, which is responsible for almost 29% of the total amount of NTL in Brazil (ANEEL (2019)).

In Brazil many of the areas with high amounts of NTL are also areas dominated by organized crime and militias. See Merenfeld (2017) for more on that relation.

3 Data

We draw information from multiple data sets: (i) long time series of prices, formal aggregated consumption (billed consumption), number of clients, total number of households within municipalities, and NTL; (ii) a cross-section of the utility's residential clients; (iii) a panel of feeders. This section describes the data sets and explains how key variables were defined.

3.1 Data Preparation

Prices

We use data on electricity prices obtained from the Brazilian Electricity Regulatory Agency (ANEEL, in Portuguese). The data set covers monthly ex and post-tax average[6] retail prices for the period January 2003 to July 2019. We convert all prices to January 2003 R$ using the Extended National Consumer Price Index (IPCA, in Portuguese). Figure 3 depicts how prices evolved over the years and reveals two features of our data set. First, real prices declined steadily from 2003 to 2015, except for a few isolated spikes. This decreasing real price behavior is consistent with interventions made by prior Brazilian governments that kept electricity prices artificially low. Second, taxes play a significant role on providing additional price variation that is useful for the model estimation. Our base model electricity price variable is the average post-tax retail price with one lag[7].

Households

We would like information on the number of households for each one of the 31 Rio de Janeiro's counties the utility serves. The Brazilian Institute of Geography and Statistics (IBGE, in Portuguese) is the agency responsible for reporting the official population count. It provides annual estimates of the total population for each one of those counties, but not for the total number of households. We recover the latter by proceeding in the following way: first, we

[6]Average ex-tax prices are calculated by dividing the total revenue from electricity distribution by the total amount of billed electricity. Average post-tax prices are calculated by dividing the total revenue, including taxes, from electricity distribution by the total amount of billed electricity

[7]We believe the lag best describes how households react to price changes: they only become aware of the change after receiving their utility bill. We also used a price variable without a lag and the results were very similar.

use the 2000 and 2010 Census obtained from IBGE to find the average household size in each county in both years. Next, we use linear interpolation to obtain estimates for the period 2003 to 2009. From 2011 to 2019, we assume the average household size is constant as of 2010. Our proxy to the total number of households is obtained by dividing the IBGE's population estimate in each county and year by its respective average household size.

Formal Consumption and Number of Clients

We obtained information on formal residential electricity consumption and number of clients from two sources. The first data set contains monthly information on (i) the aggregated amount of electricity the utility sold to residential clients, and (ii) the number of residential clients it had. This piece of information come from ANEEL and comprises the period January 2003 to July 2019. The second data set is a cross-section of the utility's residential clients on November 2016, with information on each household consumption and spatial coordinates for that month. Figure 6 shows the histogram of consumption using this cross-sectional data. We find that approximately 18.4% of all residential clients had zero consumption. This could be due to different reasons, such as households that own one or more vacation houses, vacant properties that are available for rent, etc. We dropped these clients from our analysis by assuming that the number of clients with non-zero consumption was 18.4% less than that reported by ANEEL at each month[8]. We are left with approximately 3.2 million clients that consume on average 194.47 kWh/month, as shown in Table 1. The aggregate formal consumption per household variable we use on the base model is the ratio of the aggregated amount of electricity sold to residential clients to the number of residential clients with non-zero consumption.

Non-Technical Losses (NTL)

We obtained time series data on NTL from the utility firm. The data set covers information from January 2008 to January 2015. The aggregate informal consumption per household variable we use on the base model is the ratio of the amount of NTL to the difference between the total number of households and the number of formal residential clients with non-zero consumption[9].

Feeders

We use proprietary data on feeders provided by the utility. The data set is a panel of feeders containing feeder-month level information for the year of 2017. The variables available are the amount of electricity (i) generated, (ii) billed, and (iii) used for public lightening; as

[8]This may not be a very reasonable assumption if the number of clients that do not consume any electricity changes over months and/or years, though.

[9]Our assumption is that all households consume electricity.

well as the amount of (iv) technical losses, and (v) non-technical losses. The data set does not contain any geospatial information on each feeder. Notwithstanding, we recover their latitude and longitude through an additional data set, a cross-section of feeders' smaller components, technically called *trafos*, for November 2016. This alternative data set is useful for two reasons: first, it has the latitude and longitude of each *trafo*. Second, it links each *trafo* to its respective feeder. We were then able to obtain the geospatial position of each feeder by defining its latitude and longitude as the average latitude and longitude of their respective *trafos*.

3.2 Descriptive statistics

Tables 2 and 3 provide some reduced form evidence of how consumers respond to changes in electricity prices. In Table 2 we report the results of OLS regressions of the log of aggregated formal consumption against different price variables using the time series data. All columns include a month fixed effect to control for seasonality in consumption. In column 3 we use the log of electricity price as the price variable. Alternatively, we use the log of the price of electricity with taxes in column 4. As mentioned previously, taxes provide some useful additional price variation (at least graphically). Columns 5 and 6 both use the log of the deflated price of electricity, but only the latter includes taxes. The results suggest all price coefficients have the expected negative sign and are close to each other. Since we are using a log-log specification, we can interpret the coefficients as elasticities. In column 3, for instance, a 1% increase in the price of electricity is associated to a -0.20% reduction in the aggregated formal consumption. We also found a positive and statistically significant time trend coefficient.

Table 3 reports the results of OLS regressions of the log of aggregated formal consumption and the log of NTL against the same price variables shown in Table 2, but using the panel of feeders instead of the time series. We used feeder fixed effects in columns 1 to 8. We also removed seasonality from the formal electricity consumption (billed consumption) and NTL data . Again, the price elasticities for formal consumption (columns 1 to 4) have the expected sign and are statistically significant. We note that they are bigger in magnitude than those we found in Table 2, though. Regarding NTL, the positive and statistically significant coefficients found in columns 5 to 8 suggest that an increase on electricity prices is associated to higher amounts of NTL.

Consumers respond to higher prices by migrating to electricity theft

In the beginning of 2011, ANEEL (the regulator) changed the rules of who could qualify for the social tariff. First, the criteria to qualify became stricter, and second, it stopped being automatic and started requiring additional documental evidence in order to qualify.[10] Consumers that failed to re-register for the social tariff were gradually kicked out of the program throughout the year and automatically moved into the regular tariff. Figure 1 shows the number of total clients, and number of clients with a regular tariff, before and after this change in policy (during 2011). The number of clients with a regular tariff increased dramatically during 2011, followed by a partial decrease. This is consistent with the anedoctal evidence that many people only became aware of the change after seeing the increase in their bill. The fact that this reduction only partially offset the initial increase is also consistent with the stricter criteria applied after the change. This led to a change in the number of total residential consumers. Since we do not expect any consumer to stay without power (and since consumers cannot buy electricity from any company other than Light), this effect is likely driven by consumers migrating to electricity theft. In Figure 8, we observe an increase in the share of stolen electricity in 2012 (decrease in the share of formal consumption) which is consistent with this migration to electricity theft.

Consumption seasonality and increasing variance over the years

Electricity consumption is expected to be seasonal in Brazil. Since we use data from a utility that operates in a Brazilian state where summer temperatures are high enough to justify the usage of air conditioning, but winters are not cold enough to create demand for heating systems, we would expect consumption to be above average over summer only (end of December to end of March). This electricity usage behavior is precisely what we see in Figures 4 and 5. The former shows that formal aggregate consumption spikes near January and goes down abruptly near the middle of the year. The latter is a box-plot also for formal consumption from January to December for the whole period of study (January 2003 to July 2019). It shows that consumption starts to increase consistently in October and keeps above average over January, February and March. On the other hand, it starts to decrease around April and goes beyond average from June to September, when it starts to increase again. Figure 4 also shows an interesting pattern of formal electricity consumption: it became more volatile over the years. Figure 7 suggests that NTL presents a similar pattern, in the sense that we can observe both seasonality and an increase in volatility over the years for informal consumption.

[10] To be more specific, before the new rule the Social Tariff was automatically applied for all consumers with a total quantity under 80 kWh. Families that consumed between 80 and 220 kWh could still benefit from the social tariff, but they would have to show evidence of low income. With the new policy, every single consumer between 0 and 220 kWh would only qualify if they showed evidence of low income *and* were registered in the national list of people under social programs ("Cadastro Único"). A high income family with consumption under 80 kWh would qualify for the social tariff before but not after the change.

4 Empirical Model

4.1 Framework

Household i chooses either formal electricity consumption ($j = 0$) or electricity theft ($j = 1$) in market t. In this exercise, a market is the entire region covered by the electricity company in a given month. Conditional on their choice, utility is quasi-linear on the consumption of electricity q and the numeraire good m:

$$u_{ijt}(q, m) = \phi_{ijt}(q) + m. \tag{1}$$

We work with a quadratic specification for $\phi_{ijt}(\cdot)$:

$$\phi_{ijt}(q) = \theta_{1,jt} q - \frac{1}{2}\theta_2 q^2 + \eta_{jt} + \varepsilon_{ijt}, \tag{2}$$

where individual preference shocks ε_{ijt} are iid distributed across households, markets and choices with extreme value distribution $F_{\varepsilon_{ijt}}(\varepsilon) = \exp\left(-\exp\left(-\sigma\varepsilon\right)\right)$.

Letting y_{it} denote household income and substituting (2) into (1):

$$u_{ijt} = \theta_{1,jt} q - \frac{1}{2}\theta_2 q^2 + y_{it} - p_{jt} q + \eta_{jt} + \varepsilon_{ijt}, \tag{3}$$

where p_{jt} is the price payed by household per each unit of load consumed. The price of electricity in the formal consumption choice, p_{0t}, is just the post-tax electricity retail price p_t, while $p_{1t} = 0$.

The quadratic specification for the quasi-linear utility yields a linear demand for electricity. In the case of formal consumption the linear electricity demand will be:

$$q_{it,j=0} = \frac{\theta_{1,0t}}{\theta_2} - \frac{1}{\theta_2} p_t. \tag{4}$$

Informal consumers face zero marginal electricity price and therefore consume up to satiation:[11]

$$q_{it,j=1} = \frac{\theta_{1,1t}}{\theta_2}. \tag{5}$$

[11] The existence of such point of satiation is a feature of the quadratic quasi-linear utility and its associated linear demand. Alternatively, we could generate a constant elasticity (ξ) demand with

$$\phi_{ijt}(q) = \theta_{1,jt} q^{\frac{\xi - 1}{\xi}} + \eta_{jt} + \varepsilon_{ijt}.$$

Although in this case we would need a slightly different strategy to deal with informal consumption, since there would be no satiation.

4.2 Identification

As in most electricity retail markets, price changes here are set by the regulator. These price changes are typically directly related to the availability of water in hydro reservoirs and the rain patterns in their associated basins. Most electricity that supplies the Brazilian grid is hydro generated, so the availability of electricity is sensitive to seasonal and decennial weather patterns that are closely monitored by the regulator. These price changes are therefore unrelated to current demand shocks.[12] We consider therefore prices as exogenous in formal demand equation (4), conditional on controlling for month fixed effects.

Using the parameters recovered from the demand equation and quantities consumed we are able to identify the consumer surplus (net of fixed ameneties/penalties from formal v. informal consumption) generated by formal and informal choices.[13] Finally we use variation in extensive margin choices between formal v. informal consumption and the extreme value structure of errors to recover parameters governing the fixed components of the quasi-linear utility represented by η_{jt}.

4.3 Estimation

We propose a 2-step estimation approach. In the first step, we estimate the electricity demand using formal and informal consumption data (intensive margin), which allow us to recover $\theta_{1,0t}$, $\theta_{1,1t}$ and θ_2. In the second step we estimate parameters governing the relative desirability of formal versus informal consumption η_{jt}.

1st-step: Using aggregate formal consumption data (per household), estimate:

$$q_{0t} = \delta_{0,t} - \alpha p_t + \nu_t, \tag{6}$$

where $\delta_{0,t}$ includes trend and seasonality effects. We can use the estimates from the equation above to recover:

$$\hat{\theta}_2 = \frac{1}{\hat{\alpha}},$$
$$\hat{\theta}_{1,0t} = \frac{\hat{\delta_{0,t}} + \hat{\nu}_t}{\hat{\alpha}}.$$

[12]One could argue that past demand shocks could play a role in determining current price changes, as a past demand shock could alter the stock of water in the reservoirs. So if demand shocks are correlated over time this could be a potential source of concern for the price exogeneity assumption. For the moment, we abstract from this possibility.

[13]This is the part of (2) that varies with q.

Using informal consumption q_{1t} (per capita) we can recover $\theta_{1,1t}$:

$$\hat{\theta}_{1,1t} = q_{1t}\hat{\theta}_2.$$

2nd-step: Using the estimates from the first step, compute the consumer surplus in each type of consumption:

$$\hat{\psi}_{0t} = \hat{\theta}_{1,0t}q_{0t} - \frac{\hat{\theta}_2}{2}q_{0t}^2 - p_t q_{0t}, \tag{7}$$

$$\hat{\psi}_{1t} = \hat{\theta}_{1,1t}q_{1t} - \frac{\hat{\theta}_2}{2}q_{1t}^2. \tag{8}$$

Extreme value distribution on preferences imply the usual logit regression:

$$\log(P_{0t}) - \log(P_{1t}) = \frac{1}{\sigma}\left(\hat{\psi}_{0t} - \hat{\psi}_{1t}\right) + \frac{1}{\sigma}\left(\eta_{0t} - \eta_{1t}\right). \tag{9}$$

Here we can normalize $\eta_{1t} = 0$ and estimate:

$$\log(P_{0t}) - \log(P_{1t}) = \gamma\left(\hat{\psi}_{0t} - \hat{\psi}_{1t}\right) + \gamma_t + e_t, \tag{10}$$

where γ_t includes trend and seasonality components. In order for the exercise above to be consistent we need ν_t to be independent from e_t and ε_{ijt}.

4.3.1 Welfare

We consider two sets of welfare measures. The first set is comprised by the already computed surpluses $\hat{\psi}_{0t}$ and $\hat{\psi}_{1t}$. These only take into account the surpluses (in R$) generated by the consumption of electricity, that is they are net of fixed differences between formal and informal consumption. For example, they abstract from the disutility of informal consumption associated with the threat of punishment or the utility associated with service reliability of formal consumption.

Our second measure of welfare, also measured in R$, is given by the expected utility prior to the realization of idiosyncratic shock ε_{ijt} and is standard in the discrete choice framework:

$$E_{\varepsilon_{ijt}}\left[\max_{j=0,1} u_{ijt}\right] = \frac{1}{\gamma}\ln\left(e^{\gamma\hat{\psi}_{0t}+\gamma_t} + e^{\gamma\hat{\psi}_{1t}}\right), \tag{11}$$

or alternatively,

$$E_{\varepsilon_{ijt}}\left[\max_{j=0,1} u_{ijt}\right] = \frac{1}{\gamma}\ln\left(e^{\gamma\left(\hat{\psi}_{0t}-\hat{\psi}_{1t}\right)+\gamma_t} + 1\right) + \hat{\psi}_{1t}. \tag{12}$$

4.4 Estimation with micro data

[To be completed]

4.5 Estimation Results

Table 4 reports estimates and standard errors from the linear demand equation (6) for the formal sector. As expected, formal per household consumption is sensitive to price changes. One stardard deviation increase in prices leads to a quarter of a standard deviation decrease in formal consumption. There are also relevant trend and seasonality effects.

We then use the estimated price coefficient from linear demand together with formal and informal consumption series to build consumer surplus estimates (7) and (8). Since our informal consumption measure (based on NTL) is shorter than our formal consumption series, we use predicted informal consumption for the missing months. That is, we regress our series of informal consumption on trend and time dummies and use adjusted values for the missing observations.

Table 5 reports estimates from the second step. The consumer surplus difference coefficient has the expected sign: When the surplus from formal consumption increases relative to the surplus of electricity theft, there is an increase in the share of formal consumers. The trend and seasonality controls capture differences in benefits/penalties from formal versus informal consumption that are not related to differences in the ammount of electricity consumed.

5 Counterfactual Results

We use the estimated model parameters to analyse four different counterfactual scenarios displayed in Table 6.

In the first two scenarios (CF1 and CF2), we just promote 10% electricity price changes. These price changes imply moderate changes in consumption of formal households (intensive margin), but have little effect in the share of formal consumers (extensive margin). Therefore there seems to be little space for price decreases to lead an increase in formalization that could compensate in part the loss revenue from such price drop.

In the last two scenarios, we consider an exogenous ban on theft. In CF3 we keep electricity price as in baseline, so revenue goes up by 26%. In CF4 we decrease electricity price in order to keep utility's revenue as in baseline. In this scenario prices could drop by 27%. Note that this understates the potential for price decreases in this scenario of no-theft as it does not take into account the savings for the utility company from the averted NTL.

6 Conclusion

Electricity theft is a significant phenomenon throughout the world, particularly in developing countries, with potential consequences for energy costs and the environment. In this paper, we use detailed micro data from electricity theft in Brazil to evaluate the potential benefits from different policies to reduce energy theft. Our data includes a panel of NTL at the feeder and month levels. Information at this level of disaggregation has not been used in academic research previously, to the best of our knowledge.

We find that eliminating electricity theft would increase welfare for the average consumer. The electric utility would lower its price if there were no NTL, which would benefit the consumers that are currently paying. Consumers that are stealing electricity would be worse off if theft was not an option (by revealed preference). However, this effect is smaller than the disutility of paying a positive price, as those consumers would still have additional non-pecuniary benefits from being formal clients (no risk of being caught, no cost of making an ilegal connection, etc).

The policies that we evaluate may have distributional consequences, as some consumers would gain and others would lose. The results from our paper indicate that it would be more efficient to eliminate theft and redistribute income with lump-sum transfers or through progressive income taxation. In other words, it is possible to redistribute the gains from eliminating NTL in a way that makes every consumer better off.

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.

ANEEL (2019): "Perdas de Energia Elétrica na Distribuição," .

Auffhammer, M. and C. D. Wolfram (2014): "Powering up China: Income distributions and residential electricity consumption," *American Economic Review*, 104, 575–80.

Burlig, F. and L. Preonas (2016): "Out of the darkness and into the light? development effects of rural electrification," *Energy Institute at Haas WP*, 268.

Costa, F. and F. Gerard (2018): "Hysteresis and the Welfare Effect of Corrective Policies: Theory and Evidence from an Energy-Saving Program," *National Bureau of Economic Research*.

Deryugina, T., A. MacKay, and J. Reif (2017): "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation," *NBER Working Paper*.

Dubin, J. A. and D. L. McFadden (1984): "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica: Journal of the Econometric Society*, 345–362.

EPE (2018): "Anuário Estatístico de Energia Elétrica 2018, ano base 2017," .

Galenianos, M. and A. Gavazza (2017): "A structural model of the retail market for illicit drugs," *American Economic Review*, 107, 858–96.

Gertler, P. J., O. Shelef, C. D. Wolfram, and A. Fuchs (2016): "The demand for energy-using assets among the world's rising middle classes," *American Economic Review*, 106, 1366–1401.

Ito, K. (2014): "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing," *American Economic Review*, 104, 537–63.

Jacobi, L. and M. Sovinsky (2016): "Marijuana on main street? Estimating demand in markets with limited access," *American Economic Review*, 106, 2009–45.

Jiménez, R., T. Serebrisky, and J. Mercado (2014): "Sizing electricity losses in transmission and distribution systems in Latin America and the Caribbean," *Published by Inter-American Development Bank*.

Lee, K., E. Miguel, and C. Wolfram (2017): "Electrification and Economic Development: A Microeconomic Perspective," .

Lewis, F. B. (2015): "Costly 'Throw-Ups': Electricity Theft and Power Disruptions," *The Electricity Journal*, 28, 118–135.

Lipscomb, M., A. M. Mobarak, and T. Barham (2013): "Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil," *American Economic Journal: Applied Economics*, 5, 200–231.

Magnolfi, L. and C. Roncoroni (2016): "Political connections and market structure," *Working paper.*

McRae, S. (2015): "Infrastructure quality and the subsidy trap," *American Economic Review*, 105, 35–66.

McRae, S. and R. Meeks (2016): "Price perception and electricity demand with nonlinear tariffs," .

Merenfeld, L. (2017): "O impacto do controle de território sobre o furto de energia elétrica: O caso do Rio de Janeiro," *Dissertacão de Mestrado UFRJ.*

Min, B. and M. Golden (2014): "Electoral cycles in electricity losses in India," *Energy Policy*, 65, 619–625.

Reiss, P. C. and M. W. White (2005): "Household electricity demand, revisited," *The Review of Economic Studies*, 72, 853–883.

Smith, H. (2004a): "Supermarket choice and supermarket competition in market equilibrium," *The Review of Economic Studies*, 71, 235–263.

Smith, T. B. (2004b): "Electricity theft: a comparative analysis," *Energy policy*, 32, 2067–2076.

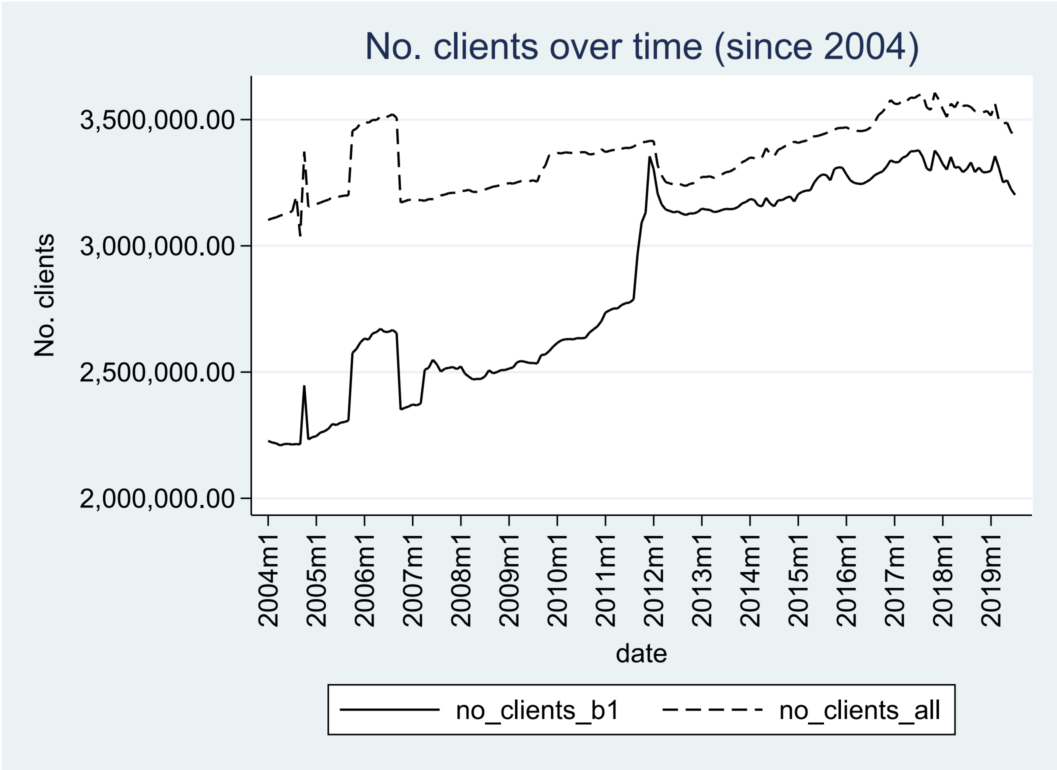
Szabo, A. (2015): "The value of free water: analyzing South Africa's free basic water policy," *Econometrica*, 83, 1913–1961.

Szabó, A. and G. Ujhelyi (2015): "Reducing nonpayment for public utilities: Experimental evidence from South Africa," *Journal of Development Economics*, 117, 20–31.

Wolfram, C., O. Shelef, and P. Gertler (2012): "How will energy demand develop in the developing world?" *Journal of Economic Perspectives*, 26, 119–38.

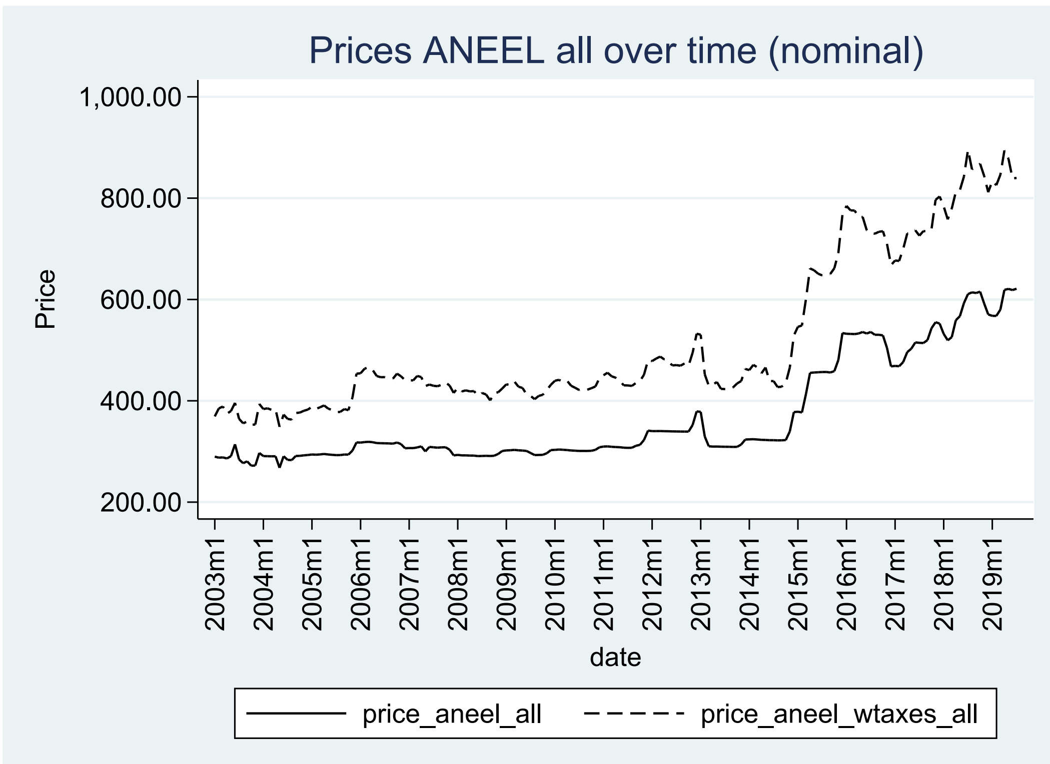
7 Figures

Figure 1: Number of (Formal) Clients Over Time



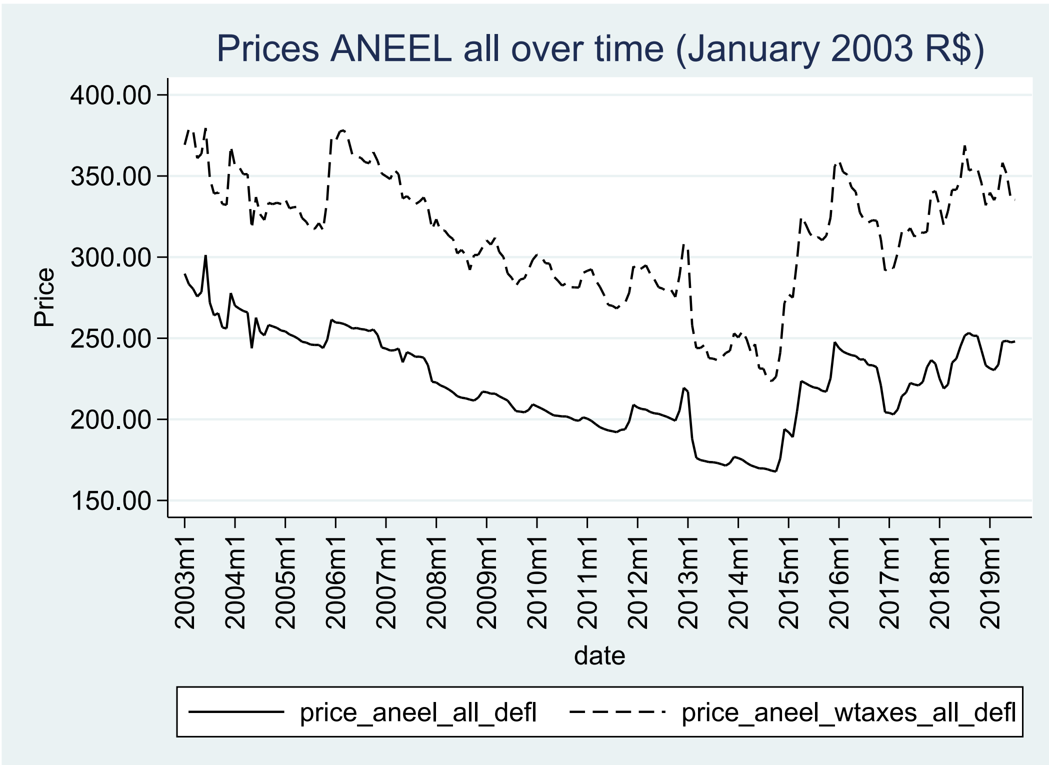
Note: This figure shows the number of the utility's formal residential clients for the period January 2004 to July 2019. The solid line represents B1 clients, who pay the regular electricity tariff scheme. The dashed line represents all the residential clients, which is the sum of B1 clients and those who pay the social tariff scheme. Data comes from ANEEL.

Figure 2: Electricity Prices Over Time (nominal)



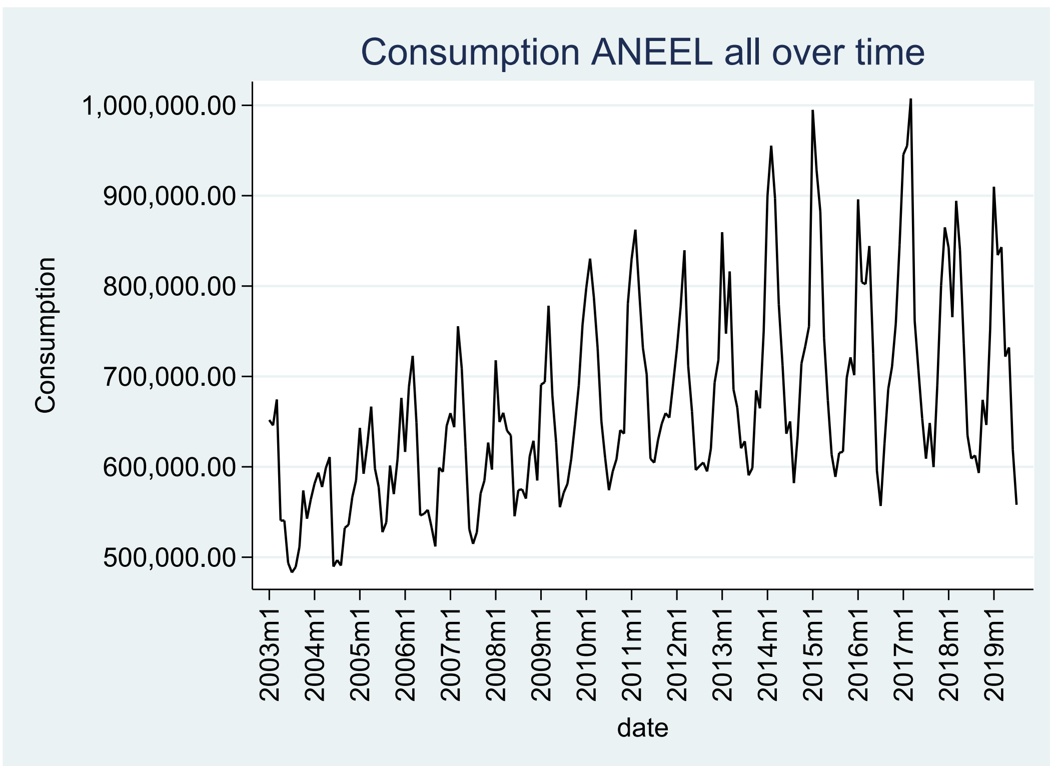
Note: This figure shows electricity nominal ex and post-tax average residential retail prices for the period January 2003 to July 2019. Average ex-tax prices are calculated by dividing the total revenue from electricity distribution by the total amount of billed electricity. Average post-tax prices are calculated by dividing the total revenue, including taxes, from electricity distribution by the total amount of billed electricity. Prices are measured in R$/mWh. Data comes from ANEEL.

Figure 3: Electricity Prices Over Time (deflated)



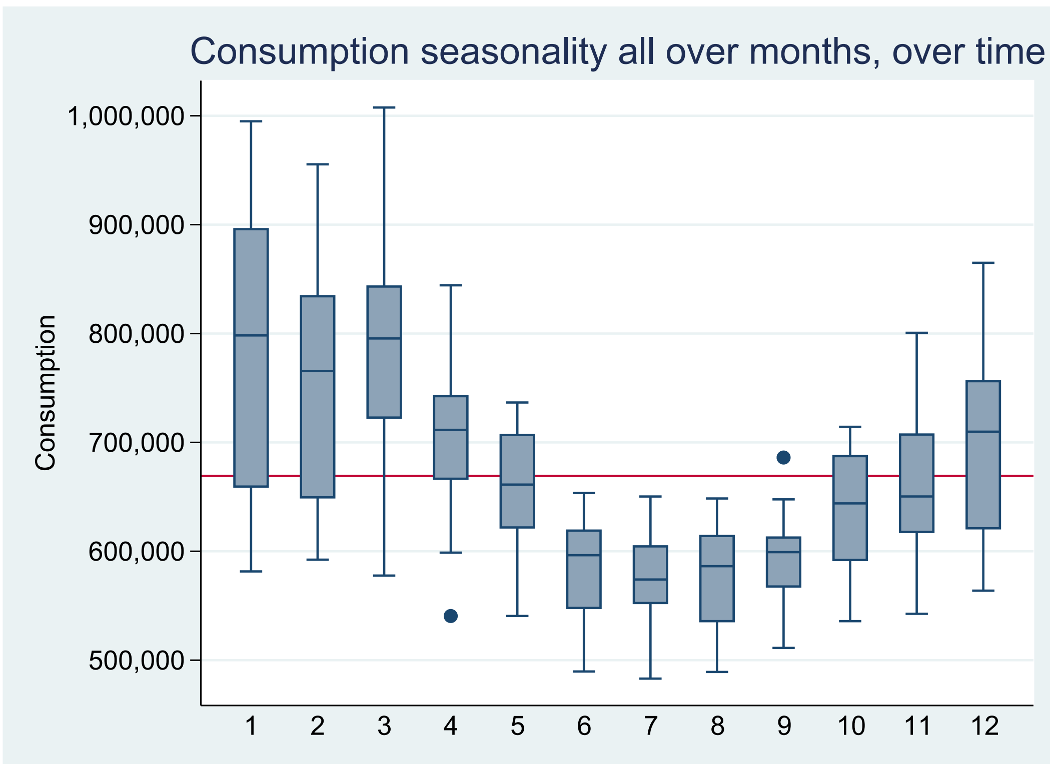
Note: This figure shows electricity real ex and post-tax average residential retail prices for the period January 2003 to July 2019. All prices were converted to January 2003 R$ using the Extended National Consumer Price Index (IPCA). Average ex-tax prices are calculated by dividing the total revenue from electricity distribution by the total amount of billed electricity. Average post-tax prices are calculated by dividing the total revenue. including taxes, from electricity distribution by the total amount of billed electricity. Prices are measured in R$/mWh. Price data comes from ANEEL and IPCA comes from IBGE.

Figure 4: Formal Electricity Consumption Over Time



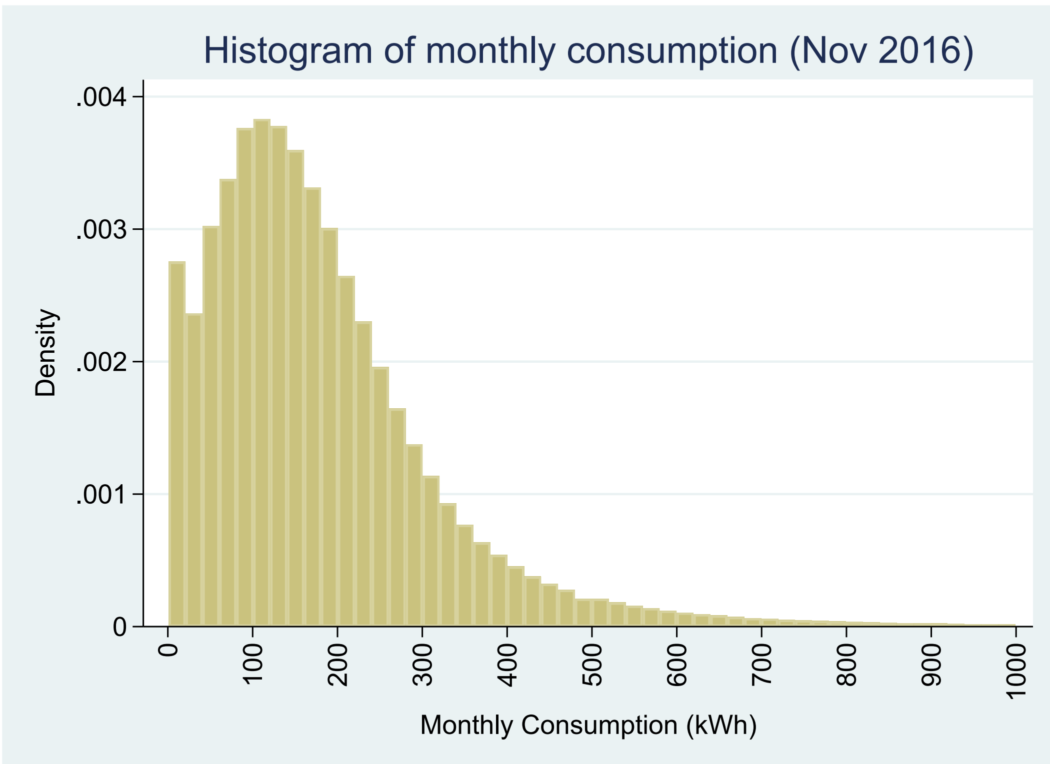
Note: This figure reports the amount of billed electricity consumed by all residential clients for the period January 2003 to July 2019. Consumption is measured in mWh. Data comes from ANEEL.

Figure 5: Consumption Seasonality (Box-Plot)



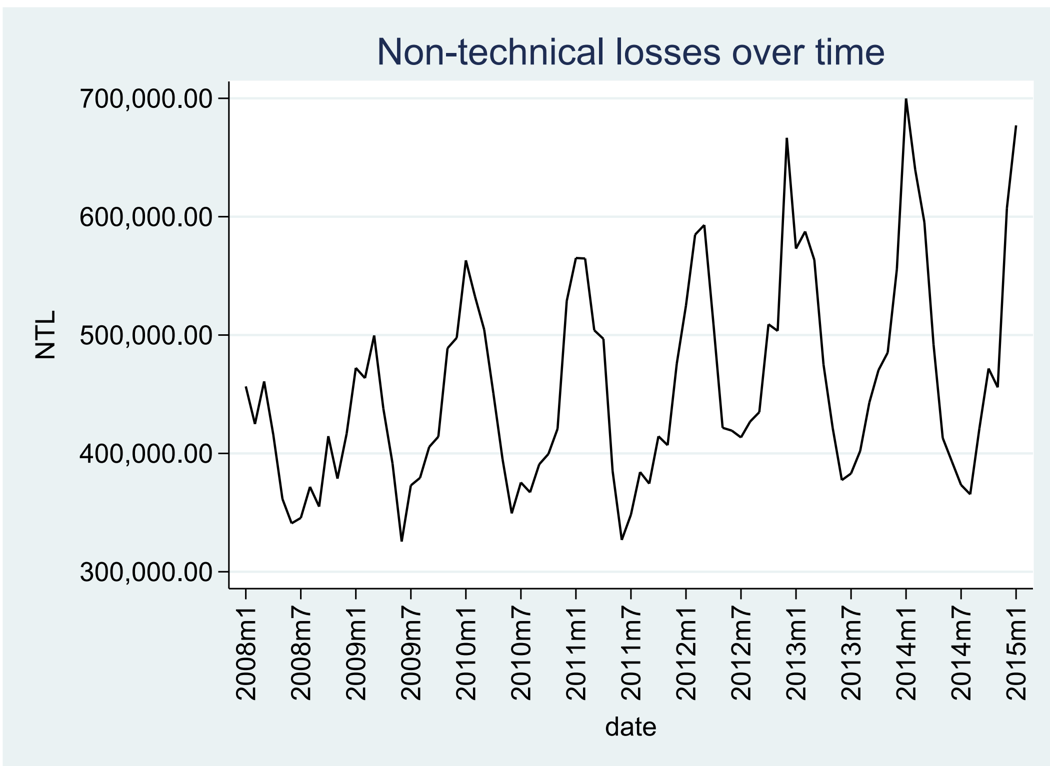
Note: This box-plot was constructed using formal residential electricity consumption (billed electricity) data for the period January 2003 to July 2019. Consumption is measured in mWh. The solid horizontal line represents the average consumption over the whole period and equals 669,230 mWh. Data comes from ANEEL.

Figure 6: Histogram of Monthly Household Consumption (kWh



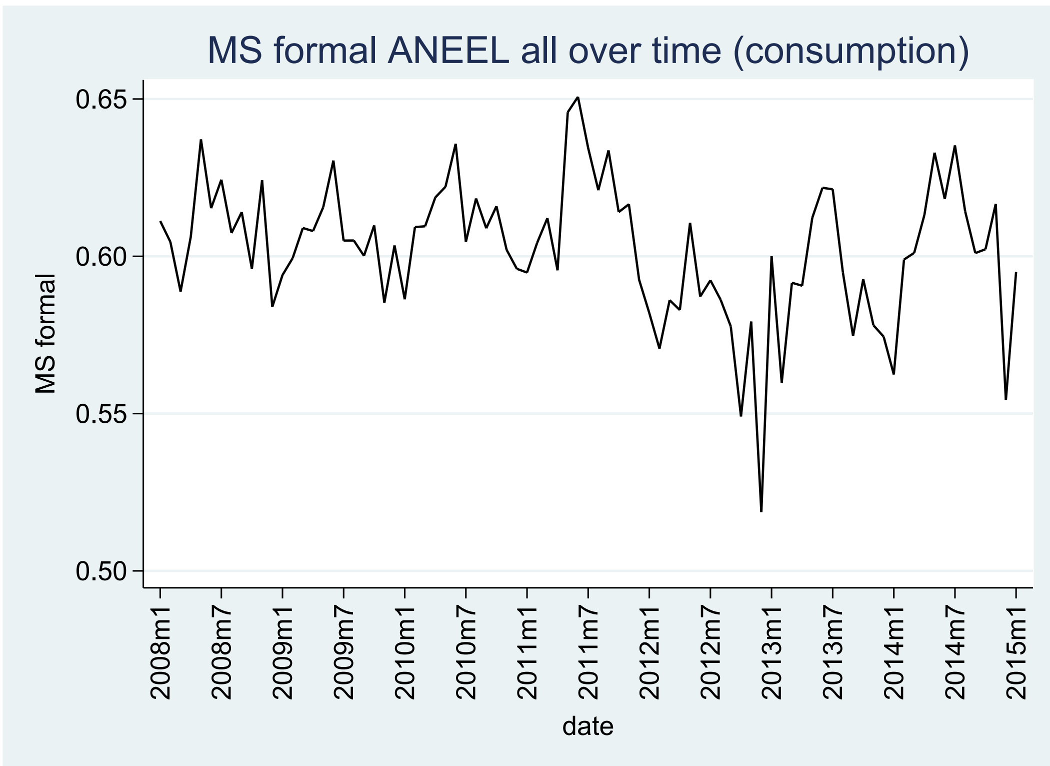
Note: This histogram was constructed using a cross-section of electricity consumption for the universe of residential clients for the month of November of 2016. We dropped those clients with zero consumption, which represent approximately 18.4% of the total. We also dropped consumers with over 1000 kWh/month, which represent less than 0.6% of the sample. Data was provided by the utility.

Figure 7: Non-Technical Losses Over Time



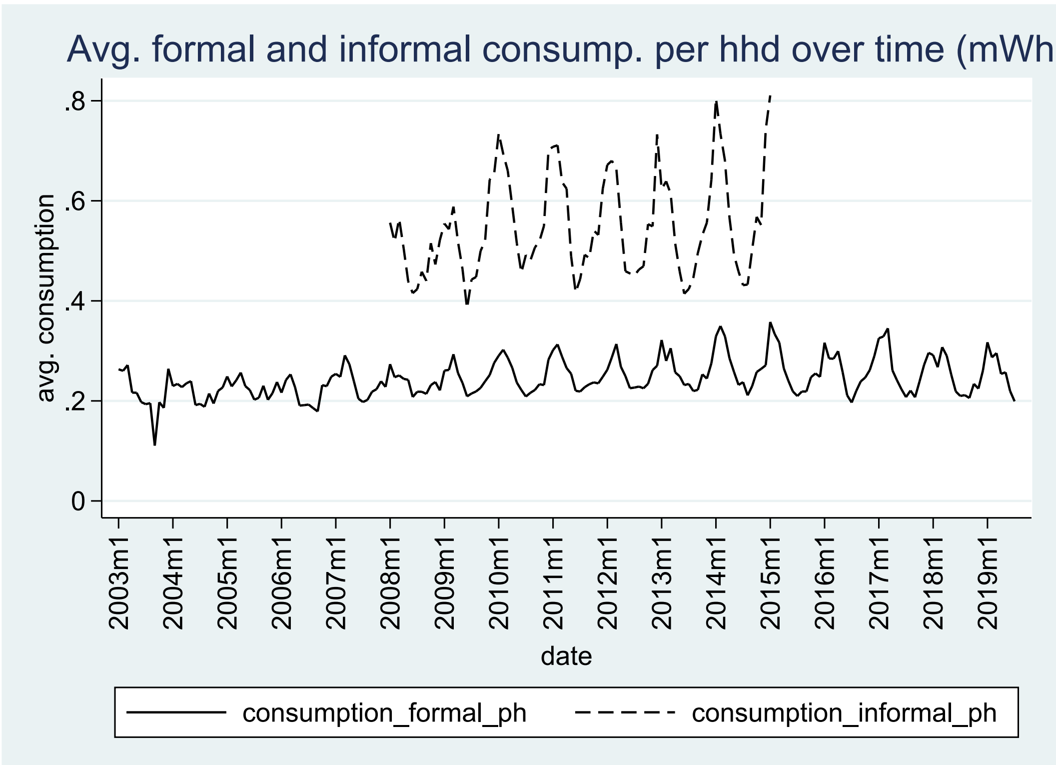
Note: This figure shows the evolution of NTL for the period January 2008 to January 2015. NTL are measured in mWh. Data was provided by the utility.

Figure 8: Market Share of total residential consumption that is billed



Note: This figure reports the market share of formal residential consumption. The market share variable was defined as the ratio of the billed consumption to the sum of the billed consumption and the amount of non-technical losses. Billed consumption data comes from ANEEL. NTL data comes from the utility.

Figure 9: HH consumption



Note: This figure shows the average formal electricity consumption per household for the period January 2003 to July 2019 and average informal electricity consumption per household from January 2008 to January 2015. The average formal consumption per household variable is defined as the ratio of the aggregated amount of electricity sold to residential clients to the number of residential clients with non-zero consumption. The average informal consumption per household variable we use is defined as the ratio of the amount of NTL to the difference between the total number of households and the number of formal residential clients with non-zero consumption. Residential clients with zero consumption accounts for approximately 18.4% of the total number of residential clients the utility had in November 2016. Number of clients data comes from ANEEL and the utility. Number of households data comes from IBGE. NTL data was provided by the utility.

8 Tables

Table 1: Statistics on cross-section of consumers (Nov 2016)

	Obs	Mean	Std. Dev.	Min	Max
HH consumption (mWh)	3,176,331	194.47	400.38	1	100452
Clients not using grid (dummy)	3,893,201	0.184	0.388	0	1

This table reports descriptive statistics at the household level for all the formal clients of the electric utility. In particular, we report on two variables: 1) "Household consumption of electricity", conditional on being positive, and 2) a dummy variable that is equal to one if the consumption was zero in that month.

Table 2: Time Series Regressions

VARIABLES	(1) Log(Cons)	(2) Log(Cons)	(3) Log(Cons)	(4) Log(Cons)	(5) Log(Cons)	(6) Log(Cons)
Log(Price)			-0.2032***			
			(0.0334)			
Log(Price w/ Taxes)				-0.1827***		
				(0.0347)		
Log(Deflated Price)					-0.2396***	
					(0.0368)	
Log(Deflated Price w/ Taxes)						-0.2104***
						(0.0383)
Time Trend		0.0190***	0.0280***	0.0279***	0.0161***	0.0174***
		(0.0010)	(0.0017)	(0.0019)	(0.0010)	(0.0010)
Constant	13.5548***	13.1372***	14.1326***	14.0775***	14.4981***	14.3850***
	(0.0280)	(0.0275)	(0.1653)	(0.1807)	(0.2104)	(0.2283)
FE Month	Yes	Yes	Yes	Yes	Yes	Yes
Observations	199	199	199	199	199	199
R-squared	0.4961	0.8276	0.8564	0.8500	0.8597	0.8518

Note: Each column reports coefficients from an OLS regression, with standard errors in parentheses. The dependent variable in all columns is the log of the formal residential electricity consumption. In column (3), we use the log of electricity price as the price variable. In column (4), we use the log of the price of electricity with taxes. In column (5), we use the log of the deflated price of electricity. In column (6), we use the log of the deflated price with taxes. Deflated prices were converted to 2003 R$ using the Extended National Consumer Price Index (IPCA, in Portuguese). All columns include a month fixed effect to control for seasonality in consumption. Sample covers the period January 2003 to December 2019. Prices and consumption data come from ANEEL. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Panel Regressions: with Feeder FE

VARIABLES	(1) Log(Cons)	(2) Log(Cons)	(3) Log(Cons)	(4) Log(Cons)	(5) Log(NTL)	(6) Log(NTL)	(7) Log(NTL)	(8) Log(NTL)
Log(Price)	-0.7068***				0.9169***			
	(0.0347)				(0.1061)			
Log(Price w/ Taxes)		-0.7857***				0.6759***		
		(0.0388)				(0.1188)		
Log(Deflated Price)			-0.7913***				1.0565***	
			(0.0393)				(0.1201)	
Log(Deflated Price w/ Taxes)				-0.8850***				0.7352***
				(0.0445)				(0.1363)
Constant	10.6412***	11.4164***	10.5009***	11.3246***	-0.5090	0.7500	-0.4881	0.9792
	(0.2162)	(0.2560)	(0.2118)	(0.2561)	(0.6612)	(0.7836)	(0.6475)	(0.7840)
FE Feeder	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Desazonalized	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,872	19,872	19,872	19,872	18,628	18,628	18,628	18,628
R-squared	0.0223	0.0220	0.0218	0.0212	0.0044	0.0019	0.0045	0.0017
Number of feeder_idn	1,671	1,671	1,671	1,671	1,672	1,672	1,672	1,672

Note: Note: Each column reports coefficients from an OLS regression, with standard errors in parentheses. In columns (1)-(4), the dependent variable is the log of the formal residential electricity consumption. In columns (5)-(8), the dependent variable is the log of NTL. In columns (1) and (5), we use the log of electricity price as the price variable. In columns (2) and (6), we use the log of the price of electricity with taxes. In columns (3) and (7), we use the log of the deflated price of electricity. In columns (4) and (8), we use the log of the deflated price with taxes. Deflated prices were converted to 2003 R$ using the Extended National Consumer Price Index (IPCA, in Portuguese). All columns include a feeder fixed effect. We removed seasonality from formal electricity consumption and NTL data. Sample is a panel of feeders containing feeder-month level information for the year of 2017. Price data comes from ANEEL. Panel data comes from the utility. *** p<0.01, ** p<0.05, * p<0.1

Table 4: First step estimation results

	q_{0t}
p_t	-0.000258***
	(3.55e-05)
$\delta_{0,t}$:	
year trend	0.00250***
	(0.000273)
month of year dummies	Yes
Observations	198
R-squared	0.800

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

Notes: This table reports first step estimates (equation 6) by OLS.

Table 5: Second step estimation results

	$\log(P_{0t}) - \log(P_{1t})$
$\hat{\psi}_{0t} - \hat{\psi}_{1t}$	0.000864**
	(0.000399)
γ_t :	
year trend	0.00663
	(0.0133)
month of year dummies	Yes
Observations	55
R-squared	0.250

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

Notes: This table reports second step estimates (equation 10) by OLS for the post 2015 period.

Table 6: Counterfactual Results

	Baseline	CF1 (price up 10%)	CF2 (price down 10%)	CF3 (no theft)	CF4 (no theft, fix rev)
Quantity formal/hh (kWh)	0.245	0.237	0.253	0.245	0.267
Share formal	79.4%	79.3%	79.5%	100%	100%
Price formal (R$)	312.1	343.3	280.9	312.1	227.2
ψ_{0t} (R$)	118.9	111.4	126.6	118.9	140.5
Expected revenue/hh (R$)	60.3	64.0	56.2	76.0	60.4
Consumer Surplus/hh (R$)	1505.6	1500.1	1511.4	1505.6	1521.6
Delta Consumer Surplus/hh (R$)	-	-5.6	5.7	0.0	15.9

Notes: This table reports counterfactual estimates for four different scenarios. Averages over time for the variables listed in the first column in all cells.