Pricing and Informality:
Evidence from Energy Theft in Brazil

Davi Resende Gabriel Richter Marcelo Sant'Anna

André Trindade *

May 15, 2024

PRELIMINARY & INCOMPLETE—PLEASE DO NOT CITE

Abstract

In certain settings, goods can be consumed outside of formal markets (e.g.: theft, counterfeit, or illegal sharing of subscriptions). When the share of informality is large, firms' pricing decisions can be substantially affected, as the extensive margin - customers migrating to informal consumption - makes demand more elastic. We study this question in the context of electricity theft in Brazil, where stolen energy can represent roughly 40% of the total formal market. We use detailed micro data from a major electric utility on consumption and theft to estimate a structural model where consumers choose if they want to be formal or informal and then, how much to consume. For identification, we leverage a natural experiment from 2011, where prices increased permanently to a set of consumers. We use the model to simulate counterfactual scenarios where: (i) electricity theft is not possible, and (ii) the firm can price discriminate across regions with high vs low theft rates. We find that the presence of informality increases the elasticity of demand from 0.21 to 0.72. Price discrimination can be an effective tool for firms pricing under informality.

*Resende: FGV EPGE, daviresende@hotmail.com Richter: Stone, gabriel.richter@outlook.com. Sant'Anna: FGV EPGE, marcelo.santanna@fgv.br. Trindade: Nova School of Business and Economics, Universidade NOVA de Lisboa, and FGV EPGE, andre.trindade@novasbe.pt. We thank Angela Gomes, Benjamin Lockwood, Corinne Low, Eduardo Coelho, Francisco Costa, Lavinia Hollanda, Rudi Rocha, Joisa Saraiva, Mike Abito, and participants from the EPGE Applied Microeconomics Workshop and the Wharton BEPP Applied Economics Workshop for useful comments and suggestions. This work was funded by Fundação para a Ciência e a Tecnologia (UIDB/00124/2020, UIDP/00124/2020 and Social Sciences DataLab - PINFRA/22209/2016), POR Lisboa and POR Norte (Social Sciences DataLab, PINFRA/22209/2016), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Grant #001.

1 Introduction

Informality[1] has a sizeable presence in the world economy. The share of global employment that is informal has been estimated to be 61% - or about two billion people ((ILO) (2018)). Moreover, the OECD (OECD/EUIPO (2019)) measured the volume of international trade in counterfeit and pirated products and suggests that it could amount to as much as USD 509 billion. Economists have noticed this pattern and recent papers have analyzed the impact of informality on trade, labor markets, firm dynamics, and housing supply (e.g. Ulyssea (2018), Dix-Carneiro et al. (2021), Gerard and Gonzaga (2021), Rocha et al. (2018), Guedes et al. (2023)). However, less attention has been given to how informality affects firms' micro decisions such as *pricing*.

When consumers have access to a good outside of the formal market (for example via theft, piracy, counterfeiting, illegal sharing of passwords, or others), the profit of the (formal) firm supplying the good changes in two ways. *First*, revenues go down because a fraction of potential consumers stop paying the firm. *Second*, the average cost per paying consumer goes up as the firm needs to produce quantities for both paying and informal consumers. Examples under this framework are diverse but typically share those two forces. Users of a subscription service (e.g. Netflix, Spotify) may decide to share their password with friends, against the rules of the platform. This will increase the costs of the platform as the average viewership per paying customer goes up. When a counterfeit good floods the market it may cannibalize sales of the corresponding luxury brands, but also harm the value of that brand and increase advertising costs. Water theft has been estimated to range between 30% and 50% of total supply. High levels of non-revenue water (NRW) are detrimental to the financial viability of water utilities as those firms have to quantities of water that account for the theft.

Importantly, the firm can incentivize consumers to move away from informality by choosing lower prices. This implies that the actual demand curve faced by the firm will become more elastic when compared to a scenario without informal consumer markets, as consumers move both along the formal demand curve and between formal and informal goods. At the end of this section, we set up a toy model to formalize this idea.

We empirically study the pricing problem of a firm operating under informality in the context of electricity theft. In most of the developing world, Non-Technical Losses (NTL)—electricity that is consumed but not billed (i.e. stolen from the grid[2])—is pervasive and can

[1] According to Dell'Anno (2022), economists typically use *informality* as a synonym for the shadow, unofficial, hidden, black, or underground economy.

[2] 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.

be a serious problem. The percentage 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)[3], but it can be much larger than that in some countries or regions. 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 is almost no literature studying electricity theft. 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.

Specifically, in this paper we ask the following questions: How does the possibility of theft in this market affect the demand curve elasticity and optimal pricing decisions by firms? What are the welfare effects from informality? Can price discrimination be a useful tool to mitigate inefficiencies in this setting?

To answer those questions, we use detailed data from a large electric utility in Brazil, one of the countries in the world where the energy theft problem is the most severe (ANEEL, 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. Our data contains detailed consumer level information on capacity, generation, and price paid, at the month level, between 2010 and 2022. Moreover, we have a panel with NTL information at the month and feeder level[4]. Since, by definition, there is no direct information available on consumers that engage in electricity theft, the disaggregated

[3]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.

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

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 providing evidence of a causal link between pricing and the consumer decision to become informal. Then, we 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 our estimates to disentangle elasticity along the demand curve of formal consumers from the elasticity along the extensive margin. With the primitives from the model, we then simulate different counterfactual scenarios. In some of these exercises, we just promote an exogenous change in electricity prices. In others, we remove the possibility of theft from the choice set of consumers and, at the same time, reduce the electricity price in a way that the revenues/profits from the utility firm do not change.

One of the empirical challenges that we face is that a consumer's decision to buy in the informal sector is mostly affected by long term pricing and not by day-to-day price variation. To address this challenge, we leverage a natural experiment from 2011 where electricity prices increased exogenously and *permanently* to a subset of consumers[5].

We find that a 10% permanent increase in the price level will result in an increase in the share of informal consumers of 1.6 percentage points. Therefore, the aggregate demand curve that the firm faces becomes significantly more elastic when accounting for the informality margin (-0.72 vs -0.21 without informality). The informality switch would effectively constrain a monopolist firm to price significantly lower. Moreover, we find that price discrimination (e.g. discount tariffs for low income households) can be an important tool for the firm as they reduce the share of informality with smaller inframarginal losses on formal consumers.

[5]Without this natural experiment, we would have to assume how customers for price expectations into the future based on past prices.

1.1 A Toy Model

We now discuss the problem of the firm in the presence of sizeable mass of informal consumers. We derive the optimal pricing equation and compare it to the traditional monopolist problem. In this toy example, we abstract away from all household heterogeneity.

Let p be the price that the firm charges formal consumers, $\sigma(p)$ be the share of formal consumers as a function of price (we assume that $\sigma'(p) < 0$), $d(p)$ is the per-household demand conditional on being formal ($d'(p) < 0$) and $\bar{d}$ the per-houshold demand of informal consumers, which does not depend on price ($d(p) \leq \bar{d}$ for all p). Finally, c is the constant marginal cost. Then, firm profit is given by

$$\pi(p) = p\sigma(p)d(p) - c\sigma(p)d(p) - c(1 - \sigma(p))\bar{d}.$$

Notice that revenue is accrued only on the goods sold to formal consumers. The first cost term refers to the total cost of providing the good to formal consumers, while the second cost term refers to the cost of the informal good.

The monopolist case is a useful benchmark to compare the firm problem with and without informality. The first order condition of the monopolist problem is

$$\underbrace{\sigma(p)d(p) + p\sigma(p)d'(p) - c\sigma(p)d'(p)}_{\text{trad. FOC}} + \underbrace{p\sigma'(p)d(p)}_{\text{evasion adj.}} - \underbrace{c\sigma'(p)(d(p) - \bar{d})}_{\text{theft adj.}} = 0 \tag{1}$$

The first part of the FOC is exactly the same as faced by a traditional monopolist facing demand $d(p)$. The second term highlighted is an adjustment for the lost revenue from switchers betweem formal and informal status. The third term is an adjustment, which takes into account the difference in consumption between formal and informal households.

Passing to elasticities and writting the Lerner Index for the firm we can find the optimal pricing equation.

$$\frac{p - c}{p} = \frac{1}{|\xi_d|} + \frac{\xi_\sigma}{|\xi_d|} - \frac{\xi_\sigma m(p)c}{|\xi_d|p}, \tag{2}$$

where $\xi_d := pd'(p)/d(p)$ is the elasticity of per-houshold consumption conditional on formality, $\xi_\sigma := p\sigma'(p)/\sigma(p)$ is the elasticity of the formality share, and $m(p) := (d(p) - \bar{d})/d(p)$ is a measure of the relative size of the difference between formal and informal consumption.

The first term on the RHS is the traditional monopolist mark-up term. The second term in the RHS is a (negative) markup adjustment due to evasion of formal consumers, while the third term is a (negative) markup adjustment due to the fact that informal consumption is higher than formal consumption. Both "new" terms contribute to a lower markup than the one derived under the traditional paradigm.

1.2 Relevant Literature

The literature in economics and marketing studying informality in consumer markets is scarce, mainly due to the difficulties in obtaining good data. Notable exceptions are work on digital piracy (Lu et al. (2020), Li et al. (2021)), and counterfeits (Qian et al. (2015), Qian (2014)). The latter documents that counterfeit products can substitute branded products but also work as an advertising mechanism for them.

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 (2021) look at the case of Brazil, 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.

However, there is no work that we are aware of that looks at the welfare costs from electricity theft. The only work in economics that looks at NTL includes Smith (2004b), that does a cross-country comparison, Min and Golden (2014), which look at the relation between the political cycles and energy theft, and Burgess et al. (2020), who describe how the wide tolerance to governmental subsidies, theft, and nonpayment, in countries where electricity is treated as a right, can undermine universal access to reliable electricity.

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. (2020) 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 it is whether to steal energy or be a formal customer.[6]

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

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 institutional 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. We present our results in section 5. The recovered primitives are then used to simulate different counterfactual scenarios, which we do in section 6. Finally, in section 7, we conclude.

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[7]. 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

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

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 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.

2.1 The 2011 change in social tariffs

In Brazil, some consumers of electricity have access to a social tariff, which gives them discounted rates over the regular tariff. The magnitude of the discount is a decreasing function of the consumption: 65% discount for consumption up to 30 kWh/month; 40% discount for consumption between 30 and 100 kWh/month; 10% discount between 100 and 220 kWh/month; and no discount after that point.

While the discount brackets have been constant over time, the rules to access the social tariff changed. In 2010 and before every household with consumption below 80 kWh/month had automatic access. Above that limit, documents attesting that a household was "low income" were required. From 2011 onwards, everyone had to submit proof of low income in order to qualify for the social tariff. This, in fact, resulted in a large number of consumers being excluded from the social tariff and facing huge price increases.

3 Data

3.1 Data Sources and Preparation

We draw information from multiple data sets. This section describes the data sets and explains how key variables were defined. We organize the information in the following two groups: 1) data that only exhibits time series variation, and 2) panel data.

3.1.1 Time Series Data

Formal consumption and number of consumer units. We obtained aggregated information on billed residential electricity consumption and the number of residential customers of the utility, per month and consumption bracket. This data is provided by the Brazilian Electricity Regulatory Agency (ANEEL) and was collected for the period between December 2010 and November 2012.

Non-technical losses (NTL). We use aggregated data on non-technical electricity losses (i.e., a proxy for theft) from ANEEL for the period between 2008 and 2020. The data covers the entire service area of the utility and varies at the monthly level.

Households. We use annual estimates of the population of municipalities in Brazil produced by the Brazilian Institute of Geography and Statistics (IBGE). In particular, we consider the 31 municipalities served by the utility. Under the assumption that all households use electricity, this number serves as a proxy for the total number of households in the utility's service area. The difference between this number and the total number of formal customers reported by the utility is our metric for the quantity of informal consumers in this market.

Prices. We use aggregated electricity price data obtained from ANEEL for the period between 2010 and 2019. In Brazil, marginal prices depend on the level of consumption and eligibility to a lower "social tariff." We track monthly the full price schedule used by the utility. We use the National Consumer Price Index (INPC) to convert all prices to Brazilian reais at December 2021 values. On top of this base price, there is also an increment in the unit price of electricity due to the tariff flag, an increment that has been uniformly charged for all households in the country since January 2015 to internalize changes in the generation costs of electricity faced by utilities in different periods of the year. These data also come from ANEEL and exhibit monthly variation. In summary, our price variable is the unit price of electricity, plus the tariff flag, with variation per month, consumption bracket, and "social tariff" eligibility.

3.1.2 Panel Data

Our main source of panel data is the Geographic Database of the Utility (BDGD), which gathers information annually sent by energy utilities in Brazil to ANEEL. This dataset provides details on the operation of utilities, including information on all physical components of their infrastructure, as well as technical and commercial data, such as electricity consumption per household and volume of energy losses. ANEEL makes this data publicly available, and we obtained access to the information from the utility of interest for the years 2017 to 2022.

Formal consumption and number of consumer units. Through BDGD, we can track the monthly billed electricity consumption per household and the number of households registered in the utility's customer base throughout each year. It is also possible to identify the category in which these households (i.e. regular or social) were registered with the utility in December of each year; and the installed capacity, which is the total power capacity of all electrical devices installed and ready for use in each household. Consumers are not identified, so it is not possible to track them between different years. However, there is a component of the utility's infrastructure, called a "building point", which connects the low-voltage electrical network to households, which is reported in the BDGD and is identified by its geographical coordinates. There is a many-to-one map between households and the building points to which they are associated, which allows us to monitor the formal electricity consumption and the number of consumers of the utility over the years, at the month and building point level.

Non-technical losses (NTL). The BDGD's non-technical loss data is provided at the feeder level, another component of the utility's infrastructure, which distributes electricity from substations (where the voltage of the electricity is reduced, so it can be consumed in households) to different building points. The feeders also have an identifying code, which allows us to track them over time, and there is also a many-to-one map between building points and feeders. In summary, we have a panel of non-technical losses at the feeder-month level.

Demographics. We consider four demographic variables: i) income; ii) household size; iii) universe of households; and iv) crime. Income and household size data are obtained from the 2010 Brazilian Population Census. For the income variable, we consider the average per capita income of households in each census tract. For the household size, we consider the average number of residents per household, also at the census tract level. The universe of households data is obtained through the 2010 and 2022 Brazilian Population Census and provides the number of households that exist (i.e. potential customers of the utility) in each census tract. Finally, we have geographical information on crime from Fogo Cruzado Institute for the year 2019, which consists of geographical polygons delimiting territories under the influence of different criminal groups, such as militias, drug trafficking, and even regions of dispute between these groups.

To accommodate this information in our model, we aggregate them at the feeder level. To relate census tracts and feeders, we first identify in which census tract each building point is located. Then, for the income and household size information, we average the building points data per feeder. For the universe of households, we summed the information from the building points of each feeder. Finally, regarding crime data, we calculated the average of the information from building points (an indicator variable for whether the building point is located in a census tract with the presence of criminal groups), also at the feeder level.

3.2 Descriptive Statistics

Table 1 describes the main variables used in the empirical analysis. For each variable, we include the mean, the three quartiles (p25, p50, and p75) and the min and max values. Panel A includes information regarding the formal consumers at a month level (consumption, price paid, and an indicator equal to one if the consumer has access to the social tariff). Panel B describes the distribution of the amount of theft (i.e., Non-Technical Losses) reported across circuits. In panel C, we report the statistics for the variables that we use from the Census information: income and household size. Finally, panel D includes the variable "crime area indicator", which is equal to one in census areas that are controlled by traffic, a militia group, or are areas under dispute.

Table 2 uses the cross-section variation in the share of consumers that are formal and regresses the log of that variable on a set of covariates. This reduced form OLS regression will help us identify which variables to include in our structural model. Column (1) estimates the regression using data that is aggregated at the circuit level, while column (2) includes data at the neighborhood level. We find that income has a positive and significant impact on the share of formal consumers. This can be for two reasons: 1) high income households have lower price sensitivity and therefore do not mind paying the larger prices required to be formal consumers, or 2) high income households have in general higher disutility of being informal. Moreover, we find that traffic and militia areas seem to negatively affect the probability that a specific area has a high share of formal consumers.

3.2.1 Consumers Respond to Higher Prices by Migrating to The Informal Sector

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.[8] 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 3 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 anecdotal 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 the utility), this effect is likely driven by consumers migrating to electricity theft.

Figure 4 shows the impact from the change in policy on the number of customers with a social tariff. We find that the largest drop occurs in customers with a marginal rate of 35% and 60% of the regular prices. This is expected as those were the groups in the lower quantity brackets and were, therefore, the ones that were directly affected by the new requirement.

3.2.2 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 Figure 1. It shows that formal aggregate consumption spikes near January and goes down abruptly near the middle of the year. 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

[8] 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.

September, when it starts to increase again. Figure 1 also shows an interesting pattern of formal electricity consumption: it became more volatile over the years.

4 Model

4.1 Framework

We propose a model in which households decide on two nested margins. First, they need to decide between being formal consumers, paying for the electricity the assigned prices, or being an informal consumer, that is, stealing electricity from the grid through illegal connections. We call this the extensive margin of energy consumption in our setting. The second decision is on the intensive margin, that is, about how much energy to consume given a choice of formality status. Naturally, the two decisions are linked. A higher price will decrease formal demand and also decrease the implied indirect utility of the formal status. We first discuss the intensive margin of energy consumption, followed by a discussion about the extensive margin decision and how the two margins are explicitly connected.

4.1.1 Intensive margin: Electricity demand

Conditional on their formality choice, we assume household i utility in market[9] t is quasi-linear on the consumption of electricity, q:[10]

$$v_{it}(q, m) = \phi_{it}(q) - pq, \tag{3}$$

We work with a constant demand elasticity specification for $\phi_{it}(\cdot)$:

$$v_{it}(q, p) = \theta_{it} q^{\frac{\xi-1}{\xi}} - pq, \tag{4}$$

where ξ is the implied demand elasticity and we allow for household specific demand shifters θ_{it}. Equation 4 gives the shape of utility under the two modes of consumption, formal and theft. The difference between the two modes is that under formal consumption households must pay the postive price set by the utility. While under theft, households pay a zero price to the utility.

[9] A market here is defined as a region covered by the electricity utility in a period of time, a month or a year depending on the empirical exercise.

[10] The quasi-linearity assumption seems reasonable in our setting, because the electricity consumption makes up a small part of households income. During the period 2017–2018, for instance, the average household in the Southeast region of Brazil—where the utility we have data on provides its service—spent approximately 2.3% of their monthly income with electricity.

Formal demand. The quasi-linear utility specification in (4) implies that formal electricity demand will be given by

$$d_{it}(p) = \left(\frac{\xi}{\xi - 1} \frac{1}{\theta_{it}} \right)^{-\xi} p^{-\xi}. \tag{5}$$

Informal demand. Given this utility specification, informal demand would be infinity under a zero price with no additional restriction. We leverage on household demand capacity data and restrict household informal demand to $\bar{q}_{it}$.

4.1.2 Extensive margin: Energy theft

Now we turn to the initial extensive margin decision. Households choose their formality status: either formal electricity consumption ($j = 0$) or electricity theft ($j = 1$). We assume that the utility of a given formality status is a function of their consumer surplus (indirect utility) under that status. Specifically, the utility of formality status j is

$$u_{ijt} = \beta\psi_{ijt}(p_t) + \eta_{jt} + \varepsilon_{ijt},$$

where ε_{ijt} has the usual e.v. distribution, η_{jt} is a fixed formality utility shifter and

$$\psi_{ijt}(p_t) = \begin{cases} v_{it}(d_{it}(p_t), p_t) & \text{for } j = 0 \\ v_{it}(\bar{q}_{it}, 0) & \text{for } j = 1. \end{cases} \tag{6}$$

We thus propose a model of household decision about energy theft that is anchored in a more traditional model of energy consumption. The decision on energy theft is however flexible enough to acommodate observed and unobserved market-level shifters.

Our model specification yields typical logit conditional choice probabilities. The conditional choice of the formal status is

$$\mathbb{P}_i^t(formal) = \frac{\exp(\beta(\psi_{i0t}(p_t) - \psi_{i1t}) + \eta_{0t})}{1 + \exp(\beta(\psi_{i0t}(p_t) - \psi_{i1t}) + \eta_{0t})}, \tag{7}$$

where we normalized without loss $\eta_{1t} = 0$.

4.2 Identification

Intensive margin. 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.[11] We consider therefore prices as exogenous in formal demand equation (5), conditional on controlling for month of the year fixed effects.

Extensive margin. In principle, it is possible to derive implications for energy theft from our model by just using the short-run price variation we use to identify the intensive margin electricity demand. However, those changes are typically small, and transitory. Meanwhile the decision about energy theft is long-term and implies some irreversability. We thus seek here substantial permanent price variations to identify responses of the extensive margin that could be helpful to shed light on counterfactuals that involve also permanent policy shifts.

Therefore, in order to identify parameters governing the extensive margin, we leverage on a permanent policy shift that took place in 2011. Up to 2011, all households with consumption below 80 kwh had automatic access to the social tariff, and those between 80 and 220 had to submit proof of low income. From that point onwards the rule changed and everyone was suddenly required to submit proof of low income in order to qualify for the social tariff. This was a major tariff hike for a substantial set of households, that was perceived as permanent.

4.3 Estimation

We propose a 2-step estimation approach. In the first step, we estimate the electricity demand using formal consumption data (intensive margin). This first step is run using the panel of all formal households from 2017 to 2022. In the second step, we estimate the parameters governing the relative desirability of formal versus informal consumption, leveraging on the 2011 change in social tariffs.

[11] 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.

4.3.1 First Step: Formal electricity demand

Taking logs in the formal demand equation (5), we can write a regression equation that can be estimated from a panel regression under some mild conditions:

$$\log(q_{it}) = \gamma_i + \gamma_t - \xi \log(p_{it}) + \nu_{it}, \tag{8}$$

where p_{it} is the average household price, γ_i are individual fixed effects (in practice, building fixed effects), and γ_t are month fixed effects.

Given the demand elasticity (ξ) estimate, we can set the demand shifter, θ_{it}, in formal demand (equation 5) such that at t, demand is exactly as observed, that is, $d_{it}(p_{it}) = q_{it}$:

$$\theta_{it} = p_{it} \frac{\xi}{\xi - 1} q_{it}^{\frac{1}{\xi}}. \tag{9}$$

This first step then gives all we need to compute the monetary surpluses of being formal and informal (equation 6) for any counterfactual prices.

4.3.2 Second Step: Levaraging on the 2011 experiment for the extensive margin

The idea is to estimate model parameters governing the extensive margin, (β, η), based on empirical versions of equation (7) in pre-experiment period $t = 2011$ and post-experiment period $t + 1 = 2012$.

Since we only observe the formality share at aggregate levels, we base our estimation on an aggregate version of equation (7), for the pre- and post-experiment periods:

$$\mathbb{P}^t(formal) = \frac{N^t_{form}}{N} = \sum_{i \in I} \frac{\exp\Big(\beta(\psi_{0i}(p_t(q_{it})) - \psi_{1,i}) + \eta_0\Big)}{1 + \exp\Big(\beta(\psi_{0i}(p_t(q_{it})) - \psi_{1,i}) + \eta_0\Big)} w_i^t, \tag{10}$$

$$\mathbb{P}^{t+1}(formal) = \frac{N^{t+1}_{form}}{N} = \sum_{i \in I} \frac{\exp\Big(\beta(\psi_{0i}(p_{t+1}(q_{it})) - \psi_{1,i}) + \eta_0\Big)}{1 + \exp\Big(\beta(\psi_{0i}(p_{t+1}(q_{it})) - \psi_{1,i}) + \eta_0\Big)} w_i^t, \tag{11}$$

where N^t_{form} is the number of formal households in t, N is the total number of households from the Census, I is a set of observed households, and w_i^t is the unconditional population weight of household $i \in I$ in the experiment periods, which is assumed to be the same across experiment periods.

We now briefly discuss the expressions we use for the consumer surplus in each period (pre- and post-experiment) and formality status. First, we fix the demand shifter θ_{it} at the pre-experiment level.[12] We thus look at surplus variations implied purely by experimental price shifts. Second, as discussed in previous sections, in our setting the average price depends on current demand, thus the price used to compute the formal consumer surplus is a function of pre-experiment demand. The assumption here is that households see the tariff hike and choose formality status based on the impact of the tariff hike for their baseline consumption.

In order to use the system of equations (10) and (11) to recover the extensive margin parameters, we need to overcome two related challenges. First, we only observe aggregate data for the pre- and post-experiment periods. All our household level datasets are for later $T > t+1$ periods. Second, our observation of a household in the panel of formal consumers is by definition conditional on choosing to be formal.

In order to address the first challenge, since we do not observe a panel of households for the experiment periods, we need to leverage on the household panel for the later periods. First, we discuss how to recover individual weights using the observed distribution of individuals from a later period T. In the observed household dataset in T, conditional on formality, each individual weight is

$$\mathbb{P}[i|formal, T] = \frac{1}{N^T_{form}}.$$

By the Bayes' rule:

$$\mathbb{P}[i|formal, T] = \frac{\mathbb{P}[formal|i, T]\overbrace{\mathbb{P}[i|T]}^{=w_i^T}}{\underbrace{\mathbb{P}[formal|T]}_{=\frac{N^T_{form}}{N}}} = \frac{\mathbb{P}[formal|i, T]w_i^T N}{N^T_{form}}.$$

Therefore we can express the unconditional weight at T as

$$w_i^T = \frac{1}{N \times \mathbb{P}[formal|i, T]}.$$

This is helpful because although we do not observe $\mathbb{P}[formal|i, T]$ directly, we can compute this for a given parameter pair (β, η).

We are not over because in (10) and (11) we need the unconditional weight at t and not at T. We propose adjusting the unconditional weights from T to t such that we have an exact match for each consumption category at t.[13] Therefore, we assume:

[12] That is why we omit ψ's t subscript in (10) and (11).

[13] At the experiment periods, we observe aggregate number of households in different consumption categories as well as total energy demand in each category.

$$w_i^t = \gamma_c w_i^T \text{ for all } i \in c, \tag{12}$$

where c denotes a given consumption category and γ_c is a category specific adjustment factor.

In order to match the exact distribution of categories conditional on formality we observe in t, we must have that for each category c:

$$\sum_{i \in c} \mathbb{P}[formal|i,t] w_i^t = \frac{N_{form,c}^t}{N}, \tag{13}$$

where $N_{form,c}^t$ is the number of formal households in consumption category c at t. Therefore, from (12) and (13) we can write the unconditional weights at t as

$$w_i^t = \frac{N_{form,c}^t}{N} \frac{w_i^T}{\sum_{j \in c} \mathbb{P}[formal|j,t] w_j^T} \text{ for all } i \in c. \tag{14}$$

We also need adjusted quantities for the experiment baseline period t, q_i^t, for all $i \in I$. We propose a simple adjustment that matches aggregate formal consumption:

$$q_i^t = \frac{Q_t}{Q_T} q_i^T,$$

where Q_t and Q_T are formal aggregate consumption respectively at the baseline experiment period and the later period T, and q_i^T is observed household consumption at T. We can use the same adjustment factor for adjusting capacities to compute informal consumption.

This method of recovering unconditional weights introduces an additional challenge since those weights depend on the very parameters we are trying to recover (β, η) through the conditional choice probabilities of being formal.

We propose an EM algorithm to recover (β, η) from (10) and (11), using consistently updated weights. The algorithm starts with *ad-hoc* weights, recovers the parameters by solving the system of equations (10) and (11). These newly found parameters update the weights consistently with the choice model. It then iterates until convergence. We describe the algorithm below in detail.

Algorithm. Set the counter to zero, $k = 0$.

At each step k:

1. At $k = 0$, we initialize the weights for w_i^t, for each category c, set:

$$w_i^{t,k=0} = \frac{N_{form,c}^t}{N_{form,c}^T} \frac{1}{N_{form}^T} \text{ for all } i \in c.$$

2. Estimate (β^k, η^k) using equation (10) for $t = 2011$ and $t+1 = 2012$ and previous recorded weights $w_i^{t,k}$.[14]

3. Use estimated (β^k, η^k) to compute the probability of being formal for each individual at T and t:

$$\mathbb{P}^k[formal|i,T] = \frac{\exp\left(\beta^k(\psi_0(p^T(q_i^T)) - \psi_{1,i}) + \eta^k\right)}{1 + \exp\left(\beta^k(\psi_0(p^T(q_i^T)) - \psi_{1,i}) + \eta^k\right)}.$$

$$\mathbb{P}^k[formal|i,t] = \frac{\exp\left(\beta^k(\psi_0(p^t(q_i^t)) - \psi_{1,i}) + \eta^k\right)}{1 + \exp\left(\beta^k(\psi_0(p^t(q_i^t)) - \psi_{1,i}) + \eta^k\right)}.$$

4. Use the probability of formality above to compute new weights for T:

$$w_i^{T,k+1} = \frac{1}{N \times \mathbb{P}^k[formal|i,T]}.$$

5. Update weights for baseline periods for each category:[15]

$$w_i^{t,k+1} = \frac{N_{form,c}^t}{N} \frac{w_i^{T,k+1}}{\sum_{j \in c} \mathbb{P}^k[formal|j,t] w_j^{T,k+1}} \text{ for all } i \in c.$$

6. Update $k \leftarrow k+1$.

7. Repeat 2-6 until convergence of (β^k, η^k) and $\{w_i^{t,k}\}_{i \in I}$.

Spatial heterogeneity. The procedure above recovers the pair (β, η) using aggregate data from the experiment periods. We now discuss how we can use these estimated parameters together with disaggregated data from later periods to study the effect of neighborhood-level formality shifters, such as gang influence.

We introduce neighborhood characteristics into the formality shifter η_{0t} in (7) for later periods T, when household level data is available. That is we let

$$\eta_{0\ell,T} = \gamma' x_{\ell,T} + \tilde{\eta}_{\ell,T}, \tag{15}$$

[14] The weights are the same for $t = 2011$ and $t+1 = 2012$.

[15] Note that the model generated probabilities of formality are different in steps 4 and 5. In step 4 the probability is conditional on prices and quantities at T, while in step 5 we use prices and quantities at the baseline period t.

where $x_{\ell,T}$ and $\tilde{\eta}_{\ell,T}$ are, respectively, observable and unobservable neighborhood ℓ characteristics. Fixing β, we can write the model generated formality share for neighborhood ℓ as

$$\sigma_{\ell,T}(\eta_{0\ell,T};\beta) = \frac{1}{N^T_{form,\ell}} \sum_{i=1}^{N^T_{form,\ell}} \frac{\exp\left(\beta(\psi_{0i}(p_T(q_{iT})) - \psi_{1it}) + \eta_{0\ell,T}\right)}{1 + \exp\left(\beta(\psi_{0i}(p_T(q_{iT})) - \psi_{1it}) + \eta_{0\ell,T}\right)}. \tag{16}$$

For each neighborhood ℓ we can solve for $\eta_{0\ell,T}$ that rationalizes the observed formality share of the neighborhood in T, that is,

$$\mathbb{P}^T_\ell(formal) = \sigma_{\ell,T}(\eta_{0\ell,T};\beta).$$

We then project $\eta_{0\ell,T}$ on neighborhood characteristics as in (15).

5 Estimation Results

We begin with our estimates from the first step. Table 3 reports our OLS estimates of Equation (8), which is the log-log electricity demand in the formal sector. Standard errors are in parenthesis. There are 4 specifications in this table. All of them include a constant, month and year fixed effects, and a dummy for households that have access to the social tariff. On top of that, column (2) and (3) includes feeder fixed effects and building fixed effects, respectively. Column (4) adds a control for the installed capacity of the household. As expected, we find that formal consumers are sensitive to price changes. The main coefficient of interest (ξ) can be interpreted as a demand elasticity and the point estimates are consistently in the range -0.16 and -0.198. In our favourite specification (4), we find that this elasticity to be -0.16. Moreover, we find that installed capacity significantly explains the heterogeneity of consumption across households.

Next, we use our Equation (8) coefficient estimates to build the surplus estimates, $\hat{\psi}_{0t}$ and $\hat{\psi}_{1t}$. These objects will be fed into the extensive margin problem (represented by equations 10, 11, and 16, which we then estimate. Our second step results are reported in Table 4. In all columns, the estimate for β is 0.029, indicating a negative (positive) relation between prices (ψ_0) and the probability of being formal. In other words, when the surplus from formal consumption increases relative to the surplus of electricity theft, there is an increase in the share of formal consumers. Columns (1) and (2) exhibit no spatial heterogeneity, i.e., conditional on prices being the same, the probability of being formal does not vary across consumers. The difference is that in (2) the parameter η is estimated using the first stage model with capacity as a control, while that is not the case in (1). Columns (3) and (4) exhibit dispersion in the decision to be formal along the following set of covariates: income, size of the household, presence of drug traffic, presence of militia groups, and area in crime dispute. Column (3) allows for dispersion at the circuit level while column (4) at the neighborhood level. We find that the likelihood of being formal increases with (the log of) income and decreases with the presence of traffic, militia or in disputed areas. There is no statistically significant effect from household size.

Figure 5 plots the distribution of the probability of being formal across neighborhoods, for both the model in specification (4) and the actual data. We see that our model is able to capture well the heterogeneity that is present in the data.

With the structural parameters in hand, we compute price elasticities (table 5). We find that, conditional on being formal, demand elasticity is only 0.21. This shows the response along the demand curve for a customer that is always formal. However, when we introduce the informality margin, we find that the elasticity for the aggregate demand is 0.72, i.e., 3.4 times higher.

6 Counterfactuals

We use the parameter estimates from the model to analyze six different counterfactual scenarios. Table 6 displays our results.

In CF1 and CF2, we simulate a 10% decrease and increase in electricity prices, respectively. All consumers face the change in price, independently of the price that they were paying. These price changes imply a change in the average consumption of formal households (intensive margin), but they also have an impact in the share of households that choose to be formal (extensive margin). In CF1 for example, the 10% price reduction results in an increase in the share of formal consumers of 1.6 percentage points. The same price redution decreases firm revenues by 2.6% (from 196 MM to 191 MM). Given that demand from formal consumers is relatively inelastic, the fact that the revenue falls substantially less in percentage terms than prices is explained by the informality response. The effects in CF2 (a price increase) are mostly symmetric to CF1.

In CF3, we consider an exogenous ban on theft—i.e., we set the share of formal households to 100% – while keeping electricity prices constant, relative to the baseline. In this scenario, quantity consumed by formal consumers does not change but the revenue of the firm goes up to 401 MM. In CF4 we allow prices to change to the point where the revenues from the firm go back to their baseline level. In this case, prices substantially goes down, which is accompanied by an increase in the average quantity consumed.[16]

[16]We are currently trying to obtain information on the average costs of the firm in order to generate "same

The two last columns of Table 6 evaluate scenarios that explore the geographic heterogeneity in the share of informality. In particular, we simulate a world where there is no crime (as that is one of the main variables driving informality). In CF5, we keep prices constant (from baseline) while in CF6 we keep revenue constant (changing prices). We find that the impact of removing crime on the share of formal consumers is comparable to a 10% price reduction in the regular tariff.

7 Conclusion

In this paper we discuss how informal consumer markets can affect a firm´s optimal decisions. We first document a causal relation between prices and the endogenous decision by consumers to be formal. Then, we estimate the primitives of a structural model of how the consumer makes decision along the extensive and intensive margins. The model suggests that the aggregate demand elasticity that the firm faces is 3.4 times larger than what it would have been without informality. This has substantial implications for the pricing strategies of a firm.

We also find substantial heterogeneity across consumers in the propensity to be formal. This suggests that price discrimination can be an effective tool for firms to use.

Moreover, 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 suggest ways to mitigate this problem.

profit" counterfactuals. In 2017, these costs accounted for approximately 63.1% of the price of electricity charged from formal households.

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.

BURGESS, R., M. GREENSTONE, N. RYAN, AND A. SUDARSHAN (2020): "The Consequences of Treating Electricity as a Right," *Journal of Economic Perspectives*, 34, 145–69.

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 (2021): "Hysteresis and the welfare effect of corrective policies: Theory and evidence from an energy-saving program," *Journal of Political Economy*, 129, 1705–1743.

DELL'ANNO, R. (2022): "Theories and definitions of the informal economy: A survey," *Journal of Economic Surveys*, 36, 1610–1643.

DERYUGINA, T., A. MACKAY, AND J. REIF (2020): "The long-run dynamics of electricity demand: Evidence from municipal aggregation," *American Economic Journal: Applied Economics*, 12, 86–114.

DIX-CARNEIRO, R., P. GOLDBERG, C. MEGHIR, AND G. ULYSSEA (2021): "Trade and informality in the presence of labor market frictions and regulations," *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.

GERARD, F. AND G. GONZAGA (2021): "Informal labor and the efficiency cost of social programs: Evidence from unemployment insurance in brazil," *American Economic Journal: Economic Policy*, 13, 167–206.

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.

GUEDES, R., F. S. IACHAN, AND M. SANT'ANNA (2023): "Housing supply in the presence of informality," *Regional Science and Urban Economics*, 99, 103875.

(ILO), I. L. O. (2018): "Women and men in the informal economy: A statistical picture (Third edition)," .

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.

LI, X., C. LIAO, AND Y. XIE (2021): "Digital piracy, creative productivity, and customer care effort: Evidence from the digital publishing industry," *Marketing Science*, 40, 685–707.

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.

LU, S., X. WANG, AND N. BENDLE (2020): "Does piracy create online word of mouth? An empirical analysis in the movie industry," *Management Science*, 66, 2140–2162.

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.

OECD/EUIPO (2019): "Trends in Trade in Counterfeit and Pirated Goods," .

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110, 705–729.

Qian, Y. (2014): "Counterfeiters: Foes or friends? How counterfeits affect sales by product quality tier," *Management Science*, 60, 2381–2400.

Qian, Y., Q. Gong, and Y. Chen (2015): "Untangling searchable and experiential quality responses to counterfeits," *Marketing Science*, 34, 522–538.

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

Rocha, R., G. Ulyssea, and L. Rachter (2018): "Do lower taxes reduce informality? Evidence from Brazil," *Journal of development economics*, 134, 28–49.

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.

Ulyssea, G. (2018): "Firms, informality, and development: Theory and evidence from Brazil," *American Economic Review*, 108, 2015–2047.

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

8 Appendix

Table 1: Summary Statistics

	Panel A (Household Level, N=3,291,411)					
	Mean	**P25**	**Median**	**P75**	**Min**	**Max**
Consumption	164.05	77.47	148.70	233.10	0.05	500.00
Price	0.39	0.41	0.41	0.41	0.13	0.41
Social Tariff Indicator	0.07					
	Panel B (Circuit Level)					
	Mean	**P25**	**Median**	**P75**	**Min**	**Max**
Theft	249.90	79.44	145.45	266.96	1.48	6205.83
	Panel C (Census Block Level)					
	Mean	**P25**	**Median**	**P75**	**Min**	**Max**
Income	2970.00	1479.85	2038.90	3298.52	205.66	36253.00
Household Size	5.49	4.91	5.50	6.00	1.00	9.00
	Panel D (Census Block Level)					
	Mean	**P25**	**Median**	**P75**	**Min**	**Max**
Crime Areas Indicator	0.25					

Note: This table reports descriptive statistics from different variables used in the analysis. Panel A includes information at the household level, for formal consumers of the utility. Consumption is per month and measured in KWh. Panel B has the distribution of the theft quantity (in X) across the different circuits. Both the data in Panels A and B are from ANEEL. Panel C includes statistics on the distribution of household income and size, from the Census. Finally, Panel D includes statistics on the presence of crime across census block areas. This is constructed from the *Fogo Cruzado* data, and is defined as 1 if the area has a known traffic or militia group in control or if it is a disputed area by different factions.

Table 2: Regression of the share of formal consumers on covariates

	Formal Share (log)	
	Circuit	Neighborhood
	(1)	(2)
log(avg income)	0.109^{***}	0.375^{***}
	(0.019)	(0.035)
HH avg size	0.028	0.044
	(0.040)	(0.044)
Traffic Area	-0.159^{***}	-0.145^{***}
	(0.026)	(0.037)
Militia	-0.039	-0.071^{*}
	(0.026)	(0.037)
Dispute	-0.017	-0.004
	(0.035)	(0.045)
Constant	-1.355^{***}	-3.398^{***}
	(0.265)	(0.354)
Observations	1,002	822
R^2	0.111	0.148
Adjusted R^2	0.106	0.143
Residual Std. Error	0.356 (df = 996)	0.449 (df = 816)
F Statistic	24.816^{***} (df = 5; 996)	28.331^{***} (df = 5; 816)

This table reports an OLS regression of the share of the number of potential households that are formal in each location on a set of covariates. We run regressions at both the circuit and neighborhood levels. Robust standard errors are in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Table 3: Intensive Margin Results

	(1)	(2)	(3)	(4)
ξ	-0.1782	-0.186	-0.198	-0.160
	(0.12)	(0.012)	(0.002)	(0.002)
capacity				1.329
				(0.002)
Constant	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Social Tariff FE	Yes	Yes	Yes	Yes
Feeder FE	No	Yes	No	No
Building FE	No	No	Yes	Yes
Observations	230,200,302	230,200,302	230,200,302	230,200,302
R-squared	0.009	0.083	0.218	0.540

[1] This table reports estimates of Equation 8. A unit of observation is a household-month. The dependent variable is the log of the formal electricity consumption per household, in MWh. α is the coefficient on the log electricity retail price. In column (4) we also use installed capacity of each household as a regressor. Our sample covers the period 2017 to 2022. Consumption and price data come from ANEEL. Robust standard errors are in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Table 4: Extensive Margin Results

	(1)	(2)	(3)	(4)
β	0.0289	0.0289	0.0289	0.0289
η	4.704	3.914	-5.849	-10.294
			(3.852)	(1.558)
γ (log income)			0.730	1.745
			(0.277)	(0.160)
γ (avg hh size)			0.415	-0.033
			(0.579)	(0.185)
γ (traffic)			-1.872	-0.801
			(0.373)	(0.167)
γ (militia)			-0.840	-0.453
			(0.374)	(0.165)
γ (dispute)			-1.220	-0.543
			(0.517)	(0.203)
Capacity in 1st stage	No	Yes	No	No
Spatial heterogeneity	N/A	N/A	Circuit	Neighborhood
Observations			897	690
F-stat			13.758	38.84
R_squared			0.072	0.221

[1] This table reports estimates from the second step (extensive margin) - equation 16, following the algorithm detailed in the estimation section of the paper. β is the coefficient associated to the $(\psi_{0t} - \psi_{1t})$ variable, and η is a constant utility shifter for formal consumption. Finally γ is a vector of coefficients associated with the following covariates: log of income, average size of the household, presence of drug traffic, presence of a militia group, disputed territory. While columns (1) and (2) do not exhibit heterogeneity (other than price), column (3) has γ features that vary by circuit, and column (4) varies at the neighborhood level. Robust standard errors are in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Table 5: Elasticity Estimates

	No informality (1)	**With informality** (2)	**Ratio** (3)=(2)/(1)
Elasticity	0.21	0.72	**3.4x**

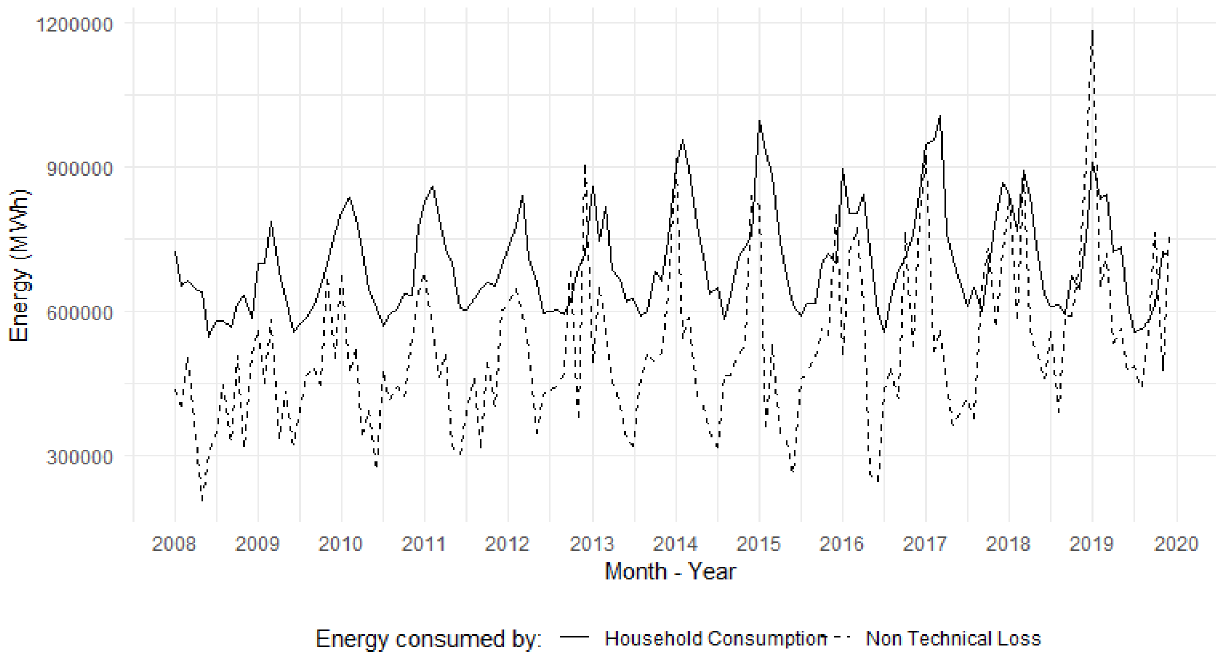
[1] In the "No Informality" scenario, we force all consumers to be formal (i.e., we drop the possibility of informality). In the "With informality" scenario that possibility is allowed.

Table 6: Counterfactual Results

	Baseline	CF1 Price Down 10%	CF2 Price Up 10%	CF3 No Theft Same Price	CF4 No Theft Same Revenue	CF5 No Crime Same Price	CF6 No Crime Same Revenue
Qt formal/hhd (kWh)	201.7	205.9	197.9	201.7	240.7	201.7	203.4
Share formal	0.808	0.824	0.791	1.000	1.000	0.826	0.833
Regular Price (R$ / kWh)	0.405	0.365	0.446	0.405	0.166	0.405	0.388
ψ_0	-83.53	-78.19	-88.50	-95.97	-46.88	-95.97	-92.64
Revenue (in MM R$)	196.0	191.0	200.5	401.1	196.0	203.0	196.0

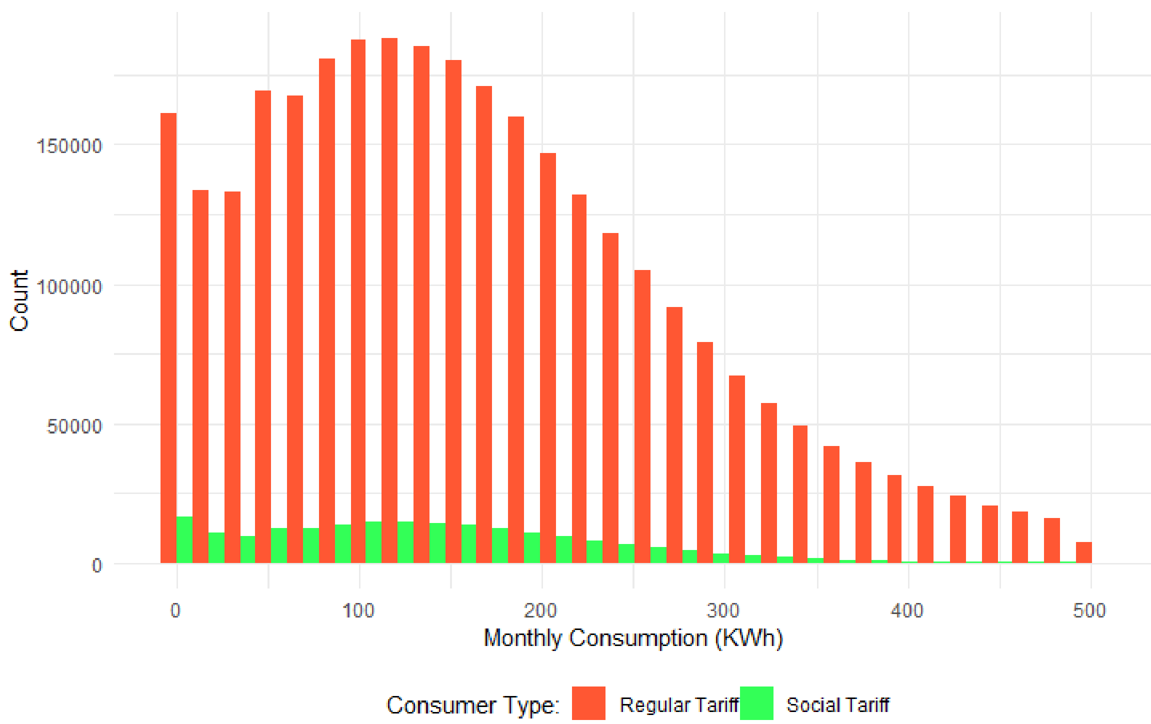
[1] In this table, we report the results of six different counterfactual scenarios, which are calculated using the parameters estimated in column (3) of table 3 and column (4) of table 4. In CF1, we exogenously decrease the electricity price by 10%. In CF2, we exogenously increase the electricity price by 10%. In CF3 and CF4, we ban the possibility of theft, (i.e, we set the share of formal households to 100.00). Finally, in CF5 amd CF6, we look at scenarios where there is no crime. Moreover, in CF3 and CF5 we keep the utility price constant (relative to baseline), while in CF4 and CF6 we keep constant the revenue of the firm (relative to baseline). *Qty. Formal* is the electricity consumption of formal households, in KWh. *Share Formal* is the percentage share of formal households. *Regular Price* is the electricity price, in BRL/KWh. ψ_0 is the surplus of formal households generated by electricity consumption only, in BRL. *Revenue* is the utility revenue measured in BRL.

Figure 1: Formal Electricity Consumption and Non-Technical Losses (in MWh), over time



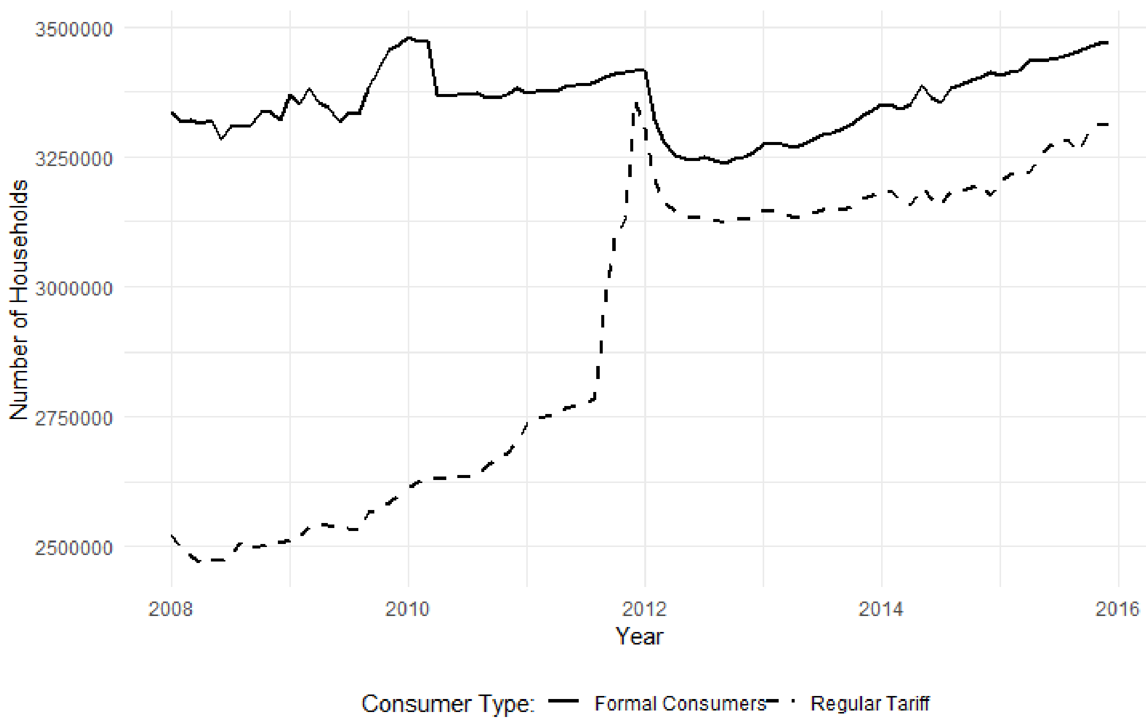
Note: This figure reports the amount of billed electricity consumed by all residential clients as well as NTL for the period January 2008 to December 2020. Consumption and NTL are measured in MWh. Data comes from ANEEL.

Figure 2: Histogram of Monthly Household Consumption (KWh)



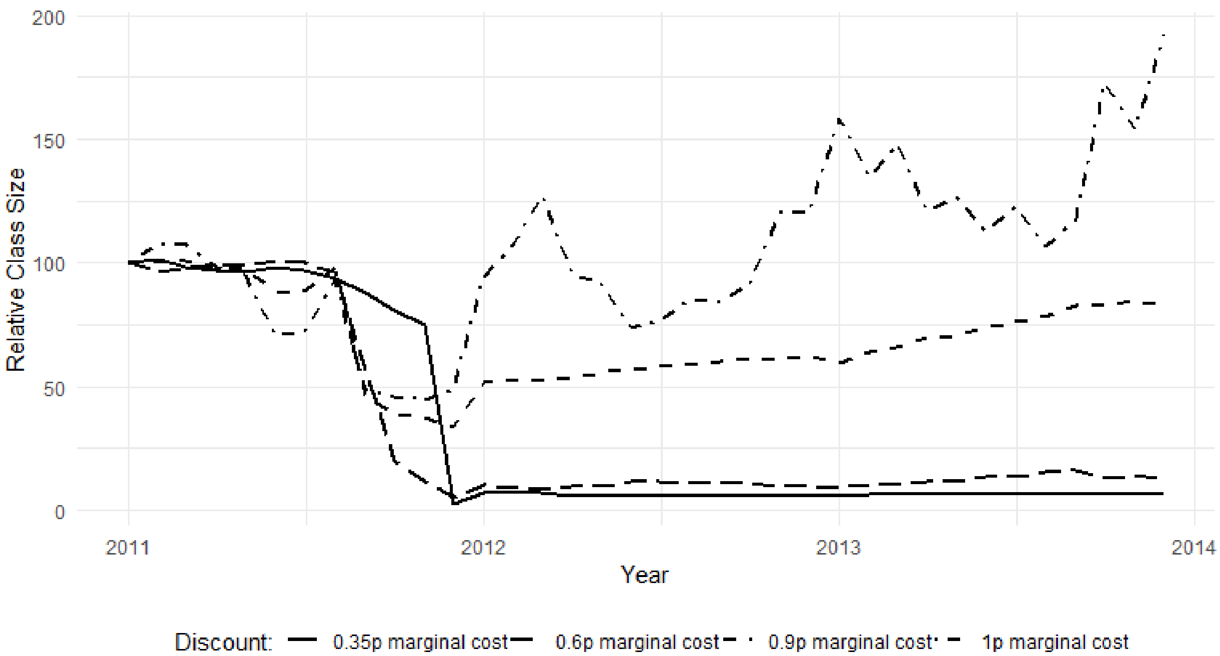
Note: This histogram was constructed using a cross-section of electricity consumption for the universe of residential clients for the the period 2017-2022. We dropped clients with zero consumptionas well as those with over 500 KWh/month. Data comes from ANEEL.

Figure 3: Number of Formal Clients (before/after the exogenous price increase)



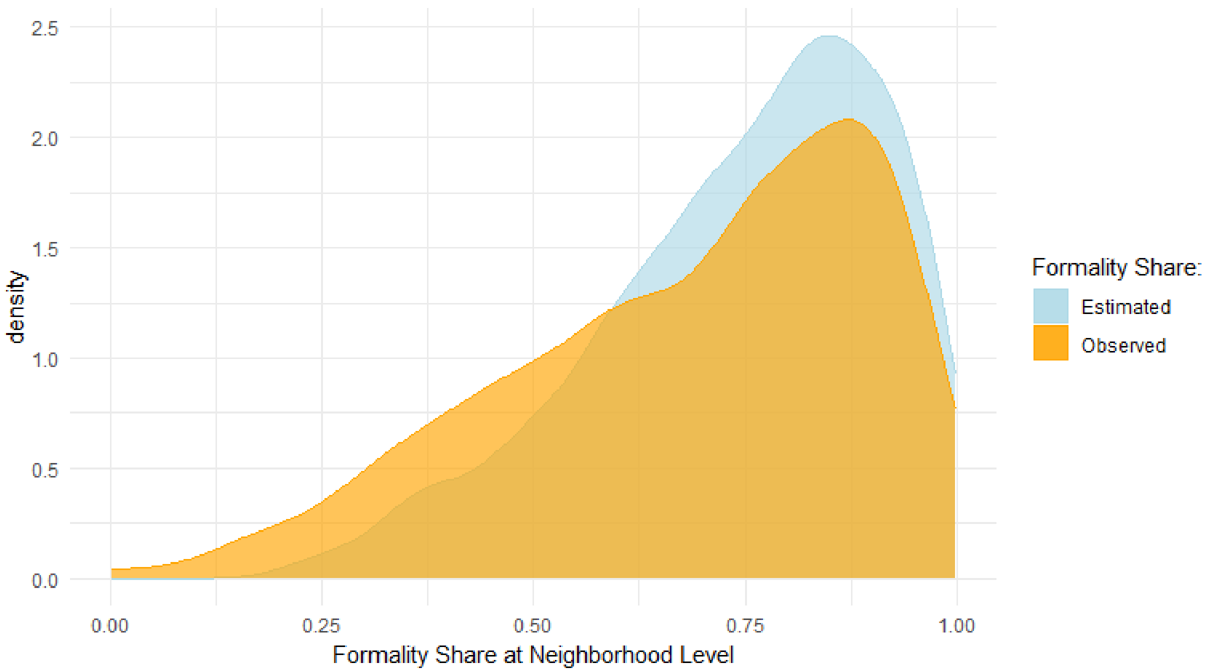
Note: This figure shows the number of the utility's formal residential clients for the period 2008 to 2016. The dashed line represents B1 clients, i.e., those who pay the regular electricity tariff schedule. The solid line represents all the residential clients, which is the sum of B1 clients and those who pay the social tariff schedule (the "social" consumers). Data comes from ANEEL.

Figure 4: Change in number of users by discount class



Note: This figure reports the evolution over time (2011-2014) of the number of clients that had access to a social tariff for electricity. Each of the lines represent a separate class, based on the discount over the regular price that each household had for their marginal consumption unit. There are 4 discount classes: 35% of the regular price, 60%, 90% and 100%. We normalize the number of customers that were in each class in 2011 to 100. Data comes from ANEEL.

Figure 5: Distribution of probability of being formal: model vs data



Note: In this figure we assess the fit of the structural model by plotting the distribution of the probability of being formal across neighborhoods for both the model and the actual data.