Pricing and Informality: Evidence from Energy Theft in Brazil

Davi Resende Gabriel Richter Marcelo Sant'Anna

André Trindade *

February 2025

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 more than 50% of the total formal market. We use detailed micro data from a major electric utility 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 where prices increased permanently to a set of consumers. We use the model to simulate counterfactual scenarios where: (i) theft is not possible, and (ii) the firm uses different pricing strategies. We find that the presence of informality increases the elasticity of demand from 0.24 to 0.39, and reduces monopoly optimal prices by 10.4%. Eliminating theft altogether would allow the firm to reduce prices by 17.7% while keeping profits constant. We also find that price discrimination is an effective tool to reduce informality rates.

JEL: L94, Q41, O17

Keywords: Informality, pricing, electricity demand, structural model.

*Resende: FGV EPGE, daviresende@hotmail.com Richter: Stone, gabrielrichter9@gmail.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, Carolina Tojal, Pedro Gardete, and participants from the EPGE Applied Microeconomics Workshop, the Wharton BEPP Applied Economics Workshop, the Toulouse Conference on the Economics of Energy and Climate, LACEA LAMES 2024, and SBE 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 recently acknowledged the importance of informality in the economy and its impacts 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 migrate to the informal market. This is in addition to the loss revenue from formal consumers implied by the Law of Demand. *Second*, especially in the case of theft or undue service utilization, 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.

Importantly, firms 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 in a context where electricity theft is common. 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 be a serious

[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.

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.

A critical consequence of Non-Technical Losses (NTL) is energy waste. When consumers steal electricity, they face zero marginal cost and thus fail to internalize generation and distribution costs, leading to excessive consumption. The environmental implications of this waste are particularly concerning given forecasts that energy demand in developing countries will double that of developed nations by 2035 (Wolfram et al., 2012). Moreover, NTL may lead to excessive prices for paying consumers as utilities pass on losses, which may have adverse distributional consequences.[4]

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 illegal markets, where access to consumer data is limited.

Specifically, in this paper we ask the following questions: How does informality (theft) in this market affect the elasticity of the demand that firms face and their optimal pricing decisions? What are the welfare effects from removing informality? Is price discrimination (e.g., social tariffs) an effective way to reduce informality in this market?

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 a household-level panel with monthly information on consumption, capacity (measured by owned appliances), and price paid for the period between 2017 and 2022.

We start by providing evidence of a causal link between pricing and the consumer decision to become informal. The empirical challenge here 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.

[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]Beyond environmental impacts, NTL creates other challenges for the energy sector: it reduces grid reliability by making demand more unstable and difficult to predict and poses safety risks through improper electrical connections that can cause injuries.

To address this challenge, we leverage a natural experiment from 2011 when electricity prices increased exogenously and *permanently* to a subset of consumers.

Then, we set up and estimate a structural demand model that captures consumer decisions in markets with informality. Our framework nests the continuous electricity consumption into a discrete choice model on the mode of consumption (formal or informal). In the model, consumers decide between being formal consumers (paying the full price) or informal consumers, stealing electricity through an illegal grid connection and avoiding the tariff. This discrete choice is taken anticipating their electricity consumption in both formal and informal alternatives. Formal consumers choose the electricity consumption optimally given the tariff, while informal consumers face zero marginal prices and consume up to their capacities. The trade-off that consumers face is clear: by moving to informality 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).

Our estimation follows a two-step approach. First, we estimate electricity demand conditional on formal status using panel data of formal clients. The demand model incorporates granular geographic and time fixed effects while accounting for capacity heterogeneity among formal consumers. For informal consumers, we assume electricity consumption is capped at a fraction of their capacity. This demand estimation enables us to calculate household-level consumer surplus for both formal and informal consumption options.

Second, we estimate a binary logit model of households' formal-informal status choice, using the previously calculated consumer surplus as an explanatory variable. To identify households' price sensitivity in this choice, we exploit the 2011 policy change that raised prices for certain consumers, thereby reducing the surplus from formal consumption. The model leverages both the resulting variation in consumer surplus and the subsequent changes in formal client numbers. We further enrich the specification with neighborhood-level controls for criminal activity, capacity, and other relevant characteristics.

A key empirical challenge is that our panel data only includes formal consumers, while the choice model requires the unconditional choice probability for the entire population of both formal and informal consumers. We address this sample selection issue by using our model to infer the population distribution of characteristics from the observed formal consumer sample. Specifically, we employ Bayes' rule to derive weights that transform the observed sample moments to match the population moments. This naturally suggests an Expectation-Maximization (EM) approach for estimation, which we implement.

We use our estimates to disentangle the price elasticity along the formal demand curve (formal consumers reducing quantity in response to higher prices) from the price elasticity along the extensive margin (consumers moving to informality in response to higher prices). Moreover, we leverage the primitives from the model to simulate different counterfactual scenarios.

In our estimates, we find that consumers significantly respond to a permanent increase in the price level both in the intensive margin (i.e. along the formal demand curve) and extensive margin (i.e., by migrating to informality). This implies that the demand curve faced by this firm becomes 1.63x larger due to presence informality (moving from -0.24 to -0.39). Consequently, an unregulated monopolist would optimally choose prices that are 10.4% higher if theft was not possible, compared with the status-quo. We also simulate scenarios where the profit of the firm is kept constant at current levels (due to regulation). In those scenarios, we find that: 1) eliminating energy theft altogether would allow regular prices to be 17.7% lower, and 2) price discrimination strategies such as a social tariff are effective ways for firm managers to curb informality.

1.1 A Toy Model

We now discuss the problem of the firm in the presence of sizable 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-household 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 consumption.

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

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

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 between 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 writing 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 articles 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.

To our knowledge, no prior research has examined the welfare costs of electricity theft. The limited economic literature on non-technical losses (NTL) includes Smith (2004b), who conducts a cross-country comparison; Min and Golden (2014), who analyze the relationship between political cycles and energy theft; and Burgess et al. (2020), who describe how widespread tolerance of 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, Deryugina et al. (2020) use a difference-in-differences matching estimator to measure quantity responses to changes in prices, Cahana et al. (2022) study real time pricing, and Burkhardt et al. (2023) leverage an experimental design to understand how peak pricing can affect consumer decisions. 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. Other examples of discrete-continuous demand models in sectors other than electricity are Smith (2004a), Magnolfi and Roncoroni (2016), and Tojal and Guidetti (2024).

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

2.1 Electricity sector in Brazil

In 2023 the total electricity consumed in Brazil was 532 TWh, making the country one of the 10 largest in the world. The total installed capacity in the same year was over 226 GW, roughly 48.6% of which was hydropower and the remaining mostly a combination of natural gas, biomass and nuclear (EPE, 2024).

There are around 50 different local regulated utilities that distribute electricity in Brazil. Most of them are privately owned but several are public (state owned). The largest 6 distributors, in terms of the number of customers served are, in order: Cemig, Enel SP, Coelba, Copel, CPFL Paulista, and Light (EPE, 2024).

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 consumers pay, 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 2023 the total electricity lost in Brazil, as a percent of the electricity injected in the system, was 14%, roughly equally divided across TL (7.3%) and NTL (6.7%). The total amount of NTL in that year was above 37 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 drug traffickers or militia groups (Merenfeld, 2017). While both types of organizations maintain territorial control and operate extortion networks, drug trafficking gangs and militia groups differ primarily in their membership composition rather than their activities (Hirata et al.,

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

2022). These criminal organizations' territorial control constrains the utility's ability to penalize electricity theft.

2.2 The 2011 change in social tariffs

In Brazil, some households have access to a social tariff, which gives them discounted rates over the regular electricity 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.[6] To relate census tracts and feeders, we first identify in which census tract each building point is located. Then, we associate the information that comes from the census tract 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 minimum and maximum values. Panel A includes information regarding the formal consumers at a month level

[6] We also aggregate and perform the exercises at the neighborhood 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. Panel D includes a "Crime Area Indicator" that equals one for census areas controlled by criminal organizations (drug traffickers or militia groups) or contested by rival gangs.

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	0	0	0	0	1

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	0.248	0	0	0	0	1

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 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 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.

Table 2: Regression of the share of formal households 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.

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.[7] 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 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.

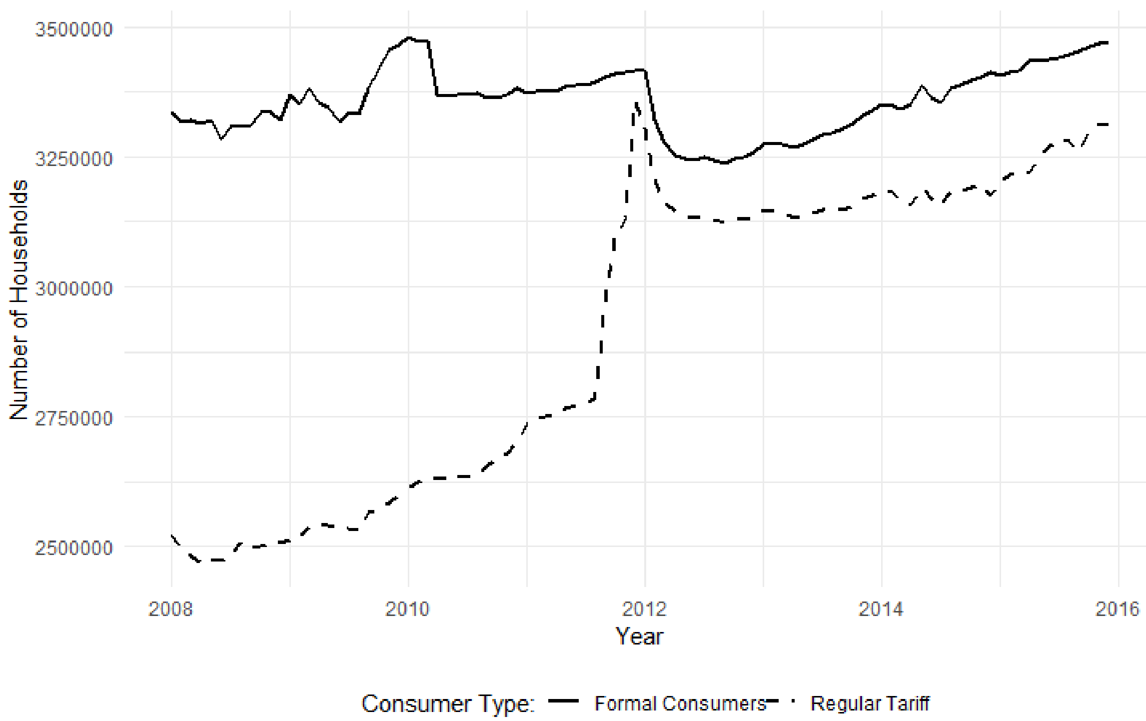
Figure 2 shows the impact from the change in policy on the number of customers with a social tariff. We find that the largest drop occurs in customers with a marginal rate of 35% and 60% of the regular prices. This is expected as those were the groups in the lower quantity brackets and were, therefore, 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 3. 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

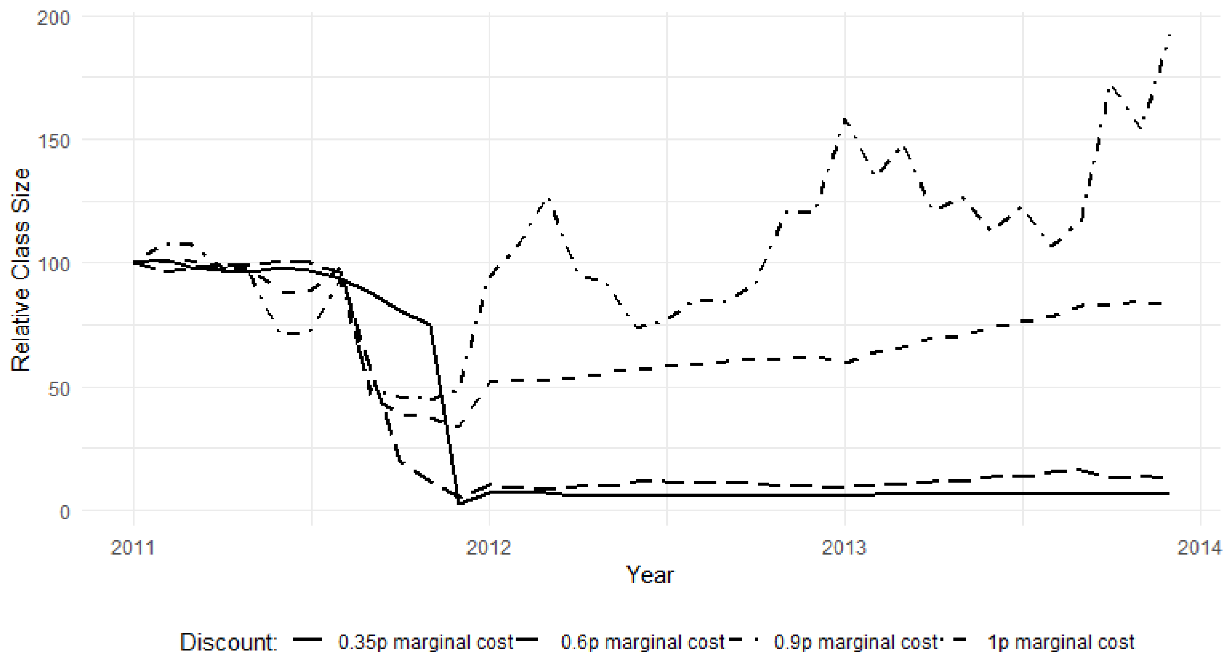
[7] To be more specific, before the new rule the Social Tariff was automatically applied for all consumers with a total quantity under 80 kWh. Families that consumed between 80 and 220 kWh could still benefit from the social tariff, but they would have to show evidence of low income. With the new policy, 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.

Figure 1: 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.

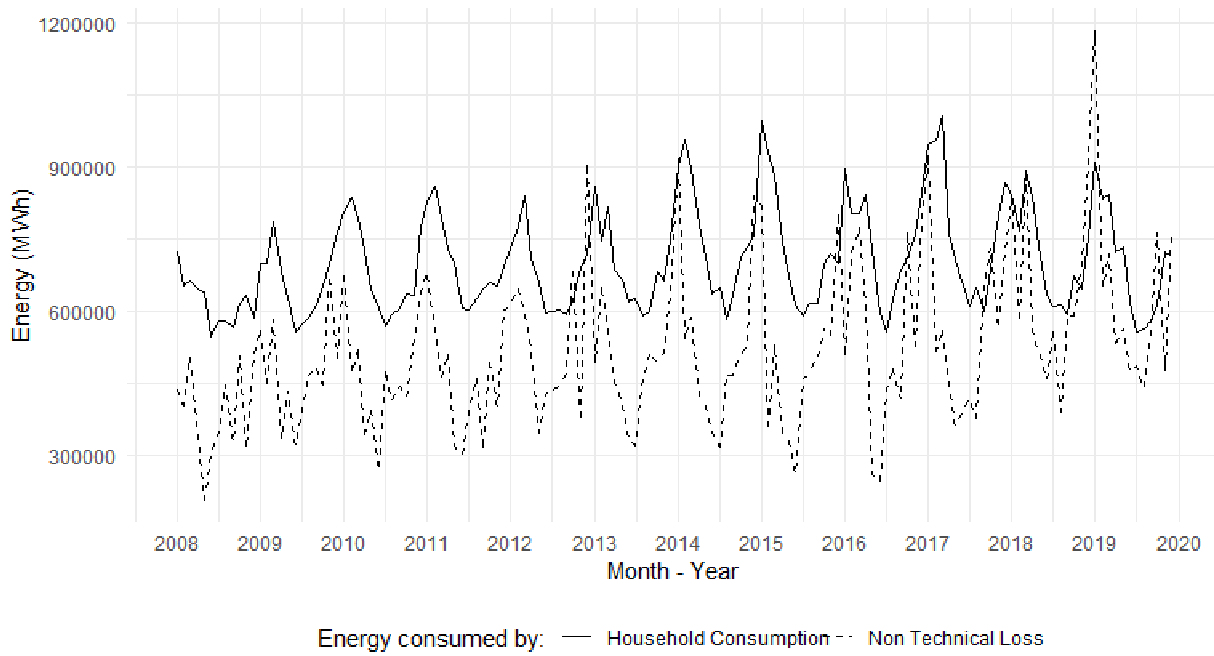
Figure 2: Change in number of users by discount class



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.

September, when it starts to increase again. Figure 3 also shows an interesting pattern of formal electricity consumption: it became more volatile over the years.

Figure 3: 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.

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[8] t is quasi-linear on the consumption of electricity, q:[9]

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

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}. This specification applies both to formal consumption and theft, with one key distinction: formal consumers pay the utility's positive price p, while those who steal electricity effectively face a price of zero.

Formal demand. From the quasi-linear utility specification in (4), we derive formal electricity demand by solving the consumer's utility maximization problem:

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

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 $\bar{q}_{it}$ – the highest consumption value that is feasible given capacity restrictions.

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

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

[8] 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.

[9] 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.

where ε_{ijt} has the usual extreme value 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.[10] 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.

[10] 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.

Therefore, in order to identify parameters governing the extensive margin, we leverage on a permanent policy shift that took place in 2011 (see section 2.2).

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.

4.3.1 First Step: electricity demand

Formal consumers

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

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

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.

Informal consumers

To complete the model, we need to compute the quantity that each consumer would have consumed if they were informal, i.e., if they paid a price of zero for each kWh of energy consumed. As with formal consumption, it is natural to expect that $\bar{q}$ also exhibits substantial heterogeneity across individuals. We want our model to allow for that dispersion, as it may be important for the results. In order to do so, we assume that $\bar{q}$ is proportional to installed capacity (IC) - a variable that we observe with significant detail (i.e., at the household level). Specifically, we estimate a linear regression of the actual quantity on installed capacity and prices, and use the coefficient of the former to predict $\bar{q}$ (zeroing out the price effect).

4.3.2 Second Step: extensive margin

The idea is to estimate the model parameters that govern 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:

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

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

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^t_i 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.[11] 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 in the immediate periods 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

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

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.[12] Therefore, we assume:

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

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:

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

where $N^t_{form,c}$ 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

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

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:

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

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

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 $w_i^{t,k}$.[13]

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

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

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

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

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:[14]

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

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

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

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).

[14] 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.

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, columns (2) and (3) include feeder fixed effects and building fixed effects, respectively. Columns (4), (5) and (6) add a control for the household's installed capacity. As expected, we find that formal consumers are sensitive to price changes.

Table 3: Intensive Margin Results

	(1)	(2)	(3)	(4)	(5)	(6)
ξ	-0.1782	-0.186	-0.198	-0.160	-0.0904	-0.2676
	(0.12)	(0.012)	(0.002)	(0.002)	(0.0029)	(0.0020)
capacity				1.329	4.617	0.6517
				(0.0020)	(0.0028)	(0.0008)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Social Tariff FE	Yes	Yes	Yes	Yes	Yes	Yes
Feeder FE	No	Yes	No	No	No	No
Building FE	No	No	Yes	Yes	Yes	Yes
Capacity Heterogeneity	No	No	No	No	Below Median IC	Above Median IC
Observations	230,200,302	230,200,302	230,200,302	230,200,302	115,157,694	115,042,608
R-squared	0.00947	0.08326	0.21762	0.53961	0.5284	0.5321

Note: 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 KWh. ξ 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. Standard errors are in parenthesis.

The main coefficient of interest (ξ) can be interpreted as a demand elasticity and the point estimates are in the range -0.16 and -0.198, without capacity heterogeneity. Our results here are consistent with other residential demand elasticity estimates from Brazil (in the range -0.13 to -0.20, Schutze, 2015) and the US (from -0.09 in the short-run to -0.27 in the long-run, Deryugina et al., 2020). Our preferred specification allows demand elasticity to vary with installed capacity. Households below the median capacity exhibit lower price sensitivity (elasticity of -0.09, column 5) compared to those above the median (elasticity of -0.27, column

6). This heterogeneity is consistent with higher-capacity households having more appliances and thus greater flexibility to adjust consumption in response to price changes.

Using our preferred demand specification, we estimate consumer surplus under both formal ($\hat{\psi}_{0t}$) and informal ($\hat{\psi}_{1t}$) status for each household. We then incorporate these surplus estimates into the extensive margin choice model (represented by equations 10, 11, and 16), which we then estimate. Our second step results are reported in Table 4. Column (1) presents baseline estimates of β and η from the algorithm described in Section 4.3.2, without spatial heterogeneity. The significantly positive β indicates that households are more likely to choose formal status when the relative surplus from formal consumption increases. Consequently, price increases, which reduce the formal surplus, decrease formal sector participation.

Column (2) of Table 4 extends the model by incorporating neighborhood-level covariates to capture spatial heterogeneity following the discussion at the end of Section 4.3.2. We find that the likelihood of being formal significantly increases with neighborhood income and installed capacity. These controls are important because households with a higher baseline consumption (which correlates with higher income) mechanically have a higher surplus benefit from stealing electricity. However, they may also face relatively higher penalties from being informal as they tend to be more connected to the formal economy overall and thus more exposed to fines and penalties. We also find that areas controlled by criminal organizations, such as traffic gangs and militia squads, show higher theft rates, likely due to utilities' limited ability to enforce penalties in these zones. Areas flagged as under dispute by criminal organizations face even higher theft rates.

5.1 Impact of Informality on Elasticity and Optimal Prices

We now leverage our empirical estimates to quantify the economic forces established in our theoretical framework from Section 1.1.

Using the full structural estimated model from our preferred specification (Table 4, column 2), we first compute aggregate demand elasticities, which we report in Table 5. We find that, conditional on being formal, demand elasticity is -0.24.[15] When accounting for switches between formal and informal status, the aggregate total formal demand elasticity increases to -0.39 – 1.63 times larger.

Following our theoretical model, total formal demand equals consumption per formal household times the share of formal households, $d(p) \times \sigma(p)$. The total elasticity thus de-

[15]This aggregate elasticity falls between the values for below- and above-median installed capacity reported in Table 3, columns (5) and (6). It is closer to the above-median value due to higher-consumption households' greater weight in aggregate formal demand.

Table 4: Extensive Margin Results

	(1)	(2)
β	0.041	0.041
	(0.008)	(0.008)
η	6.781	-7.562
	(1.193)	(1.591)
γ (log income)		0.568
		(0.1913)
γ (avg hh size)		0.198
		(0.178)
γ (traffic)		-0.782
		(0.130)
γ (militia)		-0.624
		(0.126)
γ (dispute)		-0.407
		(0.156)
γ (illiteracy)		-0.076
		(0.022)
γ (density)		0.0004
		(0.0001)
γ (avg capacity)		0.036
		(0.003)
Capacity in 1st stage	Yes	Yes
Capacity/Elasticity Heterogeneity	Yes	Yes
Spatial heterogeneity	N/A	Neighborhood
Observations		788
F-stat		98.8
R_squared (adj)		0.499

Note: 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 vector of covariates available at the neighborhood level. Robust standard errors are in parenthesis.

composes linearly into intensive margin (ξ_d) and extensive margin (ξ_σ) components:

$$\xi_{Total} = \xi_d + \xi_\sigma.$$

This decomposition reveals that changes in formal status account for 38% of the total demand response, significantly affecting optimal pricing in our context.

Table 5: Demand Elasticity Decomposition

	Total formal	Intensive margin (cond. on formal)	Extensive margin (formality share)
	ξ_{Total}	ξ_d	ξ_σ
	(1)	(2)	(3)
ξ_x	-0.39	-0.24	-0.15
ξ_x/ξ_{Total}	1.00	0.62	0.38

Note: This table reports total formal demand elasticity decomposition based on the preferred specifications (Table 3, columns 5 and 6, and Table 4, column 2). Column (1) is the total formal demand elasticity, column (2) is the demand elasticity conditional on being formal, and column (3) is the formality share elasticity.

As the demand curve faced by the firm becomes significantly different (i.e., more elastic) in the presence of informality, optimal pricing decisions will also change. Although in our setting prices are regulated (and we will take that into account in the next section), we are interested in understanding how a monopolist firm – in the absence of regulatory constraints – would optimally choose prices with and without informality. To do so, we evaluate the solution to the monopolist problem (equation 2) at the parameters that we estimate.[16] We report the optimal prices under (1) the status-quo (informality), (2) removing the evasion adjustment term. and (3) removing informality altogether (i.e., no evasion nor theft adjustment terms). We find that removing informality would increase the monopolist prices by 10.4% in our empirical setting. We also conduct comparative statics analysis by changing the elasticity of the extensive margin, while keeping the remaining parameters constant. Results are reported in Figure 4 and suggest that if the elasticity of the informality decision with respect to price was 3x higher in absolute value (i.e., -0.45 instead of the -0.15 that we estimate), removing informality would increase monopolist prices by 45%.

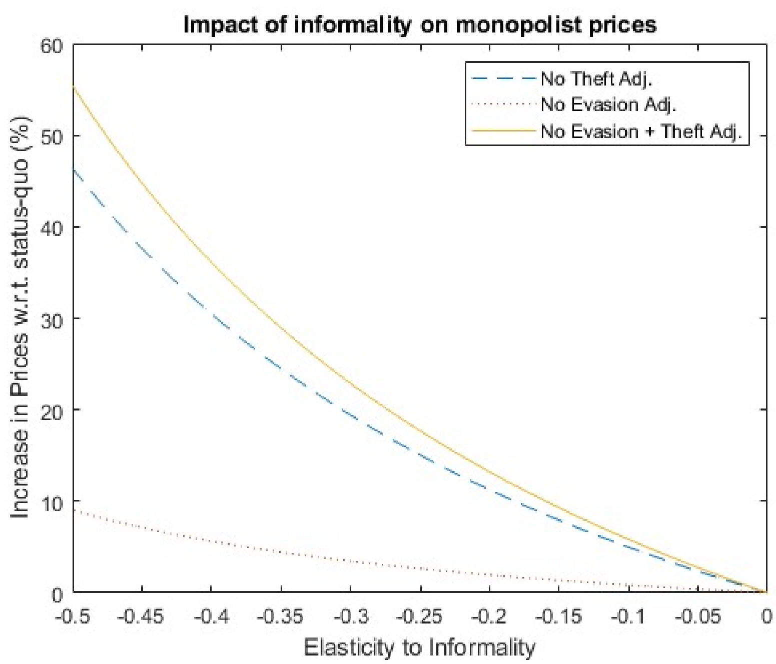
[16]One challenge that we face in doing this exercise is that there would be no solution for the monopolist problem if we were to assume that the elasticity of demand was constant for every quantity. Instead, we assume that the elasticity parameter that we estimate is locally true but varies as we increase/decrease quantity, and the demand curve is linear.

Table 6: Optimal Prices

	Price	ΔPrice
(1) Status-quo	1.42	-
(2) No informality (no Evasion adj)	1.45	2.4%
(3) No informality (no Evasion + Theft adj)	1.57	10.4%

Note: This table computes the optimal prices that a monopolist firm would choose in our empirical setting, under different informality scenarios. The first scenario that we consider (1) is the "status-quo", i.e., with informality". In the remaining two scenarios we remove informality, partially (2), and fully (3). For each scenario, we report prices and the percent deviation over the status-quo scenario. Optimal prices are computed by applying the estimated parameters from Table 4, column 2, to the FOC of the firm given by equation 2.

Figure 4: Impact of informality on monopolist optimal prices



Note: This figure shows the results of our comparative statics exercise, based on equation 2 of our toy model (section 1.1). In the x-axis we vary the elasticity of the extensive margin (i.e., decision to be formal/informal) while keeping the remaining parameters constant at the values that we estimate. In the y-axis, we compute the percentage increase in the optimal prices chosen by the monopolist if we were to remove informality from the setting.

6 Counterfactuals

Using our preferred specification (from Table 4 column (2) and Table 3 columns (5) and (6)), we evaluate three counterfactual scenarios presented in Table 7. The first scenario prohibits electricity theft (column 2) and the second eliminates the effect of organized crime (column 3), allowing us to quantify the economic burden that theft and criminal activity impose on this market. The third scenario removes the social tariff by applying the regular price to all households (column 4), helping us assess whether price discrimination can maintain formal connections among vulnerable households while preserving revenue from higher-income consumers. For each scenario, we consider two pricing regimes: Panel A maintains baseline prices, while Panel B adjusts prices to hold firm profits constant. Panel B thus helps us explore how regulators might mandate price adjustments to ensure that firm gains from each scenario flow to consumers through reduced electricity prices.

Eliminating electricity theft (column 2) would increase firm profits by almost 40% if prices remain at baseline levels (Panel A), or alternatively, could allow for a 17.7% reduction in regular prices while maintaining current profit levels (Panel B). While criminal gang activity correlates with electricity theft, eliminating organized crime (column 3) yields more modest effects: formal connections would increase by 2.1 percentage points and firm profits by 5% under baseline prices, with an additional 0.5 percentage point gain in formal participation possible if prices were adjusted downward to maintain initial profit levels.

The social tariff proves to be an effective tool for reducing informality among low-income households. Compared to a scenario with uniform pricing and equivalent firm profits, the current price discrimination scheme increases formal market participation among low-income households by 8%. This substantial gain in formal connections requires only a 2% increase in baseline prices to compensate the firm, suggesting that targeted price discrimination can efficiently expand formal access while preserving revenue stability.

7 Conclusion

In this paper we discuss how informal consumer markets can affect a firm´s optimal decisions. Through theft, counterfeit, or illegal sharing of subscriptions, consumers may be able to consume goods without revenue accruing to the firm that produces those goods. Although the focus of our empirical setting is theft, the forces that we document are more general and applicable to any type of informal consumer markets that firms may face.

With the help of a stylized model, we start by explaining how firms can be affected by informality both through lower revenues and higher costs. Second, we document a causal

Table 7: Counterfactual Results

Panel A: Holding baseline prices constant	Baseline	No Theft	No Crime	No Social Tariff
	(1)	(2)	(3)	(4)
Ave formal consumption (kWh)	116.26	129.99	118.35	115.37
Share formal	0.816	1.000	0.837	0.812
Mass formal low income	0.053	0.065	0.054	0.049
Regular price (R$ / kWh)	0.405	0.405	0.405	0.405
Revenue (in MM R$)	189.68	212.00	193.06	192.04
Profit (in MM R$)	91.75	125.45	96.54	94.30

Panel B: Adjusting prices to keep profit constant	Baseline	No Theft	No Crime	No Social Tariff
	(1)	(2)	(3)	(4)
Ave formal consumption (kWh)	116.26	135.09	119.98	116.24
Share formal	0.816	1.000	0.842	0.815
Mass formal low income	0.053	0.065	0.054	0.049
Regular price (R$ / kWh)	0.405	0.333	0.390	0.397
Revenue (in MM R$)	189.68	181.70	188.28	193.50
Profit (in MM R$)	91.75	91.75	91.75	91.75

Note: This table presents counterfactual scenarios computed using parameter estimates from Table 4 column (2) and Table 3 columns (5) and (6). Column (2) presents results when theft is prohibited in the model; column (3) removes organized crime by setting crime dummies to zero in the extensive margin model; column (4) eliminates the social tariff by applying the regular tariff to all households. Panel A maintains baseline prices, while Panel B adjusts prices downward to hold firm profits constant across scenarios. 'Share Formal' represents the percentage of households with formal connections, and 'Mass formal low income' represents formal low-income households as a share of the total population.

relation between prices and the endogenous decision by consumers to be formal. We believe that we are the first to establish this relation. Then we move on to estimate a structural model of how consumers make choices under informality, which allow us to simulate alternative scenarios. We find that firms face a demand curve that is 1.63x more elastic which has substantial implications for pricing decisions. Moreover we find that if firms could eliminate theft, prices could be 17.7% lower; that reducing crime rates would reduce informality and also lower prices; and that price discrimination can be an effective way to curb informality.

Our findings have substantial implications for managers and regulators. Understanding how the shape of the demand curve is different in a context of informality, allows firms to optimally adjust their prices. It also provides information to managers on the potential gains of curbing informality or on potential solutions to do so, e.g. by offering a two-tier pricing scheme. Moreover, regulators and public officials can leverage our findings both to understand the benefits of fighting economic crime, and also to design optimal incentive schemes for utilities that operate under rate-of-return regulation.

Our paper also contribute specifically to our understanding of electricity theft, a significant phenomenon throughout the world, particularly in developing countries. This phenomenon has potential consequences for energy costs and the environment, as it forces generation capacity to be much larger with no gains for consumers. In this paper, we suggest ways to mitigate this problem.

Regulators or firms can potentially exert monitoring effort in order to reduce informality. This is an interesting topic which we leave for future research.

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.

Burkhardt, J., K. T. Gillingham, and P. K. Kopalle (2023): "Field experimental evidence on the effect of pricing on residential electricity conservation," *Management Science*, 69, 7784–7798.

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.

Cahana, M., N. Fabra, M. Reguant, and J. Wang (2022): "The distributional impacts of real-time pricing," *CEPR Discussion Paper No. DP17200.*

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 (2024): "Anuário Estatístico de Energia Elétrica 2018, ano base 2023," .

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.

Hirata, D. V., A. Cardoso, C. C. Grillo, O. A. dos Santos Junior, D. A. Lyra, and R. C. Dirk (2022): "The expansion of milícias in Rio de Janeiro. Political and economic advantages," *Journal of Illicit Economies and Development*, 4.

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," .

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," .

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.

Schutze, A. M. (2015): "A demanda de energia elétrica no Brasil," PhD thesis, Pontificia Universidade Católica, Rio de Janeiro, Brazil.

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.

TOJAL, C. AND B. GUIDETTI (2024): "Expansion of Piped Water and Sewer Networks: The Effects of Regulation," *Working Paper*.

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.