Pricing and Informality: Evidence from Energy Theft in Brazil

Davi Resende Gabriel Richter Marcelo Sant'Anna

André Trindade *

November 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 0.26 to 0.42, and reduces monopoly optimal prices by 21%. Eliminating theft altogether would allow the firm to reduce prices by 19% 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. Retail pharmacies in the US have seen an increase in theft which have pushed them to lock some of their most expensive products as an attempt to reduce costs. 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.[2]

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

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

[2] From these point onward, whenever we mention "informality," we mean "informality in consumer markets," i.e., situations where the consumer has access to the good without paying a price. Although the examples given here (piracy, counterfeit, sharing of passwords) are used for motivation and can broadly fit the topic that we study in this paper, our empirical model is mostly focused on the issue of theft.

that is consumed but not billed (i.e. stolen from the grid)[3]—is pervasive and can 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),[4] 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.[5]

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.[6]

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 37 TWh of stolen energy in 2023). 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

[3]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.

[4]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.

[5]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.

[6]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.

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. To address this challenge, we leverage a natural experiment from 2011 when electricity prices increased exogenously and *permanently* to a subset of consumers. Although, of course, we cannot directly observe customers in the informal market, we can proxy for those by identifying the fraction of total households that is not buying in the formal market.[7]

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 satiation. 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. We address potential price endogeneity concerns by using "tariff flags," a component of the retail price that proxies for generation costs, as instrument. This estimated demand 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

[7] In section 2 we discuss why alternatives to formal/informal, such as not having power, buying a diesel generator, or solar panel are not meaningful in our setting.

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 1.61x more elastic due to presence informality (elasticity moving from -0.26 to -0.42). Consequently, an unregulated monopolist would optimally choose prices that were 26.3% higher if theft was not possible, compared with the status-quo.

The focus of this paper is on the impact of informality on optimal pricing decisions. At the same time, in our empirical setting (a monopolist electricity retailer), prices are typically regulated. For this reason, we use our demand estimates to simulate different counterfactual scenarios where i) prices can be freely chosen by the firm and ii) prices are regulated. For the latter, we consider sub-scenarios where either prices or profits are fixed by the regulator. We find that: 1) the presence of informality significantly restricts the monopolist's pricing power, 2) under the current regulated regime, eliminating energy theft altogether would allow prices to go down further be 18.5%, while keeping firm profits constant, and 3) 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

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

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 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.

Our paper is one of the first to document that consumers are rational economic agents, even when deciding to engage in theft. This idea is in line with Draca et al. (2019), in another setting, who document that criminals respond to changes in the price of goods by stealing property with high value. Schnell (2017) goes further and also shows that firms (physicians) alter their behavior in the presence of informal (secondary) markets on the consumer side.

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 consequences of electricity theft for pricing and welfare. 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.

Finally, our paper decomposes the elasticity of demand between an intensive and extensive margins. This is related with the marketing literature that tries to decompose the effect of promotions in the likelihood of a purchase and the quantity chosen (see for example Chintagunta (1993) or Bell et al. (1999) using household data, or Nair et al. (2005) using aggregate data)

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 our empirical model and discuss identification and estimation. 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, Brazil consumed 532 TWh of electricity, ranking among the world's ten largest consumers. The country's total installed capacity exceeded 226 GW, with hydropower representing roughly 48.6% of generation and the remainder coming primarily from natural gas, biomass, and nuclear sources (EPE, 2024).

Around 50 different local regulated utilities distribute electricity across Brazil. Most are privately owned, though several remain state-controlled. Consumer electricity prices are regulated by ANEEL (*Agência Nacional de Energia Elétrica*). Regulated prices are allowed to vary across utilities, though not within, to reflect heterogeneity in productive efficiency, demand conditions, and other factors. Some low-income consumers qualify for a lower "social tariff" with subsidized rates.

In 2015, ANEEL introduced a system of "tariff flags" that change monthly and vary the final price consumers pay based on the flag's color (red, yellow, or green). The flag color represents general conditions of the electric generation system, with the goal of having consumers internalize differences in generation costs over time and adjust consumption accordingly.

2.2 The 2011 change in social tariffs

As noted in the previous section, Brazil offers a social tariff providing discounted electricity rates to low-income households. The discount applies to marginal rates and decreases with consumption: households receive a 65% discount for consumption up to 30 kWh/month, a 40% discount between 30 and 100 kWh/month, a 10% discount between 100 and 220 kWh/month, and no discount beyond that point.

While these discount brackets have remained constant over time, the eligibility rules changed substantially in 2011. Previously, any household consuming below 80 kWh/month had automatic access to the social tariff. Above that threshold, households needed to provide evidence of a low-income status. From 2011 onward, everyone, regardless of consumption level, needed to also qualify as low-income to access the discounts. This policy change resulted in a large number of consumers being excluded from the program and facing substantial price increases.

The main condition to qualify as "low-income" is for the household to be registered in *CadUnico* (short for *Cadastro Único para Programas Sociais do Governo Federal*) -— the Brazilian government database that gathers socioeconomic information about low-income families. The incentives for qualifying households to be registered in CadUnico are high as this is used to identify participants in every other governmental social program or benefit (direct income support, housing program, fee exemptions for universities or federal exams, among other).

2.3 Energy theft

Electricity distribution involves two types of losses: technical (TL) and non-technical (NTL). Technical losses are natural losses inherent to transporting electricity from one place to another and depend on infrastructure quality. Non-technical losses consist mostly of electricity theft or measurement error. In 2023, total electricity lost in Brazil represented 14% of electricity injected into the system, roughly equally divided between TL (7.3%) and NTL (6.7%). Total NTL that year exceeded 37 TWh. These percentages are somewhat misleading because most NTL occurs in the residential sector. Therefore, while it is natural for the denominator of TL to be the amount of electricity injected, the usual approach compares NTL with total electricity in that sector. In that case, the percentage of NTL rises to 14.3%. This figure still hides substantial heterogeneity: at least seven utilities have NTL higher than 30% of residential consumption. This includes the firm we study, which accounts for almost 29% of Brazil's total NTL (ANEEL, 2019). Understanding how this theft occurs requires examining both the technical methods employed and the actors involved.

Informal electricity consumption most commonly involves unauthorized access to the utility's distribution network through connections known locally as "gatos." This method typically uses illegal taps that draw power from overhead distribution lines before electricity reaches meters, leaving consumption unrecorded and unbilled. In many neighborhoods, especially densely populated, lower-income areas, households rely on makeshift cables extending precariously from utility poles, creating hazardous wire clusters. These connections involve relatively low costs and readily available materials, making informal access widespread. However, they carry significant safety risks: inadequate insulation and poor installation practices frequently lead to short circuits, transformer overloads, fires, and electrocutions.

The provision of illicit connections involves diverse actors. At smaller scales, local residents and unlicensed electricians perform illegal hookups for small fees. At larger scales, particularly in areas with limited government presence, criminal groups including militias and drug trafficking factions systematically manage electricity theft. Many areas with high NTL in Brazil fall under the control of these 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., 2022). These groups install and control extensive illegal distribution networks, charging households fixed monthly fees for unlimited electricity access. Operating essentially as shadow utilities, they monopolize local infrastructure, prevent utility personnel from conducting inspections or repairs.

Despite widespread awareness of theft, utilities face substantial barriers to detecting individual violators and shutting down informal consumption. Since stolen electricity is unmetered by definition, utilities can detect theft only indirectly by analyzing discrepancies between energy delivered to local feeders and total metered consumption. Identifying specific illegal connections or meter tampering requires manual, on-site home inspections, which are costly, logistically complex, and often unsafe. In areas controlled by criminal groups, utility crews frequently face intimidation, threats, or direct violence, effectively preventing inspections or interventions. Moreover, most theft-prone neighborhoods still rely on older meters requiring manual reading, providing no immediate or remote detection of irregularities. Modern smart meters, capable of real-time monitoring and remote reporting, remain rare in these locations. Consequently, utilities lack the granular, timely information needed to effectively target and disconnect illegal consumers.

When utilities successfully identify theft, Brazilian law classifies it as larceny, punishable by prison sentences ranging from one to four years plus monetary fines. Utilities can legally recover losses through retroactive billing, estimating and charging offenders for previously unmetered consumption, and imposing penalty fees. Officially, offenders must disconnect from

their illegal supply and formalize a contract before regaining electricity access. However, enforcement remains uneven and infrequent in practice, constrained by legal complexities, operational difficulties, and security concerns. While utilities periodically undertake anti-theft operations, these efforts are usually limited in scope, particularly in areas dominated by criminal groups. Moreover, informal consumption is not confined to lower-income communities: affluent areas also experience theft, typically through more sophisticated methods like meter tampering or bypassing.

2.4 Alternatives to using the electric grid

Given the challenges with enforcement described above, one might ask whether Brazilian households have viable alternatives to purchasing electricity from regulated utilities. In practice, such alternatives are quite limited. The main substitute is distributed solar photovoltaic (PV) generation, where consumers install solar panels to produce their own electricity and inject any surplus into the grid in exchange for bill reductions. Although this technology has expanded nationwide in recent years, adoption remains relatively low (1.5% of total residential units in our area as of 2025). Moreover, households with solar panels remain connected to the grid and continue as formal consumers, as they rely on grid electricity when solar generation is insufficient, such as during nighttime hours or periods of low sunlight.

Other alternatives, such as gasoline-powered generators, are even less feasible. They are noisy, costly to operate, and typically used as temporary backup rather than a structural substitute for regular electricity supply. Furthermore, entirely foregoing electricity access is generally unrealistic in urban settings. Thus, in practice, the dominant alternative to formal electricity access is informal consumption through theft.

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. We provide a summary of data sources in Appendix Table C.2.

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.

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 at each point in time.

Prices. We use aggregated electricity price data obtained from ANEEL for the period between 2010 and 2022. 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. 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. The latter is defined as the sum of power used by all the electric equipment that exists in a given household, i.e., it represents the amount of electricity that the unit would need if all the appliances, etc, were turned on at the same time. Installed capacity is a metric which the utility estimates for each household based on customer surveys. 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 six demographic variables: i) income; ii) household size; iii) universe of households; iv) crime; v) illiteracy rate; and vi) census tract density. Income, household size, illiteracy rate, and density 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 at the census tract level. The illiteracy rate is the percentage of people aged 15 and older who cannot read or write at the census tract level. The density is the number of households per square kilometer, also measured 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.

3.2 Descriptive Statistics

Table 1 presents descriptive statistics for the main variables used in our empirical analysis. For each variable, we report the mean, three quartiles (25th, 50th, and 75th percentiles), and minimum and maximum values. Panel A describes household-level monthly data from the BDGD database (2017-2022), including electricity consumption, price paid, a social tariff indicator (equal to one if the household qualifies for subsidized rates), and installed capacity. Panel B presents census block-level characteristics from the 2010 Population Census, matched to households based on building point location. These variables include average household income, household size, illiteracy rate (as percentage), and population density. Panel C reports the Crime Area Indicator from the Fogo Cruzado Institute (2019), which equals one for census blocks controlled by criminal organizations (drug traffickers or militia groups) or contested by rival gangs.

Figure 1 presents time series variation in both real base electricity prices and real total prices including tariff flags. Real base prices exhibit a predictable pattern: sharp increases at regulatory adjustments followed by gradual declines as inflation erodes nominal values. In contrast, total prices paid by households (base price plus tariff flags) display considerably more volatility. As discussed earlier, tariff flags are flexibly adjusted by regulators when system conditions require more expensive thermal generators. This mechanism is evident in the data: severe droughts in 2014-2015 triggered substantial price spikes through tariff flags in 2015-2016, while the 2021 spike reflected the most severe drought in 91 years. These drought events reduce the availability of cheaper hydroelectric power, necessitating the dispatch of costlier thermal generation sources.

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 2 shows the number of total clients, and number of clients with a regular tariff, before and after this

[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.

Table 1: Summary Statistics

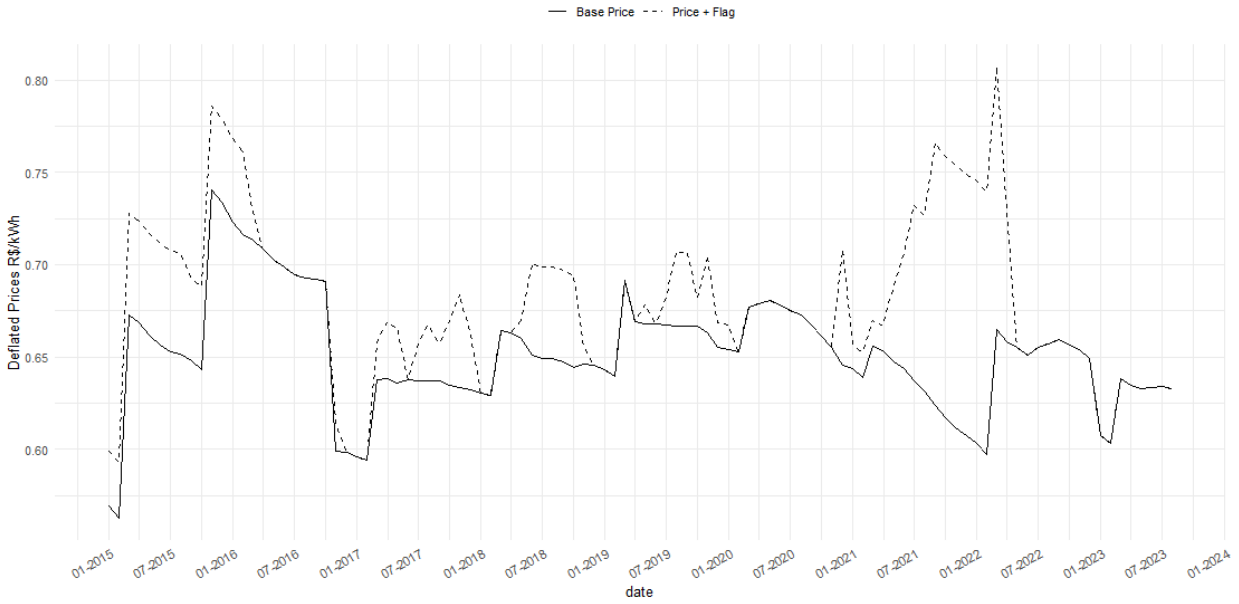
Panel A: Household Panel (BDGD, 2017-2022), N = 3,291,411						
	Mean	**P25**	**Median**	**P75**	**Min**	**Max**
Consumption (kWh/month)	201.5	90.3	161.6	257.3	0.62	1656
Final HH Price (R$/kWh)	0.729	0.704	0.757	0.767	0.208	0.913
Social Tariff Indicator	0.065	0	0	0	0	1
Installed Capacity (kW)	0.913	0.552	0.799	1.117	0.216	5.9

Panel B: Census Block Level (Population Census, 2010), N = 4,107,529						
	Mean	**P25**	**Median**	**P75**	**Min**	**Max**
Income (R$/month)	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
Illiteracy rate	3.08	0.98	2.18	4.22	0.0	56.41
Density (Hh/Km^2)	5254.6	2824.3	4764.6	6884.5	0.44	16996.6

Panel C: Census Block Level (Fogo Cruzado Institute, 2019), N = 4,107,529						
	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 household-level information on formal utility customers. All data in panel A comes from ANELL. Panel B includes statistics from the Brazilian census on household income distribution, household size, illiteracy rate, and census tract density. Illiteracy rate is measured by the number of people above 15 years old who don't know how to read and write, divided by the number of people above 15 years old within the census tract. Finally, Panel C presents statistics on crime across census block areas. This is constructed by cross-referencing the *Fogo Cruzado* data with the Census and is defined as 1 if the area has a known traffic or militia group in control, or if it is a disputed area controlled by different factions.

Figure 1: Base electricity price and price including "tariff flags"



Note: This figure displays the real base price and final price including "tariff flags," which is actually paid by households from 2015 to 2023. Real values in December 2021 levels.

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.

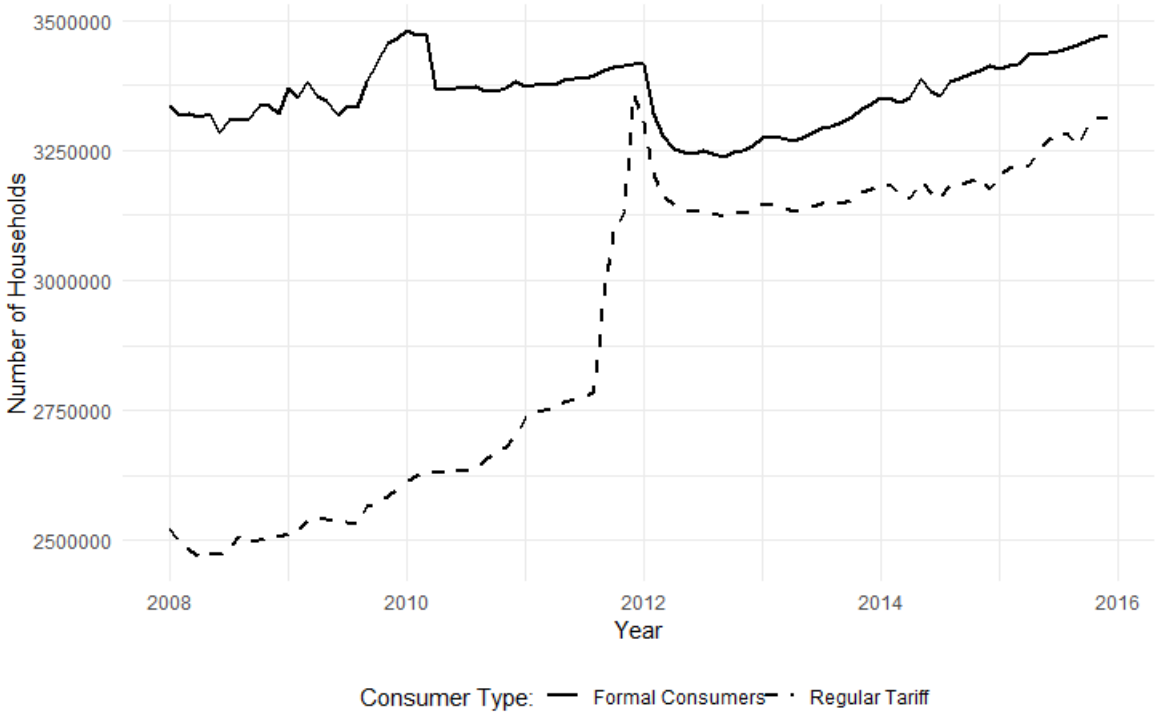
Figure 3 shows the impact from the change in policy on the number of customers with a social tariff. We find that the largest drop occurs in customers with a marginal rate of 35% and 60% of the regular prices. This is expected as those were the groups in the lower quantity brackets and were, therefore,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

4 Model

4.1 Framework

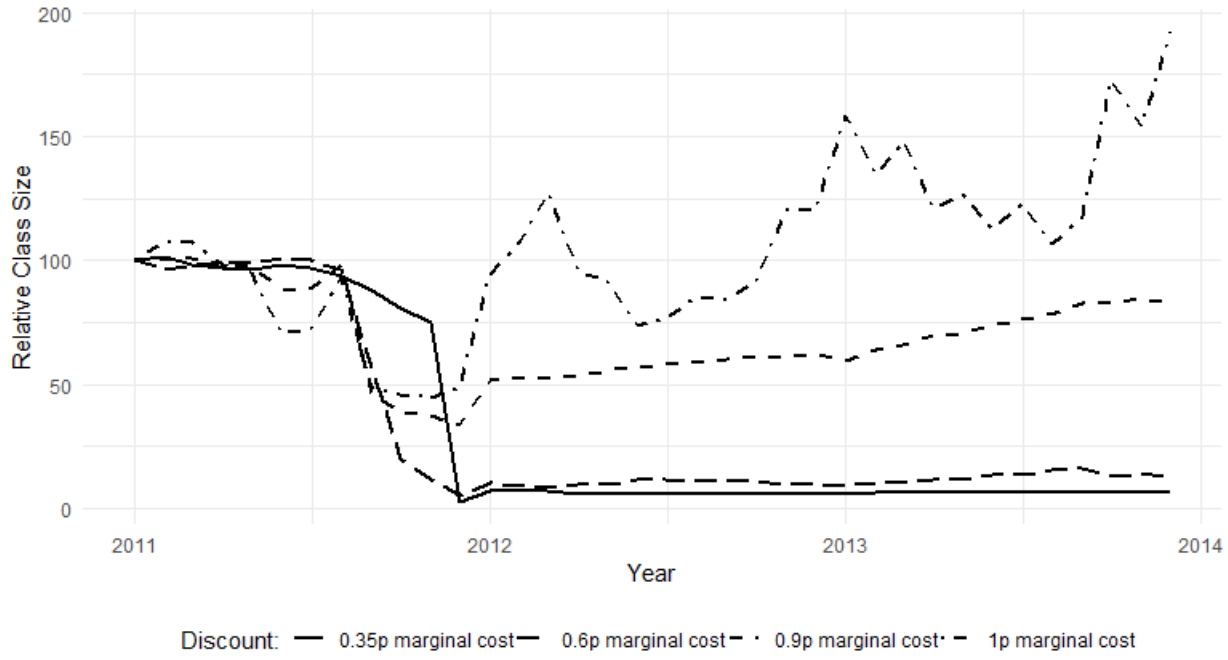
We propose a model in which households decide on two nested margins. First, they need to decide between being formal consumers, paying for the electricity the assigned prices,

Figure 2: 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 3: 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.

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. It depends on the consumption of electricity, q, and a numeraire good, m:[10]

$$U_{it}(q, m) = \theta_{it} q^{\frac{\xi-1}{\xi}} + m. \tag{3}$$

This functional form yields a constant electricity demand elasticity, ξ, while allowing for household-specific demand shifters θ_{it}. Substituting the budget constraint, we have:

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

where y_{it} is household income. This specification applies both to formal consumption and theft, with one key distinction: formal consumers pay the utility's positive price p, while households stealing electricity 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 thus assume informal demand is capped at level $\bar{q}_{it}$– the highest consumption value that is feasible given capacity restrictions. We discuss how we recover $\bar{q}_{it}$ from the data in Section 4.3.1.

[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.

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 extreme value distribution, η_{jt} is a fixed formality utility shifter and

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

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 $\eta_{1t} = 0$.

We thus propose a model of household decisions about energy theft that is grounded in consumer theory. The extensive margin decision is based on welfare comparisons derived from estimated demand parameters. In this framework, consumer surplus acts as the mechanism through which electricity prices affect informality decisions: higher prices reduce the surplus from formal consumption, increasing the welfare differential that favors informal access. The model remains flexible enough to accommodate observed and unobserved market-level shifters and household heterogeneity.

While presented in a static form, our extensive margin model can be interpreted as reflecting forward-looking household decisions. Under this interpretation, the formality utility shifter η_{0t} captures both an upfront fixed cost of converting to informality (such as the cost of theft equipment) and the present discounted value of ongoing flow differences between formal and informal status. These flow differences include the expected cost of sanctions if caught stealing and the benefits of formal connections such as service reliability. This forward-looking interpretation is consistent with our identification strategy, which relies on permanent price variation from the 2011 experiment rather than transitory fluctuations. In Appendix A, we formalize this forward-looking interpretation of our model. A limitation is that we cannot separately identify the fixed and flow components of η_{0t}. However, this

limitation is not severe for our purposes, as it does not affect our main counterfactuals of interest related to pricing.

4.2 Identification

This section discusses the main identification challenges that arise when connecting the model to the data, and how we address them.

Price endogeneity. A potential concern is that prices may be correlated with demand shocks, rendering OLS estimates of the intensive margin demand parameters inconsistent. In our setting, this price endogeneity can happen via two mechanisms, which we detail below.

The first mechanism operates at the *micro-level* and can happen if the price paid by individual consumers is endogenously chosen as a function of their electricity needs (as in Reiss and White, 2005). This occurs under non-linear pricing schedules where the marginal price varies with the quantity consumed. While this type of pricing is common in services provided by public utilities (electricity, water, gas, etc), it is less prevalent in our setting. Approximately 93% of the total number of consumers pay a constant marginal price, which varies only over time, but not across consumers. Only customers enrolled in the social tariff program (the remaining 7%) face a tiered pricing structure where marginal prices increase with consumption.

The second operates at the *macro-level* and takes place if the price is set by the firm in response to aggregate demand shocks. For example, if the temperature next month is expected to be abnormally high and the utility increases the marginal rates in order to avoid capacity constraints in the system.

To address these potential concerns, we use different strategies. First, we include building and month fixed effects in our specification, which directly absorb most of the cross-sectional (time-invariant) heterogeneity and month seasonality. To address remaining sources of potential endogeneity we use an instrumental variable approach.

Our instruments are the "tariff flags" detailed in Section 2. These flags determine a surcharge included in the final price consumers pay, reflecting the system's generation costs in a given month. When the flag is green, generation costs are low and no surcharge applies. When the flag is yellow or red, generation costs are high—typically due to low rainfall requiring dispatch of thermoelectric plants—and consumer prices increase accordingly.

These flags satisfy the requirements for valid instruments. First, they are strongly correlated with the final price paid by consumers, as the surcharge is a component of the retail price. Second, they satisfy the exclusion restriction: they are determined by system-wide

hydro-meteorological conditions and the resulting generation costs, which are plausibly uncorrelated with local demand shocks. Moreover, flags and the corresponding surcharge are set at the national level hence not affected by idiosyncratic demand conditions in the state of Rio de Janeiro.

Selection. Another potential concern in our setting is that the error terms in the intensive and extensive margins are not independent. This can introduce a bias in our estimates given that, by construction, we only observe formal consumption for those individuals that choose to be in the formal sector. For example, if a household lives in a bigger house, it may demand relatively more energy and may also have a higher probability of getting caught if it engages in electricity theft. On the contrary, imagine that some individuals are more risk averse than others. Those may be less likely to engage in energy theft (to avoid the risk of being caught) but may also be more cautious in terms of their spending on appliances and electricity. The former example would induce a positive correlation between the probability of being formal and the quantity of energy consumed, while in the latter it would induce a negative correlation. In a nutshell, a potential unobserved correlation between the intensive and extensive margins is a possibility in our setting but, unlike other discrete-continuous problems (Dubin and McFadden, 1984; Barreca et al., 2016), the direction of the bias is not clear.

To address this problem, we adapt the control function approach by Dubin and McFadden (1984). This approach assumes that the error term in the informality decision is independently distributed, and that the error term in the quantity decision is a function of the former. As such, we can use the estimated probabilities from the (binary) discrete choice as a way to directly capture a potential selection effect. Specifically, we introduce the following "Dubin-McFadden" term as an additional regressor in the intensive margin equation:

$$\lambda^{DM} \equiv \frac{\hat{P}_{1,it}\log(\hat{P}_{1,it})}{1-\hat{P}_{1,it}} + \log(\hat{P}_{0,it}).$$

where $\hat{P}_{1,it}$ and $\hat{P}_{0,it}$ are, respectively, the estimated probabilities of being formal and informal. The sign of the coefficient associated with λ^{DM} can be positive or negative, depending on the direction of the bias.

Long-run vs Short-run price variation. Given the forward-looking nature of the extensive margin decision described in Section 4.1.2, our identification strategy exploits permanent rather than transitory price variation. While in principle we could use the short-run price

fluctuations that identify the intensive margin, such transitory changes provide limited information about theft decisions that involve upfront fixed costs and depend on households' expectations about future price evolution. We therefore seek substantial permanent price variations to identify extensive margin responses, which in turn enables us to evaluate counterfactuals involving permanent policy shifts. Consistent with this forward-looking interpretation of our static model, we identify the extensive margin parameters using a permanent policy shift that occurred 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:

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

where p_t is the baseline price, γ_i are individual fixed effects (in practice, building fixed effects), and γ_t are month of the year and year fixed effects.[11] The error term ν_{it} can be written as $\bar{\nu}_t + \Delta\nu_{it}$, where the first component is structural—i.e., observed by the utility that sets the price—and the second component is an idiosyncratic deviation from the mean. To address the potential price endogeneity, we estimate equation 8 using instrumental variables, as detailed in the section above. Note as well that we include a selection correction term (λ^{DM}) when estimating this equation, as detailed in section 4.2.[12]

Given the demand elasticity (ξ) estimate, we can set the demand shifter, θ_{it}, in formal

[11] Households with access to the social tariff (6.5% of the sample) face discounted prices. Since these discounts apply as fixed percentage reductions and equation (8) uses the log of price, the discount is absorbed by household fixed effects provided that price changes do not move households into different discount brackets.

[12] In our setting, we estimate the intensive margin before the discrete decision to be formal or not. This implies that in practice we have to iterate between the two, i.e., we start by obtaining a preliminary estimate of $P_{1,it}$, which we then use as a control function in the quantity equation. We then use this corrected estimates to obtain the final estimates of $P_{1,it}$, which we report in our tables.

demand (equation 5) such that at t, demand is exactly as observed, that is, $d_{it}(p_t) = q_{it}$:

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

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 essential for the results. To do so, we assume that $\bar{q}$ is proportional to installed capacity (IC)—a variable that we observe with significant detail.[13] Specifically, we estimate a linear demand regression (10) of the actual quantity on installed capacity and prices, using the same set of controls as before:

$$q_{it} = \gamma_i + \gamma_t - \alpha p_t + \nu_{it}. \tag{10}$$

To address price endogeneity, we employ the same "tariff flag" instrument used previously. We then predict $\bar{q}_i$ from the estimated regression by setting $p_t = 0$, capturing consumption behavior when electricity is costless.

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:

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

$$\mathbb{P}^{t+1}(formal) = \frac{N^{t+1}_{form}}{N} = \sum_{i \in I} \frac{\exp\Big(\beta(\psi_{0i}(p_{t+1}(q_{it})) - \psi_{1,i}) + \eta_0\Big)}{1 + \exp\Big(\beta(\psi_{0i}(p_{t+1}(q_{it})) - \psi_{1,i}) + \eta_0\Big)} w_i^t, \tag{12}$$

[13] In Appendix B, we present an alternative approach to constructing $\bar{q}_i$ based on average non-technical losses per informal consumer and show that our main results are robust to this specification.

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.[14] 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 (11) and (12) to recover the extensive margin parameters, we need to overcome two related challenges. First, we only observe aggregate data in the immediate 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 Bayes' rule:

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

Therefore we can express the unconditional weight at T as

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

This is helpful because although we do not observe $\mathbb{P}[formal|i, T]$ directly, we can compute

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

this for a given parameter pair (β, η).

We are not over because in (11) and (12) 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.[15] Therefore, we assume:

$$w_i^t = \gamma_c w_i^T \text{ for all } i \in c, \tag{13}$$

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:

$$\sum_{i \in c} \mathbb{P}[formal|i,t] w_i^t = \frac{N_{form,c}^t}{N}, \tag{14}$$

where $N_{form,c}^t$ is the number of formal households in consumption category c at t. Therefore, from (13) and (14) we can write the unconditional weights at t as

$$w_i^t = \frac{N_{form,c}^t}{N} \frac{w_i^T}{\sum_{j \in c} \mathbb{P}[formal|j,t] w_j^T} \text{ for all } i \in c. \tag{15}$$

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 (11) and (12),[16] using consistently updated weights. The algorithm starts with *ad-hoc* weights, recovers the parameters by solving the system of equations (11) and (12). These newly found parameters update the weights consistently with the choice model. It then iterates until convergence. We describe

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

[16] For this process to work, we assume that the primitives of the model, namely β and ξ are constant throughout the earlier and later periods. Moreover, we assume that the empirical relative distribution of consumers that we observe in the later period (in terms of installed capacity, etc) would be the same in the earlier period, although the total number of individuals is allowed to vary between them.

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 (11) for $t = 2011$ and $t + 1 = 2012$ and previous recorded weights $w_i^{t,k}$.[17]

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:[18]

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

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

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

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

[18] 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.

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

$$\eta_{0\ell,T} = \gamma' x_{\ell,T} + \tilde{\eta}_{\ell,T}, \tag{16}$$

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

$$\sigma_{\ell,T}(\eta_{0\ell,T};\beta) = \frac{1}{N^T_{form,\ell}} \sum_{i=1}^{N^T_{form,\ell}} \frac{\exp\left(\beta(\psi_{0i}(p_T(q_{iT})) - \psi_{1it}) + \eta_{0\ell,T}\right)}{1 + \exp\left(\beta(\psi_{0i}(p_T(q_{iT})) - \psi_{1it}) + \eta_{0\ell,T}\right)}. \tag{17}$$

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 (16).

5 Estimation Results

We begin with our estimates from the first step. Table 2 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 7 specifications in this table. All of them include month and year fixed effects, a dummy for households that have access to the social tariff, and building fixed effects. While column (1) reports the results for a pooled regression, in the remaining columns we split the sample in those customers that are above or below the median installed capacity, to allow for additional heterogeneity along that dimension. Moreover, columns (6) and (7) include the Dubin and McFadden (1984) adjustment term. Finally, while columns (2) and (3) report OLS results, the remaining specifications use the instrumental variable detailed above. 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 estimate in the pooled regression is -0.21. This result is roughly consistent with other

Table 2: Intensive Margin Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ξ	-0.21	-0.0904	-0.267	-0.1593	-0.2681	-0.1585	-0.2685
	(0.12)	(0.0029)	(0.0020)	(0.0034)	(0.0020)	(0.0021)	(0.0021)
capacity	1.311	4.617	0.6517	4.603	0.1058	4.611	0.3119
	(0.0020)	(0.0028)	(0.0008)	(0.0027)	(0.0355)	(0.0028)	(0.0096)
Adj. Term						9e-8	5.6e-10
						(1e-5)	(1e-5)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Social Tariff FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Building FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV	Yes	No	No	Yes	Yes	Yes	Yes
Capacity Heterogeneity	No	Below Median IC	Above Median IC	Below Median IC	Above Median IC	Below Median IC	Above Median IC
Observations	230mm	115mm	115mm	115mm	115mm	88mm	88mm
R-squared	0.53961	0.5284	0.5321	0.52941	0.3193	0.5324	0.3863

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 all columns, we use the installed capacity of each household as a regressor. Column (1) presents the regression for the entire sample, excluding heterogeneity in installed capacity, but using IV. Columns (2) and (3) present the results for below and above the median of capacity, without IV. Columns (4) and (5) are the 2SLS estimation with *bandeiras* as the exogenous regressor for prices. Columns (6) and (7) add the adjustment term built following Dubin and McFadden (1984) and Barreca et al. (2016). Our sample covers the period from 2017 to 2022. Standard errors are in parentheses.

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 -0.16, column 6) compared to those above the median (elasticity of -0.27, column 7). This heterogeneity is consistent with higher-capacity households having more appliances and thus greater flexibility to adjust consumption in response to price changes. Specifically, these households are more likely to own discretionary devices such as multiple air conditioning units, whereas lower-capacity households' consumption is concentrated in essential appliances like refrigerators that are costly to adjust. This pattern aligns evidence suggesting that lower-income communities exhibit slightly lower price sensitivity (Deryugina et al., 2020).

Beyond the demand elasticity, a second key input for constructing consumer surplus under each regime is the level of electricity consumption for informal consumers, $\bar{q}_i$. We follow the approach discussed in Section 4.3.1.[19] In Appendix B, we present an alternative approach to constructing $\bar{q}_i$ based on non-technical losses data. Both the distribution of $\bar{q}_i$ and our main elasticity decomposition results remain similar under this alternative specification.

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 11, 12, and 17), which we then estimate. Our second step results are reported in Table 3. 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 3 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

[19] Appendix Figure C.1 reports the joint distribution of electricity consumption and $\bar{q}_i$. Results from the linear-in-prices demand regression used to construct $\bar{q}_i$ are reported in Appendix Table C.1.

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 parameters estimated from our preferred specification (Table 3, column 2), we first compute aggregate demand elasticities, which we report in Table 4. We find that, conditional on being formal, demand elasticity is -0.258.[20] When accounting for switches between formal and informal status, the aggregate total formal demand elasticity increases to -0.415, i.e. 1.6 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 decomposes 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.

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.[21] We report the optimal prices: (1) under the status-quo (informality), (2) without the evasion adjustment term, and (3) with the removal of informality altogether (i.e., no evasion nor theft adjustment terms). We find that removing informality would increase the monopolist prices by 26.3% 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

[20] This aggregate elasticity falls between the values for below- and above-median installed capacity reported in Table 2, columns (5) and (6). It is closer to the above-median value due to higher-consumption households' greater weight in aggregate formal demand.

[21] 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 3: Extensive Margin Results

	(1)	(2)
β	0.041	0.041
	(0.009)	(0.009)
η	6.805	-7.832
	(1.193)	(1.59)
γ (log income)		0.892
		(0.184)
γ (avg hh size)		0.120
		(0.179)
γ (traffic)		-0.809
		(0.130)
γ (militia)		-0.655
		(0.127)
γ (dispute)		-0.39
		(0.155)
γ (illiteracy)		-0.078
		(0.022)
γ (density)		-0.00001
		(0.000001)
γ (avg capacity)		0.028
		(0.003)
Capacity in 1st stage	Yes	Yes
Capacity/Elasticity Heterogeneity	Yes	Yes
Spatial heterogeneity	N/A	Neighborhood
Observations		779
F-stat		97.7
R_squared (adj)		0.497

Note: This table reports estimates from the second step (extensive margin) - equation 17, 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.

Table 4: 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	-0.415	-0.258	-0.157
ξ_x/ξ_{Total}	1.00	0.622	0.378

Note: This table reports total formal demand elasticity decomposition based on the preferred specifications (Table 2, columns 5 and 6, and Table 3, 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.

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%.

Table 5: Optimal Prices

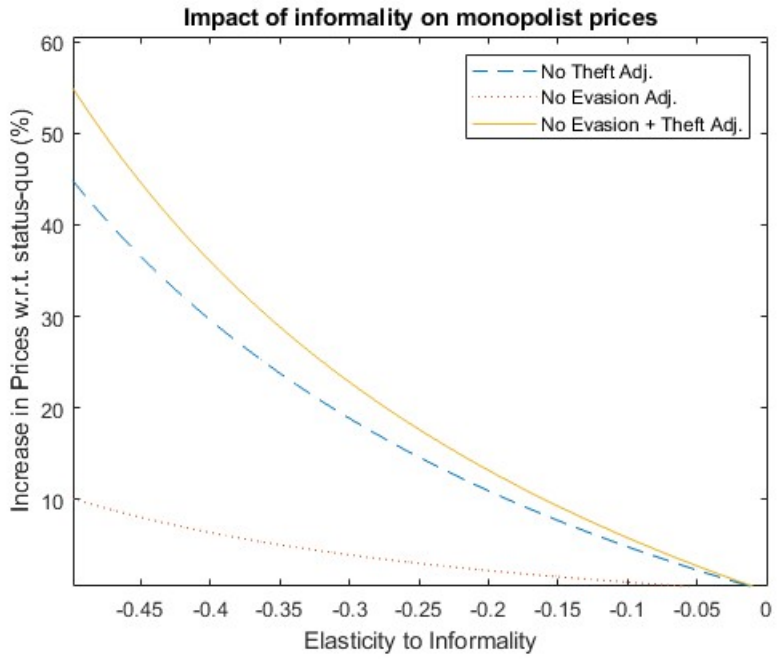
	Price	ΔPrice
(1) Status-quo	1.12	-
(2) No informality (no Evasion adj)	1.15	2.1%
(3) No informality (no Evasion + Theft adj)	1.42	26.3%

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 3, column 2, to the FOC of the firm given by equation 2.

6 Counterfactuals

Using our preferred specification (from Table 3 column (2) and Table 2 columns (6) and (7)), we evaluate three counterfactual scenarios, which we report in Table 6. In the first scenario (panel A), we keep regulated prices constant, in the second (panel B) we keep profit constant, and in the third (panel C) we compute the optimal monopolist prices with theft. We believe that all counterfactuals are policy and managerial relevant. While panels A and B provide insights under the current institutional setting, where prices are kept lower through regulation, panel C links back to the core question of the paper which is how (unrestricted)

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.

pricing decisions are affected by the presence of informal consumer markets. For each panel, we report three alternatives: i) Baseline (where informal consumer markets are present), ii) No theft (where informal consumer markets are not present), and iii) No social tariff (where price discrimination between low- and high-income is removed).

Eliminating electricity theft (column 2) increases firm profits by almost 40% if prices remain at baseline levels (Panel A). Alternatively, maintaining current profit levels constant (Panel B) allows for a 18.5% reduction in regular prices, if theft were eliminated. In panel C, we find that the optimal monopolist price when theft is present in the market is roughly 50% higher than the regulated price. The alternative "zero-theft monopoly price" would be unbounded in our setting. This occurs because the intensive-margin specification assumes a constant elasticity of demand, which we estimate to be below one (consistent with the literature). Under such conditions, a monopolist could always increase profits by raising prices. This implies that the presence of informality significantly restricts the monopolist's pricing power (in the absence of theft it would be virtually unlimited, whereas with theft it is capped at 0.61 R$/kWh).

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 (panel B), 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. Moreover, the overall share of formal consumers would also increase, suggesting that targeted price discrimination can efficiently expand formal access while preserving revenue stability.

A relevant additional question is how a large price decrease affects the overall share of formal consumers. By comparing the Baseline column in panels B and C, we observe that reducing the price from 0.61 to 0.41 increases the share of formal consumers from 69% to 81%.

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 6: Counterfactual Results

Panel A: Holding baseline prices constant			
	Baseline	No Theft	No Social Tariff
	(1)	(2)	(3)
Avg formal consumption (kWh)	116.52	164.06	115.55
Share formal	0.809	1.000	0.805
Mass formal low income	0.052	0.063	0.048
Regular price (R$ / kWh)	0.405	0.405	0.405
Revenue (in MM R$)	190.1	213.6	192.3
Profit (in MM R$)	91.6	126.4	94.1

Panel B: Adjusting prices to keep profit constant			
	Baseline	No Theft	No Social Tariff
	(1)	(2)	(3)
Avg formal consumption (kWh)	116.52	171.5	116.52
Share formal	0.809	1.000	0.808
Mass formal low income	0.052	0.063	0.048
Regular price (R$ / kWh)	0.405	0.33	0.397
Revenue (in MM R$)	190.1	182.6	193.9
Profit (in MM R$)	91.6	91.6	91.6

Panel C: Optimal baseline prices			
	Baseline	No Theft	No Social Tariff
	(1)	(2)	(3)
Avg formal consumption (kWh)	90.6	148.4	89.9
Share formal	0.69	1.000	0.68
Mass formal low income	0.049	0.071	0.04
Regular price (R$ / kWh)	0.61	0.61	0.61
Revenue (in MM R$)	222	292	225
Profit (in MM R$)	127	213	129

Note: This table presents counterfactual scenarios computed using parameter estimates from Table 3 column (2) and Table 2 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 maintain constant profits 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 1.6x more elastic which has substantial implications for pricing decisions. Moreover we find that if firms could eliminate theft, prices could be 18.5% lower; that he presence of informality significantly restricts the monopolist's pricing power; and that price discrimination can be an effective way to curb informality.

Like all empirical research, our analysis relies on specific identifying assumptions limited by data availability. We only observe aggregate shares of formal consumers for the 2011 price experiment period, which limits the heterogeneity we can incorporate into our extensive margin model. This contrasts with the intensive margin demand model, where household-level panel data allow us to specify a much more flexible framework. Moreover, we require the parameter governing the extensive margin decision to remain stable across time. Some of these limitations, however, are inherent to any study of theft or informal markets, where individual-level participation is by nature hard to observe.

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.

Barreca, A., K. Clay, O. Deschenes, M. Greenstone, and J. S. Shapiro (2016): "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century," *Journal of Political Economy*, 124, 105–159.

Bell, D. R., J. Chiang, and V. Padmanabhan (1999): "The decomposition of promotional response: An empirical generalization," *Marketing Science*, 18, 504–526.

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*.

Chintagunta, P. K. (1993): "Investigating purchase incidence, brand choice and purchase quantity decisions of households," *Marketing Science*, 12, 184–208.

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*.

Draca, M., T. Koutmeridis, and S. Machin (2019): "The changing returns to crime: Do criminals respond to prices?" *The Review of Economic Studies*, 86, 1228–1257.

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.

Nair, H., J.-P. Dubé, and P. Chintagunta (2005): "Accounting for primary and secondary demand effects with aggregate data," *Marketing Science*, 24, 444–460.

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.

Schnell, M. (2017): "Physician behavior in the presence of a secondary market: The case of prescription opioids," *Princeton University Department of Economics Working Paper*, 5, 383–410.

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.

Appendix (for online publication)

A Extensive Margin Model

In the main text we presented a static extensive margin model of the formality decision, which we interpret as a long-run choice. In this section we formalize the connection between the static model we estimate and a forward-looking dynamic model. To make the connection transparent, we adopt two strong simplifying assumptions: the formality choice is irreversible and households have an infinite horizon. Under these assumptions, the household chooses between the present value of future formal consumption and the present value of future informal consumption minus an upfront cost of theft (κ). The household problem is then

$$\max \left\{ \underbrace{\sum_{j=0}^{\infty} \rho^j \mathbb{E}\left[\beta\psi_0(p_{t+j}) + \tilde{\eta}_{0,t+j} + \tilde{\varepsilon}_{0,i,t+j} | \mathcal{I}_t\right]}_{\text{formal}}, \underbrace{\sum_{j=0}^{\infty} \rho^j \mathbb{E}\left[\beta\psi_1 + \tilde{\varepsilon}_{1,i,t+j} | \mathcal{I}_t\right] - \kappa}_{\text{informal}} \right\},$$

where ρ is the discount rate, $\tilde{\eta}_{0,t+j}$ is a flow benefit of formal consumption relative to informal, $\tilde{\varepsilon}_{a,i,t+j}$ is an unobserved idiosyncratic flow component, and $\mathcal{I}_t$ is the information set at t.

We make an extra simplifying assumption, which is in line with our identification strategy, that households see price shifts as permanent with no additional shocks, that is, $p_{t+j} = p_t$ for all $j = 1, 2, \ldots$ This assumption allows us to simplify the household problem:

$$\max \left\{ \underbrace{\frac{\beta\psi_0(p_t)}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \mathbb{E}\left[\tilde{\eta}_{0,t+j} + \tilde{\varepsilon}_{0,i,t+j} | \mathcal{I}_t\right]}_{\text{formal}}, \underbrace{\frac{\beta\psi_1}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \mathbb{E}\left[\tilde{\varepsilon}_{1,i,t+j} | \mathcal{I}_t\right] - \kappa}_{\text{informal}} \right\},$$

We can then map this forward-looking choice model into the static long-run choice model we estimate:

$$\max \left\{ \underbrace{\beta\psi_0(p_t) + \eta_{0,t} + \varepsilon_{0,i,t}}_{\text{formal}}, \underbrace{\beta\psi_1 + \varepsilon_{1,i,t}}_{\text{informal}} \right\},$$

where

$$\eta_{0,t} := \sum_{j=0}^{\infty} \rho^j \mathbb{E}\left[\tilde{\eta}_{0,t+j} | \mathcal{I}_t\right](1-\rho) + \kappa(1-\rho)$$
$$\varepsilon_{a,i,t} := \sum_{j=0}^{\infty} \rho^j \mathbb{E}\left[\tilde{\varepsilon}_{a,i,t+j} | \mathcal{I}_t\right](1-\rho) \text{ for } a = 0, 1.$$

Therefore, this forward-looking interpretation of the model we estimate only requires re-interpreting some of the parameters. Under this perspective, $\eta_{0,t}$ captures both fixed costs of conversion to informality and flow differences between formality statuses. These flow differences capture the expected effect of sanctions should the household get caught and benefits of formal connections such as service reliability–both of which are better thought of as occurring on a per-period basis.

B Alternative Specification for Informal Consumption Using Non-Technical Losses

Our baseline specification estimates household-level informal consumption ($\bar{q}_i$) by extrapolating from the linear demand model for electricity (equation 10), setting the price equal to zero. We present here an alternative formulation for $\bar{q}_i$ based on data on non-technical losses (NTL) and conduct robustness exercises for our main results using this specification.

In this alternative formulation, we assume $\bar{q}_i$ is identical for all households within a given neighborhood and equal to the average NTL per informal consumer in that neighborhood. Specifically, for each household in neighborhood ℓ:

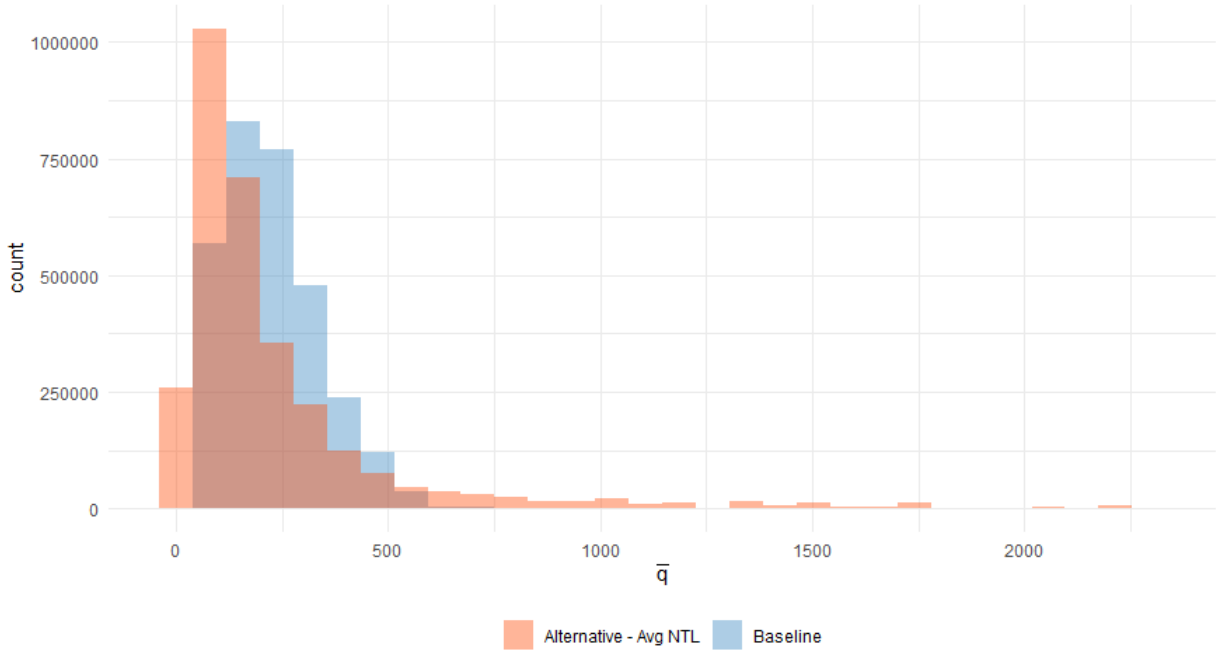
$$\bar{q}_i = \frac{NTL_\ell^T}{N_{informal,\ell}^T},$$

where NTL_ℓ^T denotes total NTL reported in period T in neighborhood ℓ and $N_{informal,\ell}^T$ is the number of informal households in that neighborhood.

This formulation yields a distribution of $\bar{q}_i$ similar to our baseline approach. Figure B.1 we display the histogram of $\bar{q}_i$ under both methods. However, the NTL-based alternative produces some extreme values of $\bar{q}_i$ that may be unrealistic. This pattern reflects a key limitation of this approach: although residential electricity theft accounts for a substantial share of total NTL, commercial sector theft is also significant. Moreover, the prevalence of commercial theft varies across neighborhoods and may correlate with demographic characteristics, potentially biasing our specifications exploring neighborhood heterogeneity.

Using this NTL-based alternative for $\bar{q}_i$, we re-estimate the full model and conduct the elasticity decomposition exercise reported in Table B.1. The elasticity decomposition produces results similar to our baseline specification, yielding a total electricity demand elasticity of -0.44, with 41% attributable to the extensive margin (changes in formality status). This consistency is noteworthy given that the two methods for constructing $\bar{q}_i$ are fundamentally different, and $\bar{q}_i$ is a key parameter in the household decision problem. The robustness of our results across these alternative specifications strengthens confidence in our main conclusions.

Figure B.1: Distribution of $\bar{q}_i$ under the baseline and NTL-based specifications



Note: This figure shows the histogram of $\bar{q}_i$ constructed under the baseline assumptions and alternative NTL-based specification.

Table B.1: Demand Elasticity Decomposition: NTL-based $\bar{q}_i$

	Total formal	Intensive margin (cond. on formal)	Extensive margin (formality share)
	ξ_{Total}	ξ_d	ξ_σ
	(1)	(2)	(3)
ξ_x	-0.44	-0.26	-0.18
ξ_x/ξ_{Total}	1.00	0.59	0.41

Note: This table reports total formal demand elasticity decomposition using a NTL-based $\bar{q}_i$. All other specifications are identical to the baseline. 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.

C Tables and Figures

Table C.1: Linear demand used to estimate informal consumption

	(1)	(2)
α	-76.48	-79.98
	(0.637)	(0.556)
IC	0.3795	0.3795
	(0.00009)	(0.00009)
Instrument for Prices	No	Yes
Month FE	Yes	Yes
Year FE	Yes	Yes
Social Tariff FE	Yes	Yes
Building FE	Yes	Yes
Capacity Heterogeneity	No	No
Observations	230,200,302	230,200,302
R-squared	0.561	0.562

Note: This table reports estimates of Equation 10. A unit of observation is a household-month. The dependent variable is the formal electricity consumption per household, in kWh. ξ is the linear coefficient on prices. Installed capacity of each household as a regressor is used to find the intercept when prices go to zero. Our sample covers the period 2017 to 2022. Standard errors are in parentheses.

Figure C.1: Joint Density of Monthly Household Formal Consumption and Theft Consumption (KWh)

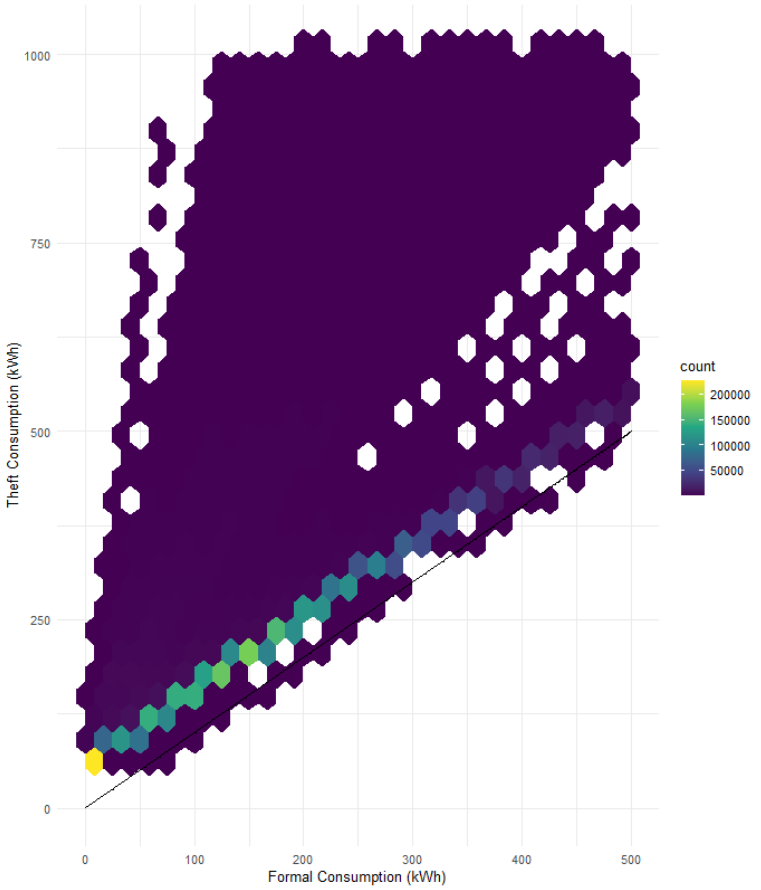


Table C.2: Summary of datasets

Database	**Description**	**Time span**	**Stage in which is used**
BDGD (ANEEL)	Data on household consumption, installed capacity, and proxy location.	2017 - 2022	Estimation of the intensive and extensive margin
Aggregate ANEEL	Data on aggregate consumption by tariff groups, on theft, and revenue	2008 - 2020	Estimation of the extensive margin
Tariffs (ANEEL)	Monthly prices per unit (kWh) consumed	2008 - 2022	Estimation of the intensive and extensive margin
Flags (ANEEL)	An additional price to be added to both regular and social tariffs, when activated	2015 - 2022	Estimation of the intensive and extensive margin
Crime (GENI)	Areas of the state controlled by organized crime	2019	Estimation of the extensive margin
Census (IBGE)	Data on demographic characteristics at the census tract level	2010	Estimation of the extensive margin