FUNDAÇÃO GETÚLIO VARGAS
ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS

DOCTORAL THESIS

Issues in the Industrial Organization of Energy Markets

Author:
Gabriel Richter de Almeida

Advisor:
André Garcia de Oliveira Trindade

Co-Advisor:
Marcelo Castello Branco Sant'Anna

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

December 17, 2020

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada peloSistema de Bibliotecas/FGV

Almeida, Gabriel Richter de
Issues in the Industrial Organization of Energy Markets / Gabriel Richter de Almeida. – 2020.
83 f.

Tese (doutorado) - Fundação Getulio Vargas, Escola Brasileira de Economia e Finanças.
Orientador: André Garcia de Oliveira Trindade
Coorientador: Marcelo Castello Branco Sant'Anna
Inclui bibliografia.

1, Energia elétrica - Consumo, 2. Energia elétrica - Aspectos ambientais. 3..Meio ambiente - Aspectos econômicos I. Trindade, André Garcia de Oliveira. II. Sant'Anna , Marcelo Castello Branco. III. Fundação Getulio Vargas Escola Brasileira de Economia e Finanças. IV.Título.

CDD – 333.79026

Elaborada por Maria do Socorro Almeida – CRB-7/4254

Gabriel Richter de Almeida

"Issues in the Industrial Organization of Energy Markets".

Tese apresentado(a) ao Curso de Doutorado em Economia do(a) EPGE Escola Brasileira de Economia e Finanças - FGV EPGE para obtenção do grau de Doutor(a) em Economia.

Data da defesa: 17/12/2020

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

Presidente da Comissão Examinadora: Prof. Andre Garcia de Oliveira Trindade

Andre Garcia de Oliveira Trindade — André Trindade

P/ Marcelo Castello Branco Sant'Anna — André Trindade

P/ Francisco Junqueira Moreira da Costa — André Trindade

P/ Shaun McRae — André Trindade

P/ Fábio Miessi Sanches — André Trindade

Em cumprimento Lei nº 13.979 de 06/02/20 - DOU nº 27 de 07/02/20, a Portaria MEC nº 473 de 12/05/20 - DOU nº 90 de 13/05//20 e ao Decreto nº.068 de 11/05/20 - Poder Executivo do Estado do Rio de Janeiro, DOE nº 082-A em 11/05/20 que dispõe sobre a suspensão temporária das atividades acadêmicas presenciais e a utilização de recursos tecnológicos (em conformidade à legislação vigente), face ao COVID-19, as apresentações das defesas de Tese e Dissertação, de forma excepcional, serão realizadas de forma remota e síncrona, incluindo-se nessa modalidade membros da banca e discente.

Lucas Jóver Maestri
Coordenador

Antonio de Araujo Freitas Junior
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação FGV

Antonio Freitas, PhD
Pró-Reitor de Ensino, Pesquisa e Pós-Graduação
Fundação Getúlio Vargas

Instrução Normativa nº 01/19, de 09/07/19 - Pró-Reitoria FGV
Em caso de participação de Membro(s) da Banca Examinadora de forma não-presencial*, o Presidente da Comissão Examinadora assinará o documento como representante legal, delegado por esta I.N.
*Skype, Videoconferência, Apps de vídeo etc

Acknowledgements

First and foremost, I would like to express my sincere gratitude and appreciation to my beloved family, especially my parents, Simonne and Luciano, and my brother, Vitor, for always being by my side. I would also like to express my deepest appreciation to my girlfriend Sara, for providing me support and affection throughout this process, and to my friend, Lucas Bastos, for inspiring me into pursuing this academic experience.

I will always be grateful to my advisors, Professors André Trindade and Marcelo Sant'Anna, for their patience, guidance, and encouragement. I could not have imagined having better advisors and mentors. I would also like to thank Professors Shaun McRae, Francisco Costa, and Fábio Miessi, for being on my thesis committee, and for their valuable comments during my Ph.D. defense. I also thank Professors Mike Abito and Joe Harrington for ensuring I had a enriching academic experience at Wharton School.

I would also thank all my fellow classmates and friends from FGV EPGE and Wharton, who made my Ph.D. experience fun and pleasant. I very much appreciate all the help provided by FGV EPGE professors and staff. In particular, I thank Andrea Machado for all the help she provided during these years.

Finally, I thankfully acknowledge the financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Resumo

Esta tese de doutorado é composta por três artigos que se encontram na intersecção entre os campos da Microeconomia Aplicada, da Economia do Meio-Ambiente e Energia, e da Organização Industrial Empírica. Nesta seção, apresento brevemente cada um destes artigos.

O primeiro artigo, *Natural Gas Demand in Brazil: Evidence from the Residential and Industrial Sectors*, é um trabalho desenvolvido em conjunto com André Trindade. Neste artigo, estimamos elasticidades de demanda por gás natural para os setores residencial e industrial no Brasil. Utilizamos dados de painel, de onze estados brasileiros, para o período entre Janeiro de 2007 e Dezembro de 2017. Os nossos resultados sugerem uma elasticidade de demanda de curto prazo por gás natural de –0,18 no setor residencial e –0,22 no setor industrial. Esses resultados são consistentes com resultados prévios da literatura econômica e com evidências anedóticas de membros da indústria.

O segundo artigo, *The Welfare Effects from Electricity Theft: Evidence from Brazil*, é um trabalho desenvolvido em conjunto com André Trindade e Marcelo Sant'Anna. Nesse artigo, perguntamos: o quanto (se algum) de excedente do consumidor os consumidores residenciais perdem em função da atividade de furto de eletricidade? Para responder a essa pergunta, estimamos um modelo estrutural de demanda por eletricidade para estudar as decisões das famílias sobre consumo e furto. Em seguida, simulamos cenários contrafactuais onde o roubo de energia não é permitido. Na versão atual do nosso modelo, a proibição do furto reduz o excedente do consumidor médio.

No terceiro e último artigo, *Electricity Tariff Flags and Consumer Behavior*, eu estudo se a implementação do sistema de bandeiras tarifárias no Brasil mudou a forma como os consumidores respondem a alterações no preço da energia elétrica. Para isso, construo um modelo econométrico de demanda por eletricidade no setor residencial no Brasil, utilizando dados agregados das principais empresas nacionais de distribuição de energia elétrica. Os meus resultados sugerem que a resposta dos consumidores às bandeiras é limitada, mas problemas relacionados à endogeneidade das variáveis do modelo devem ser considerados em versões futuras desse trabalho.

Palavras-Chave: Economia do Meio-Ambiente e Energia, Mercados de Energia, Estimação de Demanda.

Abstract

This thesis is composed of three articles lying in the intersection between Applied Microeconomics, Energy and Environmental Economics, and Empirical Industrial Organization. In this section, I briefly introduce each of these articles.

The first article, *Natural Gas Demand in Brazil: Evidence from the Residential and Industrial Sectors*, is a joint work with André Trindade. In this article, we estimate natural gas demand elasticities for the residential and industrial sectors in Brazil. We use panel data from eleven Brazilian states for the period January 2007 to December 2017. Our results suggest a short-run natural gas demand elasticity of –0.18 in the residential sector, and –0.22 in the industrial sector. These results are consistent with previous literature findings and match anecdotal evidence from the industry.

The second article, *The Welfare Effects from Electricity Theft: Evidence from Brazil*, is a joint work with André Trindade and Marcelo Sant'Anna. In this article, we ask: how much (if any) consumer surplus do households lose due to electricity theft? To address this question, we estimate a structural model of electricity demand to study households' decisions on consumption and theft. We then simulate counterfactual scenarios where theft is not possible. In the current version of our model, we find that banning theft would reduce the average consumer surplus.

In the third and last article, *Electricity Tariff Flags and Consumer Behavior*, I study whether the implementation of the tariff flags system in Brazil changed how consumers respond to electricity prices. To achieve this goal, I build an econometric model of demand for electricity in the residential sector in Brazil, using aggregated data from major electric national utilities. My results suggest that consumers' response to the flags is limited, but the issue of endogeneity in the model should be addressed in future versions of this work.

Keywords: Energy and Environmental Economics, Energy Markets, Demand Estimation.

Contents

List of Figures

List of Tables

Chapter 1

Natural Gas Demand in Brazil: Evidence from the Residential and Industrial Sectors

1.1 Introduction

The threat of climate change has increased both public and governmental awareness about the need to balance global sustainable economic growth with a reduction of carbon emissions. Natural gas is sometimes referred to as a "bridge" fuel[1], as it could aid in the transition from a fossil fuel reliant economy to a carbon-free economy grounded on the consumption of renewable energy sources. Until 2040, the global demand for natural gas is expected to grow 1.4% a year, on average, reaching 5.2 trillion cubic meters of gas annually (IGU, 2020). It is also projected to account for a quarter of the world total energy consumption, surpassing coal as the second most used source of energy globally, only behind oil (IGU, 2020).

The natural gas industry in Brazil is still maturing. The demand reached 35.8 billion cubic meters in 2019 (BP, 2020), the majority of which came from the industrial sector. However, the fuel only accounts for nearly 12% of the domestic power supply. Historically, the industry has been very concentrated, but in recent years some laws have been proposed with the aim to increase competition. Naturally, estimates of consumer demand are key resources to help policymakers understand the impacts of the industry deregulation on consumer behavior and welfare.

Table 1.1 shows some point estimates for natural gas demand elasticities in the resi-

[1] Although it is not a renewable energy source, it is the least polluting fossil fuel. For example, it produces the same amount of energy as coal, while generating half the amount of carbon dioxide (CO_2).

dential and industrial sectors, pulled from the economic literature. These studies use data from the United States. To the best of our knowledge, there is no other paper that estimates the price elasticity of demand for natural gas in Brazil. Therefore, our goal in this paper is to fill in this gap.

In this paper, we estimate an econometric demand model for natural gas in Brazil. In particular, we focus on the residential and industrial sectors. Our data comes from multiples datasets, including a panel of monthly aggregated natural gas consumption, prices, pipelines grid length, and temperature. Our data comprises eleven Brazilian states and covers the period January 2007 to December 2017.

Our identification strategy relies on the use of instrumental variables to account for the endogeneity of prices. We build on the idea of Hausman instruments (Hausman, 1996; Nevo, 2001) and use the price of natural gas in other sectors, within the same state, as instruments. These instruments are found to have a strong first stage.

The results suggest a short run natural gas demand elasticity of –0.18 in the residential sector, and –0.22 in the industrial sector. These results are consistent with previous findings from the literature on natural gas demand estimation, as shown in Table 1.1. Our results also match anecdotal evidence from the industry, in which consumers have a low price sensitivity in the short run.

One potential reason for the low price elasticity in the residential sector is that households in Brazil use natural gas mostly for water heating and cooking. For these activities, there are not many substitutes for natural gas, so consumers may have little margin for adjusting energy consumption in the short run. Liquefied Petroleum Gas (LPG), which is gas that comes in cylinders, is an alternative to natural gas—which comes in pipes—for cooking. However, LPG is more dangerous to handle due to fire hazards and the cylinders also require physical space to store, so they are usually less preferred when natural gas is available[2]. Moreover, switching to alternative fuels might be costly. For example, a household that wants to use electricity for water heating, instead of natural gas, must bear the costs of purchasing and installing an electric shower.

Another possible explanation for these results is that consumers are inattentive to prices. This connects to a growing literature on price salience (Chetty et al., 2009; Finkelstein, 2009; Bollinger et al., 2011; Sexton, 2015; Taubinsky and Rees-Jones, 2017; Seim et al., 2017).

In the industrial sector, one potential reason is that firms have limited flexibility to readapt industrial plants for alternative fuel and feedstock sources in the short run.

The remainder of this paper is organized as follows. Section 1.2 introduces relevant

[2] Low-income households typically use LPG because they live in neighborhoods where piped gas infrastructure is not available.

institutional details of the natural gas market in Brazil. Section 1.3 describes the data used in the analysis. Section 1.4 introduces the empirical model and discusses some identification challenges. Section 1.5 presents our estimation results, which we discuss in Section 1.6. Finally, Section 1.7 concludes.

1.2 Background

Natural gas accounts for approximately 12% of the domestic power supply in Brazil (BEN, 2020). Most of the production takes place offshore, where it is usually obtained as a by-product of oil extraction. When compared internationally, the Brazilian natural gas industry has still a lot to develop. In the United States, for example, this industry has its own dynamics[3], which is made possible, in part, by the existence of a robust infrastructure of gas transportation pipelines, nearly 48 times more extensive than the existing pipeline grid in Brazil (BTS, 2020).

On the demand side, almost half of the national demand for this fuel comes from the industrial sector, where it is used as an input for the production of chemicals, fertilizers, ceramics, glass, cement, among other products. The participation of the residential sector is substantially lower, representing only about one percent of the total demand. In this sector, the natural gas is mostly used for water heating and cooking.

The natural gas market has three distinct levels: the upstream, midstream, and downstream. The upstream comprises the exploration and production of the commodity, both onshore and offshore, as well as its processing, after which it becomes fit for consumption. The midstream involves the transportation of the production in pipelines to the city gates, which are stations where the distribution companies receive the gas for its subsequent distribution. Finally, the downstream corresponds to the activity of transporting the gas through distribution pipelines to the final consumers—households, businesses, factories, thermal power stations, etc.

Historically, the lack of competition has been one remarkable feature of the Brazilian natural gas industry. Until the mid-1990s, the federal government had the monopoly—through a company called Petrobrás—of both the upstream and the midstream of the oil and gas industries, while the states had market power over the downstream. In that decade, Brazil underwent an extensive privatization program which ended the State monopoly, created a national regulatory agency—the *Agência Nacional do Petróleo, Gás Natural e Biocombustíveis* (ANP)—and allowed states to grant the gas distribution service to private

[3] As it is independent from the oil sector. This was made possible in part by the large-scale implementation of the hydraulic fracturing (fracking) technology.

companies. Despite the end of Petrobrás monopoly, at least in the legal sphere, in practice the company kept its market power by owning virtually all of the gas transportation pipelines, as well as a large share in most of the distribution companies. This situation has only started to change recently, with the company's commitment to divest its infrastructure assets in the gas market. Moreover, in the recent years, some laws have been proposed—for example, the New Gas Law—with the aim to reform the regulatory framework of the gas industry and increase its competition.

The market regulation occurs at both the federal and state levels. The upstream and midstream are regulated by ANP, while the downstream follows the determinations of state regulatory agencies. The distribution activity is also a monopoly—distribution companies serve regions that do not intersect. In the downstream, there are 27 utilities, serving consumers in 23 states and in the federal district[4]. While two states have more than one distribution company each—São Paulo has three and Rio de Janeiro two—three other states[5] have none. Moreover, São Paulo and Rio de Janeiro have approximately 85% of the total number of clients, 68% of the total distribution pipelines grid length, and they distribute nearly 52% of the total amount of natural gas consumed in the country (BNDES, 2020)[6].

The prices of the natural gas distributed to final consumers are set by the regulator and follow a nonlinear scheme. Customers pay a fixed and a variable price, where the latter increases in discrete steps. Both of these prices are a function of the consumption. Since the regulation of the distribution service occurs at the state level, there is a lot of heterogeneity on how prices are set nationwide. For example, in some states, prices are updated every quarter, while in others they are adjusted only once a year. The number of steps and the consumption thresholds associated to each step are also heterogeneous. Even more interesting is how the discrete steps change as a function of the quantity consumed. The economic literature suggests that increasing-block pricing—where the marginal price is an increasing function of consumption—in the residential sector may have desirable distributional effects (Borenstein, 2012), although there is evidence that they might be only mildly progressive depending on the correlation between income and natural gas consumption (Borenstein and Davis, 2012). Despite these academic findings, some states—for example, Minas Gerais—charge residential clients a higher fixed price and a lower variable price when consumption increases. In others, like Espírito Santo, the opposite holds.

[4]Brasil has 26 states and one federal district, Brasília.

[5]Acre, Roraima, and Tocantins.

[6]Besides having the strongest economies in Brazil, these two states have the oldest distribution companies. Their pipeline infrastructure is secular, as they were used in the past to transport coal gas for public lighting.

The natural gas prices are basically composed of four elements. First, the price of the commodity itself. Second, the costs associated with the transportation of the gas. Third, federal and state taxes. Fourth, the margins of the distribution companies. The first three components are normally passed on to consumers and they represent the largest part of the final price. In 2018, they accounted for nearly 80% of the price charged from the average industrial consumer (CERI, 2019). The distribution companies margins are also usually higher for residential clients (BNDES, 2020).

1.3 Data

We draw information from multiple datasets: a panel of (i) aggregated consumption, (ii) price, (iii) pipeline length, and (iv) temperature. The data comprises eleven[7] states in Brazil: Amazonas, Bahia, Espírito Santo, Mato Grosso do Sul, Minas Gerais, Paraná, Pernambuco, Rio de Janeiro, Rio Grande do Sul, São Paulo, and Santa Catarina. Moreover, it covers two sectors, the residential and the industrial. In this section, we describe each one of these datasets, explain how key variables are defined, and report some descriptive statistics.

1.3.1 Data Sets

The natural gas consumption and pipeline length data that we use come from *Associação Brasileira das Empresas Distribuidoras de Gás Canalizado* (ABEGÁS), an association composed of piped gas distribution companies in the natural gas industry in Brazil. This information is available at the state-sector-month level and cover the period January 2007 to December 2017. The consumption data is measured in thousands of cubic meters. Ideally, we would prefer to build a consumption per client variable, which controls for changes in the number of clients served by each utility over time. Despite being released to the public, the information on the number of consumers exhibits some abrupt spikes which are hard to reconcile[8], so we do not use them.

[7]Despite the sector regulation, there was not a dataset on pricing previously available. Therefore, a major challenge was to reach out to different distribution companies and state regulators to obtain access to pricing data, and then to assemble the information into one comprehensive dataset. Moreover, even though we were just able to gather information from eleven states, they include many of the most relevant Brazilian states in terms of natural gas consumption.

[8]For example, in the state of Bahia the number of clients jumped from 44, in March 2009, to 2,251 a month later, while the aggregated consumption remained virtually unchanged. These abrupt changes in the number of clients were also verified in other states, such as Espírito Santo, Mato Grosso do Sul, Pernambuco, among others.

Next, we expand our dataset by adding information on prices, which come from multiple sources. Some information came from the distribution companies upon our request, while others were obtained by reaching out to the regulators in each state. The prices are nonlinear and composed of multiple steps. We also collect data on Liquefied Petroleum Gas (LPG) prices from ANP, the regulatory agency at the federal level. The LPG is the main substitute of the natural gas in the residential sector. While the first usually comes in cylinders, the latter is transported through pipelines. The natural gas price data is available at the state-sector-month level, whereas the LPG price information varies in each state and month. All prices are nominal and measured in Brazilian Reais (BRL).

Finally, we incorporate temperature information using hourly data from the ERA5 dataset, which is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The ERA5 combines vast amounts of historical observations with geophysical models—a procedure called climate reanalysis—to produce global estimates of atmospheric parameters. These observations come from different sources, such as satellites, radars, and weather stations. The dataset covers Earth at a 0.25º x 0.25º grid resolution—approximately 30km x 30km. The information is daily updated and is available from the year 1979 to 5 days behind real time. In our model, we use a temperature variable that varies at the state-month level, which we build in four steps. First, we use shapefiles from *Instituto Brasileiro de Geografia e Estatística* (IBGE)—the agency responsible for collecting official statistical and geoscientific data, under the supervision of the Ministry of Economy—to obtain the latitude and longitude of every Brazilian municipality centroid. Next, we use yearly population count information, also provided by IBGE, to find the population of these municipalities for the period 2007–2017. Third, we create an auxiliary temperature variable, at the state-hour level, by taking the population-weighted average of the hourly temperature of every municipality within each state. We conclude by taking the monthly average of this auxiliary variable for each state.

1.3.2 Descriptive Statistics

Table 1.2 reports summary statistics for key variables in our dataset. The average natural gas consumption is of 2.43– and 69.54 millions of cubic meters a month in the residential and industrial sectors, respectively. Figures 1.1 and 1.2 also reveal that the demand in these two sectors have very distinct patterns. The demand of industrial clients seems to be more correlated with the economic activity—for example, during the global financial crisis of 2008, there was a large drop in the average industrial consumption—but this does not seem to hold for residential clients. On the other hand, in the residential

sector, there is strong evidence of seasonality in consumption.

We also present statistics on two price variables, the natural gas baseline price and the normalized price. We define the first as the first nonzero price level in the natural gas nonlinear pricing schedule. The latter is defined as the amount charged from clients with consumption between 16 and 55 cubic meters a month in the residential sector, and between 300 thousand and 500 thousand cubic meters a month in the industrial sector. These variables were built as an attempt to deal with the natural gas price endogeneity, which we further discuss in the next section.

In Figure 1.3, we exhibit the evolution of these price variables over time. There is more volatility in prices during the beginning of the time series. This happens because our panel of prices is unbalanced[9], so the average prices respond to the inclusion of data from new utilities. This figure reveals two interesting features of our data. First, there is more (time series) variation across prices in the residential than in the industrial sector. Second, the average baseline price is lower than the normalized price for households, while higher for industrial consumers. This last point is an evidence that the average residential client faces an increasing-block pricing scheme—where the natural gas marginal and average prices increase with the amount of gas consumed—whereas the opposite holds for the average industrial consumer, as he faces a decreasing-block pricing schedule.

1.4 Empirical Model

We propose a static model to estimate the natural gas demand in the residential and industrial sectors. The model is estimated at the state(i)-sector(s)-month(t) level, according to the following equation:

$$\log(C_{ist}) = \beta_0 + \beta_1 \log(P_{ist}) + \beta_2 X_{ist} + \gamma_{is} + \lambda_{month(t),s} + \varepsilon_{ist} \tag{1.1}$$

Here, C_{ist} is the natural gas consumption, in thousands of cubic meters. P_{ist} is the natural gas ex-taxes baseline price, in BRL per cubic meter. X_{ist} is a set of covariates, including the logarithm of the pipeline grid length, cooling degrees[10], the logarithm of the price of substitute fuels—such as the LPG in the residential sector—and a linear time trend. γ_{is} and $\lambda_{month(t),s}$ are, respectively, state and month of the year fixed effects. Our parameter of interest is β_1, the own-price elasticity of demand.

[9]Unfortunately, since the price information come from multiple sources, we were not able to gather pricing data, for the whole period of this study, for all eleven states.

[10]We define $Cooling_{it} = Max\{Temperature_{it} - 18, 0\}$, where $Temperature_{it}$ is the actual atmospheric temperature. This allows for capturing nonlinear effects of temperature on natural gas consumption.

1.4.1 Identification

There are two main issues concerning the identification of our parameter of interest, β_1. The first challenge, which emerges from the nonlinear price regime, is that both the marginal and the average prices are a function of the quantity consumed, so they are correlated with unobserved demand shocks by construction. Auffhammer and Rubin (2018) recommends using what we call the baseline price, which is the first nonzero price step. They argue that this price variable is strongly correlated to the marginal and average prices, but does not suffer from the same sort of endogeneity, since it is not a function of the amount of natural gas consumed. As noted by the authors, one disadvantage of using the baseline price is that it does not capture the higher prices associated with other consumption thresholds. We believe this is less troublesome in our setting, because our price variable is in logarithms instead of levels. Therefore, regardless of the number of steps in the nonlinear pricing regime, we just need the percentage change in the price in each step to be similar across steps over time, which is usually the case. Thus, we follow the authors and use the baseline price as our price variable, P_{ist}, in Equation 1.1.

A second obstacle concerns the classical price endogeneity that results from the simultaneous determination of price as a supply and demand equilibrium. The purchase and sale agreements between the natural gas supplier and the utilities provide a greater understanding on how this sort of endogeneity arises. These contracts set very clear pricing rules. First, they establish that prices are only allowed to be adjusted at specific time intervals—usually every quarter or year, depending on the utility—to reflect exogenous changes in the acquisition costs of the commodity, such as inflation, fluctuations in the international price of oil, and variations in the US Dollar/Brazilian Real exchange rate. Second, they set fixed margins of profit for the utilities during periods ranging from one to five years. Third, they foresee a guaranteed fixed daily supply of natural gas that the utilities commit to purchase. In general, there is a surcharge mechanism which is activated when demand goes above this agreed amount. The standard practice is to increase the marginal cost of the natural gas by a certain amount for these extra units when the daily allowance is exceeded by 5% to 10%, and by an even higher amount when it surpasses 10%. Therefore, in order to keep their margins of profit constant, the utilities increase the average natural gas price charged from consumers whenever there are unobserved demand shocks that raise the costs they face to purchase the commodity.

The standard approach for dealing with this classical price endogeneity is to use instrumental variables. We build on the concept of Hausmann instruments, which are commonly used in the Industrial Organization literature on demand estimation—see,

for example, Hausman (1996) and Nevo (2001). This approach would recommend using the price of natural gas in other markets—states or sectors, for instance—as instruments. The rationale is that they are able to capture common costs shocks—for example, in the commodity price and the transportation cost, which account for much of the natural gas price to final consumers—but are independent from shocks in the demand-side.

We instrument the natural gas baseline price in the residential sector with the normalized price of this fuel in the industrial sector, within the same state[11], and the baseline price in the industrial sector with the normalized price in the residential sector, also within the same state. These instruments are found to have a strong first stage. The exclusion restriction is less straightforward, though. It requires that the natural gas normalized price in one sector is not correlated to shocks that affect consumption decisions in the other sector. This assumption would be threatened by a national or regional event, such as an adverse economic shock, that affects the demand in both sectors simultaneously. In fact, this is a common critique of Hausman instruments. Our preference for using the normalized price as the instrument—instead of the baseline price of the opposite sector, for example, which could be seen as a more natural choice—is explained by the fact that it has a very strong first-stage F-statistic.

One caveat is that, as we previously discussed, the industrial sector accounts for almost half of the natural gas consumption in Brazil. Therefore, a positive demand shock in the industrial sector—due to an unexpected economic boom, for instance—that forces utilities to buy more natural gas than they were previously expecting would likely increase the gas acquisition costs and, as a consequence, affect the price setting in the residential sector[12]. As a result, the residential price would not be a valid instrument for the industrial price, because the exogeneity condition would not hold.

A first counterpoint to this concern is that utilities have access to very granular consumption data and are thus able to make reasonable demand forecasts. Second, the concession agreements establish a "price-cap" mechanism in which the regulator sets a maximum price that utilities are allowed to charge from consumers (CERI, 2019). Therefore, even if the utilities were not able to predict consumers demand accurately, they would still have the incentive to commit with purchasing a more conservative (i.e. fix a higher upper bound) daily amount of natural gas from the gas supplier. The trade-off is clear: on the one hand, by committing to purchase more natural gas they increase their fixed costs. On

[11] A second alternative would be to use the normalized price in the industrial sector, but in another state. However, this would add a new layer of complexity to the estimation, as we would need to decide which state to use the price from.

[12] The opposite should be less an object of concern, though, because the residential sector represents only approximately 1% of the total internal natural gas demand.

the other hand, they avoid the risk of not being able to pass along the extra gas acquisition costs—that arises from the pricing surcharge mechanism—due to the price cap.

1.5 Estimation Results

Tables 1.3 and 1.4 present our estimates of Equation 1.1 for the residential and industrial sectors, respectively. Robust standard errors are in parenthesis. All specifications include a constant, a *Cooling Degrees*[13] variable that captures nonlinear effects of temperature on consumption decisions, and state fixed effects, which account for potential time-invariant unobserved heterogeneity within states. Columns 1–4 represent specifications estimated by OLS, while columns 5–6 in Table 1.3 and column 5 in Table 1.4 are estimated using 2SLS. All observations were weighted[14] based on the share of each state consumption relative to the total natural gas consumption in Brazil each year. Therefore, states with higher consumption, such as Rio de Janeiro and São Paulo, were given more weight.

Column 1 of Table 1.3 reports estimates of a specification with no additional controls. We find a positive coefficient for the natural gas price, which is inconsistent with economic theory predictions. In column 2, we add month fixed effects, which control for seasonal patterns that are not captured by the *Cooling Degrees* variable. For example, it could be the case that some months have a lower than average consumption just because they have less business days than others. The price coefficient is still positive, though. This changes when we add a linear time trend control, in column 3. This variable captures the potential development and adoption of energy-saving technologies, as well as changes in consumption behavior and inflation over the years. In column 4, we add a market access control, the logarithm of the pipeline grid length, which we include to account for mechanical increases in consumption that arises only because more households have access to the natural gas infrastructure. In fact, if we had good quality data on the number of households served in each state, we could simply use the consumption per client as our dependent variable. Unfortunately, this is not the case, as per our previous discussion in Section 1.3.

In columns 5–6, we adopt an instrumental variables strategy as an attempt to deal with the classical price endogeneity issue. We instrument the natural gas baseline price in the residential sector with the normalized price of this fuel in the industrial sector, within the

[13]Since the non-cooking use of natural gas is mostly water heating, the negative and statistically significant coefficient of the *Cooling Degrees* variable suggests that when the weather is hot, people usually do not take warm showers.

[14]The weights were applied because we believe our data is not equally reliable. In fact, we noticed that the price and consumption information for states with smaller consumption are more volatile.

same state[15]. Our first-stage F-statistics suggest a strong first stage.

In column 5, our price coefficient suggests a natural gas demand elasticity of –0.19. In column 6, we include the logarithm of the price of the LPG as a new control. The LPG is the most immediate substitute for the natural gas in the residential sector. For this variable, we find a point estimate that is positive, but not statistically significant. One possible explanation for this result is that there are non-negligible switching costs between these fuels in the short run. For example, once a household has all of their electric appliances adapted to the piped natural gas, it might be costly to readapt their technologies to use LPG cylinders. Another possibility is that the LPG price is also endogenous and its coefficient is biased towards zero. Still in column 6, we find a natural gas demand elasticity of –0.18.

Table 1.4 reports our findings for the industrial sector. The price coefficient point estimates vary less along the specifications, relative to those in the residential sector—for example, they have the expected negative sign in all five columns. The results in columns 1–2 suggest that the inclusion of month fixed effects produces virtually no change in the magnitude of the price coefficient. However, in column 3, it goes to zero when we control for a linear time trend, remaining unchanged even when we introduce, in column 4, the logarithm of the pipeline grid length as our market access control. In column 5, we instrument the natural gas baseline price in the industrial sector with the normalized price of this fuel in the residential sector, within the same state. Once again, we observe a strong first stage. Our findings suggest a natural gas demand elasticity of –0.22 in the industrial sector.

1.6 Discussion

Our results are consistent with previous findings from the literature on natural gas demand estimation. The low price elasticities are also supported by anecdotal evidence from oil and gas industry experts. We now discuss a few possible explanations for the inelastic behavior observed in both the residential and industrial sectors.

We start by characterizing the consumption in the residential sector. In Brazil, households use natural gas mostly for water heating and cooking. For cooking, the main substitute for natural gas is LPG, although low-income households may also substitute with firewood and coal. For water heating, electricity is the main alternative, but it may require the purchase and installment of electric showers which are costly. Therefore, the

[15]We also tried other instruments, such as the baseline price of the natural gas in the industrial sector, within the same state. We found results for the demand elasticity that are qualitatively identical. These results are not reported in this paper.

lack of outside options is the first explanation we offer to reconcile the low price elasticity in the residential sector.

Another point worth noting is that many low-income Brazilian households have limited access to natural gas. This happens because the pipeline infrastructure required for the natural gas transportation is frequently unavailable in low-income neighborhoods. Thus, another possible explanation for our results is that our data is composed of a biased sample of medium– and high-income households, which in turn are less sensitive to price changes[16].

A third avenue is that consumers are inattentive to their natural gas bills because the salience generated by the prices is not high enough to increase the benefits of tracking gas prices, against the cost of having to pay attention to them. Automatic bill payments that eliminate the necessity for consumers to check their bills recurrently could also decrease the price salience—see Sexton (2015), for example in the electric sector.

When it comes to the industrial sector in Brazil, natural gas is mainly used as a fuel and as a chemical feedstock. These industries demand substantial capital investments—such as in facilities and machinery—that take a long time to mature. Therefore, the low price elasticity that we find in the industrial sector is likely reflecting the limited flexibility to readapt industrial plants for alternative fuel and feedstock sources in the short run.

1.7 Conclusion

With an increase in the awareness of climate change worldwide, the reduction of carbon emissions has become a first-order problem. Natural gas is the least polluting fossil fuel and is sometimes referred to as a "bridge" fuel, because it may pose useful in the transition from a fossil fuel driven economy to one that utilizes more renewable power sources.

To the best of our knowledge, this paper is the first to measure short run natural gas demand elasticities of residential and industrial clients in Brazil. We believe our results can be extended to other warm middle-income countries with similar consumption patterns.

We find that both residential and industrial consumers exhibit a very low price sensitivity to changes in natural gas prices. Potential reasons include the lack of alternative fuels, consumers inattention to prices, among others.

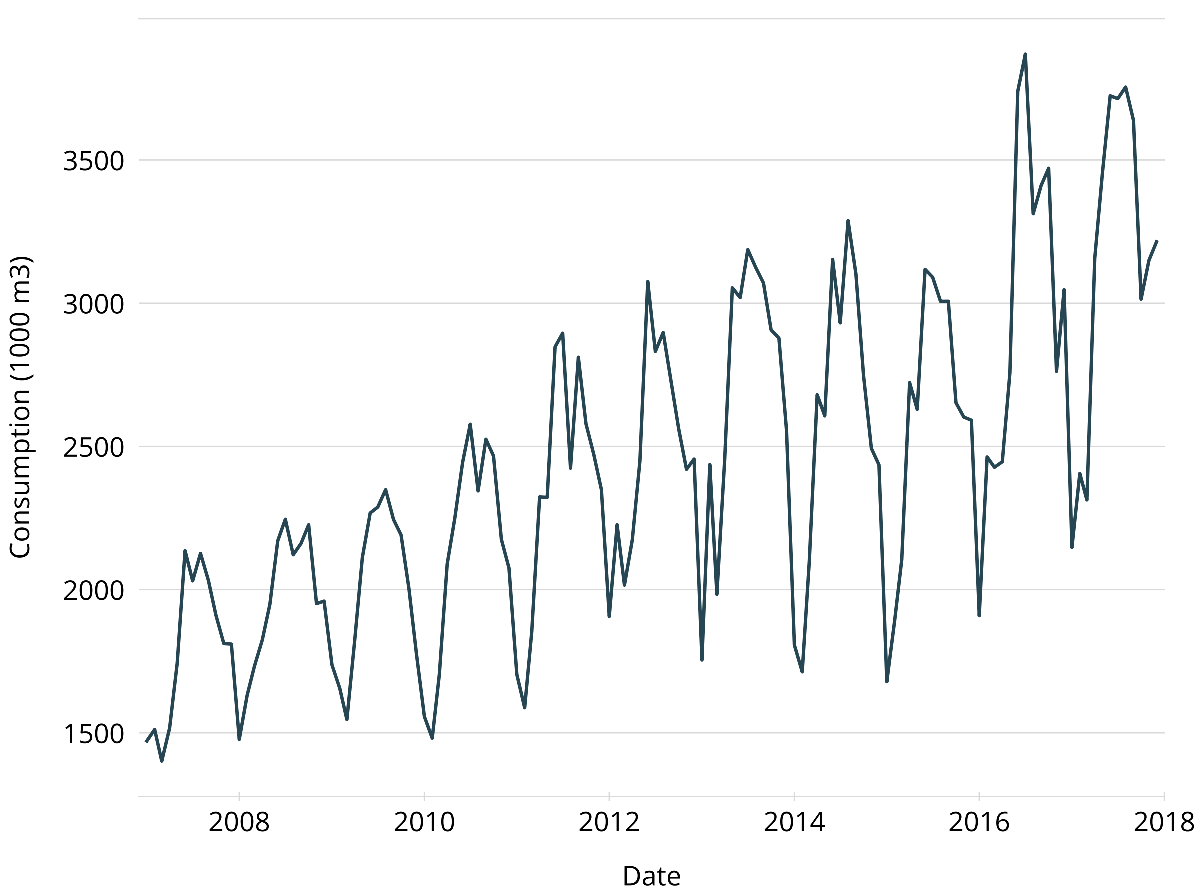
Our findings can aid policymakers in building better regulations that incentivize the development of the natural gas industry, which can in turn increase tax and royalties

[16]In fact, Auffhammer and Rubin (2018) find that low-income households appear to be twice as sensitive to price changes than high-income households.

collection for the government, the industry competitiveness through a reduction of energy costs, and also contribute to the reduction of pollutant emissions in the national territory.

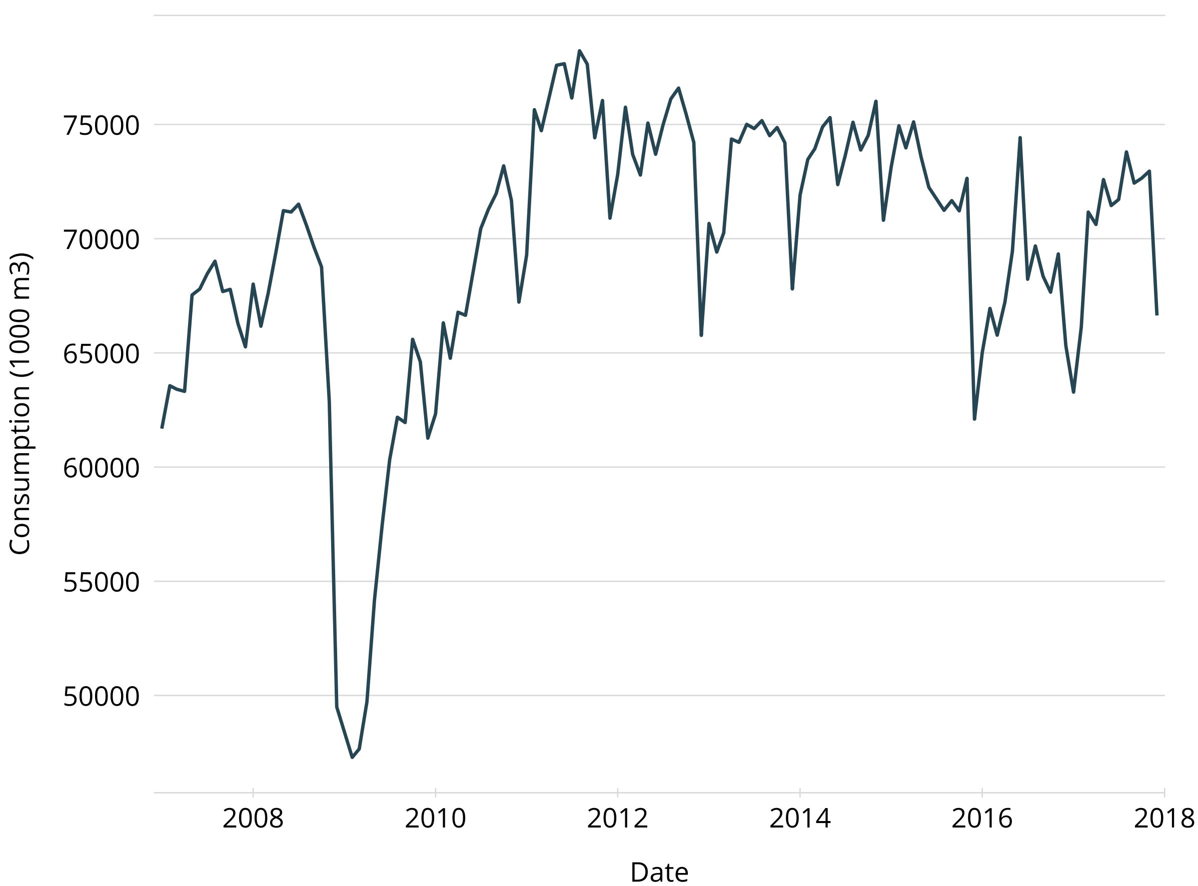
1.A Appendix

Figure 1.1: Average Residential Consumption



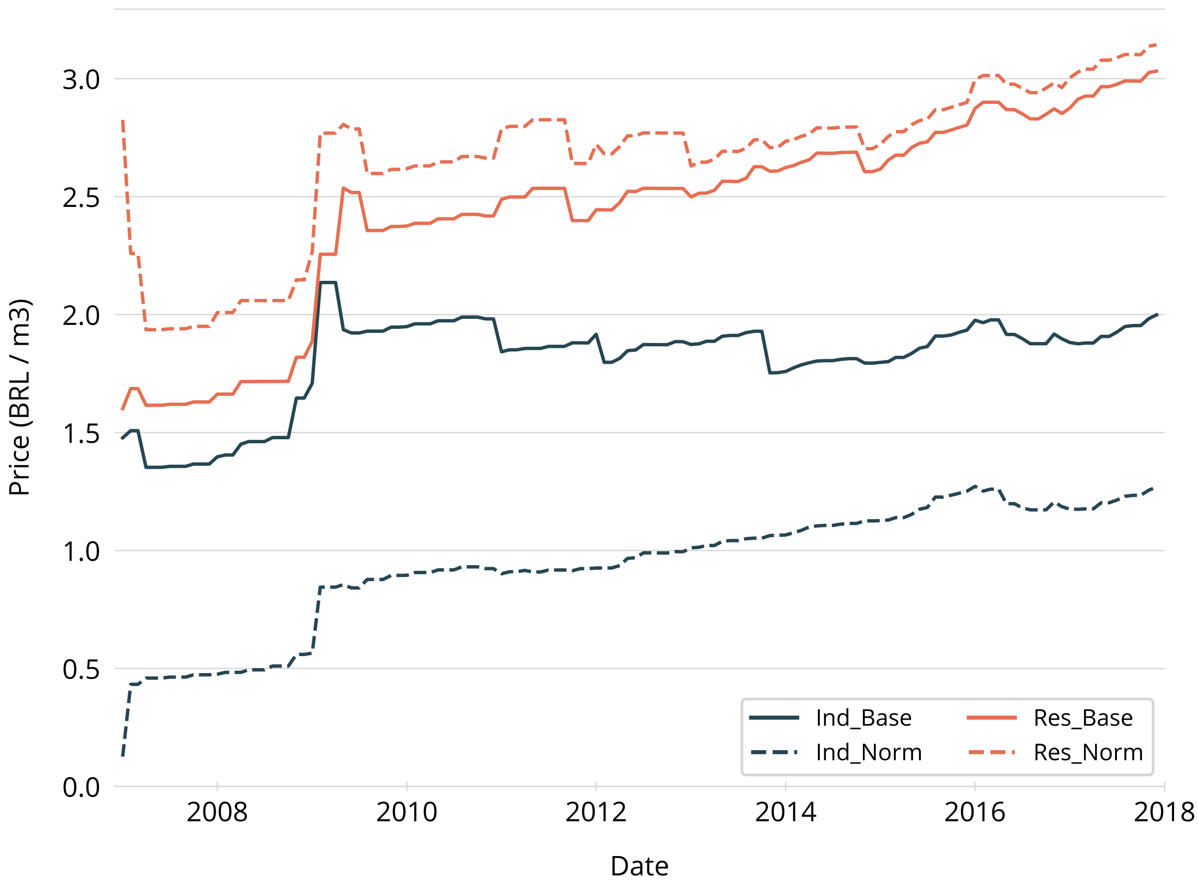
Note: This figure shows the monthly evolution of the average residential natural gas consumption over the years. Our sample covers the period January 2007 to December 2017. Consumption data comes from ABEGÁS. Consumption is measured in thousand of cubic meters.

Figure 1.2: Average Industrial Consumption



Note: This figure shows the monthly evolution of the average industrial natural gas consumption over the years. Our sample covers the period January 2007 to December 2017. Consumption data comes from ABEGÁS. Consumption is measured in thousand of cubic meters.

Figure 1.3: Average Prices



Note: This figure shows the monthly evolution of the residential and industrial average baseline and normalized prices over the years. Baseline price is the first nonzero price level—which is not a function of the consumption—in the nonlinear pricing scheme. Normalized price is defined as the amount charged from those who consume between 16 and 55 cubic meters per month in the residential sector, and between 300 thousand and 500 thousand cubic meters per month in the industrial sector. Prices data comes from both state regulators and the distribution companies. Our sample covers the period January 2007 to December 2017. The prices dataset is an unbalanced panel. All prices are nominal.

Table 1.1: Natural Gas Demand Elasticities — Point Estimates from the Literature

Article	Data	Residential	Industrial
Auffhammer and Rubin (2018)	Panel	–0.17 to –0.23	—
Davis and Kilian (2011)	Cross-Section	–0.10	—
Davis and Muehlegger (2010)	Panel	–0.28	–0.71
Hausman and Kellogg (2015)	Panel	–0.11	–0.16
Huntington (2007)	Time Series	—	–0.24

[1] Authors own elaboration.

Table 1.2: Descriptive Statistics

	Mean	Std. Dev.	Median	Min	Max	N
Pipeline Length	1,970.45	3,595.54	620.46	0.00	18,693.79	1,440
Consumption						
Residential	2.43	5.31	0.09	0.00	27.28	1,452
Industrial	69.54	94.89	41.68	0.00	396.08	1,452
Baseline Price						
Residential	2.60	0.92	2.17	1.47	5.04	953
Industrial	1.85	0.56	1.84	0.82	3.95	977
Normalized Price						
Residential	2.77	1.42	2.13	1.29	8.04	953
Industrial	1.04	0.25	1.08	0.13	1.67	977

[1] In this table, a unit of observation is a state in a month. Our sample covers the period January 2007 to December 2017, and it is an unbalanced panel. Pipeline length is in kilometers. Consumption is in millions of cubic meters. Prices are nominal and measured in BRL per cubic meter. Baseline price is the first nonzero price level—which is not a function of the consumption—in the nonlinear pricing scheme. Normalized price is the amount charged from those who consume between 16 and 55 cubic meters in the residential sector, and between 300 thousand and 500 thousand cubic meters per month in the industrial sector. Consumption and pipeline length data come from ABEGÁS. Pricing data come from both state regulators and the distribution companies.

Table 1.3: Demand Estimation Results — Residential Sector

	OLS				IV	
	(1)	(2)	(3)	(4)	(5)	(6)
Log (NG Price)	0.28***	0.24***	–0.49***	–0.20***	–0.19***	–0.18***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Log (LPG Price)	—	—	—	—	—	0.09
						(0.10)
Log (Pipeline Length)	—	—	—	0.50***	0.50***	0.50***
				(0.04)	(0.04)	(0.04)
Cooling Degrees	–0.06***	–0.03***	–0.04***	–0.05***	–0.05***	–0.05***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
First-Stage F-Statistic	—	—	—	—	5,001.33	4,537.67
State FE	Y	Y	Y	Y	Y	Y
Month FE	N	Y	Y	Y	Y	Y
Constant	Y	Y	Y	Y	Y	Y
Linear Time Trend	N	N	Y	Y	Y	Y
Observations	913	913	913	913	913	913
Adjusted R^2	0.95	0.96	0.98	0.98	0.98	0.98

[1] This table reports the estimates of Equation 1.1 using data from the residential sector. Each column presents the result of a different regression. A unit of observation is a state in a month. The dependent variable is the logarithm of the consumption of natural gas, in thousands of cubic meters. *Log (NG Price)* is the logarithm of the natural gas baseline price in the residential sector, in BRL per cubic meter. *Log (LPG Price)* is the logarithm of the LPG price. *Log (Pipeline Length)* is the logarithm of the pipeline grid length. *Cooling Degrees* is defined as $Cooling = Max\{Temperature - 18, 0\}$, where *Temperature* is the actual atmospheric temperature. In columns 1–4, the regression is run by OLS. In columns 5–6, the regression is run by 2SLS. In these two columns, we instrument the natural gas baseline price in the residential sector with the normalized price of this fuel in the industrial sector, within the same state. All columns include a constant and state fixed effects. Our sample covers the period January 2007 to December 2017, and it is an unbalanced panel. Consumption and pipeline length data come from ABEGÁS. Prices data comes from both state regulators and the distribution companies. All prices are nominal. LPG information comes from ANP. Temperature data comes from ECMWF. Robust standard errors are in parenthesis. *p<0.1, **p<0.05, ***p<0.01.

Table 1.4: Demand Estimation Results — Industrial Sector

	OLS				IV
	(1)	(2)	(3)	(4)	(5)
Log (NG Price)	–0.10***	–0.09***	–0.00	–0.00	–0.22***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)
Log (Pipeline Length)	—	—	—	0.05	0.07
				(0.04)	(0.04)
Cooling Degrees	–0.01***	–0.02***	–0.02***	–0.02***	–0.01**
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
First-Stage F-Statistic	—	—	—	—	537.24
State FE	Y	Y	Y	Y	Y
Month FE	N	Y	Y	Y	Y
Constant	Y	Y	Y	Y	Y
Linear Time Trend	N	N	Y	Y	Y
Observations	977	977	977	977	953
Adjusted R^2	0.97	0.97	0.97	0.97	0.97

[1] This table reports the estimates of Equation 1.1 using data from the industrial sector. Each column presents the result of a different regression. A unit of observation is a state in a month. The dependent variable is the logarithm of the consumption of natural gas, in thousands of cubic meters. *Log (NG Price)* is the logarithm of the natural gas baseline price in the industrial sector, in BRL per cubic meter. *Log (Pipeline Length)* is the logarithm of the pipeline grid length. *Cooling Degrees* is defined as $Cooling = Max\{Temperature - 18, 0\}$, where *Temperature* is the actual atmospheric temperature. In columns 1–4, the regression is run by OLS. In column 5, the regression is run by 2SLS. In this column, we instrument the natural gas baseline price in the industrial sector with the normalized price of this fuel in the residential sector, within the same state. All columns include a constant and state fixed effects. Our sample covers the period January 2007 to December 2017, and it is an unbalanced panel. Consumption and pipeline length data come from ABEGÁS. Prices data comes from both state regulators and the distribution companies. All prices are nominal. Temperature data comes from ECMWF. Robust standard errors are in parenthesis. *p<0.1, **p<0.05, ***p<0.01.

Chapter 2

The Welfare Effects from Electricity Theft: Evidence from Brazil

2.1 Introduction

In most of the developing world, Non-Technical Losses (NTL)—electricity that is consumed but not billed (i.e. stolen from the grid[17])—is pervasive and can be a serious problem. For example, the percent of electricity losses out of all the energy injected into the system is 14% in Africa and 17% in Latin America and the Caribbean (Jiménez et al., 2014)[18], but it can be much larger than that in some countries. There are several potential negative impacts on the energy sector as a consequence of NTL. Among them: 1) *a less reliable grid*, with more power outages, as demand becomes more unstable and difficult to predict; 2) *energy waste*, because consumers that steal energy pay a price of zero per KWh and do not internalize the generation and distribution costs; 3) *excessive prices* as the electric utilities typically pass on the cost of the stolen electricity to formal consumers; 4) *personal injuries* due to illegal connections that cause electric shocks; and so on. Moreover, there can be environmental costs due to the waste of energy. This is important to consider as it is forecasted that by 2035 the energy demand in the developing world will be twice that of the developed world (Wolfram et al., 2012). Not addressing the issue of NTL can contribute to an increase in CO_2 emissions from electricity generation worldwide.

Despite the relevance of the question, there is almost no literature studying NTL. That

[17] 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.

[18] 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.

is noted in a recent survey (Lee et al., 2017) where the authors list the issue of NTL among the key areas for future research. In particular, the authors call for a better understanding of how utilities and policymakers should respond to NTL. One of the reasons for the lack of past work is the difficulty in obtaining detailed micro data on electricity theft. See for example Jacobi and Sovinsky (2016) and Galenianos and Gavazza (2017) for other studies that discuss the difficulties of empirical work in markets with limited access to consumer data.

In this paper we try to address this hole in the literature and understand better the economics behind electricity theft. In particular we try to answer the following question: How much (if any) consumer surplus is lost due to NTL? Our focus lies in the residential sector. Moreover, we are agnostic about the positive externalities associated with a reduction of theft, including: (1) a decrease in the amount of pollution from carbon emissions; (2) an increase in the quality of the service provided by utilities, due to a reduction in the number of power outages; among others.

We obtained access to detailed data from a large electric utility in Brazil. Brazil is one of the countries in the world where the energy theft problem is the most severe (ANEEL, the sector regulator, reports over 33 TWh of stolen energy in 2018). The firm that we study provides electricity to over 10 million people, and is located in an area where power theft is particularly severe. We obtained and combined a number of different datasets: a long time series with aggregated data on billed consumption and NTL over time, a cross-section of all the "formal" consumers and, importantly, a panel with NTL information at the month and feeder level[19]. Since, by definition, there is no direct information available on consumers that engage in electricity theft, the disaggregated information at the feeder level is as good as one can get. At the feeder level, the utility knows how much electricity was distributed, and the amount of technical losses. Therefore, NTL is just the difference between the two. This is the traditional and best available method to compute NTL (Lewis, 2015).

We start by documenting patterns in our data. We show that both the amount of billed consumption and NTL are highly seasonal. Moreover, we present evidence that consumers respond to permanent price increases by migrating to informality.

In order to evaluate the impact of different policies on consumer surplus, we then set up and estimate a structural demand model. In the model, consumers make a discrete and a continuous decisions. First they decide if they want to be formal consumers of the firm (paying the full price) or steal the product (and pay zero). Then, conditional on that decision, they decide how much electricity to consume. The trade-off that consumers face

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

is clear: by moving to NTL they face a price of zero for each unit of electricity and hence are able to increase their utility from consumption. On the other hand they incur in a non-pecuniary fixed cost (which represents the costs of the illegal connection, lost benefits from not being a formal consumer, etc).

In this preliminary version of our model, we assume that the consumer is either formal or informal. However, we do not allow for an intermediate scenario in which households pay for a fraction of their total power consumption and steal the remainder. Based on anecdotal evidence, we believe this intermediate case is the most likely to occur in practice. Therefore, as of this version of the paper, we are running the risk of underestimating the number of households that steal power and, as a consequence, overestimating the average electricity consumption of informal consumers.

We use the primitives from the model to simulate different counterfactual scenarios. In some of these exercises, we just promote an exogenous change in electricity prices. In others, we remove the possibility of theft from the choice set of consumers and, at the same time, reduce the electricity price in a way that the revenues/profits from the utility firm do not change. In the current version of our model, this exogenous ban on theft leads to reductions in the consumer surplus relative to the baseline scenario, even when we repeat the counterfactual exercises for different income quintiles. We show evidence that these unrealistic counterfactuals are likely driven by idiosyncratic logit shocks and discuss the importance of obtaining extra cross-section variation to allow for the estimation of further heterogeneity in our model. Finally, we discuss avenues for improving our model.

Relevant Literature

There has been a strong interest recently in studying the electricity sector in developing countries. Examples of questions being asked are: the economic effects of electrification, the relation between the income distribution and demand for electricity, among others. For example, Lipscomb et al. (2013) and Costa and Gerard (2018) look at the case of Brazil, McRae (2015) at Colombia, Gertler et al. (2016) study Mexico, Allcott et al. (2016) and Burlig and Preonas (2016) focus on India, and Auffhammer and Wolfram (2014) on China. See Lee et al. (2017) for a recent survey on the literature on electrification in developing countries.

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

electricity is treated as a right, can undermine universal access to reliable electricity.

In this paper we estimate a discrete-continuous demand for electricity model. Several other papers have tried to empirically understand how consumers make decisions in this sector. For example, Ito (2014) uses spatial discontinuities to provide evidence that consumers respond to electricity average price and not marginal, McRae and Meeks (2016) use a survey to ilicit consumer information about price schedules, and Deryugina et al. (2017) use a difference-in-differences matching estimator to measure quantity responses to changes in prices. However, the closest papers to ours are those that estimate a structural econometric model of electricity demand, namely Dubin and McFadden (1984), Reiss and White (2005), and McRae (2015). In particular, we also estimate a discrete-continuous model like Dubin and McFadden (1984), although in their case the discrete decision is which appliances to purchase while in our case it is whether to steal energy or be a formal customer.[20]

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

In the next section we describe the relevant institutional details. In section 2.3 we detail the different datasets that we have available, and present descriptive statistics and figures. Then, in section 2.4 we introduce and estimate our empirical model. We present our results in section 2.5. The recovered primitives are then used to simulate different counterfactual scenarios, which we do in section 2.6. In section 2.7, we identify avenues to improve our model and discuss how they can be implemented in future work. Finally, in section 2.8, we conclude.

2.2 Institutional Details

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

There are around 50 different local monopolies that distribute electricity in Brazil.

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

Most of them are privately owned but several are public (state owned). The five largest distributors in terms of the number of customers served are, in order: Cemig, Eletropaulo, Coelba, Copel, and Light (EPE, 2018).

The sector is regulated by *Agência Nacional de Energia Elétrica* (ANEEL), which is supervised by the Ministry of Mines and Energy. The price of electricity is regulated. Up to 1993 there was a single electricity price for all of Brazil. From that point onwards, the regulated price was allowed to vary across utilities - but not within. The idea is that the different tariffs reflect the heterogeneity across utilities in terms of productive efficiency, demand conditions, and so on. The residential price varies with the quantity consumed[21]. Some low-income consumers qualify for a lower "social rate". The discount in that case will be a negative function of the quantity consumed, but it can go up to 65% (for low-income, low-consumption households). In 2015, ANEEL introduced a system of "tariff flags" which change each month and introduce some variation in the final price that consumers pay, depending on the color of the flag (red, yellow or green). The color of the flag represents the general conditions of the electric generation system and the goal is for consumers to internalize part of the differences in generation costs over time and adjust consumption accordingly.

There are two types of losses in the distribution of electricity: technical (TL) and non-technical (NTL). The former are just natural losses inherent to the activity of transporting electricity from one place to another, and are a function of the quality of the infrastructure. The latter mostly consists of electricity theft or measurement error. In 2018, the total electricity lost in Brazil, as a percent of the electricity injected in the system, was 14%, roughly divided equally across TL (7.5%) and NTL (6.6%). The total amount of NTL in that year was above 33 TWh. Those percentages are a little misleading because most of the NTL take place in the residential sector. Therefore, while it is natural for the denominator of TL to be the amount of electricity injected, the usual approach is to compare NTL with the total amount of electricity in that sector. In that case, the percentage of NTL goes up to 14.3%. Again, this hides some heterogeneity: at least 7 utilities have NTL higher than 30% of residential consumption. That is the case for the firm that we will study, which is responsible for almost 29% of the total amount of NTL in Brazil (ANEEL, 2019).

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

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

2.3 Data

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

2.3.1 Data Preparation

Prices

We use data on electricity prices obtained from ANEEL. The data set covers monthly ex- and post-tax average[22] retail prices for the period January 2003 to July 2019. We convert all prices to January 2003 R$ using the Extended National Consumer Price Index (IPCA, in Portuguese). Figure 2.1 depicts how prices evolved over the years and reveals two features of our data set. First, real prices declined steadily from 2003 to 2015, except for a few isolated spikes. This decreasing real price behavior is consistent with interventions made by prior Brazilian governments that kept electricity prices artificially low. Second, taxes play a significant role in providing additional price variation that is useful for the model estimation. Our base model electricity price variable is the average post-tax retail price with one lag[23].

Households

We would like information on the number of households for each one of the 31 Rio de Janeiro's counties the utility serves. The Brazilian Institute of Geography and Statistics (IBGE, in Portuguese) is the agency responsible for reporting the official population count. It provides annual estimates of the total population for each one of those counties, but not for the total number of households. We recover the latter by proceeding in the following way: first, we use the 2000 and 2010 Census obtained from IBGE to find the average household size in each county in both years. Next, we use linear interpolation to obtain estimates for the period 2003 to 2009. From 2011 to 2019, we assume the average household

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

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

size is constant as of 2010. Our proxy to the total number of households is obtained by dividing the IBGE's population estimate in each county and year by its respective average household size.

Formal Consumption and Number of Clients

We obtained information on formal residential electricity consumption and number of clients from two sources. The first data set contains monthly information on (i) the aggregated amount of electricity the utility sold to residential clients, and (ii) the number of residential clients it had. This piece of information comes from ANEEL and comprises the period January 2003 to July 2019. The second data set is a cross-section of the utility's residential clients in November 2016, with information on each household consumption and spatial coordinates for that month. Figure 2.2 shows the histogram of consumption using this cross-sectional data. We find that approximately 18.4% of all residential clients had zero consumption. This could be due to different reasons, such as households that own one or more vacation houses, vacant properties that are available for rent, etc. We dropped these clients from our analysis by assuming that the number of clients with non-zero consumption was 18.4% less than that reported by ANEEL at each month[24]. We are left with approximately 3.2 million clients that consume on average 194.47 kWh/month, as shown in Table 2.1. The aggregate formal consumption per household variable we use on the base model is the ratio of the aggregated amount of electricity sold to residential clients to the number of residential clients with non-zero consumption.

Non-Technical Losses (NTL)

We obtained time series data on NTL from the utility firm. The data set covers information from January 2008 to January 2015. The aggregate informal consumption per household variable we use on the base model is the ratio of the amount of NTL to the difference between the total number of households in the utility area and the number of formal residential clients with non-zero consumption[25].

Feeders

We use proprietary data on feeders provided by the utility. The data set is a panel of feeders containing feeder-month level information for the year of 2017. The variables

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

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

available are the amount of electricity (i) generated, (ii) billed, and (iii) used for public lighting; as well as the amount of (iv) technical losses, and (v) non-technical losses. The data set does not contain any geospatial information on each feeder. Notwithstanding, we recover their latitude and longitude through an additional data set, a cross-section of feeders' smaller components, technically called *trafos*, for November 2016. This alternative data set is useful for two reasons: first, it has the latitude and longitude of each *trafo*. Second, it links each *trafo* to its respective feeder. To define the areas covered by each feeder, we create voronoi polygons for each *trafo* and aggregate them at the corresponding feeder level.

Demographics

We use demographics data from the 2010 Brazilian Population Census. We are particularly interested in the following variables: 1) an indicator of whether the census tract lies in an urban area; 2) another indicator of whether it belongs to a risky zone (for example, a slum, where electricity theft rates are usually higher than average); the number of 3) households, 4) residents, and 5) residents by income groups[26], that live in each census tract.

To accommodate this information in our model, we aggregate them at the feeder level. But, first, we need to create a rule to establish how census tracts and feeders are linked. We assume that a census tract is serviced by a feeder if its centroid falls into the feeder service zone. This guarantees that each census tract is always serviced by a single feeder.

Next, we proceed to the aforementioned aggregation process. We start by taking the resident-weighted average of the urban indicator, which yields the share of urban area covered by each feeder. This variable is called *Urban*. We repeat this procedure for the risky zone indicator, generating a risky area share variable we denominate *Risky*. Next, we find the size of the average household served by each feeder, a variable we name *HHD Size*. To do so, we just calculate the ratio of the number of residents to the number of households living in the area covered by each feeder. Finally, we compute the *Income* variable, which is measured in 2017 minimum wages. We proceed in two steps: first, we calculate the share of residents in each income group; second, we use these shares as weights to perform a weighted average of the midpoints of the income groups lower– and upper bounds.

[26]There are ten income groups, as follows: households with no income; households with at most 1/8 minimum wage per capita; between 1/8 and 1/4 minimum wage per capita; between 1/4 and 1/2 minimum wage per capita; between 1/2 and 1 minimum wage per capita; between 1 and 2 minimum wage per capita; between 2 and 3 minimum wages per capita; between 3 and 5 minimum wages per capita; between 5 and 10 minimum wages per capita; and with more than 10 minimum wages per capita.

2.3.2 Descriptive Statistics

Tables 2.2 and 2.3 provide some reduced form evidence of how consumers respond to changes in electricity prices. In Table 2.2 we report the results of OLS regressions of the log of aggregated formal consumption against different price variables using the time series data. All columns include month of the year fixed effects to control for seasonality in consumption. In column 3 we use the log of electricity price as the price variable. Alternatively, we use the log of the price of electricity with taxes in column 4. As mentioned previously, taxes provide some useful additional price variation. Columns 5 and 6 both use the log of the deflated price of electricity, but only the latter includes taxes. The results suggest all price coefficients have the expected negative sign and are close to each other. Since we are using a log-log specification, we can interpret the coefficients as elasticities. In column 3, for instance, a 1% increase in the price of electricity is associated with a -0.20% reduction in the aggregated formal consumption. We also found a positive and statistically significant time trend coefficient.

Table 2.3 reports the results of OLS regressions of the log of aggregated formal consumption and the log of NTL against the same price variables shown in Table 2.2, but using the panel of feeders instead of the time series. We used feeder fixed effects in columns 1 to 8. We also removed seasonality from the formal electricity consumption (billed consumption) and NTL data. Again, the price elasticities for formal consumption (columns 1 to 4) have the expected sign and are statistically significant. We note that they are bigger in magnitude than those we found in Table 2.2, though. Regarding NTL, the positive and statistically significant coefficients found in columns 5 to 8 suggest that an increase on electricity prices is associated with higher amounts of NTL.

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.[27] 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.

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

Figure 2.3 shows the number of total clients, and number of clients with a regular tariff, before and after this change in policy (during 2011). The number of clients with a regular tariff increased dramatically during 2011, followed by a partial decrease. This is consistent with the anecdotal evidence that many people only became aware of the change after seeing the increase in their bill. The fact that this reduction only partially offset the initial increase is also consistent with the stricter criteria applied after the change. This led to a change in the number of total residential consumers. Since we do not expect any consumer to stay without power (and since consumers cannot buy electricity from any company other than the utility), this effect is likely driven by consumers migrating to electricity theft. In Figure 2.4, we observe an increase in the share of stolen electricity in 2012 (decrease in the share of formal consumption) which is consistent with this migration to electricity theft.

Consumption Seasonality and Increasing Variance Over the Years

Electricity consumption is expected to be seasonal in Brazil. Since we use data from a utility that operates in a Brazilian state where summer temperatures are high enough to justify the usage of air conditioning, but winters are not cold enough to create demand for heating systems, we would expect consumption to be above average over summer only (end of December to end of March). This electricity usage pattern is precisely what we see in Figures 2.5 and 2.6. The former shows that formal aggregate consumption spikes near January and goes down abruptly near the middle of the year. The latter is a box-plot also for formal consumption from January to December for the whole period of study (January 2003 to July 2019). It shows that consumption starts to increase consistently in October and keeps above average over January, February and March. On the other hand, it starts to decrease around April and goes beyond average from June to September, when it starts to increase again. Figure 2.5 also shows an interesting pattern of formal electricity consumption: it became more volatile over the years. Figure 2.7 suggests that NTL presents a similar pattern, in the sense that we can observe both seasonality and an increase in volatility over the years for informal consumption.

2.4 Model

2.4.1 Framework

Household i chooses either formal electricity consumption ($j = 0$) or electricity theft ($j = 1$) in market t. A market here is defined as the entire region covered by the electricity utility in a given month. Conditional on their choice, utility is quasi-linear on the consumption of electricity, q, and the numeraire good, m[28]:

$$\tilde{v}_{ijt}(q, m) = \phi_{ijt}(q) + m \tag{2.1}$$

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

$$\phi_{ijt}(q) = \theta_{1jt}q - \frac{1}{2}\theta_2 q^2 \tag{2.2}$$

Letting y_{it} denote the household income, we obtain:

$$v_{ijt} = \theta_{1jt}q - \frac{1}{2}\theta_2 q^2 + y_{it} - p_{jt}q, \tag{2.3}$$

where p_{jt} is the price households pay for each KWh consumed. The price of electricity in the formal consumption choice, p_{0t}, is just the post-tax electricity retail price. Naturally, $p_{1t} = 0$.

Next, we embed this specification in a random-coefficient framework:

$$u_{ijt} = \beta_i \left(\theta_{1jt}q - \frac{1}{2}\theta_2 q^2 + y_{it} - p_{jt}q \right) + \eta_{jt} + \varepsilon_{ijt} \tag{2.4}$$

$$\beta_i = \beta + \sigma\gamma_i, \qquad \gamma_i \overset{iid}{\sim} N(0, 1) \tag{2.5}$$

Here, β_i is a random-coefficient that describes how preferences vary across households. γ_i accounts for unobserved household characteristics that may be relevant in the decision of whether to be formal or informal. For example, households could respond differently to the service reliability, or to the threat of punishment associated with electricity theft. η_{jt} are unobserved (by the econometrician) choice/market-specific features that govern the

[28]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.

relative desirability of formal versus informal consumption, such as demographics[29] and seasonality. And ε_{ijt} are mean-zero individual preference shocks, which are independent and identically distributed across households, markets, and choices, with extreme value distribution, $F_{\varepsilon_{ijt}}(\varepsilon) = \exp(-\exp(-\varepsilon))$.

Intensive Margin

The intensive margin decision is determined by the first order condition of the household utility maximization problem. Note that the random-coefficient is irrelevant for this choice. The quadratic specification for the quasi-linear utility yields a linear demand for electricity. In particular, since informal consumers face a zero marginal electricity price, they consume up to satiation[30]. Conditional on alternative j, household i electricity demand is:

$$q_{ijt} = \frac{\theta_{1jt}}{\theta_2} - \frac{1}{\theta_2} p_{jt} \tag{2.6}$$

Extensive Margin

Substituting (2.4) into (2.5), we obtain:

$$\begin{aligned} u_{ijt} &= (\beta + \sigma\gamma_i)\left(\theta_{1jt}q - \frac{1}{2}\theta_2 q^2 + y_{it} - p_{jt}q\right) + \eta_{jt} + \varepsilon_{ijt} \\ &= y_i(\beta + \sigma\gamma_i) + \underbrace{\left(\theta_{1jt}q - \frac{1}{2}\theta_2 q^2 - p_{jt}q\right)}_{=\psi_{jt}}(\beta + \sigma\gamma_i) + \eta_{jt} + \varepsilon_{ijt} \\ &= y_i(\beta + \sigma\gamma_i) + \beta\psi_{jt} + \eta_{jt} + \sigma\gamma_i\psi_{jt} + \varepsilon_{ijt}, \end{aligned} \tag{2.7}$$

where ψ_{jt} is the household surplus (in BRL) generated by electricity consumption only.

Since only differences in utility matter to the extensive margin decision, we follow the standard practice in the discrete choice literature and normalize $\eta_{1t} = 0$. The extreme

[29] Demographic variables usually carry the i subscript to indicate they refer to individuals. Here we drop their subscript, though, since these individual variables are only observed aggregated at the market level.

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

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

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

value distribution on ε_{ijt} implies that household i probability of choosing to be formal is:

$$P_{i0t} = \frac{\exp\Big(\beta(\psi_{0t} - \psi_{1t}) + \eta_{0t} + \sigma\gamma_i(\psi_{0t} - \psi_{1t})\Big)}{1 + \exp\Big(\beta(\psi_{0t} - \psi_{1t}) + \eta_{0t} + \sigma\gamma_i(\psi_{0t} - \psi_{1t})\Big)} \tag{2.8}$$

In most settings, the researcher have data on the share of consumers that choose each alternative. Our case is a little different, though, because our dataset only contains the share of electricity—i.e., a share in quantities, rather than in the number of people—that is either paid for or stolen. Adjusting for that, our model predicts that the market share (in quantities) of electricity consumed formally, s_{0t}, is:

$$s_{0t} = \frac{q_{0t}P_{0t}}{q_{0t}P_{0t} + q_{1t}(1 - P_{0t})}, \qquad \text{where } P_{0t} = \int P_{i0t}dF(\gamma_i) \tag{2.9}$$

2.4.2 Identification

As in most electricity retail markets, price changes here are set by the regulator. These price changes are typically directly related to the availability of water in hydro reservoirs and the rain patterns in their associated basins. Most electricity that supplies the Brazilian grid is hydro generated, so the availability of electricity is sensitive to seasonal and decennial weather patterns that are closely monitored by the regulator. These price changes are therefore unrelated to current demand shocks.[31] We consider therefore prices as exogenous in formal demand equation (2.6), conditional on controlling for month of the year fixed effects.

2.4.3 Estimation

We propose a 2-step estimation approach. In the first step, we estimate the electricity demand using formal and informal consumption data (intensive margin), which allow us to recover θ_{10t}, θ_{11t} and θ_2. This first step uses the long time series dataset of household consumption and prices. In the second step, we estimate the parameters governing the relative desirability of formal versus informal consumption. The second step uses the panel dataset of feeders.

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

First Step

Using aggregate formal consumption data per household, we estimate:

$$\bar{q}_{0t} = \delta_{0t} - \alpha \bar{p}_{0t} + \nu_t, \tag{2.10}$$

where δ_{0t} includes month of the year fixed effects to control for seasonality, and a linear time trend to account for potential changes in consumption behavior and the advent of energy-efficient technologies over the years.

Using the estimates from (2.10), we recover:

$$\hat{\theta}_2 = \frac{1}{\hat{\alpha}}, \qquad \hat{\theta}_{10t} = \frac{\hat{\delta}_{0t} + \hat{\nu}_t}{\hat{\alpha}}.$$

Next, using aggregated informal consumption per household, $\bar{q}_{1t}$, we recover:

$$\hat{\theta}_{11t} = \bar{q}_{1t}\hat{\theta}_2.$$

There is an abuse in notation in the use of the market subindex, t, here. In this first step, equation (2.10) is estimated using aggregated data at the utility-month level, so we are actually averaging over all feeders in a given month. Therefore, in this step, t refers to a month, and not to a pair month-feeder—as elsewhere in the section. Averaging over feeders is possible because of the demand linearity, but it is not totally innocuous. It implies θ_{10t} and θ_{11t} are allowed to vary over time, but not along the cross section of feeders.

Second Step

Using $\hat{\theta}_{10t}$, $\hat{\theta}_{11t}$, and $\hat{\theta}_2$, we can compute the household surplus (in BRL) in each type of consumption:

$$\hat{\psi}_{0t}(q_0) = \hat{\theta}_{1,0t} q_{0t} - \frac{\hat{\theta}_2}{2} q_{0t}^2 - p_{0t} q_{0t}, \tag{2.11}$$

$$\hat{\psi}_{1t}(q_1) = \hat{\theta}_{1,1t} q_{1t} - \frac{\hat{\theta}_2}{2} q_{1t}^2. \tag{2.12}$$

Substituting (2.11) and (2.12) into (2.8), we obtain:

$$P_{i0t} = \frac{\exp\left(\delta_t(\hat{\psi}_{0t}, \hat{\psi}_{1t}, \eta_{0t}; \beta) + \mu_{it}(\hat{\psi}_{0t}, \hat{\psi}_{1t}, \gamma_i; \sigma)\right)}{1 + \exp\left(\delta_t(\hat{\psi}_{0t}, \hat{\psi}_{1t}, \eta_{0t}; \beta) + \mu_{it}(\hat{\psi}_{0t}, \hat{\psi}_{1t}, \gamma_i; \sigma)\right)} \tag{2.13}$$

$$\delta_t(\cdot) = \beta(\hat{\psi}_{0t} - \hat{\psi}_{1t}) + \eta_{0t}, \qquad \mu_{it}(\cdot) = \sigma\gamma_i(\hat{\psi}_{0t} - \hat{\psi}_{1t}),$$

where $\delta_t(\cdot)$ and $\mu_{it}(\cdot)$ are, respectively, the "mean utility" and an heteroskedastic disturbance of the household utility.

Our estimation algorithm is very close to that proposed in BLP (1995)[32], with some slight modifications. It can be summarized as follows:

Algorithm

Draw γ_i from a $N(0, 1)$, for a set of NS households, in each market t. Now, for each guess of σ:

1. Set a starting vector, δ^0.

2. For each $k > 0$, obtain δ^{k+1}, using:

$$\delta^{k+1} = \delta^k + \log(S_0) - \log\left(s_0(\delta^k, \sigma)\right),$$

   where

$$s_{0t}(\delta^k, \sigma) = \frac{q_{0t}P_{0t}}{q_{0t}P_{0t} + q_{1t}(1 - P_{0t})}, \qquad P_{0t} = \frac{1}{NS}\sum_{i=1}^{NS} \frac{\exp(\delta_t + \mu_{it})}{1 + \exp(\delta_t + \mu_{it})},$$

   and S_0 is a vector of observed market shares.

3. Iterate until convergence, that is, δ^{k+1} sufficiently close to δ^k.

4. Estimate:

$$\delta_t(\sigma) = \beta(\hat{\psi}_{0t} - \hat{\psi}_{1t}) + \gamma_t + e_t.$$

[32]Note that, if we had data at the consumer-level, we could estimate the model by Maximum Likelihood. However, even if we had micro-data on formal consumption, there would still be missing a key ingredient—consumer-level data on theft—which are, unsurprisingly, unavailable even to the own utility company. The use of structural models and the BLP algorithm are a very powerful combination in scenarios like this, because they allow us to model the consumer decision at the individual-level, and use the feeders aggregated data—or market-shares—to recover the primitives of the model.

This will yield the linear parameters, $\hat{\beta}(\sigma)$ and $\hat{\gamma}_t(\sigma)$, a set of demographic variables. Note that we are implicitly assuming that ν_t, from equation (2.10), is independent from e_t and ε_{ijt}.

5. Construct the residuals:

$$\hat{e}_t(\sigma) = \hat{\delta}_t(\sigma) - \hat{\beta}(\sigma)(\hat{\psi}_{0t} - \hat{\psi}_{1t}) - \hat{\gamma}_t(\sigma).$$

6. Build the sample moments conditions:

$$\bar{g}(\sigma) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z_t}\hat{e}_t(\sigma),$$

where $\boldsymbol{z_t}$ is a vector of instruments, including the own explanatory variables and their squares.

7. Minimize the GMM objective function:

$$\bar{g}(\sigma)W\bar{g}'(\sigma),$$

where W is a matrix of weights.

2.4.4 Consumer Surplus

Our measure of consumer surplus, in BRL, is given by the household expected utility prior to the realization of the ε_{ijt} shocks, which is standard in the discrete choice framework:

$$E_{\beta_i,\varepsilon_{ijt}}\left[\max_{j=0,1} u_{ijt}\right] = \int \frac{1}{\beta_i} \ln\left(\sum_{h=0,1} \exp\left(\beta_i\psi_{ht} + \eta_{ht}\right)\right) dF(\beta_i) \tag{2.14}$$

In counterfactual scenarios where theft is removed from the household choice set, equation (2.14) simplifies to:

$$E_{\beta_i,\varepsilon_{i0t}}\left[\max u_{i0t}\right] = \int \frac{1}{\beta_i} \ln\left(\exp(\beta_i\psi_{0t} + \eta_{0t})\right) dF(\beta_i) \tag{2.15}$$

2.5 Estimation Results

We begin with our estimates from the first step. Table 2.4 reports our OLS estimates of Equation (2.10), which is the linear electricity demand in the formal sector. Standard errors are in parenthesis. The regression includes a constant, month of the year fixed effects, and a linear time trend. As expected, we find that formal consumers are sensitive to price changes. In particular, a one standard deviation increase in electricity prices leads to a quarter of a standard deviation decrease in the formal consumption per household. Moreover, we find relevant trend and seasonality effects.

Next, we use our Equation (2.10) coefficient estimates to build the formal and informal household surplus estimates, $\hat{\psi}_{0t}$ and $\hat{\psi}_{1t}$. They are given by Equations (2.11) and (2.12), respectively. One challenge that we face is that the formal and informal (NTL) consumption time series have different lengths. While the former goes from January 2003 to July 2019, the latter only covers the period January 2008 to January 2015. This is illustrated in Figure 2.8. To use the maximum amount of information available, we first run an OLS regression of our informal consumption per household (NTL) variable against a linear time trend, and month of the year fixed effects; then, we replace the missing observations with out-of-sample predictions.

Our second step estimates are reported in Table 2.5. Standard errors are in parenthesis. In the *Logit* column, we exhibit the results obtained when $\sigma = 0$. In this case, the heteroskedastic disturbance term of the household utility, μ_{it}, disappears from our model, which simplifies to a standard Logit. Alternatively, in the *RC* column, we present our estimates from the Random Coefficient model. This version of the model is estimated following the algorithm previously described in this paper. Asymptotically efficient standard errors are in parenthesis.

Both specifications yield similar results. Their point estimates have the same sign and close magnitudes, although some of the coefficients from the *RC* model are not statistically different from zero, including the heterogeneity term, σ. The consumer surplus difference $(\psi_{0t} - \psi_{1t})$ coefficient, β, has the expected positive sign, meaning that when the surplus from formal consumption increases relative to the surplus of electricity theft, there is an increase in the share of formal consumers. The demographic variables also have the predicted signs. For example, we find that both wealthier and rural households are more likely to be formal. On the other hand, households living in risky areas—such as slums, which are irregular settlements usually located in poor and violent regions controlled by drug-traffickers and militias—are more likely to be informal. We can think of two possible explanations for this latter result: the first reason is that the utility employees may have

limited access to these risky areas, which reduces the cost of informality associated with the chance of being caught and fined for stealing power. The second motive is that households living in these regions could be coerced into obtaining illegal electricity services sold by some criminal groups, even if they would otherwise choose formality. Finally, we find a negative sign for the average household size variable, suggesting that larger households—that tend to consume more electricity and thus face more expensive utility bills—are more likely to steal power.

2.6 Counterfactuals

We use the parameters estimates from the Logit model to analyze five different counterfactual scenarios. Table 2.6 displays our results.

In Columns (1) and (2), we promote an exogenous 10% increase and decrease in electricity prices, respectively. These price changes imply a moderate change in the consumption of formal households (intensive margin), but they have a small effect in the share of formal households (extensive margin). Therefore, there seems to be little room for price reductions to lead an increase in formalization that could compensate, at least in part, the revenue lost from such price drops.

In Columns (3)–(5), we consider an exogenous ban on theft—i.e., we set the share of formal households to 100%. In Column (3), we keep the electricity prices constant, relative to the baseline. In this scenario, the utility revenue goes up by nearly 18%. Moreover, the utility profit almost triplicates. This happens because: (i) the utility average revenue per household increases, since there are now more consumers paying for the electricity; and (ii) the utility average electricity acquisition cost decreases[33] because the share of informal consumers, who tend to use more power, on average, becomes zero. In Column (4), we decrease the price in order to keep the utility's revenue as in the baseline. In this scenario, prices could drop by 19%. Finally, in Column (5), we reduce the price even more (by approximately 29%), in order to keep the utility's profit constant, relative to the baseline.

The two last rows of Table 2.6 measure the changes in consumer surplus relative to the baseline. The values in the penultimate row, ΔCS, come from Equations 2.14 and 2.15. In the last row, ΔCS (Without ε), we compute the component of ΔCS ex- the contribution of the idiosyncratic ε logit shocks. A similar decomposition of consumer surplus is performed in Petrin (2002). These two rows suggest, first, that the current version of our model produces unrealistic changes in consumer surplus. This becomes clear when we consider

[33] We obtain information on average costs of acquiring electricity from ANEEL. In 2017, these costs accounted for approximately 63.1% of the price of electricity charged from formal households.

the counterfactual scenarios in Columns (3)–(5). For example, in Column (5), we find that the average household loses nearly 322 BRL a month in consumer surplus, relative to the baseline. Second, these rows provide evidence that, in our model, the logit shocks accounts for an important component of the consumer surplus, which reinforces the need of better controlling for heterogeneity. The use of random coefficients is one potential solution[34], as they allow preferences to vary among consumers and reduce the dependence on the ε.

We further investigate how our counterfactual results change for different quantiles of the income distribution. Tables 2.7 to 2.11 report the same counterfactual scenarios for distinct income quintiles. The results are consistent with what we expected. First, the fraction of formal households increase monotonically with the income quintile. Moreover, in Table 2.7, there is a significant loss of consumer surplus for the average household when theft is banned—Columns (3) to (5)—relative to the world in which electricity theft is allowed. This comes with no surprise, since informal consumers are now worse-off (by revealed preference), relative to the baseline scenario. However, this loss falls monotonically and sharply as we go to the top of the income distribution. In Table 2.11, for instance, these counterfactual scenarios still yield a negative change in consumer surplus, but they are now much closer to zero.

2.7 Discussion

In this section, we identify opportunities to improve our model and discuss how they can be implemented in future work.

The first point of concern is the lack of seasonality controls in the second step estimation. Since both ψ_{0t} and ψ_{1t} are functions of the formal and informal consumption, respectively, which exhibit strong seasonal patterns (Figures 2.5, 2.6, and 2.7), it is natural to expect that the consumer surplus difference, $(\psi_{0t} - \psi_{1t})$, does also present some sort of seasonality. However, we believe that the extensive margin choice is more of a long-run and permanent decision. For example, consistent with anecdotal evidence from the industry, once incurring the costs involved in the migration towards informality, it is very unlikely that a consumer goes back to the formal sector.

In the present version of our model, we are not able to include month of the year dummies, in the estimation of the second step, to control for the seasonality that arises from $(\psi_{0t} - \psi_{1t})$. The main reason is that this variable only exhibits variation in the time

[34] Again, we do experiment a random coefficient specification, but we find that the σ coefficient, which introduces heterogeneity to the model by allowing preferences to vary among consumers and reduces the dependence on ε, is statistically zero.

series dimension (at the month level), so it is collinear with the dummies. By not including the dummies, we force the discrete choice to respond to seasonal variations of the ψ's. As this does not likely occur in practice, our estimates of β are biased towards zero.

One attempt to improve our estimates is to use differences in the share of social and regular consumers in each market to produce cross-section variation[35]. This would allow for the use of month of the year dummies to control for seasonality in the second step estimation, producing (potentially) better estimates of the β coefficient. This cross-section variation could also be useful to pin down the heterogeneity coefficient, σ, in our Random Coefficients specification. As we have previously discussed in the Counterfactuals section of this paper, this specification would likely reduce our model reliance on the ε logit shocks, thus producing changes in consumer surplus that are more realistic than those we found in Tables 2.6 to 2.11.

As an exercise, we estimate the parameters governing the extensive margin decision once again. Now we add month of the year dummies and demographic variables as controls to the logit regression. As per our explanation above, we have to omit two dummies[36]—to avoid collinearity with both the constant and the $(\psi_{0t} - \psi_{1t})$ variables. Next, using our new estimates, we calculate the change in consumer surplus, ΔCS, associated to the fifth counterfactual exercise—i.e., we exogenously ban theft and set the electricity price such that the utility profit is constant, relative to the baseline—for different values of the β coefficient, *ceteris paribus*. Figures 2.9 and 2.10 exhibit our results. While the first provides evidence on the shape of ΔCS as a function of β, the second zooms in and reveals the existence of an interval of values for β where the average household experiences gains of consumer surplus, relative to the baseline. This exercise serves more as a sanity check, in the sense that it suggests that our model and data allows for different combinations of parameters that may generate both gains and losses of consumer surplus.

[35] As we describe in the Institutional Details section of this paper, some low-income (social) consumers in the residential sector qualify for a lower "social rate", with price discounts that can go up to 65% relative to that charged from regular consumers. We do have time series data of prices, consumption and number of social clients that come from the regulator, which are the ingredients we would need to estimate the parameters governing the intensive margin decision of our model. Moreover, we could use the income demographic variable from the 2010 Census to calculate the share of households in each market that qualify for a social tariff. This variable could then be used as a proxy to the actual share of social clients in each market.

[36] Here, we omit the dummies associated with the months of January and December.

2.8 Conclusion

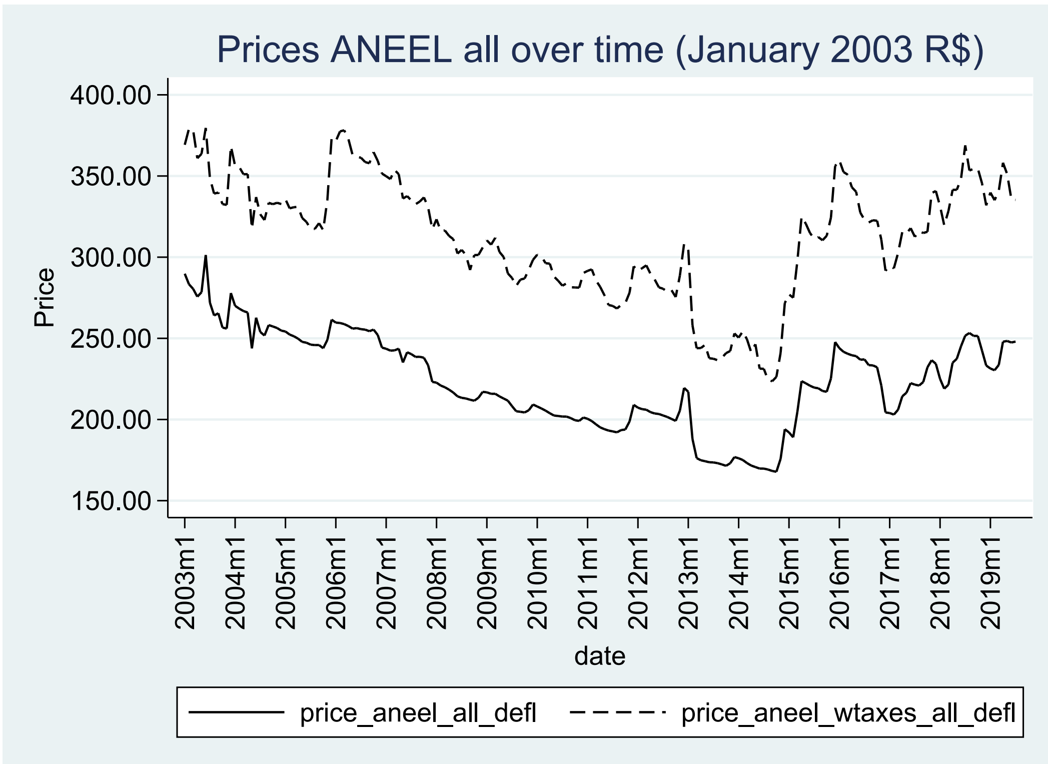
Electricity theft is a significant phenomenon throughout the world, particularly in developing countries, with potential consequences for energy costs and the environment. In this paper, we use detailed micro data from electricity theft in Brazil to evaluate the potential benefits from different policies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

In the current version of our model, we find that banning theft reduces the average consumer surplus, even when the utility reduces the electricity price so that its profit remains unchanged. The price reduction would benefit the (formal) consumers that are currently paying. Consumers that are stealing electricity would be worse-off if theft was not an option (by revealed preference). In principle, our findings could indicate that the disutility of paying a positive price is larger than the gain from nonpecuniary benefits associated with the the choice of formality (no risk of being caught, no cost of making an illegal connection, etc). However, we provide evidence that these preliminary and unrealistic counterfactual results are likely driven, instead, by idiosyncratic logit shocks.

We augment our model with the introduction of random coefficients, although the heterogeneity coefficient is found to be statistically zero. We discuss avenues for improving the parameters identification, such as the introduction of extra cross-section variation through the use of shares of social consumers.

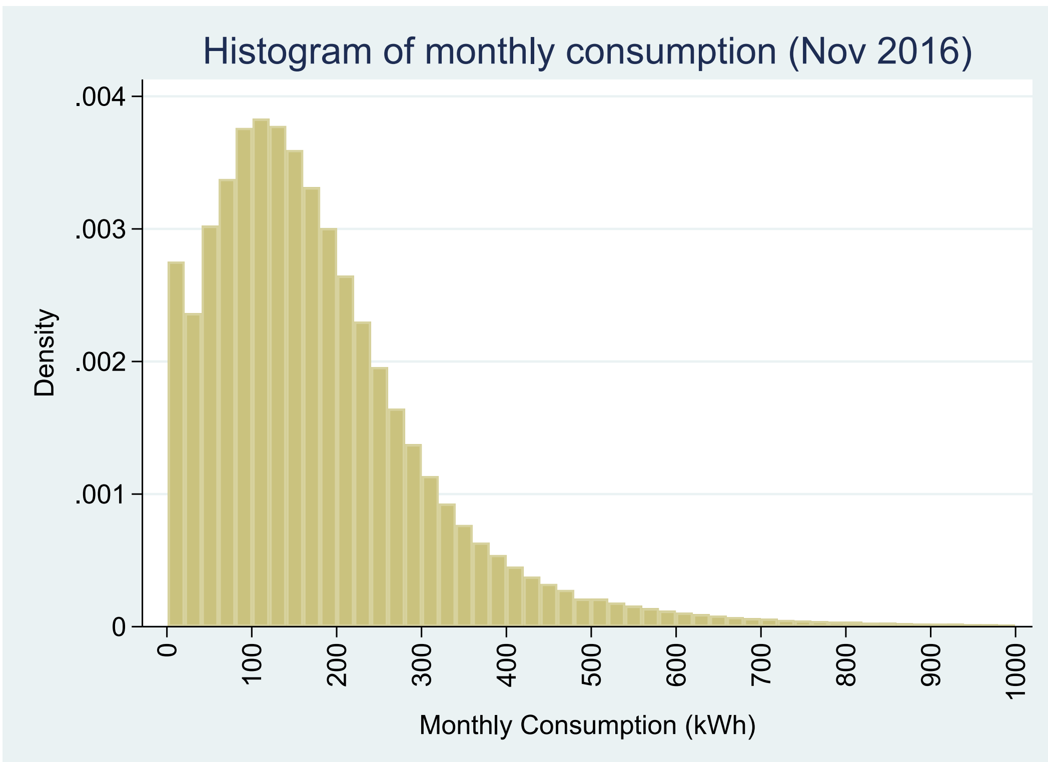
2.A Appendix

Figure 2.1: Electricity Real Prices (in January 2003 BRL)



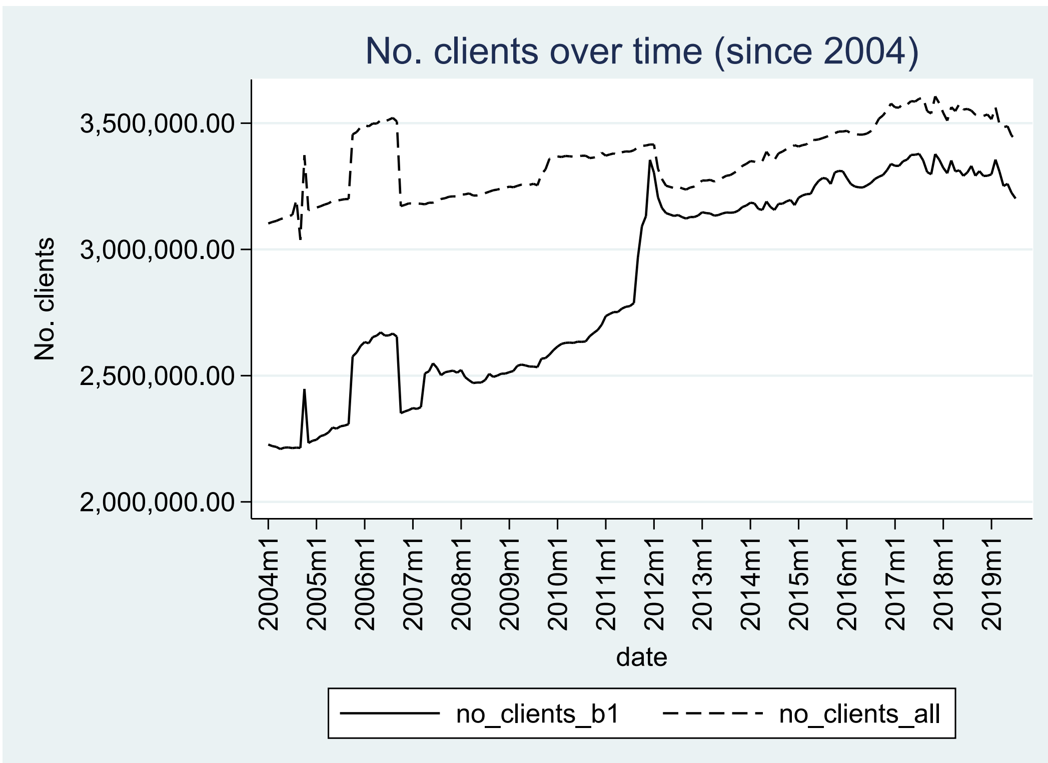
Note: This figure shows electricity real ex– and post-tax average residential retail prices for the period January 2003 to July 2019. All prices were converted to January 2003 R$ using the Extended National Consumer Price Index (IPCA). Average ex-tax prices are calculated by dividing the total revenue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

Figure 2.2: Histogram of Monthly Household Consumption (KWh)



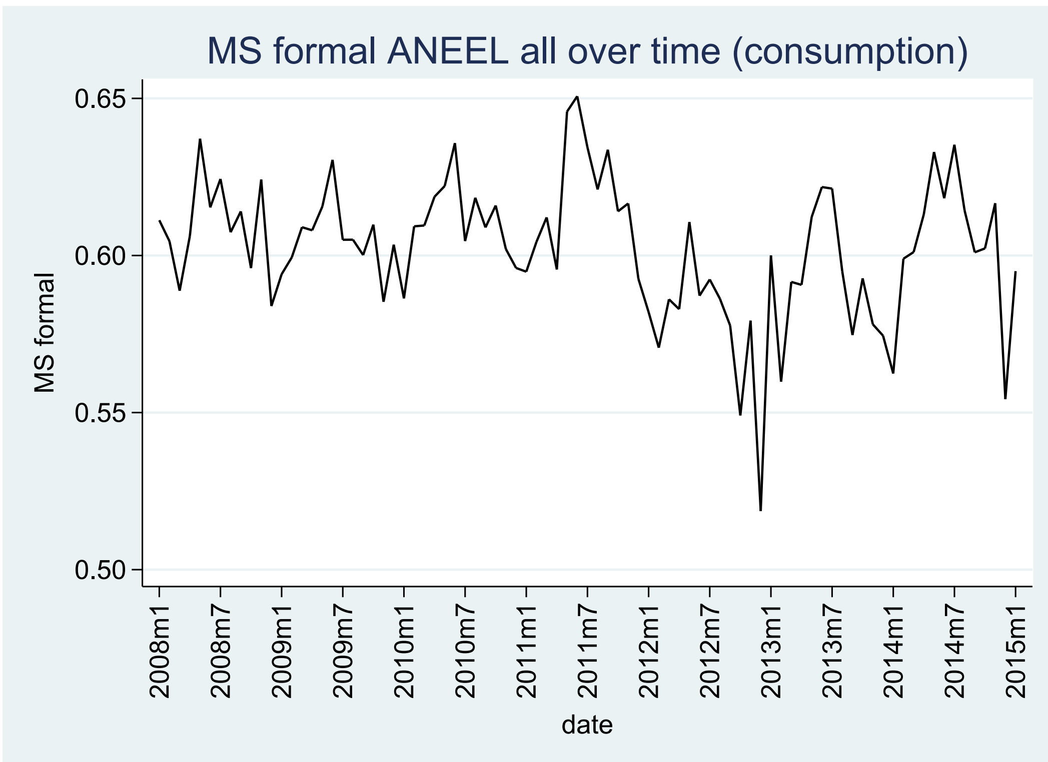
Note: This histogram was constructed using a cross-section of electricity consumption for the universe of residential clients for the month of November of 2016. We dropped clients with zero consumption, which represent approximately 18.4% of the total. We also dropped consumers with over 1,000 KWh/month, which represent less than 0.6% of the sample. Data was provided by the utility.

Figure 2.3: Number of Formal Clients



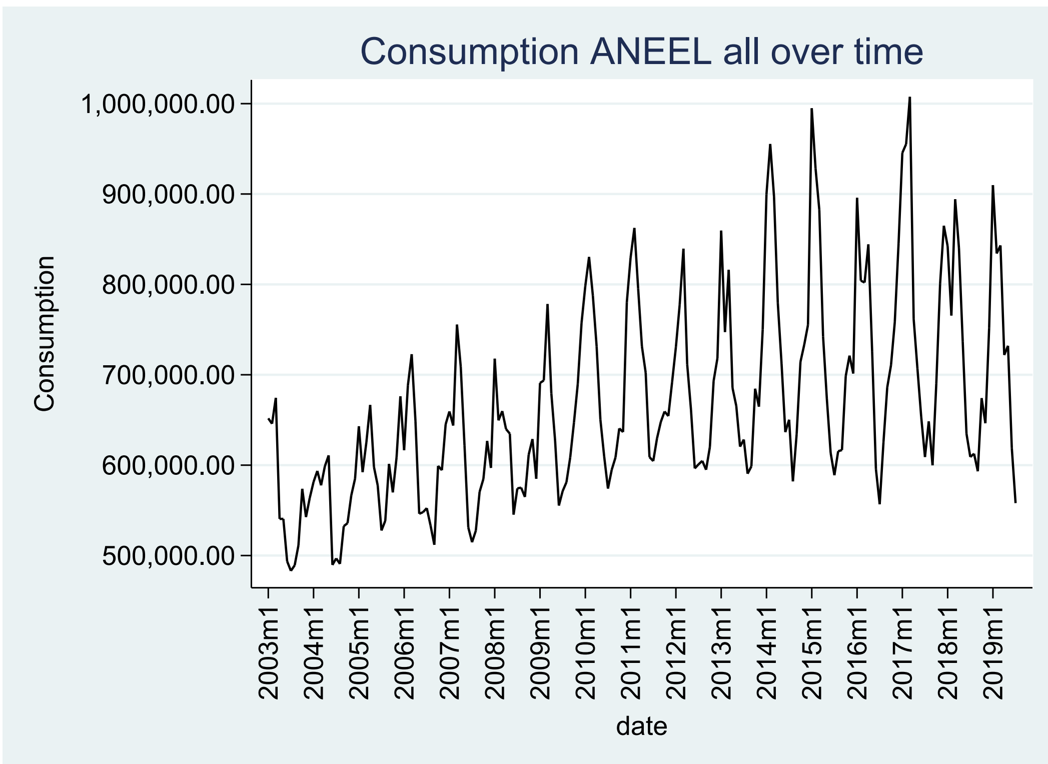
Note: This figure shows the number of the utility's formal residential clients for the period January 2004 to July 2019. The solid line represents B1 clients, who pay the regular electricity tariff schedule. The dashed line represents all the residential clients, which is the sum of B1 clients and those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

Figure 2.4: Market Share of Total Billed Residential Consumption



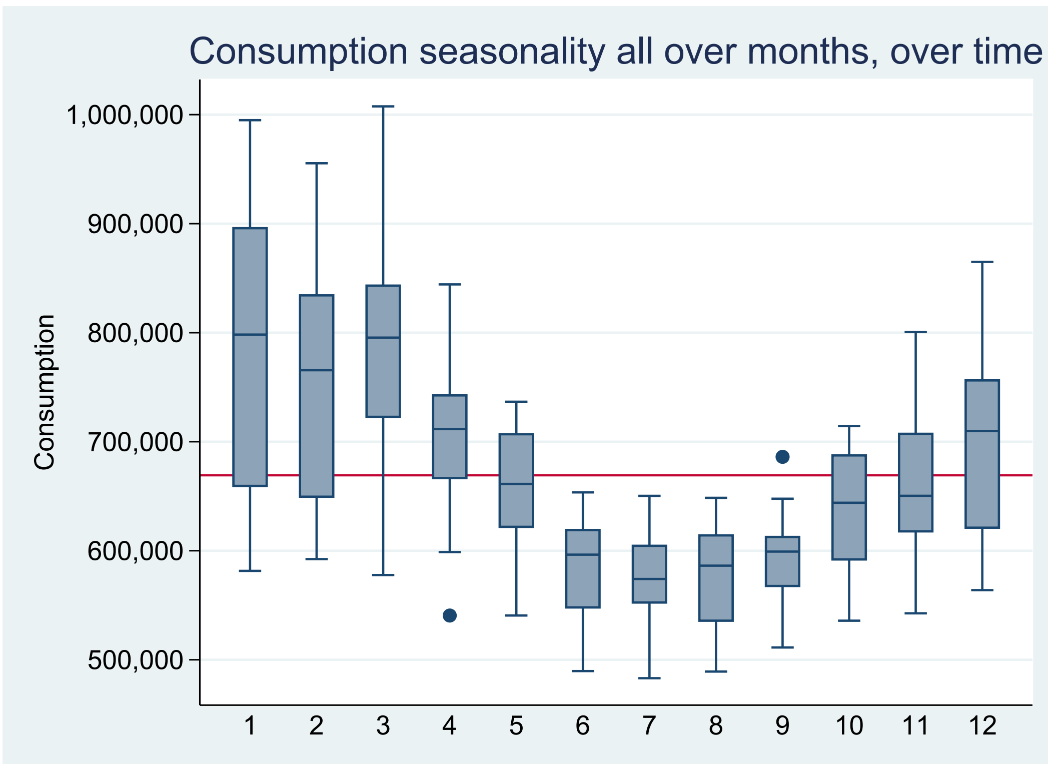
Note: This figure reports the market share of formal residential consumption. The market share variable was defined as the ratio of the billed consumption to the sum of the billed consumption and the amount of non-technical losses. Billed consumption data comes from ANEEL. NTL data comes from the utility.

Figure 2.5: Formal Electricity Consumption (in MWh)



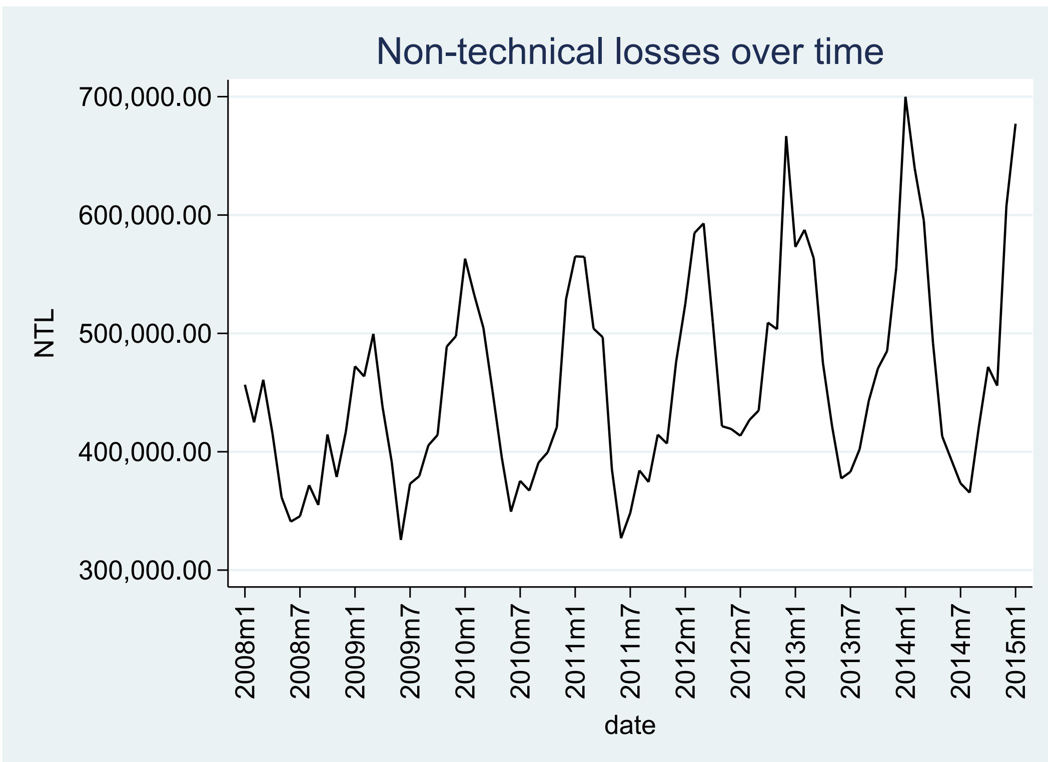
Note: This figure reports the amount of billed electricity consumed by all residential clients for the period January 2003 to July 2019. Consumption is measured in MWh. Data comes from ANEEL.

Figure 2.6: Consumption Seasonality (Box-Plot)



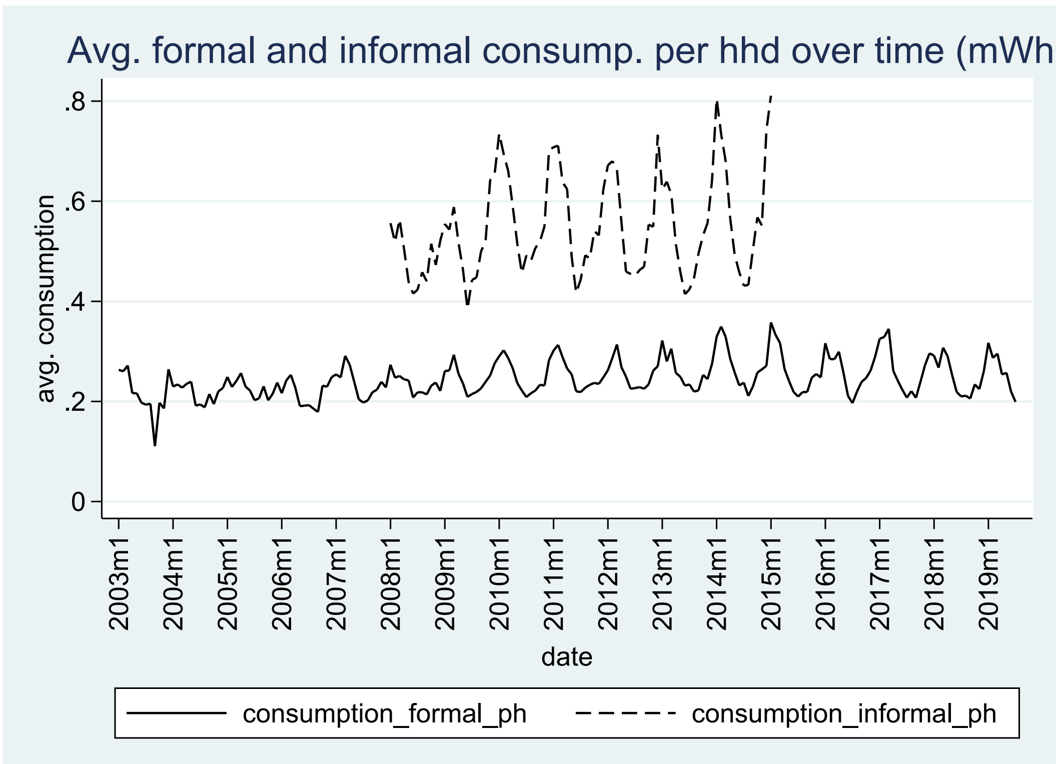
Note: This box-plot was constructed using formal residential electricity consumption (billed electricity) data for the period January 2003 to July 2019. Consumption is measured in MWh. The solid horizontal line represents the average consumption over the whole period, which is of 669,230 MWh. Data comes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

Figure 2.7: Non-Technical Losses (in MWh)



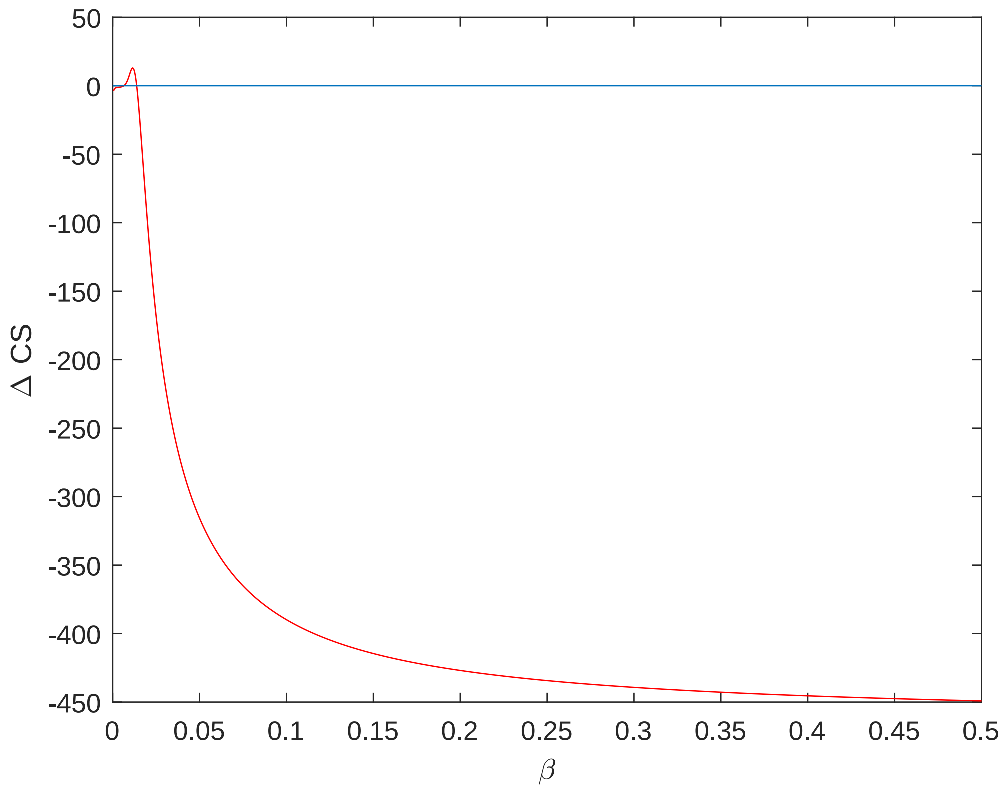
Note: This figure shows the evolution of NTL for the period January 2008 to January 2015. NTL are measured in MWh. Data was provided by the utility.

Figure 2.8: Average Consumption per Household (in MWh/Month)



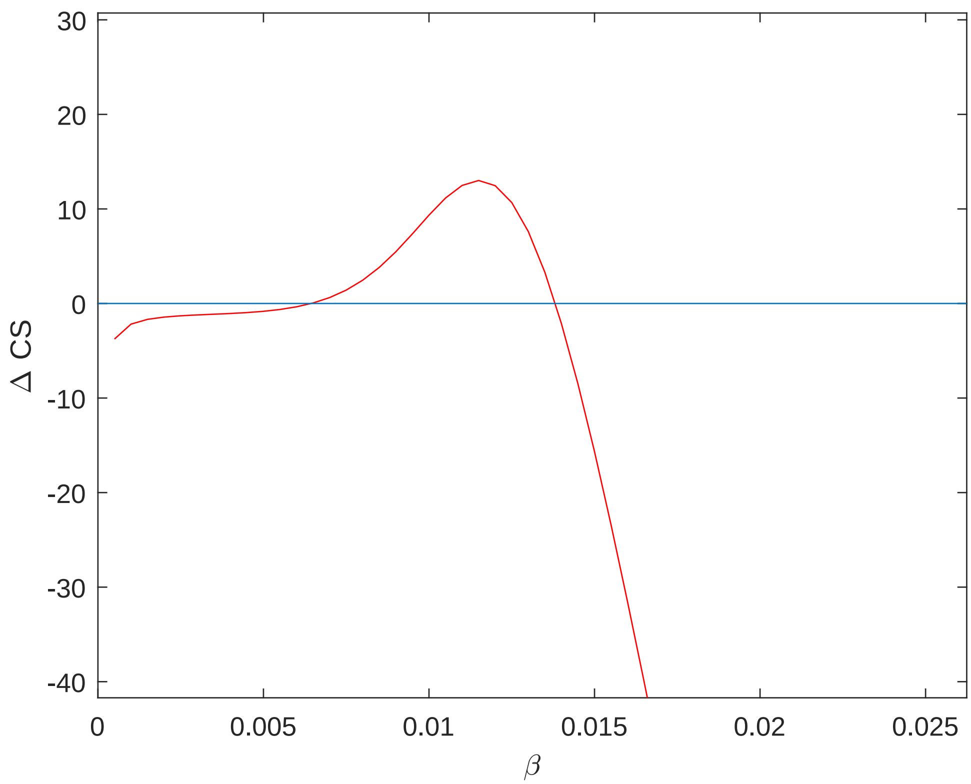
Note: This figure shows the average formal electricity consumption per household for the period January 2003 to July 2019 and average informal electricity consumption per household from January 2008 to January 2015. The average formal consumption per household variable is defined as the ratio of the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

Figure 2.9: ΔCS by β
(Counterfactual 5 — No Theft and Constant Profits)



Note: This figure shows how the average household surplus (ΔCS) changes for different values of the β coefficient. In particular, we consider the fifth counterfactual exercise, in which we exogenously ban theft and set the electricity price such that the utility profit is constant, relative to the baseline. To produce this figure, we first estimate the parameters governing the extensive margin decision. We use month of the year dummies and demographic variables as controls to a logit regression. To avoid collinearity, we omit the dummies associated to the months of January and December. Next, using our estimates, we plot the values of ΔCS associated to the fifth counterfactual exercise, for a discretized grid of 1,001 β values, ranging from 0 to 0.5, *ceteris paribus*. ΔCS is measured in BRL per month.

Figure 2.10: ΔCS by β — Zoom In
(Counterfactual 5 — No Theft and Constant Profits)



Note: This figure shows how the average household surplus (ΔCS) changes for different values of the β coefficient, zooming in the region in which ΔCS is positive. In particular, we consider the fifth counterfactual exercise, in which we exogenously ban theft and set the electricity price such that the utility profit is constant, relative to the baseline. To produce this figure, we first estimate the parameters governing the extensive margin decision. We use month of the year dummies and demographic variables as controls to a logit regression. To avoid collinearity, we omit the dummies associated to the months of January and December. Next, using our estimates, we plot the values of ΔCS associated to the fifth counterfactual exercise, for a discretized grid of 1,001 β values, ranging from 0 to 0.5, *ceteris paribus*. ΔCS is measured in BRL per month.

Table 2.1: Summary Statistics — Cross-Section of Consumers (Nov. 2016)

	N	Mean	Std. Dev.
Formal HHD Consumption	3,176,331	194.47	400.38
Clients Not Using the Grid	3,893,201	0.184	0.388

Note: This table reports descriptive statistics at the household level for all the utility formal clients. *Formal HHD Consumption* is the monthly consumption of formal households, in November 2016, conditional on being positive. It is measured in KWh. *Clients Not Using the Grid* is an indicator variable of whether the household consumption was zero in November 2016.

Table 2.2: Time Series Regressions

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

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

Table 2.3: Panel Regressions: with Feeder FE

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

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

Table 2.4: Intensive Margin Results

	Formal Consumption
Price	–0.000258***
	(3.55e–05)
Linear Time Trend	0.0025***
	(2.73e–04)
Constant	Yes
Month FE	Yes
N	198
R^2	0.80

[1] This table reports OLS estimates of Equation 2.10. A unit of observation is a month. The dependent variable is the formal electricity consumption per household, in MWh. *Price* is the electricity average post-tax retail price, with one lag, in January 2003 BRL/MWh. *Linear Time Trend* is a year time trend. Our sample covers the period January 2003 to July 2019. Consumption and price data come from ANEEL. Robust standard errors are in parenthesis. *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Table 2.5: Extensive Margin Results

	Logit	RC
Constant	6.45***	8.98***
	(0.12)	(0.40)
β	0.0005***	0.0006
	(4.95e–05)	(3.71e–04)
σ	—	0.00003
		(0.02)
Income	0.18***	0.26***
	(0.01)	(0.03)
Urban	–1.52***	–2.24***
	(0.04)	(0.17)
Risky	–0.90***	–1.25
	(0.04)	(0.08)
HHD Size	–1.05***	–1.52
	(0.03)	(0.12)
N	18,169	18,169
R^2	0.44	—

[1] This table reports estimates from the second step (extensive margin). A unit of observation is a feeder in a month. The *Logit* column presents OLS estimates of a regression of the logarithm of the ratio of the market-share (in number of people) of formal households to the market-share (in number of people) of informal households, against a constant, a consumer surplus difference variable, $(\psi_{0t} - \psi_{1t})$, and a set of demographic variables. *Income* is the average income, measured in 2017 minimum wages. *Urban* is the share of urban area. *Risky* is the share of risky areas (for instance, slums, which are irregular settlements usually located in poor and violent regions controlled by drug-traffickers and militias). *HHD Size* is the average household size. β is the coefficient associated to the $(\psi_{0t} - \psi_{1t})$ variable. Robust standard errors are in parenthesis. *p<0.1; **p<0.05; ***p<0.01. The *RC* column reports our estimates from the Random Coefficient model. σ is our heterogeneity coefficient. This version of the model is estimated following a minimum-distance procedure, according to the algorithm previously described in the Data section of this paper. Asymptotically efficient standard errors are in parenthesis.

Table 2.6: Counterfactual Results

	Baseline	(1)	(2)	(3)	(4)	(5)
Qty. Formal	264.05	256.04	272.05	264.05	279.97	286.71
Share Formal HHDs	85.11	85.06	85.15	100.00	100.00	100.00
Price	0.31	0.34	0.28	0.31	0.25	0.22
ψ_{0t}	139.60	131.56	147.90	139.60	156.35	163.73
Revenue	69.43	74.00	64.44	81.72	69.43	63.64
Profit	7.63	13.50	1.32	30.15	14.75	7.63
ΔCS	—	–6.84	7.05	–345.71	–328.97	–321.58
ΔCS (Without ε)	—	–8.37	8.63	471.79	488.53	495.92

[1] In this table, we report the results of five different counterfactual scenarios, which are calculated using the parameters estimated in the second step (extensive margin Logit regression). This regression uses the full data sample and includes demographics controls. In Column (1), we exogenously increase the electricity price in 10%. In Column (2), we exogenously decrease the electricity price in 10%. In Columns (3)–(5), we exogenously ban the possibility of theft, (i.e, we set the share of formal households to 100.00). In Column (3), we keep the electricity price constant, relative to the baseline. In Column (4), we exogenously choose the electricity price that maintains the utility revenue constant, relative to the baseline. In Column (5), we exogenously choose the electricity price that maintains the utility profit constant, relative to the baseline. *Qty. Formal* is the electricity consumption of formal households, in KWh. *Share Formal HHDs* is the percentage share of formal households. *Price* is the electricity price, in BRL/KWh. ψ_{0t} is the surplus of formal households generated by electricity consumption only, in BRL. *Revenue* and *Profit* are the utility revenue and profit, respectively, both measured in BRL. ΔCS is the change in consumer surplus, relative to the baseline, measured in BRL. If negative, then consumers are worse off, relative to the baseline. ΔCS (Without ε) is the component of ΔCS ex- the contribution of the idiosyncratic ε logit shocks. It is also measured in BRL. Again, if negative, then consumers are worse off, relative to the baseline. The unit of observation is the average household in a month. For example, in our Baseline scenario, the average formal household consumes 264.05 KWh of electricity a month.

Table 2.7: Counterfactual Results — 1th Income Quintile

	Baseline	(1)	(2)	(3)	(4)	(5)
Qty. Formal	264.03	256.02	272.04	264.03	292.70	304.63
Share Formal HHDs	71.26	71.24	71.28	100.00	100.00	100.00
Price	0.31	0.34	0.28	0.31	0.20	0.15
ψ_{0t}	139.58	131.53	147.88	139.58	170.43	184.23
Revenue	58.19	62.04	53.98	81.72	58.19	46.51
Profit	-13.03	-8.08	-18.34	30.15	0.99	-13.03
ΔCS	—	–5.73	5.91	–2,748.06	–2,717.20	–2,703.41
ΔCS (Without ε)	—	–7.02	7.24	1,744.66	1,775.52	1,789.31

[1] In this table, we report the results of five different counterfactual scenarios, which are calculated using the parameters estimated in the second step (extensive margin Logit regression). This regression only uses data from markets in the first quintile of the income distribution. It also includes demographics controls. In Column (1), we exogenously increase the electricity price in 10%. In Column (2), we exogenously decrease the electricity price in 10%. In Columns (3)–(5), we exogenously ban the possibility of theft, (i.e, we set the share of formal households to 100.00). In Column (3), we keep the electricity price constant, relative to the baseline. In Column (4), we exogenously choose the electricity price that maintains the utility revenue constant, relative to the baseline. In Column (5), we exogenously choose the electricity price that maintains the utility profit constant, relative to the baseline. *Qty. Formal* is the electricity consumption of formal households, in KWh. *Share Formal HHDs* is the percentage share of formal households. *Price* is the electricity price, in BRL/KWh. ψ_{0t} is the surplus of formal households generated by electricity consumption only, in BRL. *Revenue* and *Profit* are the utility revenue and profit, respectively, both measured in BRL. ΔCS is the change in consumer surplus, relative to the baseline, measured in BRL. If negative, then consumers are worse off, relative to the baseline. ΔCS (Without ε) is the component of ΔCS ex- the contribution of the idiosyncratic ε logit shocks. It is also measured in BRL. Again, if negative, then consumers are worse off, relative to the baseline. The unit of observation is the average household in a month. For example, in our Baseline scenario, the average formal household consumes 264.03 KWh of electricity a month.

Table 2.8: Counterfactual Results — 2^{th} Income Quintile

	Baseline	(1)	(2)	(3)	(4)	(5)
Qty. Formal	264.04	256.03	272.05	264.04	284.12	292.58
Share Formal HHDs	80.75	80.71	80.79	100.00	100.00	100.00
Price	0.31	0.34	0.28	0.31	0.23	0.20
ψ_{0t}	139.59	131.55	147.89	139.59	160.88	170.31
Revenue	65.90	70.25	61.15	81.72	65.90	58.31
Profit	1.14	6.73	–4.86	30.15	10.39	1.14
ΔCS	—	-6.49	6.69	–754.99	–733.71	–724.28
ΔCS (Without ε)	—	–8.15	8.41	894.81	916.10	925.53

[1] In this table, we report the results of five different counterfactual scenarios, which are calculated using the parameters estimated in the second step (extensive margin Logit regression). This regression only uses data from markets in the second quintile of the income distribution. It also includes demographics controls. In Column (1), we exogenously increase the electricity price in 10%. In Column (2), we exogenously decrease the electricity price in 10%. In Columns (3)–(5), we exogenously ban the possibility of theft, (i.e, we set the share of formal households to 100.00). In Column (3), we keep the electricity price constant, relative to the baseline. In Column (4), we exogenously choose the electricity price that maintains the utility revenue constant, relative to the baseline. In Column (5), we exogenously choose the electricity price that maintains the utility profit constant, relative to the baseline. *Qty. Formal* is the electricity consumption of formal households, in KWh. *Share Formal HHDs* is the percentage share of formal households. *Price* is the electricity price, in BRL/KWh. ψ_{0t} is the surplus of formal households generated by electricity consumption only, in BRL. *Revenue* and *Profit* are the utility revenue and profit, respectively, both measured in BRL. ΔCS is the change in consumer surplus, relative to the baseline, measured in BRL. If negative, then consumers are worse off, relative to the baseline. ΔCS (Without ε) is the component of ΔCS ex- the contribution of the idiosyncratic ε logit shocks. It is also measured in BRL. Again, if negative, then consumers are worse off, relative to the baseline. The unit of observation is the average household in a month. For example, in our Baseline scenario, the average formal household consumes 264.04 KWh of electricity a month.

Table 2.9: Counterfactual Results — 3th Income Quintile

	Baseline	(1)	(2)	(3)	(4)	(5)
Qty. Formal	264.07	256.07	272.08	264.07	280.44	287.35
Share Formal HHDs	84.67	84.62	84.73	100.00	100.00	100.00
Price	0.31	0.34	0.28	0.31	0.25	0.22
ψ_{0t}	139.63	131.59	147.93	139.63	156.86	164.45
Revenue	69.06	73.6	64.1	81.71	69.06	63.09
Profit	6.95	12.78	0.70	30.15	14.29	6.95
ΔCS	—	–6.80	7.01	–318.32	–301.09	–293.50
ΔCS (Without ε)	—	–8.50	8.77	472.55	489.77	497.37

[1] In this table, we report the results of five different counterfactual scenarios, which are calculated using the parameters estimated in the second step (extensive margin Logit regression). This regression only uses data from markets in the third quintile of the income distribution. It also includes demographics controls. In Column (1), we exogenously increase the electricity price in 10%. In Column (2), we exogenously decrease the electricity price in 10%. In Columns (3)–(5), we exogenously ban the possibility of theft, (i.e, we set the share of formal households to 100.00). In Column (3), we keep the electricity price constant, relative to the baseline. In Column (4), we exogenously choose the electricity price that maintains the utility revenue constant, relative to the baseline. In Column (5), we exogenously choose the electricity price that maintains the utility profit constant, relative to the baseline. *Qty. Formal* is the electricity consumption of formal households, in KWh. *Share Formal HHDs* is the percentage share of formal households. *Price* is the electricity price, in BRL/KWh. ψ_{0t} is the surplus of formal households generated by electricity consumption only, in BRL. *Revenue* and *Profit* are the utility revenue and profit, respectively, both measured in BRL. ΔCS is the change in consumer surplus, relative to the baseline, measured in BRL. If negative, then consumers are worse off, relative to the baseline. ΔCS (Without ε) is the component of ΔCS ex- the contribution of the idiosyncratic ε logit shocks. It is also measured in BRL. Again, if negative, then consumers are worse off, relative to the baseline. The unit of observation is the average household in a month. For example, in our Baseline scenario, the average formal household consumes 264.07 KWh of electricity a month.

Table 2.10: Counterfactual Results — 4th Income Quintile

	Baseline	(1)	(2)	(3)	(4)	(5)
Qty. Formal	264.05	256.04	272.05	264.05	274.48	278.94
Share Formal HHDs	90.54	90.50	90.58	100.00	100.00	100.00
Price	0.31	0.34	0.28	0.31	0.27	0.25
ψ_{0t}	139.60	131.56	147.90	139.6	150.46	155.23
Revenue	73.89	78.77	68.57	81.72	73.89	70.29
Profit	15.81	22.07	9.08	30.15	20.28	15.81
ΔCS	—	–7.27	7.50	–176.74	–165.88	–161.11
ΔCS (Without ε)	—	–8.80	9.07	369.73	380.58	385.35

[1] In this table, we report the results of five different counterfactual scenarios, which are calculated using the parameters estimated in the second step (extensive margin Logit regression). This regression only uses data from markets in the fourth quintile of the income distribution. It also includes demographics controls. In Column (1), we exogenously increase the electricity price in 10%. In Column (2), we exogenously decrease the electricity price in 10%. In Columns (3)–(5), we exogenously ban the possibility of theft, (i.e, we set the share of formal households to 100.00). In Column (3), we keep the electricity price constant, relative to the baseline. In Column (4), we exogenously choose the electricity price that maintains the utility revenue constant, relative to the baseline. In Column (5), we exogenously choose the electricity price that maintains the utility profit constant, relative to the baseline. *Qty. Formal* is the electricity consumption of formal households, in KWh. *Share Formal HHDs* is the percentage share of formal households. *Price* is the electricity price, in BRL/KWh. ψ_{0t} is the surplus of formal households generated by electricity consumption only, in BRL. *Revenue* and *Profit* are the utility revenue and profit, respectively, both measured in BRL. ΔCS is the change in consumer surplus, relative to the baseline, measured in BRL. If negative, then consumers are worse off, relative to the baseline. ΔCS (Without ε) is the component of ΔCS ex- the contribution of the idiosyncratic ε logit shocks. It is also measured in BRL. Again, if negative, then consumers are worse off, relative to the baseline. The unit of observation is the average household in a month. For example, in our Baseline scenario, the average formal household consumes 264.05 KWh of electricity a month.

Table 2.11: Counterfactual Results — 5th Income Quintile

	Baseline	(1)	(2)	(3)	(4)	(5)
Qty. Formal	264.05	256.05	272.06	264.05	269.87	272.37
Share Formal HHDs	94.89	94.85	94.93	100.00	100.00	100.00
Price	0.31	0.34	0.28	0.31	0.29	0.28
ψ_{0t}	139.61	131.56	147.91	139.61	145.62	148.24
Revenue	77.45	82.56	71.86	81.71	77.45	75.54
Profit	22.34	28.92	15.29	30.15	24.75	22.34
ΔCS	—	–7.63	7.86	–55.68	–49.67	–47.05
ΔCS (Without ε)	—	–8.74	9.01	152.71	158.72	161.34

[1] In this table, we report the results of five different counterfactual scenarios, which are calculated using the parameters estimated in the second step (extensive margin Logit regression). This regression only uses data from markets in the fifth quintile of the income distribution. It also includes demographics controls. In Column (1), we exogenously increase the electricity price in 10%. In Column (2), we exogenously decrease the electricity price in 10%. In Columns (3)–(5), we exogenously ban the possibility of theft, (i.e, we set the share of formal households to 100.00). In Column (3), we keep the electricity price constant, relative to the baseline. In Column (4), we exogenously choose the electricity price that maintains the utility revenue constant, relative to the baseline. In Column (5), we exogenously choose the electricity price that maintains the utility profit constant, relative to the baseline. *Qty. Formal* is the electricity consumption of formal households, in KWh. *Share Formal HHDs* is the percentage share of formal households. *Price* is the electricity price, in BRL/KWh. ψ_{0t} is the surplus of formal households generated by electricity consumption only, in BRL. *Revenue* and *Profit* are the utility revenue and profit, respectively, both measured in BRL. ΔCS is the change in consumer surplus, relative to the baseline, measured in BRL. If negative, then consumers are worse off, relative to the baseline. ΔCS (Without ε) is the component of ΔCS ex- the contribution of the idiosyncratic ε logit shocks. It is also measured in BRL. Again, if negative, then consumers are worse off, relative to the baseline. The unit of observation is the average household in a month. For example, in our Baseline scenario, the average formal household consumes 264.05 KWh of electricity a month.

Chapter 3

Electricity Tariff Flags and Consumer Behavior

3.1 Introduction

Understanding consumer behavior is at the heart of economic theory. Estimates of consumer demand are a powerful resource for market regulators and policymakers from numerous sectors. In the electric sector, the benefits of understanding how consumers respond to prices include: 1) increasing the reliability of power grids; 2) postponing capital-intensive investments in further generation and transmission capacity infrastructures; 3) reducing carbon emissions during peak consumption periods; etc. Papers that have studied how consumers respond to electricity prices include Reiss and White (2005), Ito (2014), Sexton (2015), Deryugina et al. (2020). Another branch of the literature discusses how consumers react to non-price interventions. For example, how an increase in the billing frequency makes them more sensitive to price changes (Jessoe and Rapson, 2014); or how they were found to reduce their electricity usage after receiving reports which described how their consumption compared to their neighbors (Schultz et al., 2007; Allcott, 2011; Allcott and Rogers, 2014; Allcott and Kessler, 2019).

In this paper, I explore the implementation of the Tariff Flags system, a price mechanism introduced in Brazil, starting in 2015, to make electricity consumers internalize changes in the power generation costs faced by utilities. The flags have three distinct colors—green, yellow, and red—that reflect the costs of generating electricity in the short run, and they are associated with an extra charge that is added on to the price paid by households for every kilowatt of electricity consumed. Before 2015, these extra generation costs were paid by utilities beforehand, and were only passed on to consumers once a year, at specific

price adjustment dates. The new system made these costs salient to consumers on a monthly-basis.

In particular, I build an econometric model of demand for electricity to answer the following questions: do consumers respond to the electricity price and to the flag mechanism? If yes, do they respond more to price or flag changes? If no, why don't they? These questions are related to a literature that investigates consumers behavioral responses towards non-salient product attributes. For example, consumers have been found to underreact to: taxes not included in posted prices in grocery shopping (Taubinsky and Rees-Jones, 2017), and alcohol purchases (Chetty et al., 2009); add-on shipping and handling costs in Ebay auctions (Hossain and Morgan, 2006; Blake et al., 2018); road tolls (Finkelstein, 2009) and electricity utility bills (Sexton, 2015) after the introduction of electronic/automatic payment systems; repetition fees in the market for driving lessons (Seim et al., 2017); the amount of calories consumed in fast-food meals (Bollinger et al., 2011); among others.

I conduct my analysis using data from the 18 largest utilities in terms of consumption in Brazil, which account for approximately 80% of all the electricity distributed to Brazilian households. The information comes from multiple data sets, including a panel of (i) aggregated billed consumption, (ii) price, (iii) temperature; and (iv) a time series of flags, with their colors and incremental charges.

My results are preliminary, but they suggest that consumers do not respond either to the electricity price in the short-run or to flag mechanism. I present four different views on why these results might occur. First, because households could be inattentive to their utility bills and the salience generated (if any) by the flags may not be enough to make the benefits of tracking electricity prices worth the costs of paying attention to them. Second, because there are not many alternatives for the electricity service provided by power utilities, and the ones that exist—for example, solar power—are relatively more expensive. Third, because households might face non-negligible fixed costs for adjusting power consumption in the short run, such as the need to purchase more efficient electric home appliances. Fourth, due to technical concerns with regard to the model I estimate—more precisely, because the electricity price and the flag might be endogenous variables.

The remainder of this paper is organized as follows. Section 3.2 introduces relevant institutional details of the Brazilian electricity market and the flag system. Section 3.3 describes the data used in the analysis. Section 3.4 introduces the empirical model and Section 3.5 presents the estimation results, which I discuss in Section 3.6. Section 3.7 concludes.

3.2 Background

Most of the electricity generated in Brazil comes from renewable sources. In 2019, they represented 83% of the total domestic supply, contrasting with a 26% average share in OECD countries and 22% worldwide (EPE, 2020). Moreover, Brazil heavily relies on hydro power plants. In 2018, nearly two thirds out of the 636.4 TWh national electricity consumption came from hydro sources (EPE, 2019).

The electric sector is regulated by the *Agência Nacional de Energia Elétrica* (ANEEL). There are currently 53 distribution concessionaires ("utilities"). They acquire electricity from multiple power stations through the *Sistema Interligado Nacional* (SIN), a national electric grid with nearly 141 thousand kilometers of transmission lines that accounts for approximately 99% of all the electricity supplied to residential, commercial, and industrial consumers. This large network is coordinated by the *Operador Nacional do Sistema Elétrico* (ONS) under the supervision of ANEEL.

The prices are set by the regulator and follow a nonlinear scheme. Households pay a marginal price given by the sum of a baseline price, a flag increment, and taxes. The taxes are an increasing step function of the consumption relative to a baseline level. In addition, the baseline price is lower for social clients, which are low-income households who qualify for price discounts ranging from 0 to 65%.

Given the relevance of hydro power to the Brazilian energy supply profile, the volume and frequency of rains are major determinants of electricity acquisition costs. When rains are scarce and water reservoirs low, the stability of supply becomes more dependent on thermoeletric power plants, which are more expensive to operate. For many years, utilities had to pay for these extra generation costs beforehand. Due to the regulatory framework, the pass-through to consumers could only occur once a year, when tariffs were adjusted. Naturally, this rule had downsides. First, households had no financial incentives to reduce consumption and conserve energy when power supply constraints could lead to blackouts. Second, utilities were frequently required to take out loans in order to mitigate cash flow shortfalls.

The Tariff Flags system was designed to address these issues. Every month, the flag can assume three different colors: green, yellow, or red, in ascending order of costs. These colors reflect the costs of generating electricity in the short run, and they are associated with incremental charges that are added up to the marginal price. These extra values are the same across utilities nationwide and client categories—i.e., regular or social. Both the colors and incremental amounts are determined by the regulator. His decision, which includes temperature and rain forecasts, weights the benefits of using hydro power plants

to generate electricity at a low cost against supply risks associated with water shortages in the hydro reservoirs. The system was officially introduced in January 2015 by ANEEL. However, between July 2013 and December 2014, every utility had to display the color of the next month flag in their bills as a simulation exercise. Households were not charged, but the training allowed them to learn about how the new system would work.

3.3 Data

3.3.1 Data Sets

I draw information from multiple data sets: a panel of i) aggregated billed consumption, ii) prices, and iii) temperatures; a time series of iv) flags. I use data from the 18 largest utilities in terms of consumption, which account for nearly 80% of all the electricity distributed to Brazilian households. Moreover, I restrict my analysis to regular households—i.e., those who do not benefit from the social tariff. I choose to do so because the consumption and price data for the social clients presented pathological patterns which are unlikely to hold in reality. This section describes each one of these data sets and explains how key variables are defined. I also present some descriptive statistics and discuss some of the patterns observed in the data.

I obtain data on electricity consumption from ANEEL. This information is available at the utility-month level, from January 2015—when the flags system was introduced in Brazil— to December 2019. This variable is not robust to changes in the average household size over time, given the aggregated structure of the data. I handle this issue by further using information on the number of (regular) households served by each utility and month, in order to build an average consumption variable, measured in kilowatts per hour (KWh). This additional information also comes from ANEEL, and covers the period January 2015–December 2019.

To these data I add a baseline prices data set. Whenever a price is readjusted, ANEEL releases a public report with information on the new baseline price. The rationale used to perform its calculation is also revealed to the public. I collect these documents to build a novel electricity baseline price data set. In Section 3.2, I explain that the marginal price that households pay is given by the sum of a baseline price, a flag increment, and taxes. The historical information on flags also comes from ANEEL. This data set displays the incremental charges faced by consumers each month in their utility bills, since January 2015. The tax component is more challenging to work with, though, given the nonlinearity of the price schedule. On the one hand, the tax rates and brackets per utility and month can be

recovered from the same documents used to find the baseline prices. On the other hand, I am not able to create an average price[37]—that would better reflect the price experienced by the average consumer—since I cannot observe how households are distributed within each "step" of the demand curve. While recognizing some of the limitations of this approach, I use the baseline price excluding the flag as the price variable in my model. I express this price in January 2015 Brazilian Reais (BRL), adjusting for inflation using the *Índice de Preços ao Consumidor Amplo* (IPCA), the country's official price index.

Finally, I incorporate temperature information using hourly data from the ERA5 data set, which is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The ERA5 combines vast amounts of historical observations with geophysical models—a procedure called climate reanalysis—to produce global estimates of atmospheric parameters. These observations come from different sources, such as satellites, radars, and weather stations. The data set covers Earth at a 0.25º x 0.25º grid resolution—approximately 30km x 30km. The information is daily updated and is available from the year 1979 to 5 days behind real time. In my model, I use a temperature variable that varies at the state-month level[38], which I build in four steps. First, I use shapefiles from IBGE to obtain the latitude and longitude of every Brazilian municipality centroid. Next, I use yearly population count information—also provided by IBGE—to find the population of these municipalities, for the period 2015–2019. Third, I create an auxiliary temperature variable, at the state-hour level, by taking the population-weighted average of the hourly temperature of every municipality within each state. I conclude by taking the monthly average of this auxiliary variable for each state.

3.3.2 Descriptive Statistics

Table 3.1 reports summary statistics for the key variables in my data set. The average household consumes nearly 170KWh of electricity per month. There is also evidence of seasonality, which is better observed in Figure 3.1. These seasonal patterns are expected, since most states are located in tropical, equatorial, and semi arid climate zones. During the summer—i.e., from the end of December to the end of March—households experience very high temperatures, which are eased by the cost of a spike in the use of air-conditioners, and as a consequence, the consumption of more electricity.

The average baseline price increased by approximately 0.05 BRL a year in nominal

[37] The economic literature suggests that when faced with nonlinear prices, consumers usually respond to the average price (Shin, 1985; Wichman, 2014; Ito, 2014).

[38] This implies that clients from different utilities within the same state are assumed to perceive the same average temperature.

terms since the flags were introduced in 2015, with the exception of the year 2017. Figure 3.2 compares the real and nominal prices evolution over the years. The deflated price surged in the beginning of 2015, to correct for interventions made by prior Brazilian governments that kept electricity prices artificially low. It then remained roughly constant until the middle of 2016, when it started to decline steadily. The V-shaped format observed in the beginning of 2017 reflects a punctual discount, applied to all households' April utility bill, that aimed to compensate for undue charges that occurred during the year of 2016. From there on, the real price experienced a consistent growth path ranging from the second semester of 2017 to nearly the end of 2019.

Figure 3.3 depicts how the flags colors and increments changed since the introduction of the flags system, in January 2015. During the first year of its implementation, the flag color was red in every single month. As suggested by Table 3.1, the average flag increment to the baseline price was 0.05 BRL/KWh, but they showed some variation—ranging from 0.03- to 0.05 BRL/KWh. In January 2016, ANEEL announced that the red flag would split into two levels, 1 and 2, where the former would be associated with a lower increment relative to the latter. The effective date of this red flag decomposition was announced to be February 2016.

3.4 Empirical Model

I estimate the following equation at the utility(i)-month(t) level:

$$\log(C_{it}) = \beta_0 + \beta_1 P_{it} + \beta_2 Flag_t + \beta_3 X_{it} + \gamma_i + \lambda_{month(t)} + \varepsilon_{it} \tag{3.1}$$

Here, C_{it} is the average electricity consumption, in KWh. P_{it} is the electricity baseline price excluding the flag, in January 2015 BRL/KWh. $Flag_t$ is the monthly flag increment, also in January 2015 BRL/KWh. X_{it} is a set of covariates, including cooling degrees[39] and a linear time trend. γ_i and $\lambda_{month(t)}$ are, respectively, utility and month of the year fixed effects. The coefficients of interest are β_1 and β_2.

[39] I define $Cooling_{it} = Max\{Temperature_{it} - 18, 0\}$, where $Temperature_{it}$ is the actual atmospheric temperature. This allows for capturing nonlinear effects of temperature on electricity consumption.

3.5 Estimation Results

I estimate Equation 3.1 by Ordinary Least Squares. The results are presented in Tables 3.2 and 3.4. All specifications include a constant, and robust standard errors are in parenthesis. In both tables, column 1 reports estimates of a specification with no additional controls. In column 2, I add a *Cooling Degrees* variable that captures nonlinear effects of temperature on consumption decisions. In columns 3 and 4, I include utility and month of the year fixed effects, respectively. While the first controls for potential time-invariant unobserved heterogeneity within the area serviced by each utility, the second accounts for seasonal patterns not captured by the *Cooling Degrees* variable—for example, some months could have a lower than average consumption just because they have less business days than others. Finally, in column 5, I add a linear time trend as an attempt to control for potential changes in consumption patterns and the advent of energy-efficient technologies over the years.

In Table 3.2, I find a negative and statistically significant *Price* coefficient along all five specifications. This result suggests that households are sensitive to electricity prices in the short run, which goes in the opposite direction of a large literature on electricity demand estimation that documents an inelastic consumption behavior in the short-term. Moreover, I cannot reject the null hypothesis that the *Flag* coefficient is statistically zero. In other words, I do not find any evidence that consumers respond to changes in prices that arise from the flags mechanism. Furthermore, with the exception of column 2, the *Cooling Degrees* coefficient is positive and significant, as one would expect: Brazil is a tropical country and when temperature goes up, especially during the summer—December to March—households use more air conditioners and thus increase their power consumption.

A natural question that arises is whether the difference between the *Price* and *Flag* coefficients is statistically different from zero. In Table 3.3, I employ a two-sided t-test to address this question. The coefficients I use in this test were pulled from column 5, which is the specification with the greatest number of controls in Table 3.2. The results provide evidence that consumers respond less to a one BRL increase in the flag than in the price of electricity. This finding seems counter-intuitive, because the flags mechanism was designed to increase the price salience and thus make consumers pay more attention to changes in electricity prices. Therefore, although one would expect households to become more responsive to price changes when more informed, this is not what my preliminary analysis suggests.

A conjecture for this finding is that the *Price* variable may be capturing a long-run price trend. The idea is that the *Price* variable, which represents the current month (short run)

price, could be correlated with the price of electricity from previous months. Therefore, it is as if there were omitted variables in the regression which bias the coefficient downwards. To investigate this hypothesis, I build a new price variable, *Price Trend (EMA)*, defined as the exponential moving average of the prices from the twelve last months, including the current one. The rationale is that consumers are expected to respond to a medium to long run average price, but recent prices are given more weight. The *Price Trend (EMA)* should thus be interpreted as an "average" long run price, similar to the idea of including lags of the *Price* variable as additional controls in the regression.

In Table 3.4, I report the results of a regression very similar to that in Table 3.2, but with the inclusion of the *Price Trend (EMA)* variable as a control. The idea is that after controlling for this long run price trend, the *Price* coefficient yields the short run response to a one BRL increase in the price of electricity. Once again, I focus on column 5, which is the most complete specification. Now, the *Price* and *Flag* coefficients are statistically zero. The former indicates that households do not respond to the electricity price in the short run. The latter suggests that there is no evidence of a salience effect arising from the flags mechanism. In contrast, the *Price Trend (EMA)* coefficient is negative and significant, as expected.

3.6 Discussion

My results suggest that consumers' responses to the flags are limited. The flags are a short run price mechanism, but I find no evidence that consumers respond to electricity prices in the short run. In this section I discuss some potential explanations for these findings.

The first possible reason connects with a literature on price salience. The main idea is that households are inattentive to their utility bills and need time to learn that prices have changed. In this case, if the salience generated by the flags are not high enough to increase the benefits of tracking the prices relative to the costs of paying attention to them, then the implementation of the flags system did not accomplish its goal of making consumers internalize the seasonality in power generation costs.

Another possibility is that there are not many substitutes for the electricity service provided by power utilities, so households cannot do much when prices go up. In recent years, solar power has become more affordable and accessible, with the development of cheaper and more efficient solar panels. However, buying and installing these panels are still expensive relative to the price of purchasing electricity directly from utilities.

It could also be argued that even with a lack of substitutes, consumers are still allowed

to buy new and more efficient electric home appliances—one could think of a new fridge or air conditioner, for instance—but they must bear the costs of purchasing these items. Therefore, a third reason might be that consumers face non-negligible fixed costs for adjusting power consumption in the short run.

A fourth avenue concerns the identification of the coefficients associated with the *Price* and *Flag* variables, which could be endogenous. One could argue that past demand shocks could alter the stock of water in reservoirs and thus play a role in determining current price and flag changes. So if demand shocks are correlated over time, the price and flag exogeneity assumptions would be threatened. In fact, as I discuss in Section 3.2, the regulator set the electricity price and the flag color using temperature and rain forecasts, as well as the current level of water in the hydro reservoirs. Therefore, I cannot rule out the possibility of endogeneity and given that I use OLS to estimate the model, both *Price* and *Flag* coefficients could be biased towards zero. In the current version of this paper I abstract from this possibility, which I plan to address in future work. One way to address this issue is to use instrumental variables. If both variables are endogenous, then I would need at least two instruments. One instrument could be the marginal cost of electricity generation, which is revealed every week by ONS. Another instrument could be the hydro reservoir levels or inflows, which are also provided by ONS but on an hourly basis.

3.7 Conclusion

This paper provides a preliminary approach to examining the implementation of the Tariff Flags system in Brazil, starting in 2015. It consists of a 3-tier color code in the electricity bill, which is associated with a price adjustment for every KWh of power consumption. This mechanism was designed to make consumers internalize changes in power generation costs faced by utilities.

Understanding if the system achieved its goal is a key resource for Brazilian regulators, policymakers and utilities, as they can improve the electric grid stability, reduce (massive) unnecessary investments in grid infrastructure, decrease pollution from carbon emissions, among others.

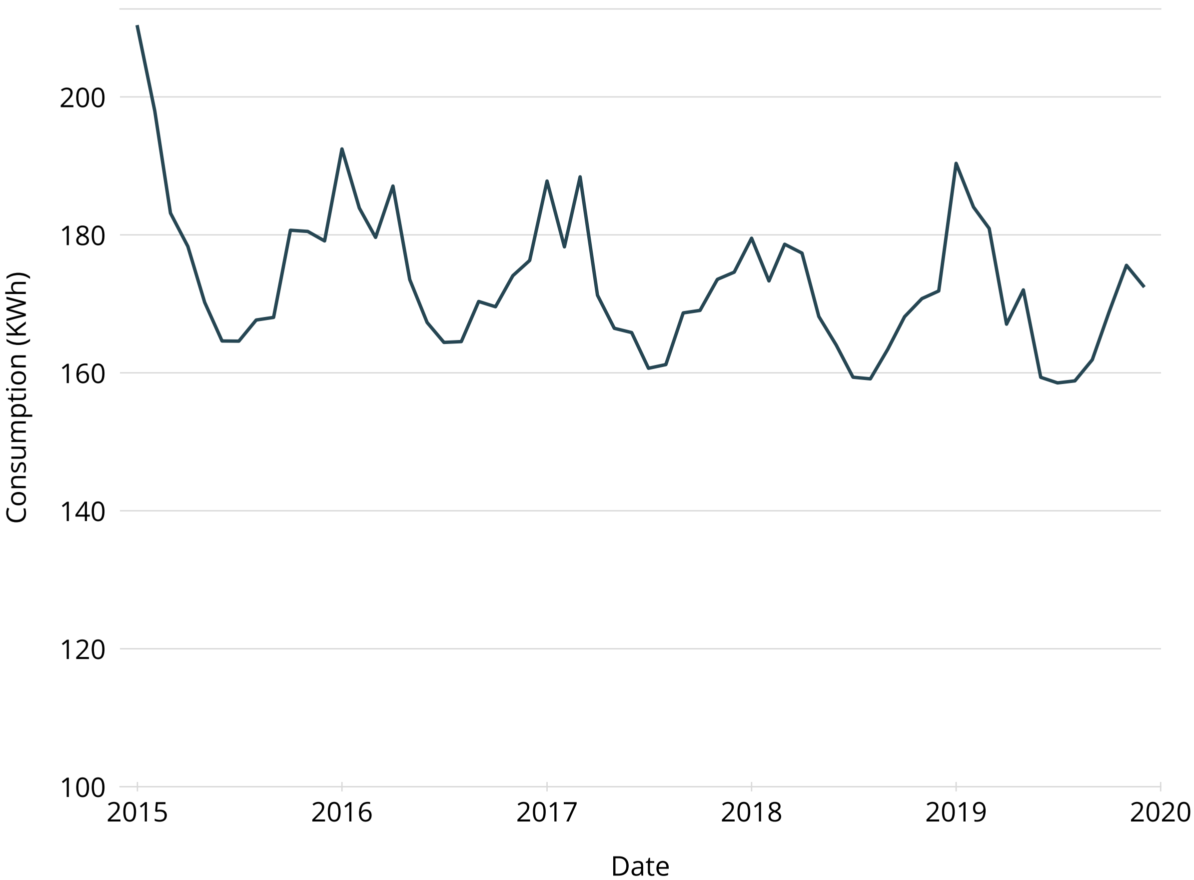
My results suggest that consumers' response to the flags is limited. The flags are a short run price mechanism, but I find no evidence that households respond to electricity prices in the short run—although they seem to respond to an "average" long run price.

I present four different potential explanations for these findings: (1) the salience generated by the flags are not sufficient to make households pay attention to changes in the prices of electricity; (2) the lack of alternatives for the electricity provided by power utilities;

(3) the existence of non-negligible fixed costs for adjusting energy consumption in the short run; and (4) the endogeneity concerning both the electricity price and flag variables, which is not being addressed.

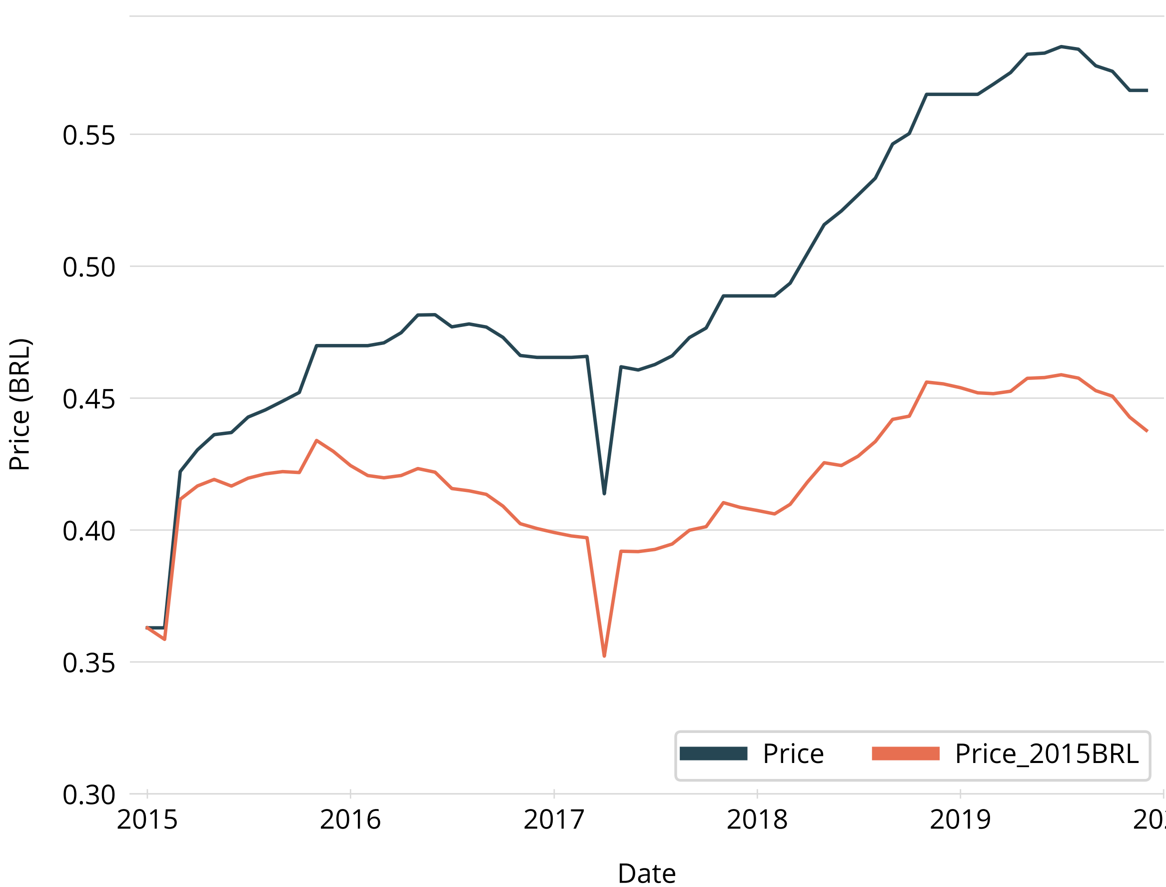
3.A Appendix

Figure 3.1: Average Consumption/HHD



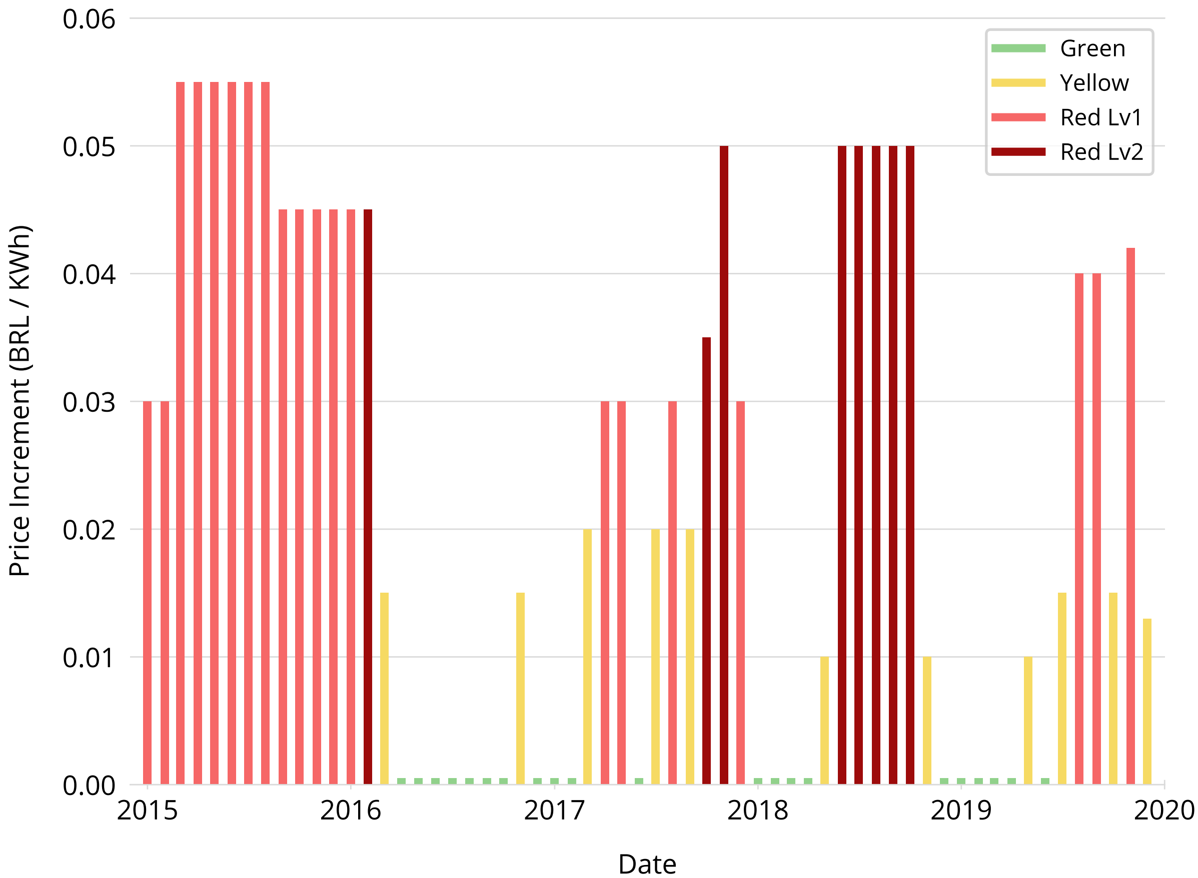
Note: This figure shows the monthly evolution of the average electricity consumption over the years. My sample covers the period January 2015 to December 2019. I use data from the 18 largest utilities in Brazil in terms of consumption, which account for nearly for nearly 80% of all the electricity distributed to Brazilian households. The analysis is restricted to regular households. Consumption and number of households data come from ANEEL. Consumption is measured in KWh.

Figure 3.2: Average Baseline Price



Note: This figure shows the monthly evolution of the nominal (*Price*) and real (*Price_2015BRL*) average baseline prices over the years. My sample covers the period January 2015 to December 2019. Prices data comes from ANEEL. Nominal prices are measured in BRL/KWh. Real prices are measured in January 2015 BRL/KWh.

Figure 3.3: Flags



Note: This figure displays the monthly price increments (in BRL/KWh)—associated with the tariff flags—that are added on to the electricity baseline price. Flags data comes from ANEEL.

Table 3.1: Descriptive Statistics

	2015		2016		2017		2018		2019	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Avg. Consumption	178.75	40.56	175.26	35.61	172.15	35.80	169.47	33.25	170.84	35.11
Baseline Price	0.43	0.06	0.47	0.04	0.47	0.06	0.52	0.06	0.57	0.06
Flag Increment	0.05	0.01	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.02
Utilities per State	1.38	1.12	1.38	1.12	1.38	1.12	1.38	1.12	1.38	1.12
Temperature	25.89	1.25	25.35	1.73	25.54	1.47	25.06	1.63	25.97	0.95

[1] In this table, a unit of analysis is an utility in a month. My sample covers the period January 2015 to December 2019. *Avg. Consumption* is the average electricity consumption, measured in KWh. *Price* is the baseline price, in BRL/KWh. *Flag Increment* is the flag price increment, in BRL/KWh. *Utilities per State* is the number of utilities per state. *Cooling Degrees* is defined as $Cooling = Max\{Temperature - 18, 0\}$, where *Temperature* is the actual atmospheric temperature.

Table 3.2: OLS Regression – Dependent Variable: Log (Average Consumption / HHD)

	(1)	(2)	(3)	(4)	(5)
Price	–0.71***	–0.61***	–0.62***	–0.57***	–0.49***
	(0.12)	(0.13)	(0.06)	(0.06)	(0.06)
Flag	–0.23	–0.30	0.03	0.17	–0.03
	(0.33)	(0.33)	(0.11)	(0.11)	(0.12)
Cooling Degrees	—	–0.01***	0.03***	0.03***	0.03***
		(0.00)	(0.00)	(0.00)	(0.00)
Utility FE	N	N	Y	Y	Y
Month FE	N	N	N	Y	Y
Constant	Y	Y	Y	Y	Y
Linear Time Trend	N	N	N	N	Y
Observations	1,080	1,080	1,080	1,080	1,080
Adjusted R^2	0.03	0.05	0.90	0.90	0.91

[1] This table reports estimates from Equation 3.1. Each column presents the result of a different regression. A unit of observation is a utility in a month. The dependent variable is the logarithm of the average electricity consumption, in KWh. *Price* is the baseline price, in January 2015 BRL/KWh. *Flag* is the flag increment, in January 2015 BRL/KWh. *Cooling Degrees* is defined as $Cooling = Max\{Temperature - 18, 0\}$, where $Temperature$ is the actual atmospheric temperature. My sample covers the period January 2015 to December 2019. Robust standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Table 3.3: Two-Sided T-Test for Difference in Means

Price	Flag	**Diff**
–0.49***	–0.03	–0.46***
(0.06)	(0.12)	(0.14)

[1] *Price* is the baseline price and *Flag* is the flag increment. These coefficients were pulled from column 5 of Table 3.2. *p<0.1; **p<0.05; ***p<0.01.

Table 3.4: OLS Regression – Dependent Variable: Log (Average Consumption / HHD)

	(1)	(2)	(3)	(4)	(5)
Price	–0.42	–0.41	–0.33***	–0.32***	–0.10
	(0.31)	(0.31)	(0.10)	(0.10)	(0.10)
Flag	–0.14	–0.24	0.14	0.26**	0.04
	(0.34)	(0.34)	(0.11)	(0.11)	(0.12)
Price Trend (EMA)	–0.35	–0.24	–0.42***	–0.35***	-0.52***
	(0.34)	(0.35)	(0.12)	(0.12)	(0.12)
Cooling Degrees	—	–0.01***	0.03***	0.03***	0.03***
		(0.00)	(0.00)	(0.00)	(0.00)
Utility FE	N	N	Y	Y	Y
Month FE	N	N	N	Y	Y
Constant	Y	Y	Y	Y	Y
Linear Time Trend	N	N	N	N	Y
Observations	1,080	1,080	1,080	1,080	1,080
Adjusted R^2	0.03	0.05	0.90	0.90	0.91

[1] This table reports estimates from Equation 3.1. Each column presents the result of a different regression. A unit of observation is a utility in a month. The dependent variable is the logarithm of the average electricity consumption, in KWh. *Price Trend (EMA)* is the exponential moving average of the price (i.e., the sum of the baseline price and the flag increment) of the twelve last months (i.e., the current and the 11 previous months), where all prices are measured in January 2015 BRL/KWh. *Price* is the baseline price, in January 2015 BRL/KWh. *Flag* is the flag increment, in January 2015 BRL/KWh. *Cooling Degrees* is defined as $Cooling = Max\{Temperature - 18, 0\}$, where *Temperature* is the actual atmospheric temperature. My sample covers the period January 2015 to December 2019. Robust standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Bibliography

ALLCOTT, H. (2011): "Social norms and energy conservation," *Journal of Public Economics*, 95, 1082–1095.

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.

ALLCOTT, H. AND J. B. KESSLER (2019): "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, 11, 236–76.

ALLCOTT, H. AND T. ROGERS (2014): "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *American Economic Review*, 104, 3003–37.

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

AUFFHAMMER, M. AND E. RUBIN (2018): "Natural Gas Price Elasticities and Optimal Cost Recovery Under Consumer Heterogeneity: Evidence from 300 million natural gas bills," NBER Working Papers 24295, National Bureau of Economic Research, Inc.

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

BEN (2020): "Balanço Energético Nacional," .

BLAKE, T., S. MOSHARY, K. SWEENEY, AND S. TADELIS (2018): "Price Salience and Product Choice," NBER Working Papers 25186, National Bureau of Economic Research, Inc.

BNDES (2020): "Gás para o Desenvolvimento," .

BOLLINGER, B., P. LESLIE, AND A. SORENSEN (2011): "Calorie Posting in Chain Restaurants," *American Economic Journal: Economic Policy*, 3, 91–128.

BORENSTEIN, S. (2012): "The Redistributional Impact of Nonlinear Electricity Pricing," *American Economic Journal: Economic Policy*, 4, 56–90.

BORENSTEIN, S. AND L. W. DAVIS (2012): "The Equity and Efficiency of Two-Part Tariffs in U.S. Natural Gas Markets," *Journal of Law and Economics*, 55, 75–128.

BP (2020): "Statistical Review of World Energy," .

BTS (2020): "Natural Gas Pipeline Profile," .

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

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

CERI (2019): "Distribuição de Gás Natural no Brasil: Dados e Aspectos Regulatórios," .

CHETTY, R., A. LOONEY, AND K. KROFT (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–77.

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

DAVIS, L. W. AND L. KILIAN (2011): "The Allocative Cost of Price Ceilings in the U.S. Residential Market for Natural Gas," *Journal of Political Economy*, 119, 212–241.

DAVIS, L. W. AND E. MUEHLEGGER (2010): "Do Americans consume too little natural gas? An empirical test of marginal cost pricing," *The RAND Journal of Economics*, 41, 791–810.

DERYUGINA, T., A. MACKAY, AND J. REIF (2017): "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation," *NBER Working Paper*.

——— (2020): "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation," *American Economic Journal: Applied Economics*, 12, 86–114.

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

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

——— (2019): "2019 Brazilian Energy Balance - Baseline Year 2018," .

——— (2020): "2020 Statistical Yearbook of Electricity - Baseline Year 2019," .

FINKELSTEIN, A. (2009): "E-ztax: Tax Salience and Tax Rates*," *The Quarterly Journal of Economics*, 124, 969–1010.

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

GERTLER, P. J., O. SHELEF, C. D. WOLFRAM, AND A. FUCHS (2016): "The demand for energy-using assets among the world's rising middle classes," *American Economic Review*, 106, 1366–1401.

HAUSMAN, C. AND R. KELLOGG (2015): "Welfare and Distributional Implications of Shale Gas," *Brookings Papers on Economic Activity*, 71–125.

HAUSMAN, J. (1996): "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, National Bureau of Economic Research, Inc, 207–248.

HOSSAIN, T. AND J. MORGAN (2006): "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay," *The B.E. Journal of Economic Analysis and Policy*, 5, 1–30.

HUNTINGTON, H. G. (2007): "Industrial Natural Gas Consumption in the United States: An Empirical Model for Evaluating Future Trends," *Energy Economics*, 29, 743–759, modeling of Industrial Energy Consumption.

IGU (2020): "Global Gas Report 2020," .

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.

JESSOE, K. AND D. RAPSON (2014): "Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use," *American Economic Review*, 104, 1417–38.

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

LEE, K., E. MIGUEL, AND C. WOLFRAM (2017): "Electrification and Economic Development: A Microeconomic Perspective," .

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

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

MAGNOLFI, L. AND C. RONCORONI (2016): "Political connections and market structure," *Working paper*.

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

MCRAE, S. AND R. MEEKS (2016): "Price perception and electricity demand with nonlinear tariffs," .

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

MIN, B. AND M. GOLDEN (2014): "Electoral cycles in electricity losses in India," *Energy Policy*, 65, 619–625.

NEVO, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

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

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

SCHULTZ, P., J. NOLAN, R. CIALDINI, N. GOLDSTEIN, AND G. V (2007): "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 18, 429–434, pMID: 17576283.

SEIM, K., M. A. VITORINO, AND D. MUIR (2017): "Drip Pricing When Consumers Have Limited Foresight: Evidence from Driving School Fees," *SSRN Electronic Journal*.

SEXTON, S. (2015): "Automatic Bill Payment and Salience Effects: Evidence from Electricity Consumption," *The Review of Economics and Statistics*, 97, 229–241.

SHIN, J.-S. (1985): "Perception of Price When Price Information Is Costly: Evidence from Residential Electricity Demand," *The Review of Economics and Statistics*, 67, 591–598.

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.

SZABÓ, 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.

TAUBINSKY, D. AND A. REES-JONES (2017): "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *The Review of Economic Studies*, 85, 2462–2496.

WICHMAN, C. J. (2014): "Perceived price in residential water demand: Evidence from a natural experiment," *Journal of Economic Behavior & Organization*, 107, 308–323.

WOLFRAM, C., O. SHELEF, AND P. GERTLER (2012): "How will energy demand develop in the developing world?" *Journal of Economic Perspectives*, 26, 119–38.