The Distributional Impacts of Real-Time Pricing

Michael Cahana Natalia Fabra Mar Reguant Jingyuan Wang*

February 15, 2022

PRELIMINARY AND INCOMPLETE.
PLEASE, DO NOT CITE OR CIRCULATE.

Abstract

We study the distributional impacts of real-time pricing (RTP) in the Spanish electricity market, where RTP was rolled out as the default tariff for a large share of residential customers. We complement aggregate patterns of distributional effects with a method to infer individual households' income using zip-code income distributions. We identify three channels for the distributional impacts of RTP: consumption profiles, appliance ownership, and location. The first channel makes the switch from monthly to hourly prices progressive since high-income households consume disproportionately more at peak times when real-time prices are higher. However, the other two channels make the switch from annual to monthly prices regressive: low-income households, who tend to have more electric heating, benefit from the price insurance provided by time-invariant prices during winter, when prices tend to be higher and more volatile. Given that price differences are greater across months than within months, the regressive effect dominates in our application. Using counterfactual experiments, we find that RTP makes low-income households particularly vulnerable to adverse weather shocks during winter. In the future, the wider adoption of enabling technologies (including storage and demand response devices) by the high-income groups might worsen the distributional impacts of RTP.

*Michael Cahana: Energy Policy Institute at the University of Chicago (EPIC), cahana@uchicago.edu. Natalia Fabra: Department of Economics, Universidad Carlos III and CEPR, natalia.fabra@uc3m.es. Mar Reguant: Department of Economics, Northwestern University, CEPR and NBER, mar.reguant@northwestern.edu. Jingyuan Wang: Department of Economics, Northwestern University, mar.reguant@northwestern.edu. The initial work on this project was part of Cahana's senior thesis at Northwestern University. Fabra and Reguant have generously benefited from the support of the BBVA Foundation. Fabra acknowledges support from the European Research Council (Grant Agreement No 772331) and Comunidad de Madrid (MIMA-CM). Reguant acknowledges the support of NSF grant SES-1455084 and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101001732-ENECML).

Contents

List of Tables

List of Figures

1 Introduction

Economists have traditionally favoured the use of dynamic prices in electricity retail tariffs to more accurately reflect changing marginal production costs, which can fluctuate substantially during a day or a week (see Borenstein (2005), Borenstein and Holland (2005), Jessoe and Rapson (2014), Burger et al. (2019), Faruqui et al. (2009), Wolak (2011), Allcott (2011), among others). Dynamic pricing creates incentives for energy conservation during high-priced hours, as well as for load shifting from high-priced to low-priced hours. This improves productive and investment efficiency, mitigates market power (Poletti and Wright, 2020) and might have positive environmental impacts (Holland and Mansur, 2008). The energy transition will strengthen the benefits of dynamic pricing since the intermittency of renewable energy will enlarge the marginal cost and price savings that can be achieved through demand response. The same applies to the deployment of electric vehicles, batteries or other forms of electricity storage for which dynamic pricing will provide more efficient signals.

Despite these well-known benefits, policy makers have traditionally been reluctant to broadly implement dynamic pricing. While industrial households often face some form of dynamic pricing (Blonz, 2018), extending it to residential households is often highly controversial. One potential reason is that the benefits for residential households are expected to be small. Indeed, the magnitude of such benefits critically depends on the elasticity of residential demand to short-run price changes, which most studies report to be limited (in the range 0.0 to -0.1).[1] It is reasonable to expect higher demand elasticities in the longer-run, but not until the technology for price-responsive demand improves, and the broader deployment of renewables and batteries makes it worthwhile for households to invest in automated devices for demand adjustment (Bollinger and Hartmann, 2020; Joskow and Wolfram, 2012).

Another reason for why policy makers have opposed dynamic pricing is the widespread fear that dynamic pricing will have adverse distributional impacts across households (Joskow and Wolfram, 2012). Worldwide, regulators are increasingly concerned about the potential adverse effects that electricity tariff design might have on low income households,[2] particularly so in countries with large shares of energy-poor households.[3] A neat illustration is provided by Levinson and Silva (2021), who show how preferences over income redistribution affect electricity rate design.

We quantify the distributional impacts of dynamic pricing in the Spanish electricity market. We analyze the bill and socio-economic impacts of implementing real-time pricing (RTP) across a large population of households relative to time-invariant prices. We show that uncovering households' income heterogeneity is key to properly estimating the distributional impacts of real-time pricing. For this purpose, we propose a novel method for inferring households unobserved income distribution from zip-code-level data. This method should also prove useful in a variety of contexts beyond electricity markets.

Our analysis relies on hourly smart meter consumption data of over 2 million households in Spain for a period of eighteen months. Having access to the Spanish data is particularly valuable because the Spanish energy regulator decided to make RTP the default option for a vast majority of households. Since no other

[1] In Fabra et al. (2021) we show that the short-run elasticity of Spanish households to changes in real-time prices (RTP) was not significantly different from zero. Borenstein and Holland (2005) show that, even if price elasticities are small, the overall benefits of broadly implementing dynamic pricing -beyond residential households- can be substantial. Harding and Sexton (2017) survey the experimental studies on the effects of dynamic pricing.

[2] Another concern is that RTP will increase bill volatility. However, while RTP certainly affects hourly price volatility, bill volatility is mitigated given that bills are computed over one or two months periods.

[3] In Europe, it is estimated that there are between 50 and 125 million people facing energy poverty (see for instance, the European Commission's energy poverty observatory https://www.energypoverty.eu/). Similar concerns often arise in the US as well (Wang et al., 2021)

country has ever implemented residential RTP as broadly,[4] the Spanish experience is a unique opportunity to test the effects of RTP.

Understanding the distributional consequences of RTP in Spain provides policy relevant insights for other jurisdictions as well. First, it allows regulators to better identify and hence target those who lose from dynamic pricing. By better informing them about the possibilities offered by dynamic pricing, they are more likely to become price responsive and mitigate the bill impacts. Furthermore, addressing the needs of those most impacted would also allow to mitigate social and political resistance to a broader implementation of dynamic pricing.

Whether a switch from time-invariant rates to RTP makes households worse or better off mainly depends on their consumption profile over the day and across the year. Under time-invariant prices, households with low consumption at low-priced hours implicitly cross-subsidize those households with high consumption at high-priced hours.[5] Hence, assuming away potential differences in elasticity and risk aversion across households, the distributional impacts of dynamic pricing critically depend on who consumes when and how that correlates with income. However, there is limited, and often mixed evidence, regarding the correlation between income and consumption patterns. For instance, Faruqui et al. (2010) find that low-income customers have less peaky demand than other customers, while Borenstein (2012b) finds that the load profiles of low-income customers are no flatter or more peaky on average than those of other customers. For this reason, the analysis of the distributional impacts of RTP has to be performed over a large sample of representative households, which is one of the strengths of this paper. Since the system of RTP in Spain has been implemented as the default option,[6] the mass of households under RTP rate is large and representative of the whole population.

We compute the bills of households in our sample under real-time and time-invariant prices. We consider time-invariant prices that are set at the average RTP price of the month or of the year in order to capture different sources of seasonal price variation. Using information on the households' locations, we can match each household's bill impacts with socio-demographic factors to understand who gains and who loses from RTP.

We first follow the standard approach of assigning each household the observed distribution of income at the zip-code level. We find that the policy's impacts are not correlated with income, or just very modestly, with low-income households benefiting on average from the switch to RTP. However, these zip-code-level regressions miss important within-zip-code heterogeneity, potentially biasing the analysis, as also shown by Borenstein (2012b).

To better capture within-zip-code income heterogeneity, we propose a procedure in two steps to assign a probabilistic income distribution to each household. In a first step, we use flexible classification algorithms to assign households to representative types. In addition to these flexible types, we also classify households depending on their amount of contracted power, which is typically correlated with income. In a second step, and once we have classified households into types, we estimate the probability that each of such types has a certain income level. We do so by imposing that the distribution of income based on our household types

[4]For instance, Borenstein (2013) states that "I'm aware of no place in the U.S. that time-sensitive rates are the default for residential customers.". Also, according to the European Commission European Commission (2009), "The case of Spain with a regulated default dynamic price contract is unique ". In some countries, such as Norway or New Zeland, RTP is offered by the competitive retailers but it is not, as far as we are aware of, a default option.

[5]In a recent paper, Hahn and Metcalfe (2021) show that energy subsidies can have important welfare consequences, beyond their distributional impact.

[6]Fowlie et al. (2021) document the relevance of the default effect in a field experiment of residential electricity price choices. They show that a significant fraction of households who were defaulted into RTP remained into RTP, despite being allowed to opt-out.

matches the (observed) probability of income distribution at the zip-code level using a generalized method of moments (GMM). This provides us with a probability of certain types having a particular level of income.

The key assumption to identify the distribution of households' income is that there are shared unobserved types across zip-codes, which in turn have a fixed distribution of income. This allows us to combine the distribution of income across a set of zip-codes to estimate the probability distribution of the discrete types that rationalizes the observed data. Once the income distribution of types has been estimated, one can obtain the household-level distribution by combining it with the probability that a given household belongs to a certain type. Because the classification algorithm in the first step is a choice of the researcher, we show the validity of our approach using a Monte Carlo and we provide sensitivity analysis to researchers' choices both in the Monte Carlo and in our application.

We find that inferring households' unobserved income distribution is important to uncover the distributional impacts of real-time pricing. In fact, it can reverse the predictions made with aggregate income distributions at the zip-code level. We find that accounting for income heterogeneity within zip-codes shows that real-time pricing is slightly regressive as compared to an annual flat price. In turn, this effect can be decomposed in two effects with opposite distributional implications: while the switch from annual to monthly prices is regressive as low-income households lose the price insurance during winter, the switch from monthly to hourly prices is progressive as high-income households consume disproportionately more at peak times within the day/month. The former effect dominates given that price differences are wider across months than across hours of the day or month in our sample.

We also explore the main channels that explain these findings: appliance ownership and households' locations. Electric heating and AC, which account for almost 30% of an average household's annual consumption, vary widely across regions depending on their average weather conditions as well as on the availability of gas infrastructure. Furthermore, electric heating and AC are negatively and positively correlated with income, respectively. Since electricity prices in Spain are significantly higher during winter and lower during summer, the use of electric heating by the low income households and the use of AC by the high income households explain the adverse distributional implications of exposing households to the monthly price variation.

In any event, the overall effects remain economically small, suggesting that real-time pricing does not lead to concerning levels of redistribution across income groups, at least given the relatively small price volatility of the Spanish electricity market during our sample period. However, patterns could change going forward if there is an increasing incidence of extreme price events across the year and/or greater price volatility within the day. Furthermore, if high income households are the most able to respond to price spikes and to benefit from price volatility, e.g., by investing in batteries, solar panels or electric vehicles, the magnitude of the regressive impacts of the policy could be enlarged.

1.1 Related Literature

There is an increasing policy and academic interest regarding the distributional impacts of electricity tariff design. A hotly debated issue is whether the fixed costs of electricity supply should be recovered through fixed fees of through volumetric charges. For instance, Burger et al. (2019) analyze the distributional impacts of moving towards two-part tariffs in which the fixed costs of electricity supply are recovered through the fixed fee instead of volumetric charges.[7] They find that this would hurt low income households more, but argue that two-part tariffs can be designed so as to mitigate such adverse impacts while preserving most of the efficiency gains. Borenstein (2012a), Borenstein (2013), and more recently Brolinson (2019), have also

[7] See also Borenstein (2012b), and Borenstein and Davis (2012), among other.

analyzed the distributive implications of increasing block pricing, which is often used in order to promote energy conservation. Even though these issues are related with the distributional impacts of electricity tariff design, the questions addressed are distinct from the one analyzed in this paper, which concerns the shift from time-invariant to dynamic prices.

Most of the studies that have analyzed the distributional impacts of dynamic pricing have focused on the effects of Critical Peak Pricing (CPP), probably the most commonly used form of dynamic pricing. CPP combines standard fixed rates (or TOU) during most part of the year, with occasional price increases (e.g. 10-15 over a year) when the supply/demand margin is particularly tight. Borenstein (2012b) shows that CPP would have a modest impact on most residential bills, with low-consumption households seeing their bills decline, high-consumption households seeing their bills rise, and low-income households seeing almost no change in their electricity bills. Instead, Faruqui et al. (2010) find that low-income households benefit from CPP both because they tend to have flatter household profiles and because they tend to be more responsive to dynamic prices as well. The evidence reported in Faruqui et al. (2010) comes from pilot programs with voluntary participation of a small, potentially unrepresentative, set of households. An advantage of our analysis is that it relies on real data of a broad population of users who were defaulted into RTP.

Beyond their differences in terms of efficiency impacts, the distributional effects of CPP and RTP can also be quite different. First, the distributional impact of a switch from time-invariant rates to CPP is only limited to differences in consumption during the critical peaks, but has no differential effects across households otherwise, even across households with very different consumption profiles. Furthermore, the distributional impacts of CPP also depend on the household's ability and incentives to adjust its consumption after a price increase. This is less relevant in the case of RTP given that price changes are milder and more frequent, thus reducing the household's ability to avoid the potential adverse impacts of RTP on its bill. Price changes of CPP are also more salient that RTP changes. As households tend to be more aware of price changes under CPP, they are typically better equipped to mitigate such potential adverse effects by reducing their load at critical times.

An exception is Horowitz and Lave (2014), who also assume mandatory RTP, and their results differ from those above. Their analysis uses hourly load data from Commonwealth Edison residential households to determine which households would save money when moved from a time-invariant rate price to RTP. Larger households are found to save money under RTP, while smaller households, and disproportionately, low-income households, are found to lose money under RTP. To the contrary, Burger et al. (2019) find that transitioning towards more time-varying rates tends to make low income households better off. More recently, Leslie et al. (2021) have analyzed the distributional implications that a move to RTP would have in Victoria (Australia). They match substation electricity consumption data with demographic data to identify the characteristics of households who would benefit from RTP. They find that RTP would mostly benefit households in areas with low house prices, high levels of renters and elderly residents.

Another set of papers simulate the distributional impacts of an RTP system with opt-in. Borenstein (2007) addresses this question in an analysis of industrial and commercial households in Southern California. His analysis shows that if households switched into RTP, and exhibited price elasticities of -0.1, their surplus would move in a positive direction, yet a substantial share of them would still be worse off. Only with much higher elasticities would such households be better off under RTP. Horowitz and Lave (2014) find similar results in their analysis of the Northern Illinois residential market. They show that when households do not respond to real-time prices, only 35% of them would save money from the switch. Moreover, even with a presumed elasticity of -0.2, 50% of households in their study would still lose money when switching from time-invariant rates to RTP, underscoring the fact that elasticity is not always a silver bullet to the wealth transfers posed by an RTP system.

The structure of our paper is as follows. Section 2 provides a framework to assess the effects of RTP, describes the background of the Spanish RTP system, and provides an overview of the data. Section 3 describes the methodology used to infer households' income. Section 4 details the results of our analysis, and Section 5 explores the channels. Section 6 performs counterfactual analysis and Section 7 concludes.

2 Background and Data

2.1 Dynamic Pricing in the Spanish Electricity Market

In 2015, the Spanish regulator decided to make real-time pricing (RTP) the default option for all households that had previously not switched away from their default provider.[8] This means that, instead of paying a traditionally flat retail price, most residential households were defaulted into a retail tariff that varies hourly, according to the changes in wholesale electricity prices. Households that had previously switched away from their default provider were given the choice to opt-in into RTP, while households who were defaulted into RTP were given the choice to opt-out to a competitive retail supplier, most of which offer time-invariant tariffs. Given the high inertia in retail choice (Fowlie et al., 2021; Hortaçsu et al., 2017), the fact that RTP was introduced as the default option (with the possibility to opt in and out) implied that it affected a large fraction of the residential sector.

The default Spanish electricity tariffs are comprised of two components: the price of electricity in the wholesale market that varies on an hourly basis, and a regulated access charge that covers other system costs (such as the costs of transmission and distribution, among others). Since the wholesale electricity market operates at the national level, all Spanish residential electricity households, regardless of their geographic location, face the same price tariffs. The hourly prices are published one day-ahead at the System Operator's web page. Regarding the access charge, households are defaulted into a time-invariant rate. However, they can opt out and choose a Time-of-Use tariff for the access charge, which essentially implies lower charges at night and during weekends. In our sample, a fair amount of households are subject to the night-time tariff ($\sim$13.68%).

In order to be subject to RTP, households must have a smart meter installed. By the end of 2015, almost 12M smart meters had been installed in Spain, of which around 10.19 million were successfully integrated into the information and telecommunication systems of electricity suppliers. By the end of 2018, all residential households in Spain (28.02 million) had a smart meter installed.

Figure 1 shows the extent to which real-time prices moved over time during our sample period. There is daily, monthly and seasonal variation in prices due to changes in wholesale demand and supply conditions which are immediately transmitted to the retail prices. This is unlike other types of retail tariffs that also fluctuate over time but for which the pass-through rate tends to be lower. One would expect households to be able to shift their consumption across the hours of they day, but not much beyond that. Within the day, the peak vs. non-peak price differences are modest, i.e., peak prices exceed non-peak prices by approximately 30% (excluding the first part of the sample period, when price differences reached 80%). This within day price variation is much smaller than the one analyzed in the experimental literature, in which peak prices are

[8] Also, the new pricing scheme only applied to households with peak demand below 10KWh. This only excludes the household with very high consumption, which have to contract with a competitive retailer.

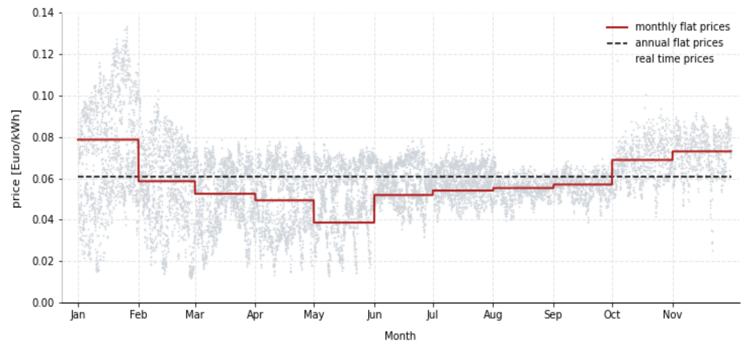


Figure 1: Price fluctuations over time (real-time, monthly and annual prices)

increased by 200-600% (Harding and Sexton, 2017). The variation in monthly prices is more pronounced, with higher prices during the winter months.

2.2 Hourly Electricity Consumption

Our dataset contains information for close to four million Spanish households from January 1st, 2016 to May 31st, 2017. It was provided to us by Naturgy, which is one of the largest Spanish utility companies. Its households tend to reside most densely in Madrid, although they are also scattered throughout Spain.[9] After treating for outliers with overly zero consumption observations or missing zip-code data,[10] as well as households outside of the regulated territories of the utility ,[11] the sample size is reduced to 1,294,357 households, covering 750 zip-code regions. We further drop December 2016 and May 2017 observations for data quality reasons, which leaves 15 months in our sample period (January 2016 to November 2016, and January 2017 to April 2017). We thus have 18,079,824 household-month pairs in total.

The dataset also includes hourly consumption information (in kWh) for each household served by the utility, leading to more than 13 billion data-points of hourly consumption data. The dataset also specifies the type of tariff each household has; the hourly prices corresponding to the tariff identifier (in €/kWh); each household's contracted capacity; and its postal code information.

2.3 Annual Bills under RTP and Time-Invariant Prices

We compare households' electricity bills under RTP versus time-invariant prices. In order to do so, we construct a revenue-neutral alternative to dynamic pricing, assuming a zero price elasticity.[12]

[9]The geographic distribution of households is shown in Figure E.1.

[10]The algorithm for cleaning outliers drops a household from the sample if more than 25% of its consumption observations are zero, or if more than 5% are null.

[11]The default geographic provider is the one in charge of offering the default RTP tariff. Hence, households outside of the utility's regional regulated territory can never be part of the RTP scheme.

[12]Assuming that households do not change their consumption depending on their tariff might bias the distributional impacts if households have a significant demand response to short-run price changes, and if such elasticities vary across income groups. We analyze this issue in Section 6.2.

For each individual household i and each month m, we compute the following expressions:[13]

$$Bill_i^{RTP} := \sum_{hdm} p_{hdm} \cdot kWh_{i,hdm} \tag{1}$$

$$\overline{Bill}_i := \overline{p} \sum_{hdm} kWh_{i,hdm} \tag{2}$$

$$\overline{p} := \frac{\sum_{hdm} p_{hdm} \cdot (\sum_i kWh_{i,hdm})}{\sum_{hdm} (\sum_i kWh_{i,hdm})} \tag{3}$$

where p_{hdm} is the real time electricity price in hour h of day d in month m and $kWh_{i,hdm}$ is the consumption of individual household i in that hour. Equations (1) and (2) give the bills under RTP and under an annual time-invariant price, respectively, which is defined in equation (3).[14] We will use the difference $Bill_i^{RTP} - \overline{Bill}_i$ to compute the loss (or saving, if the difference is negative) from being at RTP, relative to time-invariant prices.

Last, we compute another measure for the time-invariant prices but at the monthly level, i.e., a constant revenue-neutral price for each month, and the resulting annual bill,

$$\overline{Bill}_i^m := \sum_{hmd} \overline{p}_m \cdot kWh_{i,hdm} \tag{4}$$

$$\overline{p}_m := \frac{\sum_{hd} p_{hdm} \cdot (\sum_i kWh_{i,hdm})}{\sum_{hd} (\sum_i kWh_{i,hdm})}. \tag{5}$$

This allows us to decompose the loss under RTP as the loss across months, $\overline{Bill}_i^m - \overline{Bill}_i$, and the loss within months, $Bill_i^{RTP} - \overline{Bill}_i^m$. In more detail, the across months loss reflects the bill impacts due to the monthly price variation across the year, while the within months loss reflects the bill impacts due to the hourly price variation within the month.

Table 1 reports the summary statistics. The average annual consumption of a household is around 2,647 kWh, for which on average they pay 156 Euro/year.[15] There is heterogeneity in households' energy bills, due to differences in consumption levels as well as the timing of their consumption. A move from time-invariant prices to RTP implies that some households lose (on average, losers face 3.88% higher bills) while other households gain (on average, winners enjoy from a 2.9% bill reduction). The losses due to the across months price variation are highly heterogeneous across households, ranging from a 1.63% bill increase for the 75% percentile to a 2.30% bill reduction for the 25% percentile. The within month effect is smaller and more homogeneous across households.

2.4 Demographic Data

To examine whether demographics give rise to differences in the socio-economic impacts of RTP, we have also collected demographic data from two sources: the Spanish National Institute of Statistics (INE) and a private data provider, MB Research. The former provides demographics at the census district level (population, age,

[13] In practice, electricity bills also include other cost components, such as network charges, which are independent of consumption and/or energy prices. Introducing variable taxes (e.g., VAT) in our analysis would enlarge the magnitude of the distributional implications but would not change sign of the impact (i.e., whether a household loses or gains from RTP).

[14] Since the sample period includes the months of January, February, March, and April for 2006 and 2007, the observations for those months are each weighted by 0.5 in order get a measure of the annual average bills.

[15] Recall that these amounts do not include other cost components, such as network costs or taxes. Depending on the household's contracted power and tariff choice, these additional costs can multiply the household's annual electricity bill by approx. 2.

Table 1: Summary Statistics (household-annual level)

	Mean	Std	25%	50%	75%
kWh_i	2647.01	2230.29	1330.63	2094.15	3220.75
$Bill_i^{RTP}$	155.74	130.68	77.91	123.16	189.84
$\overline{Bill}_i$	155.57	131.08	78.21	123.08	189.30
$\overline{Bill}_i^m$	155.74	131.29	77.78	122.75	189.40
Loss in RTP [%]	-0.01	4.86	-2.42	-0.60	2.01
Loss in RTP (losers) [%]	3.88	4.23	1.08	2.61	5.29
Loss in RTP (winners) [%]	-2.90	2.87	-3.63	-2.13	-1.11
Decomposition:					
Loss within months [%]	0.13	1.97	-0.61	0.15	0.91
Loss across months [%]	-0.14	4.47	-2.30	-0.91	1.63

Notes: This table reports household level statistics. There are 1,294,357 observations. All units are measured in Euro, except for kWh_i which is measured in KWh. Annual bills (annual consumption [kWh]) are 11-month bills (consumption) from January to November, because we do not observe December data. All percentages are computed with the bills under an annual time-invariant price, $\overline{Bill}_i$, in the numerator. By construction, the loss under RTP is 0 when expressed in Euros, but it is different from zero when expressed in %. The reason is that we first compute the Loss for each household and we then take the average across households.

sex, education, dwelling types, and income distribution data), while the latter provides income distribution data at the zip-code-level.[16]

As we know the zip-code of each household, but not its census, we match census districts and postal codes, and then aggregate the census district data at the postal code level. We classify all the zip-codes into 5 categories based on the distribution of income per capita, from low to high. Whereas 90% of households in the poorest zip-codes have an adjusted income below 20,000 Euro/year, around half of the households in the richest zip-codes have an income above that level.

As a first step towards understanding the impacts of RTP and their correlation with income, we regress total average electricity consumption, average peak electricity consumption, and the bill impacts on income, using a cross-sectional sample at the zip-code-level:

$$\ln(kWh)_j = \ln(Median\ Income)_j + HH\ size_j + \phi_j + \epsilon_j, \tag{6}$$

$$\ln(kWh)_j^{peak} = \ln(Median\ Income)_j + HH\ size_j + \phi_j + \epsilon_j, \tag{7}$$

$$\%Loss\ under\ RTP_j = \ln(Median\ Income)_j + HH\ size_j + \phi_j + \epsilon_j, \tag{8}$$

where j indexes the zip-code and ϕ_j are province-age group-income group fixed effects. These regressions measure the correlation between median income per household at the zip-code-level and consumption, consumption at peak times (11am-10pm) and the loss from RTP. Positive (negative) coefficients would reflect that households in higher (lower) income zip-codes consume more in total, consume more at peak times and lose more from RTP.

Table 2 reports the results. Intuitively, column (1) suggests that households' electricity consumption is positively correlated with income, after controlling for household size. Column (2) suggests that peak electricity consumption is also correlated with income. Columns (3) shows a positive correlation between income and losses from RTP, thus suggesting that households in lower income zip-codes are slightly better off under RTP relative to the higher income zip-codes. In all cases, the relationship with income are noisy

[16]The Appendix provides a more detailed description of these data sources.

Table 2: Zip-code monthly-level regression results

	ln(KWh)	ln(KWh peak)	Loss under RTP [%]
ln[IncPerHH]	0.076	0.102	0.325
	(0.055)	(0.064)	(0.439)
HHsize	0.317***	0.329***	-2.576*
	(0.040)	(0.035)	(0.832)
R-squared	0.584	0.696	0.313
N	680	680	680

Notes: All regressions include province-age group-income group fixed effects. $IncPerHH$ stands for median income per household, and $HHsize$ gives the mean number of people at the household.

and statistically insignificant, in part due to limited signal in aggregate zip-code data.

3 Inferring the Household-Level Distribution of Income

The results from the reduced-form analysis face an important limitation as they overlook the existing heterogeneity within zip codes. There can be substantial income heterogeneity across households within the same zip code, which in turn could be correlated with income. Therefore, the aggregate results are likely to underestimate the extent of distributional impacts from the policy.

To get a more precise estimate of who loses and who wins from RTP, we develop a structural methodology to infer a given household's distribution of income. Let us assume that a household's allocation of their hourly electricity consumption during the day (denoted kwh_{ih} and suppressing day index),is determined by a set of variables such as hourly electricity prices and temperature (denoted x_h) and their life style (represented by their type θ_i), and some random shocks ϵ_{ih}.

$$kwh_{ih} = f(x_h, \epsilon_{ih}|\theta_i) \tag{9}$$

Allowing the household's type θ_i to be correlated with their income helps us identify how income affects electricity consumption, and therefore study the distributional impacts of RTP.

The proposed methodology follows in two steps. In the first step, we classify households into different types based on their contracted power capacity,[17] their typical consumption patterns, and HVAC ownership, which we infer from their hourly electricity consumption. Based on these results, we construct the aggregate probabilities of types and income brackets at the zip-code level. In the second step, we assume that each type has a fixed distribution of income, which is unknown. We estimate the probability distribution by exploiting aggregate moments: the implied income distribution from the types within a zip code should match the observed zip-code-level income distribution. These aggregate moments help us identify the income probability for each household type.

More formally, our objective is to uncover the income distribution of discrete household types, $\theta \in \Theta = \{\theta_1, \ldots, \theta_N\}$. To define the income distribution, we partition the income domain into bins $inc_k \in \{1, .., K\}$. We use national income quintiles of the household income distribution, so K=5. Let $\eta_k^n = Pr(inc_k|\theta_n)$ denote this discrete probability of income conditional on household type θ_n. The goal is to estimate η_k^n for each income bin k and type θ_n, which we then apply to each household based on their types to infer their expected unobserved distribution of income.

[17] Contracted power capacity is the maximum power a household can use.

The estimation assumes the income distributions of the same type θ_n from different zip codes are the same, equal to $\eta^{\theta_n} = \{\eta^n_k\}^K_{k=1}$. This assumption would be too strong if we estimated the model combining all the zip codes in Spain. We instead assume that each type's income distribution is the same across zip codes within a group of zip-code regions and estimate a set of $\{\eta^\theta\}_{\theta\in\Theta}$ for each group of zip-code regions separately. We explore several ways in which to aggregate zip-codes. Our main results assume that all zip-codes within a region share the same types, but not across regions.

We explain each step in detail and the results in the following paragraphs.

3.1 Step 1: Identifying Household Types

We define household types based on their consumption and contracted tariff. From households' tariffs, we observe their contracted power capacity, which is the maximum consumption allowed at any point in time and tends to be highly correlated with income, as it determines a significant portion of the fixed charges of the bill. We also identify each households' HVAC ownership status from the correlation of their hourly consumption and temperature across seasons. Additionally, we use the K-means cluster algorithm to classify households into flexible types based on observables or summary statistics of households' hourly consumption patterns.

3.1.1 Classification by contracted power capacity

The electricity price is a two-part tariff, including a term for maximum contracted power and an hourly price that is charged at the margin, i.e., as a function of the total electricity consumed. The contracted power is the maximum allowed hourly consumption for each household and it can vary from 1 to $10kW$. Most households chose 3.3-3.5kW. We classify households into two groups: 51% households who have a contracted power less than or equal to $4kW$ are in the low contracted power group (L), the other 49% households with a contracted power larger than $4kW$ are assigned to the high contracted power group (H).

3.1.2 Classification by appliance ownership

We identify appliance ownership by testing the seasonal correlation of hourly consumption and hourly temperature. We then match the macro moment of appliance ownership rate at regional level. Details on the identification procedure are included in Appendix C. Because our sample covers mostly the north part of Spain where people rarely use air conditioning (AC), we focus more on electric heating (EH) and AC types are not included in the income estimation model.

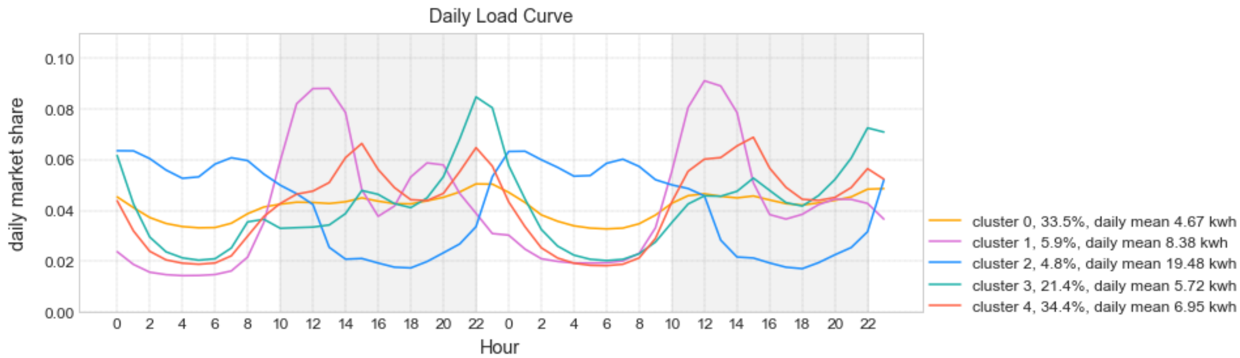
3.1.3 Classification by consumption patterns

We conduct the estimation separately for different zip code groups, as explained above. Thus, we first group zip codes based on province and population in this step. Then we classify households within zip code groups and get zip code group-specific type spaces Θ^g with size N^g.

We use K-means cluster algorithm to classify households based on high-dimensional hourly consumption data. 198 variables are generated to capture daily and seasonal consumption patterns for each household. We then apply K-means cluster algorithm to all households in the same zip code group. The 198 variables include:

- weekday average daily consumption and weekend average daily consumption in kWh.

Figure 2: K-means types in Madrid



- mean and standard deviation of hourly market share for each of the 24 hours by weekday and weekend.
- 4 variables capturing seasonal pattern: the ratio of winter consumption to annual consumption, the ratio of summer consumption to annual consumption, standard deviation of monthly consumption, and correlation of monthly consumption and monthly flat price

The first 2 sets of variables (194 variables) are explanatory variables for within month losses, because these variables capture lifestyle or type of a household and explain the correlation of consumption pattern and real time price. The last set of variables are explanatory variables for cross month losses, as they captures seasonality of electricity consumption.

For all households in the same zip code group, we first classify them into 4 subgroups based on contracted power and electric heating ownership. We then use K-means clustering algorithm to further classify households within each subgroup into θ_n types. We chose the number of K-means types to make sure number of households in each type (and within the zip code group) is greater than 100. That means, if a subgroup contains less than 100 households, we would not further classify households in this subgroup using K-means.

One example results of the K-means classification is Figure 2. This figure shows the average daily consumption pattern of all consumers within non-EH and high contracted power in Madrid. One can see that the algorithm picks up a variety of consumption patterns: people with a relatively flat profile (cluster 0), people who consume in the morning (cluster 1), people who consume mostly at night (cluster 2), people who consume in the evening (cluster 3), and people who consume at lunch time and in the evening (cluster 4).

3.2 Step 2: Identifying the Income Distribution of each Type

From step 1, we get distinct household types θ^i. We only have data for some central and northern regions in Spain, where the impact of AC is limited. Therefore, we focus on electric heating. The type space for zip code group g is $\Theta^g = \{\theta_1^g, \theta_2^g, ...\theta_{N^g}^g\}$, where θ_n^g contains information of a consumer's contracted power $\in \{L, H\}$, EH ownership $\in \{EH, noEH\}$, and his K-means type. In our main specification we set $N^g = 12$ for all zip code groups and have 3 K-means types within each tariff-EH subgroup,[18] and we include robustness test for alternative specifications. We do the same estimation for all zip code groups, thus superscript g is suppressed in all the following paragraphs.

[18] $N^g = 12$ is for zip code groups with sufficient population. As explained above, we make sure number of households within each type is great than 100. When there is too few households in a type, we merge it into other types.

We denote the share of types in zip code j by P^j_θ. To get P^j_θ for each zip-code j, we aggregate from the estimated individual types:

$$P^j_\Theta(\theta_n) = \frac{1}{HH_j} \sum_i \mathbb{1}(\theta^i = \theta_n), \tag{10}$$

where HH_j are the number of households in zip-code j.

Once we have a distribution of types at the zip-code-level, we can uncover the unknown probabilities of types having a certain income by using cross-zip-code restrictions in the share of types. For example, if a zip-code has relatively high income and also a relatively large presence of households with high contracted power, the algorithm will conclude that the likelihood of high income for the high contracted power type is larger.

Assuming the underlying income distribution of a type θ_n across zip codes are the same, we get the following moment conditions by matching the observed and predicted zip-code-level income distribution:

$$\min_\eta \quad \sum_j \omega_j \sum_k (Pr^j_{inc,k} - \sum_{\theta_n \in \Theta} \eta^{\theta_n}_k P^j_\Theta(\theta_n)) \tag{11}$$

$$s.t. \quad \sum_{k=1}^{K} \eta^{\theta_n}_k = 1 \qquad \forall \theta_n \in \Theta. \tag{12}$$

The objective function 11 uses a set of $(K-1)\times$ *Number of zip codes* moments to identify the $(K-1)\times N^g$ unknown probabilities of income η, where K is the number of income bins and N is the number of types (12 in main specification). Thus, we need at least N^g zip codes to identify η. We thus make sure that the number of zip codes in each group g is greater than or equal to N^g.

Using the estimated income for each type $\{\eta^\theta\}_{\theta\in\Theta}$ for each state, we calculate the aggregate distribution of income by contracted power and tariff choice (Figure 3). As expected, Figure 3 shows a positive correlation between income levels and contracted power, with higher (lower) income households being more likely to have high (low) contracted power. Being on the default RTP plan is negatively correlated with income. This is intuitive and can be driven by two mechanisms. First, households who receive a subsidy for energy poverty need to be on the RTP tariff to become eligible. Second, low income households have been estimated to search less in similar contexts (Byrne and Martin, 2021).

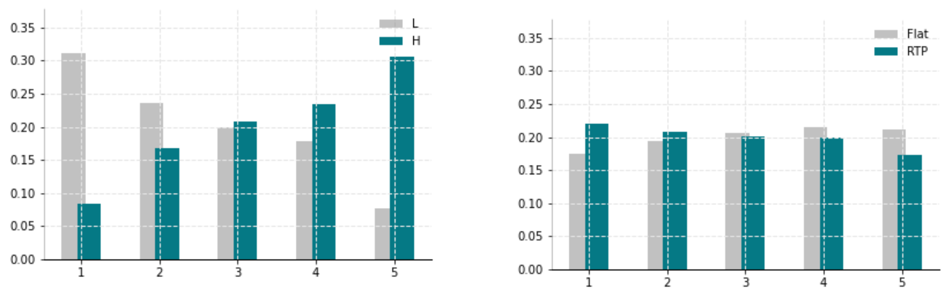
4 Quantifying the Distributional Impacts of RTP

Our aim is to identify who wins and who loses from the move to RTP, using our estimated income distribution at the household level. We explore two dimensions of the distributional impacts: across and within income groups.

4.1 Heterogeneity across Income Bins

We start by analyzing the heterogeneity in bill impacts across income groups. Figure 4 classifies households in five national income quintiles, and plots the bill impacts following a switch from a time-invariant annual price to real time prices. Results depend on whether one uses our estimated household level income distribution (GMM approach) or the zip-code level income distribution (naive approach). Under our proposed approach, the move towards real-time pricing is regressive, as it benefits the high income households while making

Figure 3: Estimated income distribution across tariff choices



(a) Estimated income distribution by contracted power capacities

(b) Estimated income distribution by tariff

Notes: 51% of households in our data are classified to low contracted power capacities, their contracted power capacities are below 4 kW. 57% of households in our data are in RTP, which is the default tariff.

the low income households worse off. Neglecting the within-zip-code income heterogeneity would deliver the opposite conclusion, with low income groups paying slightly less under RTP. Furthermore, the naive approach would also miss an important part of the distributional implications, as the predicted bill impacts would be almost flat across income groups. In any event, the magnitudes of the effects are modest under both approaches.

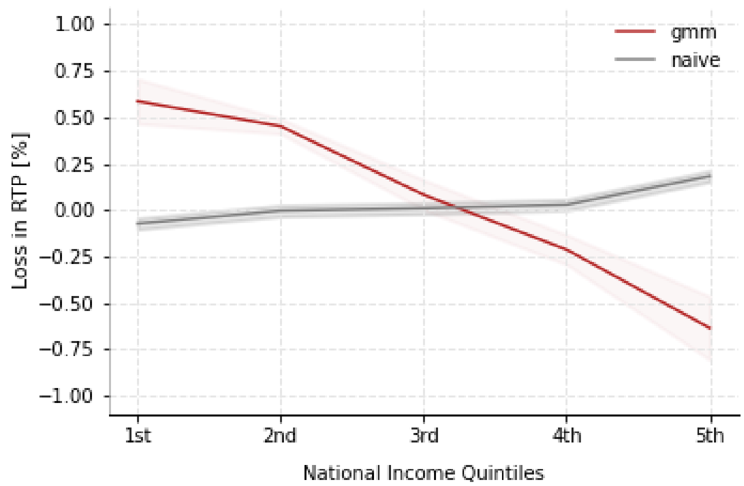
Figure 5 decomposes the bill impacts in changes within and across months. Panel (a), which relies on our estimated income distribution, uncovers a fundamentally different distributional impact depending on the source of price variation. While a move from an annual price to monthly prices is regressive (across months channel), the move from monthly to hourly prices is progressive (within month channel). The larger magnitude of the former explains why the move from an annual price to RTP is regressive overall. In the next section we explore the channels that explain these patterns. As shown in panel (b), these effects would again be hidden if we used the zip-code level income distribution rather than our estimated household-level income distribution, in which case both bill impacts would appear to be slightly progressive and very small in magnitude.

4.2 Heterogeneity within Income Bins

The relatively small average bill impacts at the income group level hide substantial heterogeneity within income groups. Panel (a) in Figure 6 plots the distribution of the percentage bill impacts from moving from an annual time-invariant price to RTP for the 1st, 3rd and 5th quintiles. Whereas most consumers gain or lose at most 2%, the gains or losses can reach +/- 6% for some households. As it can be seen, the right tail of the 1st quintile shows slightly higher bill losses.

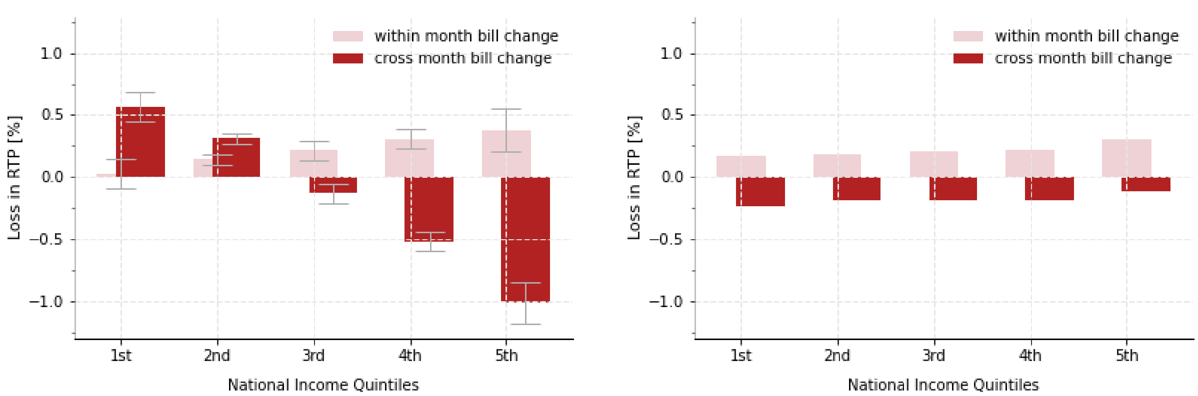
The similarities across these distributions also hide another source of heterogeneity: location and its implications for appliance ownership, an issue on which we will elaborate further below. Panels (b) and (c) split the distributions between those regions where electric heating is prevalent ("electric heating regions"), from those where it is not ("non electric heating regions"). The comparison of both plots shows that the low income households are relatively more negatively impacted in the electric heating regions, while the reverse is true for the high income households. This finding suggests that the distributional impacts are not so much driven by income differences but rather by households' locations and appliance ownership. To explore this

Figure 4: Bill increase (=loss) in RTP [%]



Notes: This figure represents the bill loss in % when moving from an annual time-invariant price to RTP. Results are reported for the five national income quintiles, with households' income classified according to our estimated income (GMM) or to the zip-code income (naive).

Figure 5: Distributional impact: decomposition

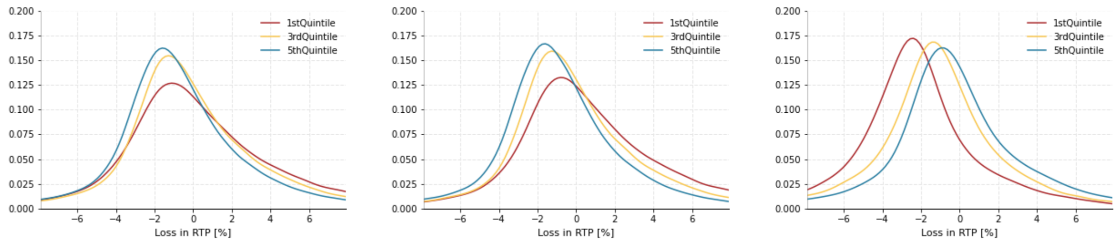


(a) Bill increase (=loss) in RTP [%]

(b) Bill increase (=loss) in RTP [%]: naive approach

Notes: These figures decompose the bill loss in % when moving from an annual time-invariant price to monthly prices (pink vars) and from monthly prices to RTP (red vars), for the five national income quintiles. Panel (a) classifies households according to our estimated income (GMM approach), while panel (b) relies on the zip-code level income (naive approach). The bars would sum up to zero if expressed in Euro, but this is not the case when expressed in %. Also note that these figures represent the national average, which hides the heterogeneity in the bill impacts across regions. See Figure 11 for the regional decomposition.

Figure 6: Bill increase (=loss) in RTP [%]



(a) All regions (b) Electric heating regions (c) Non-electric heating regions

Notes: These figures show the distribution of losses under RTP in the first, third and fifth income quintiles. Panel (a) shows the distributions at the national level, while panels (b) and (c) distinguish between regions with a high and a low prevalence of electric heating, respectively. Together, they show that (i) there are large heterogeneities within income groups, and (ii) the low income households are particularly hurt in the electric heating regions.

in more detail, the next section is devoted to disentangling these channels.

5 Channels

This section uncovers the channels by which income affects the bill impacts of RTP. We focus on differences in consumption patterns during the day, appliance ownership and regional differences.

5.1 Consumption Profile

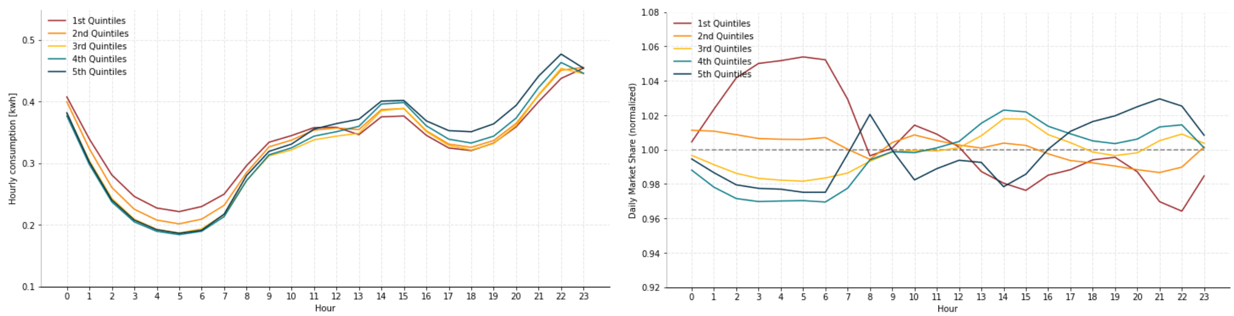
Our previous results show that a move from time-invariant monthly prices to RTP is progressive, i.e., low income groups gain from this switch. This is explained by their daily consumption patterns, as documented below. Figure 7 (a) plots the average hourly consumption patterns of households in each of the five income groups. It shows that high income households tend to consume slightly more electricity across the day, but relatively more so during the peak hours. While the differences across income groups seem small,[19] there is large heterogeneity in their consumption shares across the day. To uncover these, Figure 7 (b) plots the normalized hourly consumption share of each income group, i.e., the share of the household's consumption at a given hour, over the share at the national level. A positive number implies that the household concentrates a greater share of its consumption at that hour relative to the national share. This figure makes it clear that the high income group consumes more at peak times relative to the national average, while the low income income group consumes relatively more at off-peak times. In other words, the consumption of high (low) income households' tend to be procyclical (countercyclical).

Beyond this graphical evidence, we can use simple regressions to understand how income affects electricity consumption patterns during the day and how that leads to within month gains and losses.

We start by computing the price coefficient for each household ($price\ coeff_i$) by regressing the household's hourly consumption on hourly prices, plus a constant. This coefficient captures whether a household's consumption pattern is positively or negatively correlated with real-time prices. We then measure the

[19]These differences would look larger if we did not include Madrid, which is the only region where the high income households tend to consume less electricity. This might be explained by the high prevalence of natural gas in Madrid.

Figure 7: Load curve by income quintiles



(a) Hourly consumption patterns [kWh] (b) Normalized consumption shares [%]

Notes: Panel (a) shows the average consumption patterns over the day for the five national income quintiles. Panel (b) depicts the normalized hourly consumption shares, defined as the share of the household's consumption at a given hour, over the share at the national level. Together, this shows that while consumption levels are not very different across income groups, their distribution across time is highly heterogeneous.

correlation between the price coefficients and the income levels (regression (13))[20] as well as the extent by which the price coefficient explains the within month losses under RTP (regression (14)):

$$\textit{price coeff}_i = \sum_{k=2}^{5} \beta_k \mathbb{1}(Inc_k)_i + Z_i + \alpha_z + \epsilon_i \tag{13}$$

$$Loss_i^m = \gamma \ \textit{price coeff}_i + \alpha_z + \epsilon_i \tag{14}$$

where α_z is a zip code level fixed-effect and Z_i are control variables including households' tariff choices and appliance ownership.[21]

The estimated results are reported in Table 3, columns (1) and (2). As it can be seen, higher income households tend to have higher price coefficients, after controlling for zip-code fixed-effects and a set of individual-level control variables. In turn, the price coefficients explain 32% of the variation in within month losses, according to the R^2 in column (2). Put together, these two pieces of evidence explain why the move from monthly prices to RTP tends to benefit the low income groups.

Our two last regressions confirm that income affects bill losses mainly through the correlation between consumption and real time prices, as captured by the price coefficients:

$$Loss_i^m = \sum_{k=2}^{5} \beta_k \mathbb{1}(Inc_k)_i + \alpha_z + \epsilon_i \tag{15}$$

$$Loss_i^m = \gamma \ \textit{price coeff}_i + \sum_{k=2}^{5} \beta_k \mathbb{1}(Inc_k)_i + \alpha_z + \epsilon_i \tag{16}$$

Indeed, as shown in column (4) of Table 3, income is correlated with the within month loss. However, if controlled for the price coefficient, the direct effect of income on the within month loss becomes very small, as shown in column (4). Furthermore, the estimates for the price coefficient are very similar when controlling or not controlling for income, i.e., the first line in columns (2) and (4). This highlights that the channel for

[20] Note that in equation (13), the first income bin is omitted. Hence, β_k reflects how much more correlated are income and the price coefficients in group k, relative to the lowest income group.
[21] Details about appliance ownership variables are explained in the next subsection.

Table 3: Income, hourly consumption patterns, and within month losses

	Price coeff.	Loss^m [%]	Loss^m [%]	Loss^m [%]
Price coeff.		0.171***		0.116***
		(0.000)		(0.000)
2nd quintile	0.783***		0.109***	-0.003
	(0.018)		(0.006)	(0.004)
3rd quintile	1.075***		0.175***	-0.002
	(0.018)		(0.006)	(0.004)
4th quintile	1.371***		0.232***	0.000
	(0.018)		(0.006)	(0.004)
5th quintile	1.651***		0.269***	0.019***
	(0.020)		(0.006)	(0.004)
R^2	0.100	0.323	0.026	0.458
N	1183216	1189373	1189373	1189331
FE	zip code	zip code	zip code	zip code

Notes: This table reports regression results for equations (13), to (16) in columns (1) to (4), respectively. A zip-code fixed effect is included in all regressions.

the distributional impact runs through the correlation of households' consumption patterns and real-time prices, which differs across income levels.

Price volatility amplifies this channel. Figure 8 shows the relationship between price volatility, defined as the variance of hourly prices within a month, and the monthly redistribution effect, defined as the total loss (or total gain) including all households. Since winter months depict higher price volatility, the distributional impact becomes greater, as shown in panel (b) of Figure 8.

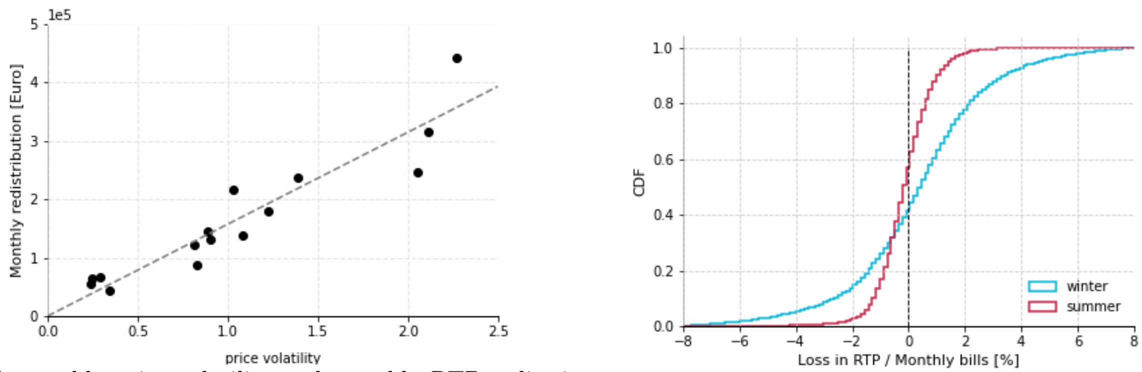
5.2 Appliance Ownership

Appliance ownership (mainly, electric heating and AC) has a strong impact on electricity consumption, both regarding the levels as well as the consumption patterns over time. Panels (a) and (b) in Figure 9 plot the average consumption patterns of households with and without electric heating, during the day and across the year, respectively; panels (c) and (d) do the same for AC. As it can be seen, there are substantial differences in the consumption patterns of households with different appliance ownership. Households with electric appliances consume significantly more across all hours of the day than those households without them. Also, their consumption tends to be peakier, particularly so in the case of heating. Furthermore, there are strong seasonal effects: as expected, households with electric heating consume more during the winter months (October through April), while households with AC consume more during the summer months (June through September). In the case of heating, these effects are more pronounced for the high than for the low income households. In contrast, in the case of AC, the effects are fairly similar across income groups.

In general, higher income households are more likely to have AC, while lower income households are more likely to have electric heating. This fact is explained by the high costs required to install other heating systems (e.g. such as gas, or central heating) relative to electric heating that commonly relies on low cost plug-in radiators. Indeed, 20% of the 5th quintile and only 15% of the 1st have AC. In contrast, 26% of the 1st quintile and 8% of the 5th have electric heating.[22] Since prices tend to be higher during the winter

[22]These results are reported in Figure E.3a in the Appendix. For AC, these differences are stronger conditional on location. In Spain, the lower income regions tend to be warmer which implies that lower income households tend to have more AC. Indeed, there are only small differences within regions.

Figure 8: Impact of price volatility on within month redistribution



(a) monthly price volatility and monthly RTP redistribuion

(b) winter and summer within month bill change in RTP

Notes: Panel (a) shows that there is more redistribution during moths of higher price volatility. The dots with the highest price volatility correspond to January 2017, January 2016, and February 2017. The higher price volatility during winter months explains why the CDF of the losses during winter is flatter than in the summer, as shown in panel (b).

Table 4: Appliance ownership and loss under RTP

	kWh	Loss [%]	Loss^m [%]	Loss^a [%]
AC	204.034***	-1.102***	0.127***	-1.229***
	(4.392)	(0.013)	(0.005)	(0.012)
Heating	1006.578***	2.949***	0.913***	2.037***
	(4.082)	(0.012)	(0.005)	(0.011)
R^2	0.177	0.129	0.066	0.122
N	1169740	1162746	1162746	1162746
FE	zip code	zip code	zip code	zip code

Notes: Column (1) reports the regression results from equation (17), columns (2)-(4) report the results from equation (18), with the dependent variables being the total loss under RTP, the within month loss under RTP, and the across months loss under RTP, respectively. A zip code fixed effect is included in all regressions. One can see that households with AC (electric heating) tend to gain (lose) under RTP. This gain (loss) is mainly driven by the across months effect.

months when electric heating is used, it follows that a move from an annual price to RTP tends to hurt the low income households relatively more. The across months effect is quantitatively strong and offsets the within month effects we documented in the previous subsection.

Similar evidence is obtained through the following regressions, which capture the impact of appliance ownership on electricity consumption on the one hand and the losses from RTP on the other:

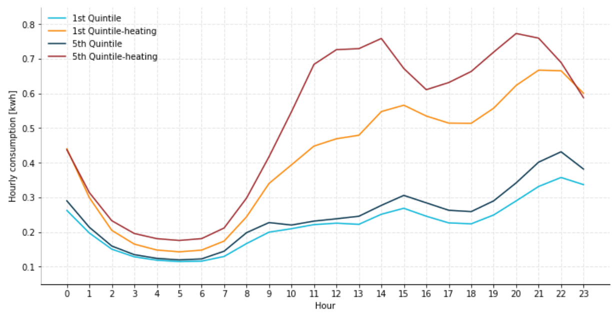
$$kWh_i = \beta^{AC}\text{AC}_i + \beta^{EH}\text{EH}_i + Z_i + \alpha_z + \epsilon_i \tag{17}$$

$$Loss_i = \beta^{AC}\text{AC}_i + \beta^{EH}\text{EH}_i + Z_i + \alpha_z + \epsilon_i \tag{18}$$

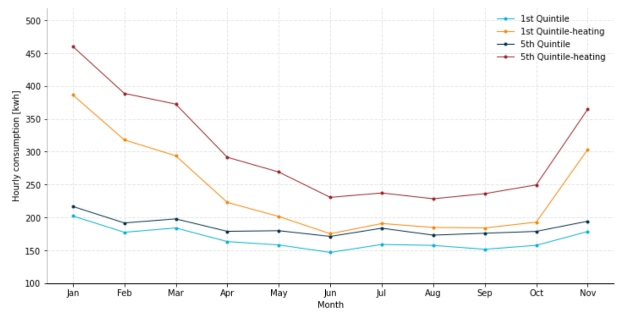
where α_z is a zip-code fixed-effect and Z_i includes household level control variables. The coefficients β^{AC} and β^{EH} capture the effect of AC and electric heating ownership on either the household's electricity consumption or on the losses under RTP.

The estimates show that AC increases annual electricity consumption by 204.03 kWh, i.e., 9.6% of a median household's annual consumption. For electric heating, the increase in annual consumption is five

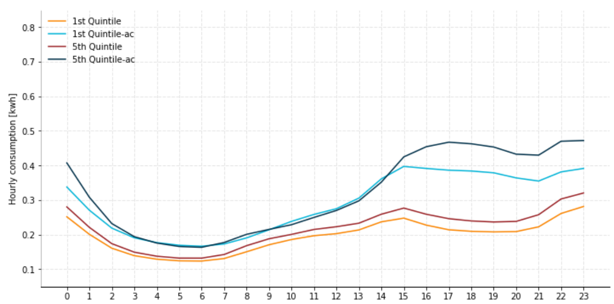
Figure 9: Load curves by appliance ownership and income



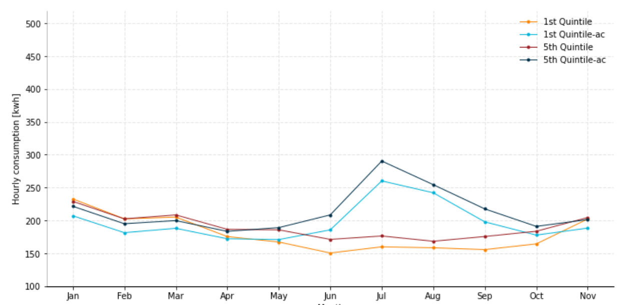
(a) Hourly consumption pattern [kWh] - Heating



(b) Annual consumption pattern [kWh] - Heating



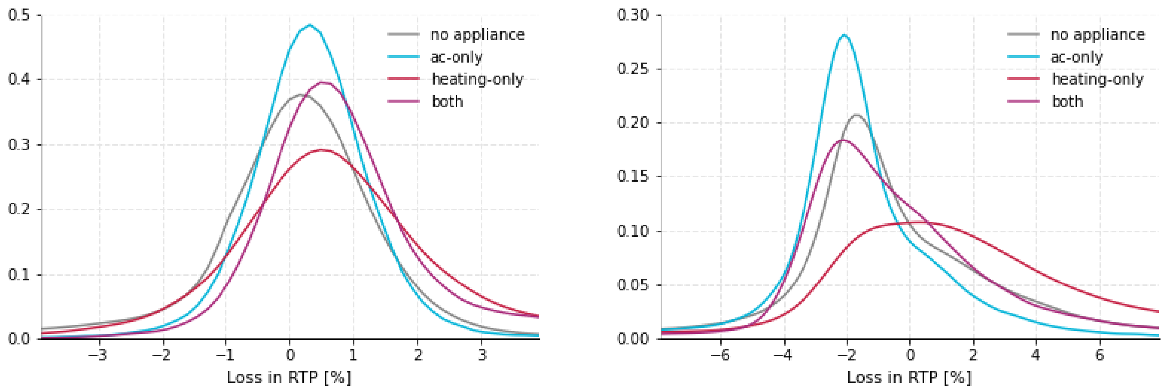
(c) Hourly consumption pattern [kWh] - AC



(d) Annual consumption pattern [kWh] - AC

Notes: These figures show consumption patterns over the day (the left panels) and the year (the right panels) for households with electric heating (the upper panels) and AC (the lower panels). Results are reported for low (1st quintile) and high income households (5th quintile).

Figure 10: Bill increase (=loss) in RTP [%] by appliance ownership



(a) Within month effects

(b) Across months effects

Notes: These figures plot the distribution of the losses under RTP for households with no appliance, with AC only, with electric heating only, or with both. The within month and across months effects are shown in panels (a) and (b), respectively. The bigger losses are suffered by households with electric heating due to the across months effect.

times higher, i.e., 1,006.58 kWh, representing 47.2% of a median household's annual consumption. Through their effects on households' consumption patterns, electric heating increases the loss under RTP by 2.9%, but AC leads to a 1.1% gain. These opposite signs have a simple explanation: electric heating (AC) increases consumption during winter (summer) months, when prices are higher (lower).

These results are consistent with the evidence reported in Figure 10, which decomposes the bill impacts in the within and across months channels, distinguishing according to appliance ownership. Regarding the within month effects shown in panel (a), both AC and electric heating owners would tend to lose on average as they consume more. The fact that the winter price volatility is ten times larger than during the summer amplifies the bill impacts of electric heating. Regarding the across months effects shown in panel (b), AC users gain from being under RTP while electric heating users lose for reasons explained above.

5.3 Location

Another key driver of the distributional implications of RTP is location heterogeneity. Consumption patterns have a lot to do with local weather conditions, which in turn affect appliance ownership, even when controlling for income. Moreover, there are regional differences in the availability of heating infrastructure, mainly gas, which affect the prevalence of electric heating in the region. For instance, whereas the availability of heating systems reaches 90.4% in Madrid, it is only 59.9% in the more rural Galicia. Castilla y Leon is the region where electric heating is least common (where only 8.6% of households have electric heating, as compared to the national average, 18.6%, because they rely more on gas and oil heating).[23]

Figure 11 decomposes the distributional effects of RTP in three dimensions: across regions (represented by the four lines), and within months in panel (a) and across months in panel (b). As it can be seen, the across months price variation is the main driver of the distributional implications of RTP, both across income groups as well as across regions. Furthermore, whereas these seasonal effects make RTP regressive in the electric heating regions (in the figure, Castilla la Mancha, Galicia, and Madrid), they make them progressive

[23]See Table E.1 in the Appendix for details.

Table 5: Average loss under RTP by region

	Loss [%]	Loss^m [%]	Loss^a [%]
Castilla y Leon	-1.250***	-0.144***	-1.106***
	(0.015)	(0.006)	(0.014)
Castilla-La Mancha	-0.486***	0.148***	-0.634***
	(0.010)	(0.004)	(0.009)
Galicia	0.454***	0.070***	0.384***
	(0.007)	(0.003)	(0.006)
Madrid	0.159***	0.332***	-0.173***
	(0.008)	(0.003)	(0.007)
R^2	0.012	0.005	0.012
N	1189373	1189373	1189373

Notes: The reported coefficients result from regressing the bill losses on the regional dummies, without any additional controls. The coefficients thus represent the mean loss at each region. Households in Castilla y Leon gain (across months) under RTP because they are less likely to have electric heating relative to other regions.

in the non-electric heating region (Castilla y Leon). The within month channel is slightly progressive, but its magnitude is small. This evidence is consistent with the results from regressing the losses on regional dummies, with the coefficients capturing the mean bill loss under RTP for each region. Results are reported in Table . Overall, we conclude that appliance ownership is a key driver of the distributional implications of RTP due to its effects on the levels and patterns of consumption across the year.

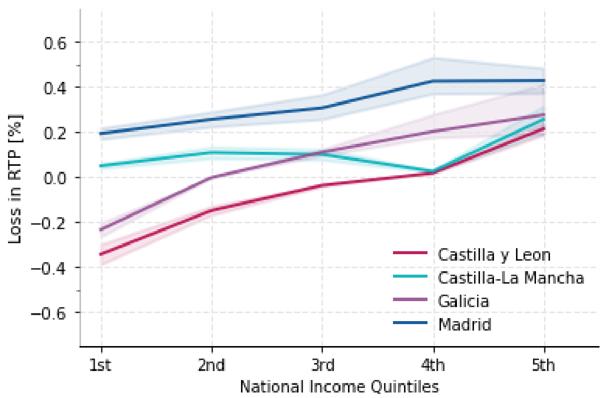
6 Counterfactual Experiments

We assess the counterfactual implications of two recent phenomena: (i) the increased incidence of price spikes and price volatility, and (ii) changes in households' equipment, due to investments in demand response devices, batteries, or solar panels.

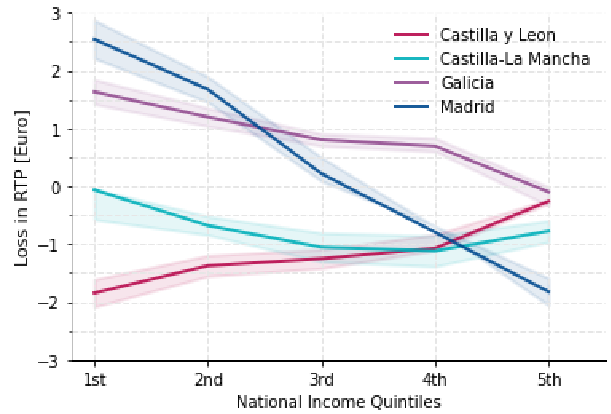
6.1 Commodity Risk and Energy Poverty

The economic magnitude of the reported distributional impacts is minor due to the small price variation during our sample period. However, the impacts would be enlarged if prices within or across months became more volatile, as it has been the case after our sample period. Indeed, during 2021 the average price in the Spanish electricity market tripled with respect to the average price in previous years. Several reasons made this price shock particularly harmful for the low income households. First, there are more low income households under the default real-time pricing policy, relative to high income households, since they are entitled to a social tariff as long as they do not opt out (see Figure 3). Second, the price shock was particularly strong during the winter months, which hit the low income households harder because they tend to have relatively more electric heating. And third, price levels increased overall without affecting the within day price variation much, which remained limited. This is explained by prices in the Spanish electricity wholesale market being set by the CCGTs, whose marginal costs depend on the daily gas and CO2 prices. However, in the future, as renewable energies start setting the market price during some hours of the day, the within day price volatility might become larger, potentially allowing the low income households to benefit from their less peaky consumption patterns.

Figure 11: Distributional impact: geographical heterogeneity and decomposition



(a) Estimated loss: within month loss



(b) Estimated loss: across months loss

Notes: These figures decompose the distributional effects of RTP in the within month and the across month effects in panel (a) and (b), respectively, for four regions. The within month channel is slightly progressive, but its magnitude is small relative to the across months channel, which is regressive in all regions (except for Castilla y Leon, in which there is little electric heating).

In order to quantify the potential distributional implications of higher and more volatile prices, we simulate market prices from August 2021 to March 2022 using actual prices from August 2020 to March 2021, as shown in Figure 12. We consider three scenarios with low, medium and high price trajectories in panel (a), and with low, medium and high price volatility in panel (b). For the mean price increases, we simply add the same constant to all prices in the month. For the volatility increases, we enlarge the departure of each price from the monthly mean.

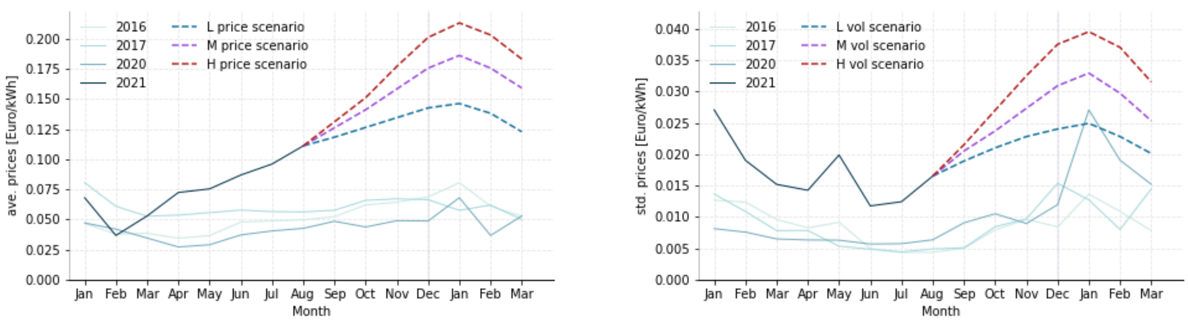
Keeping households' demand patterns fixed, Figure 13 reports the distributional effects of RTP using the actual and simulated prices for 2021-2022. As it can be seen, the effects of RTP are regressive, and the magnitude of the distributional impact is much greater than the one reported in the previous section. The within month effect is progressive, in line with our previous results. However, its magnitude is so small that it does not matter for the overall effect, which is almost fully explained by the across months effect. These effects would have to be qualified if demand responded to the price shock. We turn to this question next.

6.2 Accounting for Potential Responses: Demand Elasticity

So far, we have assumed that demand is inelastic, i.e., households' electricity consumption remains unchanged after the move to RTP. However, it is reasonable to suspect that electricity demand will depict some price elasticity to short-run price changes in the future. For instance, this will be the case if households install devices that allow them to automatically adjust their consumption in response to price changes (Bollinger and Hartmann, 2020). Another source of demand response could well come through the deployment of electric vehicles and batteries, which typically allow households to benefit from arbitraging within-day price differences. The deployment of rooftop solar installations–which can be understood as a medium-run response to price increases–could also have important distributional implications to the extent that they allow households to reduce their consumption from the grid and thus have more stable energy costs.

These investments, together with the use of more electricity-intensive equipment (e.g., electric heat pumps), enhance the possibility and incentives for more active demand management. However, it is likely that

Figure 12: Simulated large price shock

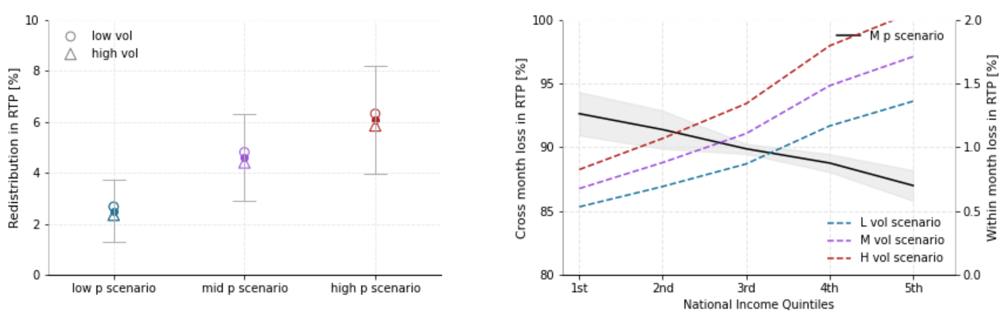


(a) Simulated prices (mean changes)

(b) Simulated prices (volatility changes)

Notes: Panel (a) plots the actual prices in the Spanish electricity market from 2016 to August 2021, and the simulated prices from August 2021 until March 2022 with low, medium and high price trajectories. Panel (b) plots the actual price volatility (measured by the standard deviation) from 2016 to August 2021, and the simulated volatility from August 2021 until March 2022 with low, medium and high volatility assumptions.

Figure 13: Distributional implications of RTP under a large price shock

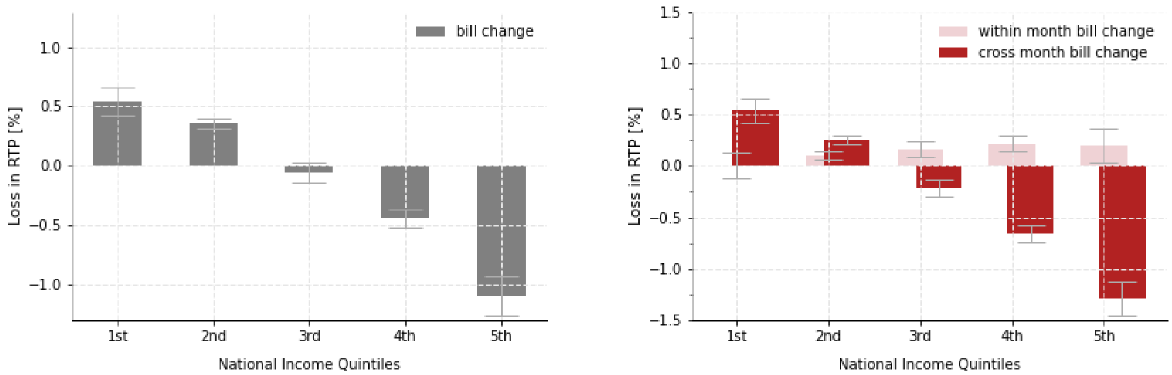


(a) Distributional implications

(b) Distributional implications

Notes: This figure illustrates the distributional implications of price increases and increased price volatility. Panel (a) shows that there is more redistribution with higher prices (as this enhances the across months channel, which makes the low income households relatively worse off) and lower volatility (as this mitigates the within month channel, that would otherwise benefit the low income households). Panel (b) shows the distributional impact due to the across months channel (scale on the left axis) and due to the within month channel (scale on right axis). The across months loss leads to a much stronger regressive effect than in previous sections. The within month effect is so mild that not even a low price volatility scenario is able to revert the regressive effect.

Figure 14: Distributional Implications of RTP under Demand Elasticity



(a) Aggregate impact (b) Decomposition

Notes: This figure illustrates the distributional implications of RTP when rich households have higher price elasticity. Panel (a) shows the aggregate loss for each income bin, and panel (b) decomposes it between the within month and the across months channel. The regressive effect of RTP is now larger.

these investments, and hence the scope for demand elasticity, will be positively correlated with households' income.

To explore these issues, we recompute households' electricity bills under the assumption that they adjust their consumption to price changes using the following parametrization:

$$kWh^{e}_{i,hdm} = kWh_{i,hdm} \times \left[1 + \frac{p_{hdm} - \overline{p}_m}{A + \overline{p}_m} \times \tau_i\right]$$

$$\tau_i = -\alpha \hat{inc}_i$$

where $kWh^{e}_{i,hdm}$ and $kWh_{i,hdm}$ denote household i's adjusted and actual consumption, respectively; p_{hdm} is the real-time price, $\overline{p}_m$ is the monthly average price, A is the bill's fixed fee, τ_i is a negative parameter, α is a scale factor to adjust the elasticity to the [0, 0.2] interval, and $\hat{inc}_i$ is the household's estimated income. Households reduce (increase) their consumption if the real-time price is higher (lower) than the monthly time-invariant price. The magnitude of the change depends on the value of τ_i, which is positively correlated with income (in absolute terms).

Figure 14 shows the distributional results with elastic consumers. As expected, demand elasticity reduces the losses under RTP given that households can adjust their consumption to the price changes. However, the high income households are made relatively better off because of their higher elasticity. High income households particularly benefit from the across months effect: because prices and price volatility are greater during winter, the higher price elasticity of high income households leads them to reduce consumption at times when prices are higher.

7 Conclusions

We have evaluated the distributional implications of the switch from time-invariant to real-time electricity prices in the Spanish electricity market, which became the first country to broadly implement RTP as the default option for residential households. While Fabra et al. (2021) shows that this regulatory change had little impact on households' electricity consumption, the question of whether households were asymmetrically impacted by it remained unanswered. This issue is particularly important as the fear of adverse distributional implications might have delayed a broader implementation of RTP elsewhere.

Access to hourly electricity consumption data at the household level for a large sample of representative Spanish households has allowed us to obtain meaningful conclusions as to how their electricity bills have changed under RTP. Access to detailed socio-demographic data has further allowed as to understand the distributional implications of those changes.

An important step of our analysis is the estimation of households' income. Working with our estimated distribution of income, rather than with the zip-code level income distribution, has allowed us to uncover distributional effects of RTP than would have otherwise remained hidden. The electricity consumption data has also allowed us to infer the households' ownership of electric heating or air conditioning, which are key determinants of the gains and losses from RTP.

The analysis reveals that, in the context of the Spanish electricity market, the move to RTP has been regressive as lower income groups have been made worse off relative to the higher income groups. Interestingly, this overall effect can be decomposed in two channels: the bill impacts due to the within month and the across months price variation. We have found that the daily consumption patterns of the low income households tend to be relatively countercyclical, i.e., they consume relative more when prices are lower, which implies that the move from time-invariant prices to RTP tends to benefit them. However, because of their dependence on electric heating, the low income households tend to consume relatively more during winter months when prices are higher. The magnitude of this latter channels explains the overall regressive effect of RTP. However, the overall impact of RTP remained small and not of concern during the period of study, thanks to the relatively stable prices and limited volatility. An increase in price levels and in price volatility (as experienced in 2021) can further worsen the distributional implications.

References

Allcott, H. (2011). Social norms and energy conservation. *Journal of public Economics 95*(9-10), 1082–1095.

Blonz, J. A. (2018). Making the best of the second-best : Welfare consequences of time-varying electricity prices.

Bollinger, B. K. and W. R. Hartmann (2020, January). Information vs. Automation and Implications for Dynamic Pricing. *Management Science 66*(1), 290–314.

Borenstein, S. (2005). The long-run efficiency of real-time electricity pricing. *The Energy Journal 26*(3).

Borenstein, S. (2007). Wealth transfers among large customers from implementing real-time retail electricity pricing. *The Energy Journal 28*(2).

Borenstein, S. (2012a). The redistributional impact of nonlinear electricity pricing. *American Economic Journal: Economic Policy 4*(3), 56–90.

Borenstein, S. (2012b, August). The Redistributional Impact of Nonlinear Electricity Pricing. *American Economic Journal: Economic Policy 4*(3), 56–90.

Borenstein, S. (2013). Effective and equitable adoption of opt-in residential dynamic electricity pricing. *Review of Industrial Organization 42*(2), 127–160.

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*(1), 75–128.

Borenstein, S. and S. Holland (2005, Autumn). On the Efficiency of Competitive Electricity Markets with Time-Invariant Retail Prices. *RAND Journal of Economics 36*(3), 469–493.

Brolinson, B. (2019, September). Does Increasing Block Pricing Decrease Energy Use? Evidence from the Residential Electricity Market. Working paper, Georgetown University, Department of Economics.

Burger, S. P., C. R. Knittel, I. J. Prez-Arriaga, I. Schneider, and F. v. Scheidt (2019, February). The efficiency and distributional effects of alternative residential electricity rate designs. Working Paper 25570, National Bureau of Economic Research.

Byrne, D. P. and L. A. Martin (2021). Consumer search and income inequality. *International Journal of Industrial Organization 79*, 102716.

European Commission (2009, 11). Energy prices and costs in europe. https://ec.europa.eu/transparency/regdoc/rep/10102/2019/EN/SWD-2019-1-F1-EN-MAIN-PART-5.PDF.

Fabra, N., M. Reguant, D. Rapson, and J. Wang (2021). Estimating the Elasticity to Real Time Pricing: Evidence from the Spanish Electricity Market. *AEA Papers and Proceedings, American Economic Association fortcoming*.

Faruqui, A., R. Hledik, and J. Tsoukalis (2009). The power of dynamic pricing. *The Electricity Journal 22*(3), 42–56.

Faruqui, A., S. Sergici, and J. Palmer (2010). The impact of dynamic pricing on low income customers. *Institute for Electric Efficiency Whitepaper*.

Fowlie, M., C. Wolfram, C. A. Spurlock, A. Todd, P. Baylis, and P. Cappers (2021, June). Default Effects and Follow-On Behavior: Evidence from an Electricity Pricing Program. Technical Report forthcoming, Review of Economic Studies.

Hahn, R. W. and R. Metcalfe (2021). Efficiency and equity impacts of energy subsidies. *American Economic Review 111*(5), 1658–88.

Harding, M. and S. Sexton (2017, October). Household Response to Time-Varying Electricity Prices. *Annual Review of Economics 9*(1), 337–359.

Holland, S. P. and E. T. Mansur (2008). Is real-time pricing green? the environmental impacts of electricity demand variance. *The Review of Economics and Statistics 90*(3), 550–561.

Horowitz, S. and L. Lave (2014). Equity in residential electricity pricing. *The Energy Journal 35*(2).

Hortaçsu, A., S. A. Madanizadeh, and S. L. Puller (2017). Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market. *American Economic Journal: Economic Policy 9*(4), 192–226.

Jessoe, K. and D. Rapson (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review 104*(4), 1417–38.

Joskow, P. L. and C. D. Wolfram (2012, May). Dynamic Pricing of Electricity. *American Economic Review 102*(3), 381–385.

Leslie, G., A. Pourkhanali, and G. Roger (2021). Identifying consumption profiles and implicit cross-subsidies under fixed-rate electricity tariffs. Working paper, Monash University.

Levinson, A. and E. Silva (2021, October). The electric gini: Income redistribution through energy prices. *American Economic Journal: Economic Policy* (forthcoming).

Poletti, S. and J. Wright (2020). Real-time pricing and imperfect competition in electricity markets. *The Journal of Industrial Economics 68*(1), 93–135.

Wang, Q., M.-P. Kwan, J. Fan, and J. Lin (2021). Racial disparities in energy poverty in the united states. *Renewable and Sustainable Energy Reviews 137*, 110620.

Wolak, F. A. (2011). Do residential customers respond to hourly prices? evidence from a dynamic pricing experiment. *American Economic Review 101*(3), 83–87.

Online Appendix for "The Distributional Impacts of Real-Time Pricing"

A Demographics: Data Sources

In this appendix, we provide further details about the demographic data that we use in our analysis. These data re provided by the Spanish National Institute of Statistics, Instituto Nacional de Estadística (INE), and correspond to the most recent census (2011). The data contain information at the census level on population, age, sex, education, dwelling types (main dwelling, secondary dwelling, or empty dwelling), number of rooms per dwelling, and net surface area of dwellings for each census district in Spain. We have also collected detailed information on the distribution of income at the district (and sometimes section) level.[24] The income distribution information reports the percentage of households below a certain threshold of normalized median income used to determine poverty levels. The normalization of household income is meant to reflect economies of scale within a household, a standard pricedure in the European Union.[25] We complement the data from the INE with data from MB Research at the postal code level.[26]

We know the zip-code of each household, but not its census. To create a crosswalk between postal codes and census districts, we use shapefiles of Spanish postal codes and census districts provided by the INE. Census districts are matched to postal codes with which they have significant intersection.[27] On average, postal codes are matched to around 7 census districts. Once census districts and postal codes are matched, census district data are aggregated a the postal code level. We find that some zip-codes are not present in the shapefiles. To complement the map between zip-codes and districts, we use data with latitude and longitude for the universe of street addresses in the postal code system ("callejero").[28] A district section and a zip-code are matched if the latitude and longitude of the address lays inside that section.

We classify all the zip-codes into 5 categories based on the distribution of income per capita, from low to high income. Whereas 90% of households in the poorest zip-codes have an adjusted income below 20,000 Euro/year, around half of the households in the richest zip-codes have an income above that level. Table A.1 shows the median income, median household size, and mean monthly electricity consumption at the zip-code-level. It also shows the zip-code average winning or losing measures: mean savings from moving from a flat rate to RTP and mean ln[Flat ratio]. The richest regions have the highest electricity consumption, as expected, but surprisingly also do the poorest ones. Household size is quite homogeneous across regions, only slightly higher in the poor regions.

Table A.1 shows the median income, median household size, and mean monthly electricity consumption at the zip-code-level. It also shows the zip-code average winning or losing measures: mean savings and relative saving from moving from a time-invariant rate to RTP. The richest regions have the highest electricity consumption, as expected, but surprisingly also do the poorest ones.

[24]For confidentiality reasons, sections are often not reported as they are a fairly small geographical small unit. For small to medium sized municipalities, data are often only available at the municipality level, which often coincides with the postal code. Very small municipalities might not have their data reported.

[25]XX ADD A REFERENCE OR MORE DETAILED EXPLANATION IN APPENDIX XX

[26]There are two major differences between the two sources of income data. First, the income distribution reported by MB research is not normalized. Second, the cutoffs are different and are meant to be representative of the quintiles in the national distribution of income, instead of thresholds related to poverty lines. Therefore, these two distributions of income complement each other at different parts of the support.

[27]The matching algorithm is as follows: if 90% or more of a census district's area is contained within a postal code, or if 90% or more of a postal code's area is contained within a census district, then the census district is matched to the postal code.

[28]This information can be obtained at https://www.ine.es/prodyser/callejero/.

Table A.1: Summary Statistics (monthly zip-code-level)

	poor	medpoor	medium	medrich	rich
Demographics					
ln[inc per c] (med)	8.99	9.17	9.27	9.43	9.71
	(8.95, 9.04)	(9.13, 9.20)	(9.23, 9.30)	(9.37, 9.48)	(9.59, 9.83)
ln[inc per hh] (med)	9.96	10.11	10.22	10.37	10.68
	(9.93, 10.02)	(10.05, 10.15)	(10.17, 10.28)	(10.33, 10.41)	(10.57, 10.83)
HH size (med)	2.60	2.53	2.58	2.53	2.58
	(2.49, 2.75)	(2.39, 2.64)	(2.43, 2.71)	(2.41, 2.66)	(2.41, 2.68)
Consumption					
ln[kwh]	4.99	4.86	4.90	4.90	4.96
	(4.88, 5.13)	(4.77, 4.98)	(4.75, 5.08)	(4.79, 4.96)	(4.91, 5.06)
ln[kwh] (peak)	4.24	4.08	4.14	4.15	4.22
	(4.10, 4.39)	(3.95, 4.17)	(4.00, 4.32)	(4.05, 4.23)	(4.16, 4.34)
ln[kwh] (off-peak)	4.31	4.20	4.22	4.21	4.27
	(4.21, 4.44)	(4.11, 4.34)	(4.07, 4.38)	(4.11, 4.27)	(4.21, 4.36)
Win or loss					
Flat cost [Euro]	15.07	13.16	13.13	12.77	13.05
	(13.67, 16.61)	(11.52, 14.53)	(11.58, 15.33)	(10.97, 13.35)	(12.06, 14.05)
Saving in RTP [Euro]	0.08	0.08	0.06	0.03	-0.00
	(0.03, 0.11)	(0.02, 0.14)	(0.01, 0.10)	(-0.01, 0.08)	(-0.03, 0.01)
ln[Flat ratio]	0.01	0.00	0.00	0.00	-0.00
	(0.00, 0.01)	(0.00, 0.01)	(0.00, 0.01)	(0.00, 0.00)	(-0.00, 0.00)

Notes: footnotes .

B Alternative parametric estimator (preliminary and incomplete)

Parametric approach Same as our main approach, we estimate the demand system for each province separately, because there are huge heterogeneity across regions as shown in Section 5. All the parameters are province specific. To save notation, we suppress subscripts for province and will not emphasis province specific in the following paragraphs.

Assume each household is optimizing daily consumption across 24 hours. Thus, we can think each household-day as a "market" and we observe this market repeatedly everyday. A household's consumption is correlated with his income $inc_i \in [1, 2, 3, 4, 5]$, indicating his income falls in which quintile of Spanish national income distribution. Let K denote the total number of income bins, namely, K=5. In day d, household i from zip code z's consumption at hour h is:

$$kWh_{ih,d} = g(inc_i) \times s_h(inc_i, \{p_{hd}\}_{h=1}^{24}, \{temp_{zhd}\}_{h=1}^{24}) \times \xi_{zhd} \times \varepsilon_{ihd} \tag{B.1}$$

$$g(inc_i) = exp(\eta^\alpha \mathbb{1}(inc_i = k) + \sigma^\alpha \nu_t^i) \tag{B.2}$$

Our zip code level income data has not only the mean income but also the distribution of household income within a zip code. We know the proportion of household that are in each national income quintiles. Therefore, unlike most papers in the literature which assumes inc_i is normal distributed, we assume inc_i follows discrete distributions $inc_i \in [1, 2, 3, 4, 5]$, representing the 1st to the 5th national income quintile. We observe the PMF of income in each zip code.

Where $g(inc_i)$ depicts the level of consumption for households from each income bin, and s_h indicates the share of kWh in this hour. ξ_{zhd} is a zip code-hour-day level error term, and ε is a household-hour-day level error term, both of which follow a normal distribution with a zero-mean.

We assume daily household consumption follows a log-Normal distribution. η^α is a vector with length K, where the kth element is the mean of kWh for income bin k, and σ^α is the scale parameter for this log-Normal distribution. s_h is a function of hourly price p_{hd}, hour-zip code specific temperature $temp_{zhd}$, and household income inc_i. The detailed form is as follows.

$$s_h(inc_i, \{p_{hd}\}_{h=1}^{24}, \{temp_{hd}\}_{h=1}^{24}) = \frac{exp(u_{ih,d})}{1 + \sum_{h=1}^{21} exp(u_{ih,d})} \qquad \forall h \in [4, 23] \tag{B.3}$$

$$u_{ih,d} = [p_{hd} \quad temp_{zhd} \quad 1]\beta_t^i + \epsilon_{ih,d} \qquad \forall h \in [4, 23] \tag{B.4}$$

$$\beta_t^i = \begin{bmatrix} \beta_{t,1}^i \\ \beta_{t,2}^i \\ \beta_{t,0}^i \end{bmatrix} \qquad \forall t \in [1, 2, 3, 4, 5] \tag{B.5}$$

$$= \begin{bmatrix} \beta_{t,1}^0 \\ \beta_{t,2}^0 \\ \beta_{t,0}^0 \end{bmatrix} + \begin{bmatrix} \eta_{t,1}^\beta \\ \eta_{t,2}^\beta \\ \eta_{t,0}^\beta \end{bmatrix} inc_i + \begin{bmatrix} \sigma_{t,1}^\beta \nu_t^i \\ \sigma_{t,2}^\beta \nu_t^i \\ \sigma_{t,0}^\beta \nu_t^i \end{bmatrix} \tag{B.6}$$

Where h is index for an hour and t is index for a time window and t = 1-5 indicates 4am-7am, 8am-11am, 12pm-3pm, 4pm-7pm, and 8pm to 11pm, correspondingly. 0-4 am are defined as outside options, as most electricity consumption is passive during that time window.

We assume in each day, a household's utility of allocating 1 kWh into a certain hour h is (B.4), where β_t^i is explained by both a set of random draws μ_t^i and his income inc_i. The first, second, and third elements of

β_t^i are coefficients for prices, temperature, and constant, correspondingly. We assume the coefficients follow normal distributions, of which the variances are σ^β and the means are $\beta^0 + k \times \eta^\beta \ \forall k \in [1, 2, 3, 4, 5]$. We allow the coefficients to be different across time windows, and the coefficients for different time windows are correlated only through household income.

To simplify the notation, let $\theta^\alpha = (\sigma^\alpha, \eta^\alpha)$ denote all consumption level parameters, and let $\theta^\beta = (\beta_t^0, , \eta_t^\beta, \sigma_t^\beta{}_{t=1}^{t=5})$.

Notice that the levels of household consumption is independent from kWh allocation parameters $\theta^\beta = (\beta_p, \beta_t, \beta_0, eta^\beta)$. Therefore, we can identify consumption level parameters $\theta^\alpha = (\alpha_z, \eta^\alpha)$ separately.

We first identify θ^α using OLS and then use a simulated moment method to estimate θ^β in this demand system. We draw inc_i from the income distribution of a zip code and draw random draws μ, then aggregate the implications for $\overline{kWh}_{zh}$ in zip code, match moments that fit kWh patterns at zip code.

$$\sum_z \omega_z \sum_d \Big(kWh_{zh,d} - kWh_{hd}(\theta^\alpha, \theta^\beta, inc_z) \Big)^2 \quad \forall z, h \tag{B.7}$$

$$kWh_{hd}(\theta, inc_z) = \int_{\theta, inc_i} kWh_{hd}(\theta, \nu, inc_i) dF_{\nu, inc_i} \tag{B.8}$$

The second equation above is the integral of equation (B.1). As individual income only affect the mean of his parameters distributions, the above moment condition is equivalent to:

$$\sum_d \Big(\overline{kWh}_{zh,d} - \sum_k Pr_z(inc_k) kWh_{hd}(inc_k) \Big)^2 \quad \forall z, h \tag{B.9}$$

$$kWh_{hd}(inc_k) = \int_\theta kWh_{hd}(\theta) dF_\theta^k \tag{B.10}$$

Where F_θ^k is the distribution of demand parameters $\theta = (\beta_p, \beta_t, \beta_0)$.

In the parametric approach, correlation between income and types is parameterized by η^β, and it is the same across all zip code. Also the distribution of types $\theta = (\beta_p, \beta_t, \beta_0)$ is the same for all zip code regions conditionally on zip code demographics. We need these assumptions to give us identification power.

Fixed grid nonparametric approach and the connection with our two-step estimator . Fixed grid nonparametric approach To simplify the computation and to better utilize the individual data we have, we consider the BFR approach.

Notice that with individual level data, we can estimate β^i for each individual. Instead of matching aggregate consumption patterns, we can also match the distribution of coefficients.

$$\sum_r^R \Big(Pr(\beta \in B_r)_{zh} - \sum_k Pr_z(inc_k) Pr(\beta \in B_r | inc_k, \theta^\beta) \Big)^2 \quad \forall z \tag{B.11}$$

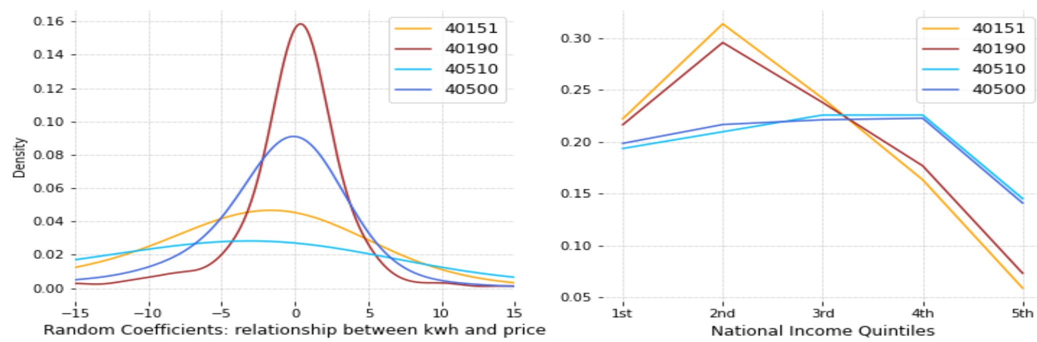
$$Pr(\beta \in B_r | inc_k, \theta^\beta) = \int_{\beta \in B_r} dF(\beta^i | inc_k, \theta^\beta) \tag{B.12}$$

where r is a type index and B_r indicates the range of β^i. If equation (B.12) is linear in θ^β, then the estimator is a simple OLS following the spirit of BFR. (The difference is that BFR knows the choice of each β type, and estimates the distribution of β to match the distribution of consumption. Here we do not know

Table B.1: Estimated distributional impact of Madrid (Loss by income)

	Naive	BLP	BFR	Two-step estimator
1st Quintile	0.54	0.93	0.98	2.43
2nd Quintile	0.46	0.72	0.85	1.77
3rd Quintile	0.38	0.42	0.47	0.38
4th Quintile	0.35	0.13	0.03	-0.22
5th Quintile	0.49	0.07	-0.02	-1.17

Figure B.1: Evidence of heterogeneous distribution of beta, even conditional on zip code income distribution



the choice of each income bin, estimate parameters in this choice, and match it with the distribution of β) Because we assume normal distributed random coefficients, it is effectively a NLS.

If we further relax the normal distribution assumption. The system becomes fully flexible and is almost the same as our two-step estimator, except for the parametric way to classify household types.

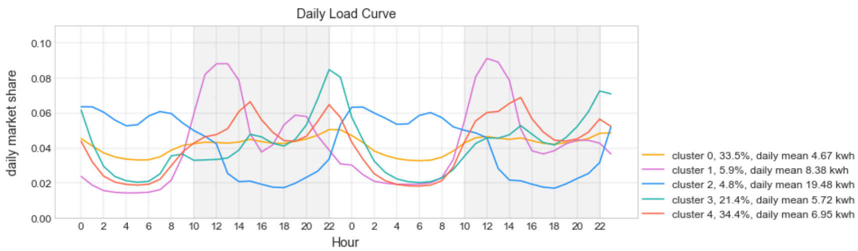
The distributional impact of the BLP and BFR approaches is more closed to the naive approach results, and we cannot identify the correct income for the households through these two approaches. Several reasons lead to the poor performance of parametric approaches in electricity consumption, and we state them one by one in the following section.

Problem of the parametric approach: Evidence 1. According to our data, different zip code regions have different distribution of the random coefficients: the correlation of each household's consumption and income. (even if we conditional on zipcode demographics).

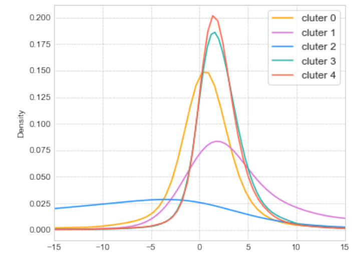
One might think we can make the utility function more flexible, in order to better suit the actual electricity consumption pattern. However, as shown in the following equations, being fully flexible does not work, we will underidentify the parameters. The only parametric way that might work is to do the market shares at household-day level, which treat each household as a "market" and each "household-day" is a "household or choice maker". However, this is computationally impossible.

If we try to modify moment conditions (B.9) and (B.10) into a more flexible way, a natural way would be the following 2 equations. The following 2 equations allow different zip code have different distribution of type θ, denoted by $Pr_z(\theta)$. The relationship between income and household types is now characterized

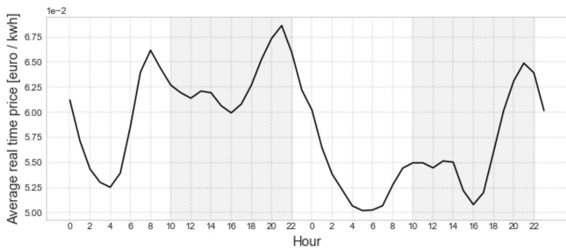
Figure B.2: Evidence of agnostic consumption patterns of different types of households



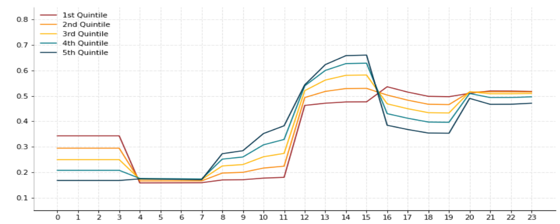
(a) Kmeans types



(b) price coefficient by types



(c) hourly average prices



(d) estimated load curve from BFR

by η^k.

$$\sum_z \omega_z \sum_h \left(\overline{kWh}_{zh} - \sum Pr_z(\theta_n) kWh_h(\theta_n) \right)^2$$
$$\sum_z \omega_z \sum_k \left(Pr_z(inc_k) - \sum_{i \in z} \sum_n \eta_n^k Pr_z(\theta_n) \right)^2$$
$$\text{s.t. } \sum_k \eta_n^k = 1, \forall n.$$

- Being fully flexible does not work here, system greatly underidentified without structure.
- E.g., if one allows as many types as zip codes, assign only one type to a zip code with probability one to perfectly match aggregate moments.

2. The second problem with the parametric approach is that electricity consumption is a highly dimensional choice and price or temperature variation is insufficient to explain it. The β is not a proper primitive to model the consumption, and equation (B.3) is not a proper model for our purpose.

Use a too simplified model would ignore unobserved heterogeneity in the consumption pattern. Two households of different consumption type might have very similar price correlation. If we ignore this, we income estimates would be biased. Figure B.2 gives an example.

The green type and the red type in Figure B.2 have very similar β distribution, but the green type is more peaky in the afternoon and the red type is more peaky in the night. They clearly have different lifestyles (types) and probably have different income distribution. The BLP approach will ignore this unobserved heterogeneity. According to the above figure, it is hard to use any parametric functional form to describe the market share choices of different types. Comparing the consumption pattern and the price variation, we know the β in equation (B.6) is not a sufficient statistics for household types.

C Inferring appliance ownership

In this appendix, we infer appliance ownership by exploiting the richness of the smart meter data. For each individual, we first run the following regression to get their correlation with temperature in winter and summer.

$$\begin{aligned} kWh_{i,hdm} =& \beta^i temp_{hdm} + \beta^i_w temp_{hdm} \times \mathbb{1}(summer \times daytime) + \beta^i_s temp_{hdm} \times \mathbb{1}(winter \times daytime) \\ &+ \alpha^i_{hm} \mathbb{1}(hour \times month) + \alpha^i \mathbb{1}(weekends) + \epsilon_{i,hdm} \end{aligned} \tag{C.1}$$

where $kWh_{i,hdm}$ is hourly consumption of household i in hour h on day d in month m, and $temp_{hdm}$ is the corresponding temperature at the time. α^i, α^i_{hm} are weekends fixed effect and hour-month fixed effect. We include these fixed effects to control for the unobserved consumption heterogeneity. The coefficients of interest are β^i_w and β^i_s, and we include the term $\beta^i temp_{hdm}$ to control for the general trend of each household. β^i_w is how much more a household responds to temperature change in winter than in other times, and β^i_s captures how much more the household responds in summer.

First, if a household has AC, then their hourly consumption would respond to temperature increases in the summer. Therefore, β^i_s should be significantly larger than 0. We only account for daytime responses because according to ... , most people turn off their AC at night. Similarly, if a household has electric heating, their consumption would increase when the temperature decreases. So β^i_w would be significantly smaller (more negative) than 0 during the winter. Second, we look at the mean consumption level for different seasons for all the households: α^i_s, α^i_w, and α^i. If a household has AC (or an electric heating system), then their summer (or winter) average hourly consumption should be higher than at other times, that is, α^i_s (or α^i_w) should be larger than α^i

After getting the seasonal consumption differences in slope and in levels for each household, we estimate the criteria of HVAC ownership by matching the estimated state level market share of HVAC with the surveyed market share. To identify household AC ownership, we do the following optimization:

$$\min_{\underline{\alpha}_s, \underline{\beta}_s} \quad \sum_s (s^{AC}_s - \hat{s}^{AC}_s(\underline{\beta}_s, \underline{\alpha}_s))^2 + \lambda(\hat{s}^{AC}_L(\underline{\beta}_s, \underline{\alpha}_s) - \hat{s}^{AC}_H(\underline{\beta}_s, \underline{\alpha}_s)) \tag{C.2}$$

$$\text{s.t.} \quad \hat{s}^{AC}_s(\underline{\beta}_s, \underline{\alpha}_s) = \sum_{i \in s} \mathbb{1}(\alpha^i_s - \alpha^i > \underline{\alpha}_s) \times \mathbb{1}(\beta^i > \underline{\beta}_s + 1.96\sigma^i_s)) \tag{C.3}$$

$$\hat{s}^{AC}_L(\underline{\beta}_s, \underline{\alpha}_s) = \sum_{i \in LowContractedPower} \mathbb{1}(\alpha^i_s - \alpha^i > \underline{\alpha}_s) \times \mathbb{1}(\beta^i > \underline{\beta}_s + 1.96\sigma^i_s)) \tag{C.4}$$

$$\hat{s}^{AC}_H(\underline{\beta}_s, \underline{\alpha}_s) = \sum_{i \in HighContractedPower} \mathbb{1}(\alpha^i_s - \alpha^i > \underline{\alpha}_s) \times \mathbb{1}(\beta^i > \underline{\beta}_s + 1.96\sigma^i_s)) \tag{C.5}$$

The first term in the objective function is the difference between the estimated AC market share at given criteria for state s $\hat{s}^{AC}_s(\underline{\beta}_s, \underline{\alpha}_s)$ and the surveyed AC market share for state s s^{AC}_s. The second term is a penalizing term that prevents the low contracted power group from having a higher share of AC ownership than households from the high contracted power group. We add this penalty because an AC or an electric heating system requires higher power usage, thus households with these appliances will tend to contract for higher maximum power.

The parameters $(\underline{\beta}_s, \underline{\alpha}_s)$ define the threshold for having an AC or not. We assume that, if a household has an AC, the correlation with temperature at summer should be significantly larger than $\underline{\beta}_s$ and their

Figure C.1: Load curve by appliance ownership

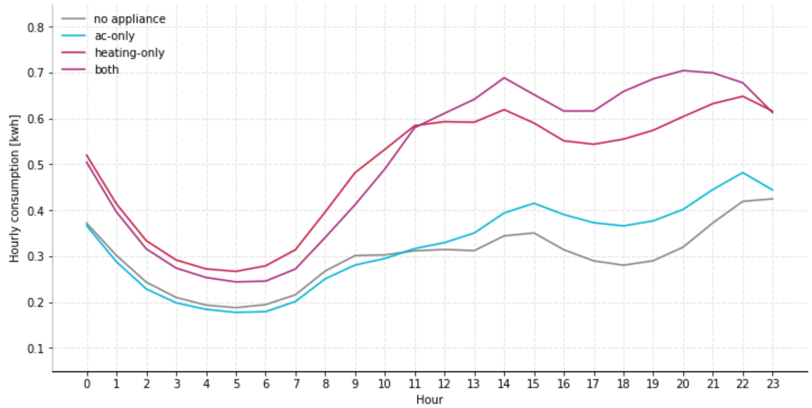


Figure C.2: Hourly consumption pattern [kWh]

average consumption in the summer should be higher than at other times by at least $\underline{\alpha}_s kWh$.

Following the same argument, we do an analogous estimation for the electric heating ownership criteria.

To show that our classification of household types is informative both about their consumption behavior, we plot consumer daily load curve by identified HVAC ownership in Figure C.1. EH owners have relatively high consumption because electric heating devices are in general more energy consuming than AC.

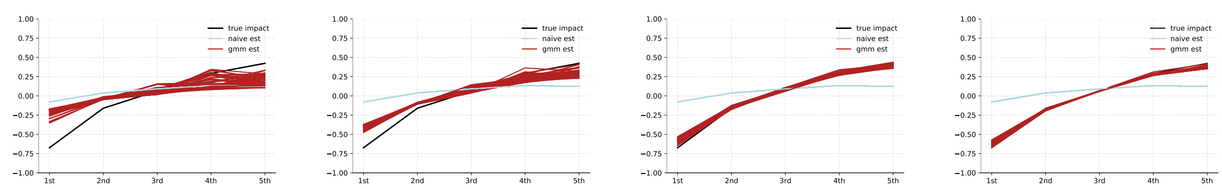
D Monte Carlo for inferring households' income

In order to understand the performance of our estimator in small samples and under misspecification, we perform a Monte Carlo simulation.

- We summarize smart-meter household level data (one entry per household) and assign each household to a group.
- We classify households into 5 types for each group using a kmeans algorithm based on their hourly market shares and total consumption (kmedian algorithm).
- For each group of zip codes, we sort types based on their peak market share (hours 8 - 23).
- To mimic the distributional patterns that we observe in the data, we assign the more "peaky" types to a higher income distribution. This distribution, conditional on type, is the same across all zip codes, but subject to noise. XX I ADD A UNIFORM DRAW TO THE PROB, e.g., $PR(inc_1|\theta_1) = 0.1 + U[-0.03, 0.03]$. This draw is specific to each group of zip codes XX
- Using the assigned types, we assign a zip code number to each household based on a pre-established $Pr(ZIP|THETA)$.
- These zip code numbers and their income distribution, together with the household-level consumption patterns, is what is observed for the estimation.

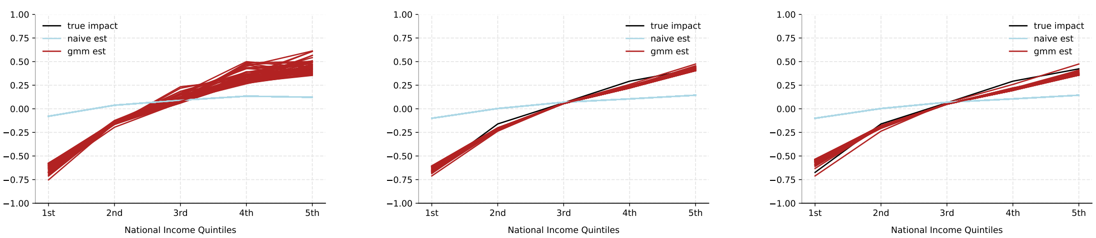
We show that first, the results improve as number of types in the estimation changes from 3 to 5. When the specified type number is smaller than true number of types, the estimated results are biased towards the naive results (Figure D.1). And as expected, accuracy improved with sample size increases. Second, the results are robust to zip code group choice. Results from random assigned zip code groups are in Figure D.2. The results can still converge to the true impact by improving sample size or adding more types in the estimation.

Figure D.1: Simulation Results: different number of types



(a) 3 types, sample size 200 (b) 4 types, sample size 200 (c) 5 types, sample size 200 (d) 5 types, sample size 500

Figure D.2: Simulation Results: wrong zipcode groups



(a) 5 types, wrong zipcode group, sample size 200

(b) 5 types, wrong zipcode group, sample size 1000

(c) 6 types, wrong zipcode group, sample size 500

E Additional Figures and Tables

Table E.1: Statistics on energy availability household heating

State	Heating availability	Electric heating			
		Total	(1)	(2)	(3)
Castilla y León	90.8	8.6	2.0	6.9	0.4
Castilla -La Mancha	86.2	15.3	1.9	13.5	–
Galicia	59.9	14.8	4.1	10.9	0.4
Madrid	90.4	15.6	8.3	8.3	0.5

Notes: The data is from 2008. (1) Individual electric boiler (2) Electric radiators and accumulators (3) Radiant wire. Source: INE

Figure E.1: Geographic Distribution of households

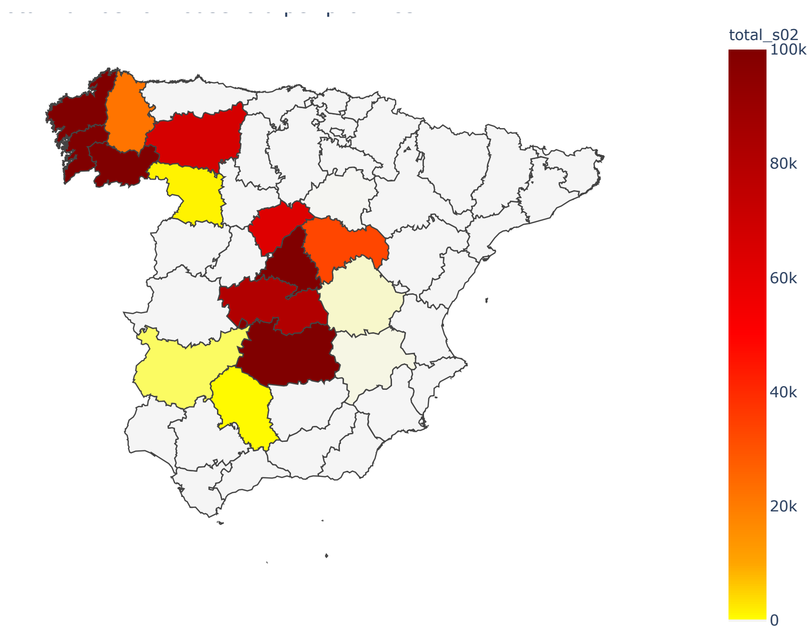


Figure E.2: Aggregate, within month, and cross month effect by different specifications

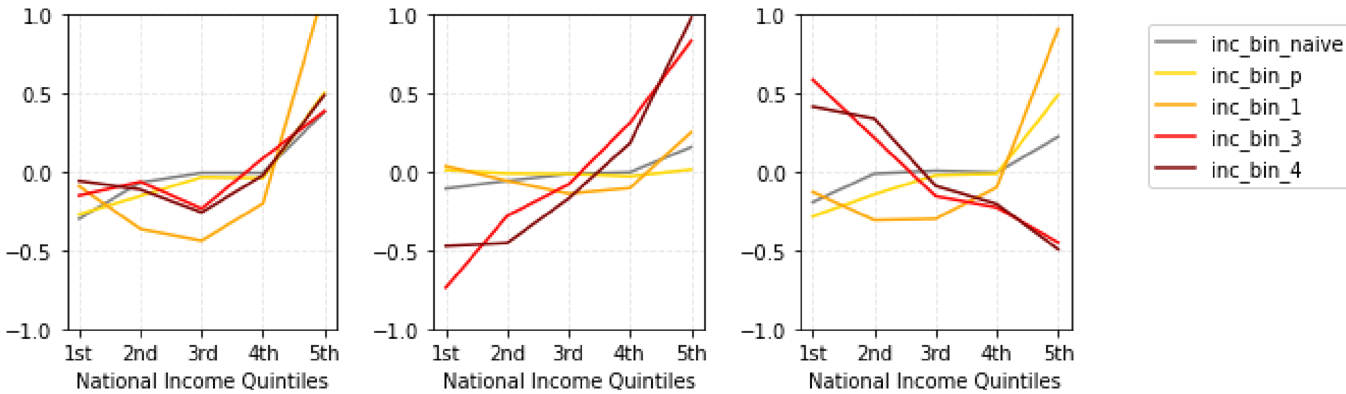
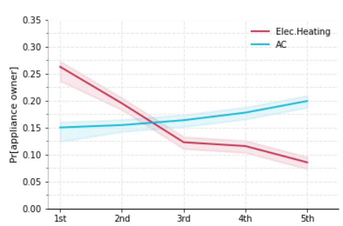
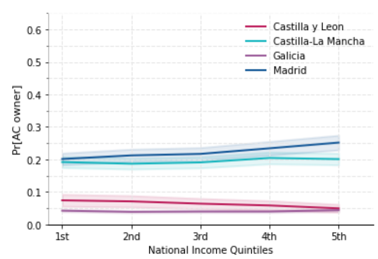


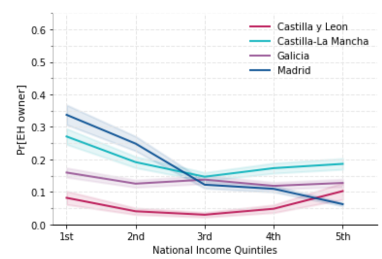
Figure E.3: Distributional impact: geographical heterogeneity and decomposition



(a) Share of electric heating owners and AC owners



(b) AC ownership by state



(c) EH ownership by state