ENERGYECOLAB

The Distributional Impacts of Real-Time Pricing

Michael Cahana[1], Natalia Fabra[2], Mar Reguant[3], and Jingyuan Wang[4]

April 2022

EEL Discussion Paper 113

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 novel method for inferring individual households' income using zip code income distributions. We identify three channels for the distributional impacts of RTP: consumption profiles, appliance ownership, and locations. 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, in the Spanish context, the other two channels make the switch from annual to monthly prices regressive. In particular, since low income households tend to have more electric heating, they benefit from the price insurance provided by time-invariant prices during winter, when prices are 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 price shocks during winter. In the future, the wider adoption of enabling technologies (including storage and demand response

[1] Quora
[2] EnergyEcoLab, Universidad Carlos III and CEPR, natalia.fabra@uc3m.es
[3] Department of Economics, Northwestern University, CEPR and NBER, mar.reguant@northwestern.edu
[4] Department of Economics, Northwestern University, jingyuanwang@u.northwestern.edu

http://energyecolab.uc3m.es/

uc3m | Universidad Carlos III de Madrid
DEPARTAMENTO DE ECONOMÍA

devices) by the high income groups might worsen the distributional impacts of RTP. Our findings should allow to design an equitable real-time pricing system while retaining its efficiency properties.

Keywords: dynamic pricing, electricity, distributional effects, generalized method of moments, clustering.

ENERGYECOLAB

Non-Technical Summary

The Distributional Impacts of Real-Time Pricing

Spain is the only country in the world where, by default, households are subject to real-time prices for retail electricity. Even though they can opt out to the time-invariant prices offered by the unregulated retailers, many households remain under the real-time pricing regime.

For economists, this is generally viewed as good policy. Dynamic pricing creates efficient incentives for energy conservation and for demand shifting across time. If a consumer derives less utility from consuming one additional unit of electricity than the costs of producing it, she should optimally not consume it. Otherwise, she could shift her consumption to a time when production costs are lower, possibly because of lower overall demand, or because of more availability of renewable energy. However, she will only make these efficient decisions if she faces electricity prices that reflect the changing marginal costs of her consumption. Therefore, facing consumers with real-time prices is a necessary condition for efficient consumption. The benefits from this regime are expected to become even stronger with the widespread deployment of renewable energy sources, which are expected to increase cost and price differences over time.

Despite these well-known benefits, policymakers have been reluctant to implement dynamic pricing beyond industrial consumers. One reason might be the fear that low-income households will be hit harder than others, particularly in countries with large shares of energy-poor households.

In this paper, we analyze the distributional implications of the adoption of real-time pricing in Spain using Smart meter hourly consumption data for over 2 Million Spanish households. In particular, we have quantified how their bills change when switching from time-invariant to real-time prices, and have studied how these changes correlate to socio-demographics. Rather than using zip code level income data, we have developed a novel method that leverages the hourly smart-meter consumption data to estimate households' income. We demonstrate that this is a critical step, as most of the distributional implications can attributed to income heterogeneity within each zip code.

We have found that real-time pricing is slightly regressive compared to an annual flat price, a conclusion that can be traced back to two main effects: the "across months" and the "within month" price effects. First, the switch from annual to monthly prices hurts low-income households, who tend to have more electric heating, as they lose the price insurance provided by time-invariant prices during winter, when prices tend to be higher and more volatile. However, the second effect moves in the opposite direction, as a switch from monthly to hourly prices hurts the high-income households, who consume disproportionately more at peak times when real-time prices are higher. The former effect dominates, given that price differences are wider across months than within months (and within days) in our sample.

ENERGYECOLAB

Non-Technical Summary

Figure below illustrates these two effects by plotting the bill changes when moving from an annual time-invariant price to monthly prices (pink bars) and from monthly prices to real-time pricing (red bars), for the five national income quintiles (using our estimated household-level incomes). As the figure shows, the across month effect implies that the first quantiles (i.e., the lower income households) face higher bills under real-time prices, while the bills of the higher quantiles go down. The within month effect hurts the higher quantiles, but this effect is relatively small compared to the bill savings due to the across month effect.

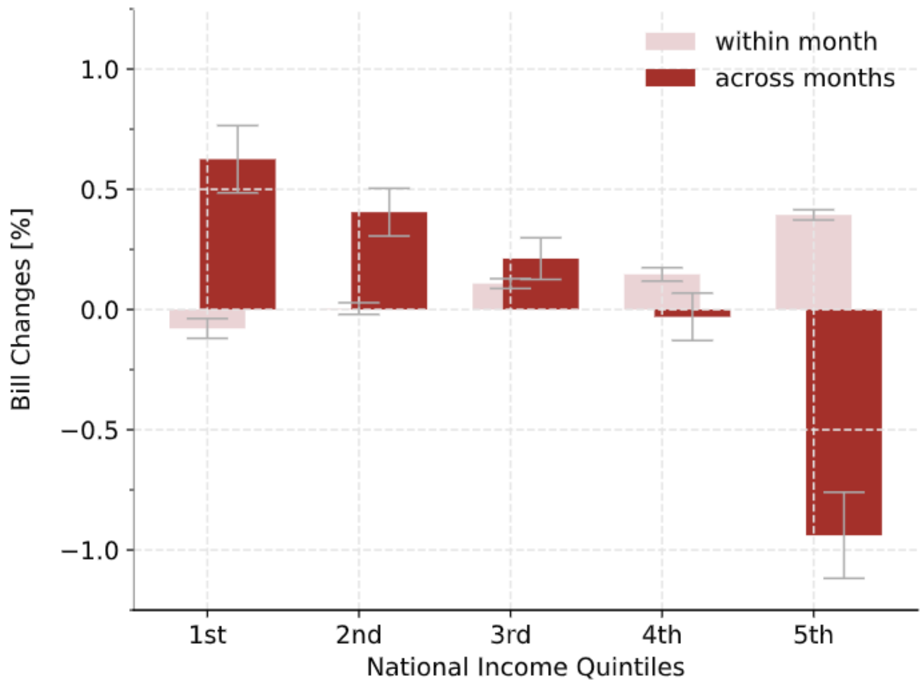


Figure: Bill changes from time-invariant prices to real-time prices per income quintile

Our studies reveal that the switch to real-time pricing in Spain gave rise to a slight regressive effect. In any event, since prices were relatively stable during our sample period, the overall effects were small. The take-away is that real-time pricing did not lead to concerning levels of redistribution. However, patterns could change if price levels and price volatility go up, as it has been the case over recent months. If high-income households are the most capable to respond to price spikes and to benefit from price volatility – for example, by investing in batteries, solar panels or electric vehicles – the regressive impacts of the policy could potentially grow.

These findings are not general condemnations of real-time pricing as a useful policy tool, but rather inform about necessary conditions for it to be successful. First, it might be required to

Non-Technical Summary

put in place public campaigns to increase awareness, and to deploy technology for automatic demand adjustments. This, together with the increasing price differences that will likely arise between hours with scarce and abundant renewables, would help induce measurable changes in consumption patterns in ways that increase overall efficiency. Furthermore, it is important to mitigate the potentially adverse distributional implications of real-time pricing, at least in contexts in which low-income households rely more on electric heating during the high-priced winter months, as is the case in Spain. One option would be to demean retail prices with the annual average, while still maintaining the hourly price signal. This could be achieved by reducing the volumetric charges (that are added to the real-time prices to cover the network costs) during the winter while raising them during the summer. Another option is to put in place policies to mitigate energy poverty; for instance, through consumption subsidies or investments in energy efficiency and solar panels in low-income households, which could be financed through windfall taxes on the inframarginal technologies, as the European Commission is currently proposing.

There is no doubt that a successful energy transition will require smart technologies just as much as smart pricing policies.

The Distributional Impacts of Real-Time Pricing

Michael Cahana Natalia Fabra Mar Reguant Jingyuan Wang*

April 2022

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 novel method for inferring individual households' income using zip code income distributions. We identify three channels for the distributional impacts of RTP: consumption profiles, appliance ownership, and locations. 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, in the Spanish context, the other two channels make the switch from annual to monthly prices regressive. In particular, since low income households tend to have more electric heating, they benefit from the price insurance provided by time-invariant prices during winter, when prices are 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 price 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. Our findings should allow to design an equitable real-time pricing system while retaining its efficiency properties.

Keywords: dynamic pricing, electricity, distributional effects, generalized method of moments, clustering.

JEL Classification: L94, H23, C55.

*Michael Cahana: Quora. 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, jingyuan-wang@u.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 (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 772331-ELECTRIC CHALLENGES) 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).

1 Introduction

Economists have traditionally favoured the use of dynamic prices for electricity retail pricing to more accurately reflect changing marginal costs, which can fluctuate substantially during a day or a week and across the year (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, policymakers 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 policymakers 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.

In this paper we quantify the distributional impacts of dynamic pricing in the Spanish electricity market. We do so by analyzing the bill and socio-economic impacts of a switch from time-invariant prices to real-time pricing (RTP) for a large population of households. We show that uncovering

[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).

households' income heterogeneity is key to properly estimating the distributional impacts of RTP. For this purpose, we propose a novel method for inferring individual households' unobserved income from zip code income distributions. This method should also prove useful in a variety of contexts beyond electricity markets.

Our analysis relies on hourly smart meter electricity consumption data of over 1 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.[4] This implies that the mass of households under RTP is large and representative of the whole population. Since no other country has ever implemented residential RTP as broadly,[5] 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 customers 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.[6] 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 high income customers. Given this mixed evidence, it is important to uncover the heterogeneity across households regarding the link between consumption profiles and income, as well as to perform the analysis over a large sample of representative households. These are two of the strengths of this paper.

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 real-time price of the month or of the year in order to capture different sources of seasonal price variation. Using information on the

[4] 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.

[5] 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. Pébereau and Remmy (2022) examine the barriers for the adoption of RTP in New zeland.

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

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 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 types across zip codes, which in turn have a given 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 income 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 sensitive to choices made by the researcher, we examine the validity of our approach using a Monte Carlo simulation. We perform sensitivity analysis, both in the Monte Carlo as well as in our application, and show that our results are robust to those choices.

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 into 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 RTP 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.

The structure of our paper is as follows. We next discuss the related literature in relation to our findings and methodology. Section 2 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 analyses and Section 7 concludes.

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 or 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 analyzed the distributive implications of increasing block pricing, which is often used in order to promote energy conservation. Even though these issues are related to 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 RTP.

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

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

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 consumption 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.

Instead of analyzing CPP, Horowitz and Lave (2014) use 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.

Our analysis thus contributes to the study of dynamic pricing by performing a detailed analysis of the distributional implications of RTP in the only country where it has been broadly implemented so far. Our results are in line with those of previous studies, as they highlight potential harm to low income households, which very much depend on the elasticity of their consumption to real-time price changes. By identifying the channels that drive these results, our analysis can be informative about the potential effects of RTP in other jurisdictions, which can help mitigate the adverse distributional impacts of RTP before it is implemented.

From a methodological point of view, we present a novel procedure to impute a distribution of income to each individual household using zip code level income distribution data. We follow a two-step estimator and take advantage of the repeated data from every individual. In the first step, using *kmeans* clustering methods, we cluster households and discretize households' group heterogeneity based on their individual-level moments. The second step estimates the households' group-specific income distribution using zip-code-level income moments. In terms of the first step, the energy engineering literature (e.g. Haben et al. (2015), Al-Wakeel et al. (2017), Melzi et al. (2015), and Tureczek and Nielsen (2017)) has used machine learning models to analyze electricity load curves but, as far as we are aware of, ours is one of the first papers using similar techniques to infer income. Furthermore, as we discuss below, we contribute to the literature that has developed methods for clustering observations and combine this with the idea of backing out individual primitives (income) from outcome variables, as in demand estimation models (Berry et al. (2004), Fox et al. (2011), and Bajari et al. (2007)).

The literature on finite mixture models proposes a series of models for inference and clustering. Most papers use a parametric approach with Normal distribution assumptions (McLachlan et al. (2019)). Bonhomme et al. (2016) proposes a nonparametric approach in which repeated data provides identification power. The literature on discrete heterogeneity using clustering methods models grouped heterogeneity differently from the finite mixture models. Bonhomme and Manresa (2015) propose a grouped fixed-effect (GFE) estimator which uses clustering methods to capture time-varying grouped heterogeneity. They show the consistency properties of estimators using clustering methods like *kmeans* when the true heterogeneity is discrete.

Bonhomme et al. (2021) proposes a two-step grouped fixed-effects estimator that uses *kmeans* clustering methods in the first step, and then estimates a model with group fixed-effects in the second step. They show the advantage of GFE over the classic FE model and prove the asymptotic properties of the GFE estimator with panel data. They show that the asymptotic properties also hold when the heterogeneity is continuous. It follows that the discrete heterogeneity is a dimension reduction device rather than a substantive assumption about population unobservables. While our first step is similar to Bonhomme et al. (2021), as we both use individual-level moments to identify groups, our second step differs from theirs. Our second step matches the aggregate moment of the primitives (the zip code level income distribution) and identifies group-specific parameters (the group-specific distribution of income), while theirs is a systematic approach to heterogeneity in a fixed-effects model regression, maximizing the likelihood of observing individual-level outcome variables in the second step.

We also contribute to the literature on demand system estimation with partial microdata. Whereas we have rich individual-level consumption data and repeated observations for every individual, we do not have any income information at the individual level except for the zip code distribution of income. This does not enable us to use micro-moments, as in Berry et al. (2004).

A classic structural approach would estimate a parametric individual consumption function with income as a covariate. Market level demographics could help identify the relationship between income and consumption, as first shown in Berry et al. (1995). However, there are two limitations. First, the inversion aggregating over individual hourly consumption data is computationally burdensome. We simplify the methodology by using our two-step estimator, as the second step's computational burden is almost the same as a constrained OLS estimator. Part of our spirit of discretizing household types is similar to the fixed grid nonparametric approach (Fox et al. (2011) and Bajari et al. (2007)), as we also simplify the computation by discretizing individual types and converting a big structural model into an OLS-style estimator.

Additionally, specifying a parametric model for electricity consumption can be complicated as the parametric relationship between income and electricity consumption can be quite varied and heterogeneous. Such simplification would load many of the variations on the impacts of RTP to the error term.[8] Therefore, an parametric approach can not fully utilizing the high-dimensional repeated observations for each individual. Our proposed two-step estimator allows for household-grouped heterogeneity, and we do not restrict any parametric functional form on the group-outcome variable relationship (e.g., consumption measured in kWh) which we allow to be fully flexible. We further relax how the income covariates affect households' types and allow the type/group distribution to be different across zip codes (even if it is conditional on demographics).

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.[9] 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 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

[8] Appendix B presents the evidence and explains the connections with the parametric estimator and the fixed grid nonparametric estimator in more detail. We relax the parametric assumptions step-by-step to derive our two-step estimator. We also include results using the parametric estimator and the fixed grid estimator.

[9] 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.

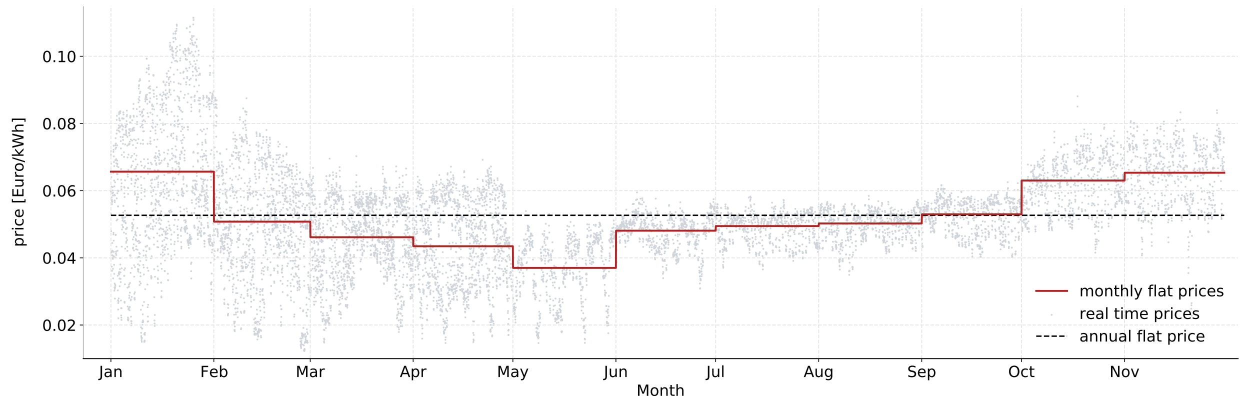


Figure 1: Price fluctuations over time (real-time, monthly and annual prices)

wholesale electricity market operates at the national level, all Spanish households under RTP face, regardless of their geographic location, the same hourly prices, which 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 (∼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 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.[10] After treating for outliers with overly zero consumption observations or missing zip code data,[11] as well as households outside of the regulated territories of the utility,[12] the sample size is reduced to 1,246,783 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 17,371,296 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.[13]

For each individual household i and each month m, we compute the following expressions:[14]

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

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

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

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).[15] We will use the difference $Bill_i^{RTP} - \overline{Bill}_i$ to compute the bill change 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.,

[10] The geographic distribution of households is shown in Figure A.1.

[11] 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.

[12] 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.

[13] 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.

[14] 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).

[15] 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.

a constant revenue-neutral price for each month, and the resulting annual bill,

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

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

This allows us to decompose the bill changes due to the switch from an annual time-invariant rate to RTP as the sum of the "across months" and the "within month" effects. These are respectively captured by the first and second terms of the equation below:

$$\Delta Bill = [\overline{Bill}_i^m - \overline{Bill}_i] + [Bill_i^{RTP} - \overline{Bill}_i^m]. \tag{6}$$

The across months effect reflects the bill impacts due to the monthly price variation across the year, while the within month effect reflects the bill impacts due to the hourly price variation within the month.

Table 1 reports summary statistics for our main variables of analysis. The average annual consumption of a household is around 2,572 kWh, for which on average they pay 135 €/year.[16] There is heterogeneity in households' energy bills, due to differences regarding consumption levels and the timing of their consumption. A move from time-invariant prices to RTP implies that some households lose (on average, losers face 4.52% higher bills) while other households gain (on average, winners enjoy a 3.12% bill reduction). The bill impacts due to the across months price variation are highly heterogeneous across households, ranging from a 1.76% bill increase for the 75% percentile to a 2.16% 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 explain 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, sex, education, dwelling types, and income distribution data)[17], while the latter provides income distribution data at the zip-code-level.[18]

As a first step towards understanding the impacts of RTP and their correlation with income, we regress the logs of average electricity consumption, average peak electricity consumption, and

[16] 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 approximately 2.

[17] 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.

[18] Appendix A provides a more detailed description of these data sources.

Table 1: Summary Statistics (household-annual level)

	Mean	Std	25%	50%	75%
kWh_i	2572.54	1964.24	1314.50	2079.06	3209.26
$Bill_i^{RTP}$	135.47	102.54	69.22	109.84	169.50
$\overline{Bill}_i$	135.47	103.44	69.22	109.48	169.00
$\overline{Bill}_i^m$	135.48	102.98	69.12	109.49	169.16
$\Delta Bill$ [%]	0.13	5.21	-2.63	-0.59	2.34
$\Delta Bill$ (losers) [%]	4.52	4.28	1.13	3.28	6.74
$\Delta Bill$ (winners) [%]	-3.12	2.96	-4.26	-2.24	-1.09
Decomposition:					
$\Delta Bill$ within month [%]	0.09	1.93	-0.79	0.11	0.99
$\Delta Bill$ across months [%]	0.04	4.84	-2.16	-0.85	1.76

Notes: This table reports household level statistics. There are 1,246,783 observations. All units are measured in €, 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 denominator. By construction, the mean changes under RTP are 0 when expressed in Euros, but they are different from zero when expressed in %. The reason is that we first compute the bill change for each household and we then take the average across households.

the bill impacts on income, using a cross-sectional sample at the zip-code-level:

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

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 bill change from the switch to RTP. Positive (negative) coefficients would reflect that households in higher (lower) income zip codes consume more in total, consume more at peak times and pay more under 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 positively correlated with income. Columns (3) shows a positive correlation between income and bill changes, 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 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

Table 2: zip code monthly-level regression results

	ln(kWh)	ln(kWh peak)	$\Delta Bill$ [%]
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.

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 the individual households' income distributions. Let us assume that households' 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},

$$kWh_{ih} = f(x_h, \epsilon_{ih}|\theta_i). \tag{8}$$

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 two steps. In the first step, we classify households into different types based on their contracted power capacity,[19] 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 K 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.

[19] Contracted power capacity is the maximum consumption allowed at any point in time.

The estimation assumes that the income distributions of the same type θ_n from different zip codes are the same, equal to $\eta^{\theta_n} = \{\eta_k^n\}_{k=1}^K$. 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. 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 data and tariff. We observe households' contracted power capacity, which is the maximum consumption allowed at any point in time. It thus tends to be highly correlated with income, as it is a function of the household size and installed electrical equipment. We also identify households HVAC ownership status from the correlation of their hourly consumption and temperature across seasons. Additionally, we use the *kmeans* cluster algorithm to classify households into flexible types based on household-level moments of hourly consumption patterns.

3.1.1 Classification by contracted power capacity

Households face a two-part tariff, including a term for contracted power capacity and an hourly electricity price, i.e., only the second term is as a function of electricity consumption. Contracted power is the maximum allowed hourly consumption for each household. It can vary from 1 to 10kW, but most households in our sample chose 2.5-5 kW. We classify households into two groups: 52% households who have a contracted power less than or equal to 4 kW are in the low contracted power group (L), the other 48% households with a contracted power larger than 4 kW are assigned to the high contracted power group (H).

Classifying along contracted power is powerful in our setting because this is a variable that we get to observe at the household level. Additionally, it tends to correlate well with income, as it is cause of a substantial part of the electricity bill (about a third) and correlated with appliance ownership and house size. Another important aspect affecting contracted power is household size, which does not tend to correlate with income. The classification implicitly captures different aspects of households with high contracted power.[20]

3.1.2 Classification by appliance ownership

We identify appliance ownership by testing the seasonal correlation of hourly consumption and hourly temperature. Intuitively, households who use a relatively high amount of electricity during

[20] In an earlier version of our estimator, we explicitly modelled household size and combined zip-code-level moments of both the income distribution and the household size distribution (mean size and number of single households). Given that it did not change our results substantially, and that we face limitations in the number of moments that we can identify, we only use the distribution of income in this version of the paper.

cold spells are identified as using electric heating, while households who use a disproportionately high amount of power during hot days are identified as using air conditioning.[21] To calibrate the thresholds to assign electric heating or air conditioning, we use a GMM estimator that matches the macro moment of appliance ownership rate at regional level. Details on the identification procedure are included in Appendix C.

The output of the procedure is a household-level indicator on whether they use AC, electric heating, or both, creating therefore a generated variable that allows us to classify households individually. Because our sample covers mostly the north part of Spain where people rarely use air conditioning (AC), and given that we are limited in the number of types that we can allow, we focus on electric heating (EH) for classification in the estimation.

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[22]. We then classify households within zip code groups and get zip code group-specific type spaces, Θ^g with size N^g.

We use a *kmeans* clustering algorithm to classify households based on household-level moments based on their hourly consumption data. In total, 198 variables are generated to capture daily and seasonal consumption patterns for each household. We then apply a *kmeans* clustering 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.
- mean and standard deviation of hourly market share for each of the 24 hours by weekday and weekend.
- 4 variables capturing seasonal patterns in consumption: 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 the monthly flat price.

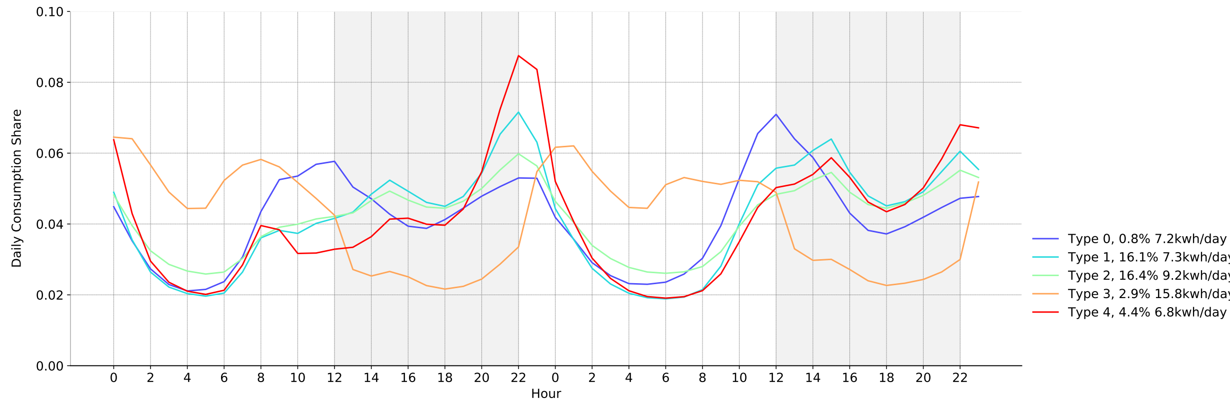
The first two sets of variables (194 variables in total) are explanatory variables for the within month effect. They capture households' lifestyles and explain the correlation of consumption patterns and the real-time price. The last are explanatory variables for the across months effect, as they capture the seasonality of electricity consumption.

For all households in the same zip code group, we first classify them into 4 subgroups based on their contracted power and electric heating ownership. We then use the *kmeans* clustering

[21] See Westermann et al. (2020) and Dyson et al. (2014) for other examples of algorithms that use high-frequency data to infer air conditioning ownership.

[22] We first classify zip codes based on province. The 750 zip codes are from 9 provinces. For the provinces with more than 100 zip codes, we further classify the zip codes into more groups based on population and estimate the model accordingly. If estimation results suggest little heterogeneity across the within-province groups, we keep the whole province as one group in our main specification.

Figure 2: An example of Kmeans types in Madrid



Notes: This figure provides an example of the *kmeans* classification of households in Madrid with high contracted power and no electric heating. The five clusters group households according to their electricity consumption profiles over the day. The first 24 hours are for weekdays and the last 24 hours are for weekends.

algorithm to further classify households within each subgroup into θ_n types. We chose the number of Kmeans types to make sure that the number of households in each type (and within the zip code group) is greater than 1,000 to avoid small sample issues. This implies that if a subgroup contains less than 1,000 households, we would not further classify households in this subgroup using the *kmeans* clustering algorithm.

Figure 2 illustrates an example of the *kmeans* classification. It shows the average daily consumption pattern for weekdays and weekends of all households high contracted power and no electric heating in Madrid. One can see that the algorithm picks up a variety of consumption patterns: households who consume mostly during weekends at brunch time and in the evening (cluster 0), households who consume at lunch time and in the evening (cluster 1 and 2), households who consume in the evening (type 3), and households who consume mostly at night (type 4).

3.2 Step 2: Identifying the Income Distribution of each Type

From step 1, we get distinct household types, θ^i. As we only have data for some central and northern regions in Spain, where AC ownership is uncommon, 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 on whether the household's contracted power is low (L) or high (H), on whether it owns electric heating (EH) or not, and on its Kmeans type. In our main specification we set number of types $N^g = 12$ for all zip code groups, with 3 Kmeans type types within each tariff-EH subgroup,[23] and we include robustness test for alternative specifications. We follow the same estimation procedure for all zip code groups. From now onward, we suppress the superscript g for clarity.

We denote the share of types in zip code j by P_θ^j. To get P_θ^j for each zip code j, we aggregate

[23] $N^g = 12$ is for zip code groups with sufficient population. As explained above, we make sure that the number of households within each type is greater than 1000. When there is too few households in a type, we merge it into other types.

from the estimated individual types:

$$P^j_\Theta(\theta_n) = \frac{1}{HH_j}\sum_i \mathbb{1}(\theta^i = \theta_n), \tag{9}$$

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

$$\min_{\eta} \quad \sum_j \omega_j \sum_{k=1}^{K}(Pr^j_{inc,k} - \sum_{\theta_n\in\Theta} \eta^{\theta_n}_k P^j_\Theta(\theta_n)), \tag{10}$$

$$s.t. \quad \sum_{k=1}^{K} \eta^{\theta_n}_k = 1 \qquad \forall\theta_n \in \Theta. \tag{11}$$

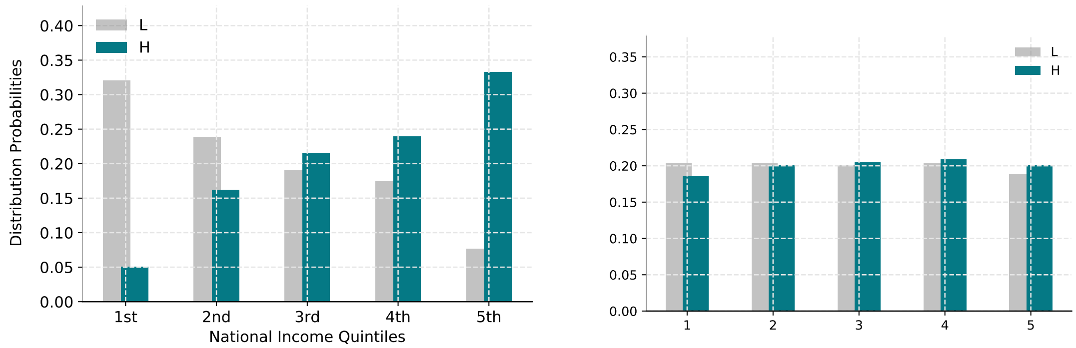
The above objective function (10) uses a set of $(K-1)\times$ *Number of zip codes* moments to identify the $(K-1)\times N$ unknown probabilities of income η, where K is the number of income bins and N is the number of types (12 in the main specification). Thus, we need at least N zip codes to identify η. For this reason we make sure that the number of zip codes in each group is greater than or equal to N.

Using the estimated income for each type $\{\eta^\theta\}_{\theta\in\Theta}$ for each state, we calculate the implied income distribution along several margins. As a sanity check, we show the aggregate distribution of income by contracted power and tariff choice in Figure 3. As expected, Panel (a) shows a positive correlation between income levels and contracted power, with higher (lower) income households being more likely to have high (low) contracted power. Panel (b) shows that one would miss a very substantial part of this correlation if using zip code level income data. Therefore, it highlights the value of our approach, which further classifies households within a zip code into income bins based on a flexible function of their observables.

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.

Figure 3: Estimated income distribution and contracted power



(a) Estimated income distribution by contracted power using a two-step method

(b) Estimated income distribution by contracted power using a naïve approach

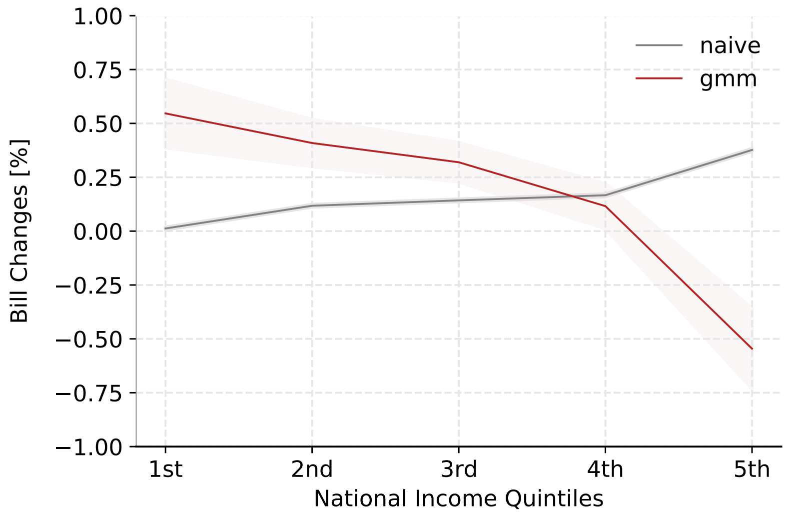
Notes: These figures depict the estimated income distribution by contracted power (low or high). Panel (a) depicts the income distribution using our two step method while Panel (b) depicts uses the zip-code-level income distribution alone. In our dataset, 52% of households are classified as having low contracted power, i.e., below 4 kW. Contracted power is strongly negatively correlated with income, as shown in Panel (a).

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 (naïve approach). Under our proposed approach, the move towards real-time pricing is regressive, as it benefits the high income households while making 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 naïve 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. We show sensitivity analysis of the results in Appendix D.

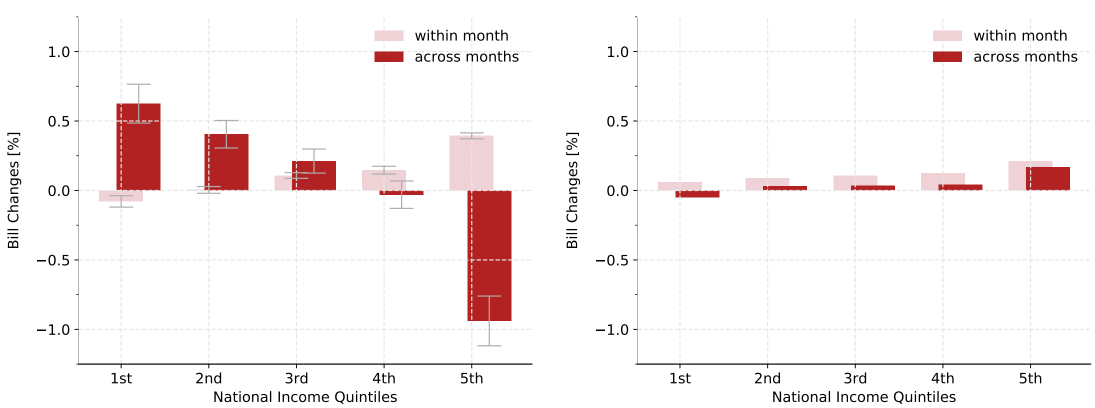
Figure 5 decomposes the bill impacts in changes within and across months. Panel (a), which relies on our estimated income distribution, uncovers a fundamentally different distributional impact depending on the source of price variation. While a move from an annual price to monthly prices is regressive (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.

Figure 4: Bill changes due to the switch to RTP [%]



Notes: This figure represents the bill increase 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 (naïve).

Figure 5: Decomposition of the distributional impact

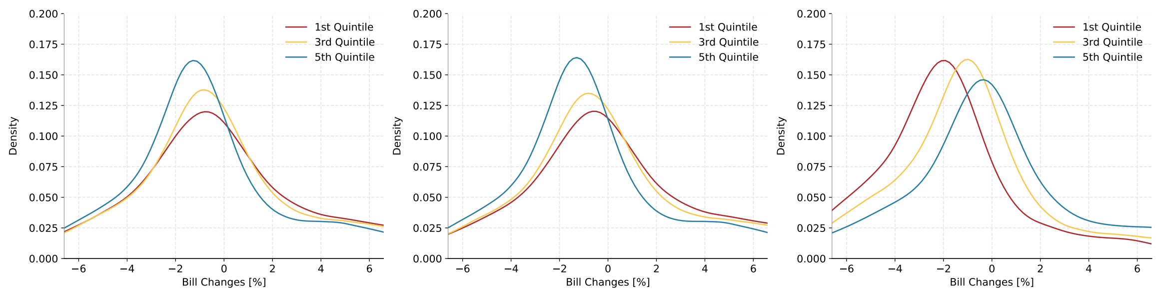


(a) Bill change [%]: GMM approach

(b) Bill change [%]: naïve approach

Notes: These figures decompose the bill change in % when moving from an annual time-invariant price to monthly prices (pink bars) and from monthly prices to RTP (red bars), 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 (naïve 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 changes due to the switch to RTP [%]



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

Notes: These figures show the distribution of the bill changes due to the switch to 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.

4.2 Heterogeneity within Income Bins

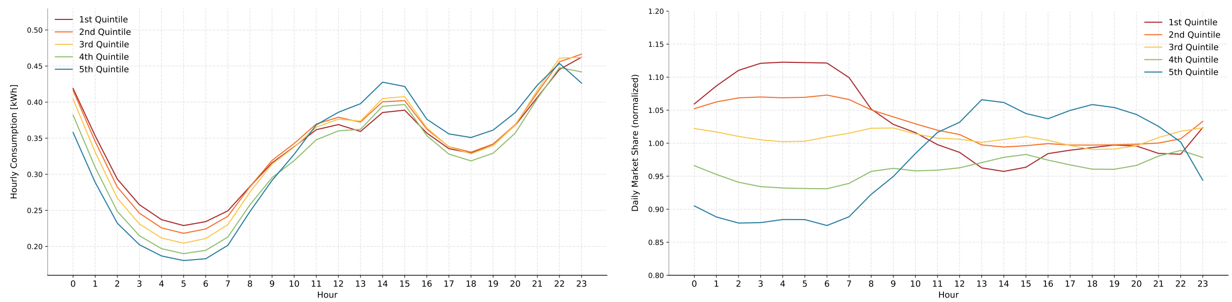
The relatively small average bill impacts at the income group level hide substantial heterogeneity within income groups. Panel (a) in Figure 6 plots the distribution of the percentage bill impacts from moving from an annual time-invariant price to RTP for the 1st, 3rd and 5th quintiles. Whereas most 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 bills.

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 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 across households regarding their consumption profiles, their appliance ownership and their locations.

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.

5.1 Consumption Profiles

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 profiles 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,[24] 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 group consumes relatively more at off-peak times. In other words, the consumption profiles 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 from RTP.

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 correlation between the price coefficients and the income levels (regression (12))[25]

[24]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.

[25]Note that in equation (12), 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.

Table 3: Income, hourly consumption patterns, and within month bill changes

	Price coeff.	$\Delta Bill^m$ [%]	$\Delta Bill^m$ [%]	$\Delta Bill^m$ [%]
Price coeff.		5.482***		5.452***
		(0.040)		(0.040)
2nd quintile	0.082***		0.091	0.042**
	(0.016)		(0.127)	(0.020)
3rd quintile	0.149***		0.312	0.057
	(0.036)		(0.338)	(0.043)
4th quintile	0.347***		0.709***	0.132***
	(0.021)		(0.118)	(0.022)
5th quintile	0.353***		1.149***	0.280***
	(0.018)		(0.134)	(0.026)
R^2	0.203	0.765	-0.133	0.767
N	1135047	1148786	1148890	1148786
FE	zip code	zip code	zip code	zip code

Notes: This table reports regression results for equations (12), to (15) in columns (1) to (4), respectively. A zip code fixed effect is included in all regressions. The number of observations is slightly lower than in Table 1 because we have dropped some households for which a good identification cannot be obtained.

as well as the extent by which the price coefficient explains the within month bill effects (regression (13)):

$$price\ coeff_i = \sum_{k=2}^{5} \beta_k \mathbb{1}(Inc_k)_i + Z_i + \alpha_z + \epsilon_i, \tag{12}$$

$$\Delta Bill_i^m = \gamma\ price\ coeff_i + \alpha_z + \epsilon_i, \tag{13}$$

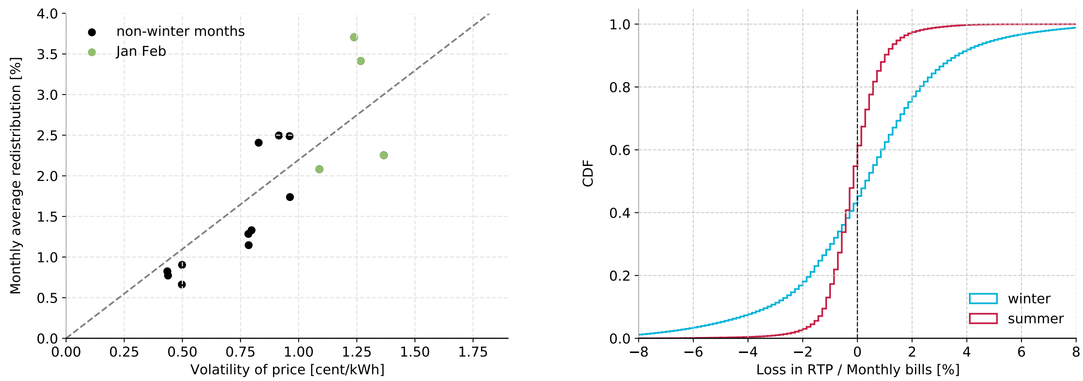
where α_z is a zip code level fixed-effect and Z_i are control variables including households' tariff choices and appliance ownership.[26]

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 bill changes, 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 changes mainly through the correlation

[26] Details about appliance ownership variables are explained in the next subsection.

Figure 8: Impact of price volatility on within month redistribution



(a) monthly price volatility and 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 3 dots with the highest price volatility correspond to January 2017, January 2016, and February 2017. Y-axis of Panel (a) is the average absolute within month bill changes at household level, and this is highly correlated with the monthly price volatility. The higher price volatility during winter months explains why the CDF of the bill changes during winter is flatter than in the summer, as shown in Panel (b).

between consumption and real time prices, as captured by the price coefficients:

$$\Delta Bill_i^m = \sum_{k=2}^{5} \beta_k \mathbb{1}(Inc_k)_i + \alpha_z + \epsilon_i, \tag{14}$$

$$\Delta Bill_i^m = \gamma \ price \ coeff_i + \sum_{k=2}^{5} \beta_k \mathbb{1}(Inc_k)_i + \alpha_z + \epsilon_i. \tag{15}$$

As shown in column (3) of Table 3, income is correlated with the within month bill increase. However, if controlled for the price coefficient, the direct effect of income on the within month impact 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 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 standard deviation of hourly prices within a month, and the monthly redistribution effect, defined as the sum of bill changes (in absolute value) including all households. In months with more price volatility, bill changes can go up to 2.5-3.5%, but the changes remain low at many other times of the year. 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, 23% of the 5th quintile and only 18% of the 1st have AC. In contrast, 32% of the 1st quintile and 11% of the 5th have electric heating.[27] Since prices tend to be higher during the winter 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 either electricity consumption or the bill changes due to switch to RTP:

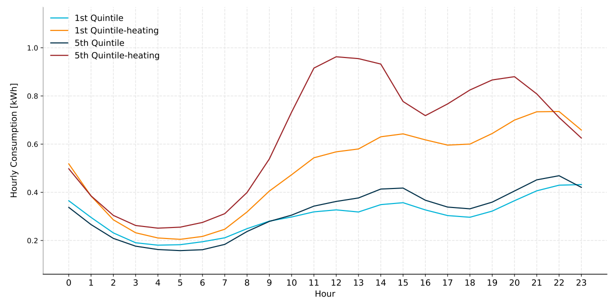
$$Y_i = \beta^{AC}\mathrm{AC}_i + \beta^{EH}\mathrm{EH}_i + Z_i + \alpha_z + \epsilon_i, \tag{16}$$

where Y_i is either kWh_i or $\Delta Bill_i$, 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 bill changes.

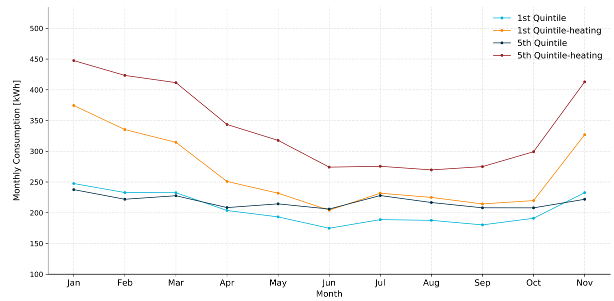
The estimates show that AC increases annual electricity consumption by 204.03 kWh, i.e., 9.6% of a median household's annual consumption. For electric heating, the increase in annual consumption is five times higher, i.e., 1,006.58 kWh, representing 47.2% of a median household's annual consumption. Through their effects on households' consumption patterns, electric heating increases bills under RTP by 2.9%, but AC leads to a 1.1% lower bills. These opposite signs have a simple explanation: electric heating (AC) increases consumption during winter (summer) months, when prices are higher (lower).

[27] These results are reported in Figure C.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.

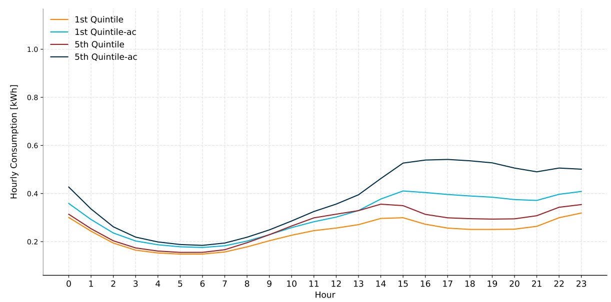
Figure 9: Load curves by appliance ownership and income



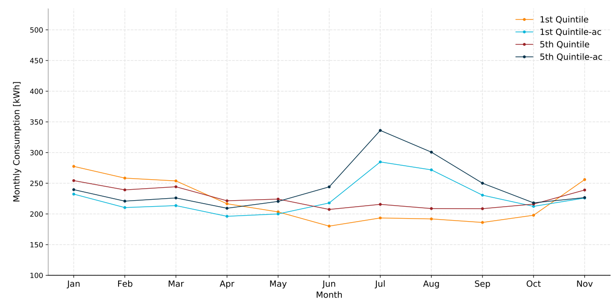
(a) Hourly consumption pattern [kWh] - Heating



(b) Annual consumption pattern [kWh] - Heating



(c) Hourly consumption pattern [kWh] - AC



(d) Annual consumption pattern [kWh] - AC

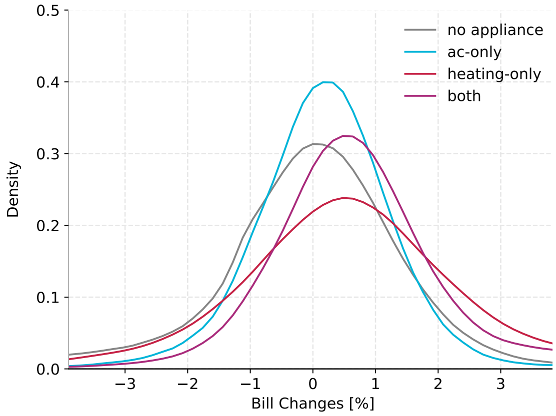
Notes: These figures show consumption profiles over the day (the left panels) and the year (the right panels) for households with electric heating (the upper panels) and AC (the lower panels). Results are reported for low (1st quintile) and high income households (5th quintile). The lines are mean hourly consumption for each group of consumers, truncating the top 1 percentile kWh observations.

Table 4: Appliance ownership and bill changes due to the switch to RTP

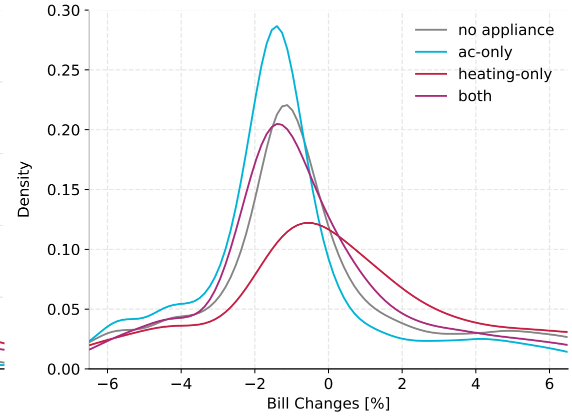
	kWh	$\Delta Bill$ [%]	$\Delta Bill^m$ [%]	$\Delta Bill^a$ [%]
AC	323.546***	-0.638***	0.194***	-0.832***
	(5.141)	(0.014)	(0.005)	(0.013)
Heating	970.297***	2.800***	1.367***	1.434***
	(4.954)	(0.013)	(0.005)	(0.012)
R^2	0.178	0.170	0.111	0.165
N	1135098	1135098	1135098	1135098
FE	zip code	zip code	zip code	zip code

Notes: Column (1) reports the regression results from estimating equation (16) for consumption as the dependent variable, and columns (2)-(4) for the total bill change, the within month bill change, and the across months bill change, respectively. A zip code fixed effect is included in all regressions. One can see that households with AC (electric heating) tend to pay less (more) under RTP. This gain (loss) is mainly driven by the across months effect.

Figure 10: Bill changes [%] by appliance ownership



(a) Within month effects



(b) Across months effects

Notes: These figures plot the distribution of the bill changes due to the switch to 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 bill increases are suffered by households with electric heating due to the across months effect.

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).[28]

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

[28]See Table C.1 in the Appendix for details.

Table 5: Average bill increase by region

	$\Delta Bill$ [%]	$\Delta Bill^m$ [%]	$\Delta Bill^a$ [%]
Castilla y Leon	-0.450***	-0.143***	-0.306***
	(0.015)	(0.006)	(0.014)
Castilla-La Mancha	-0.603***	0.131***	-0.734***
	(0.011)	(0.004)	(0.010)
Galicia	0.645***	0.016***	0.630***
	(0.007)	(0.003)	(0.007)
Madrid	0.129***	0.269***	-0.139***
	(0.009)	(0.003)	(0.008)
R^2	0.009	0.005	0.012
N	1227302	1227302	1227302

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

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 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 bill changes on regional dummies, with the coefficients capturing the mean bill increase under RTP for each region. Results are reported in Table 5. 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.

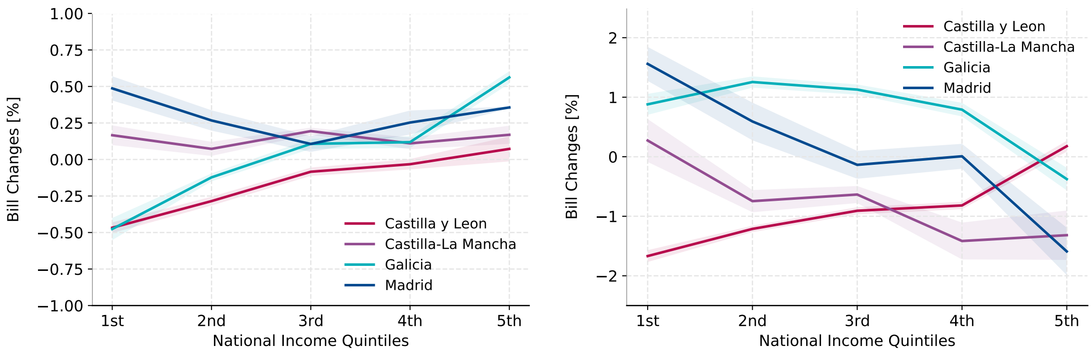
6 Counterfactual Experiments

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 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

Figure 11: Geographical heterogeneity and decomposition of the distributional impact



(a) Estimated bill change: within month

(b) Estimated bill change: across months

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).

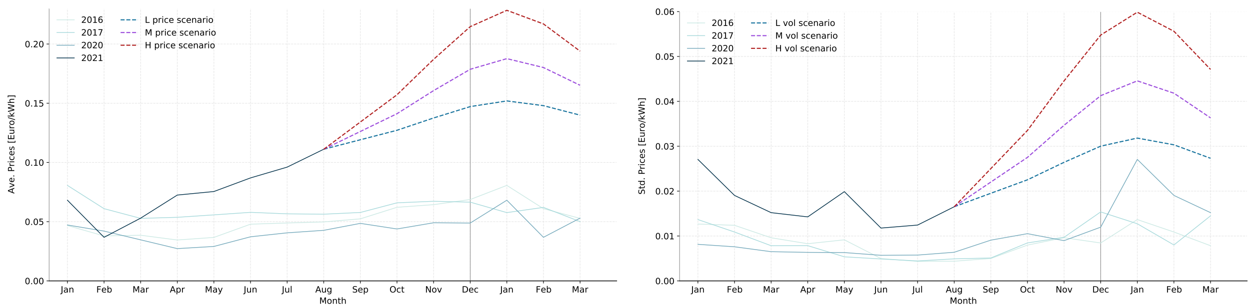
opt out (see Figure 3). Second, the price shock was particularly strong during the winter months, which hit the low income households harder because they tend to have relatively more electric heating. And third, price levels increased overall without affecting the within day price variation much, which 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.

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. Specifically, we simulate prices according to the following equation:

$$\widehat{p}^{21}_{hdm} = (p^{20}_{hdm} - \overline{p}^{20}_{m}) \times \frac{\sigma^{21}_{m}}{\sigma^{20}_{m}} + p^{21}_{m},$$

where $\widehat{p}^{21}_{hdm}$ is the 2021 simulated price for hour h in day d and month m; p^{20}_{hdm} is the actual price in 2020 at that same date; and p^{20}_{m} and σ^{20}_{m} are the mean prices and standard deviation of prices in month m in 2020. Last, p^{21}_{m} and σ^{21}_{m} are the factors by which we scale prices and the standard deviation of prices, as plotted in Figure 12. We simulate the distributional impacts under nine scenarios with high, medium, and low monthly average prices and with high, medium, and low

Figure 12: Simulating a 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 for the low, medium and high price scenarios and the middle volatility. 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 for the middle price scenario. The actual monthly mean prices from September to December 2021 are close to our high price scenario, and the actual volatility is even higher than our high volatility scenario.

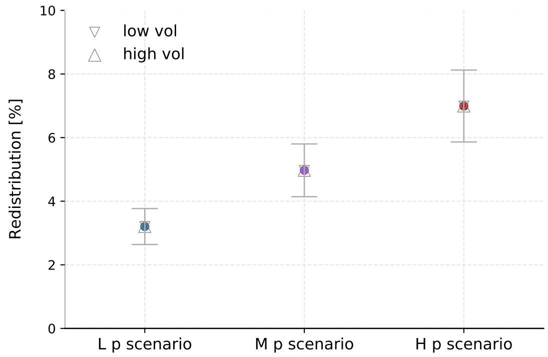
monthly price volatility.[29] The actual monthly mean prices from September to December 2021 are close to our high price scenario, and the actual volatility is even higher than our high volatility scenario.

Keeping households' demand patterns fixed, Figure 13 reports the distributional effects of switching to RTP using the actual and simulated prices for 2021-2022 under the nine scenarios described above. As it can be seen, the effects of RTP are regressive, and the magnitude of the distributional impact is greater than the one reported in the previous section. On average, a low-income household's bill increases 7% more than high income household. 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, a question to which we turn in the next section.

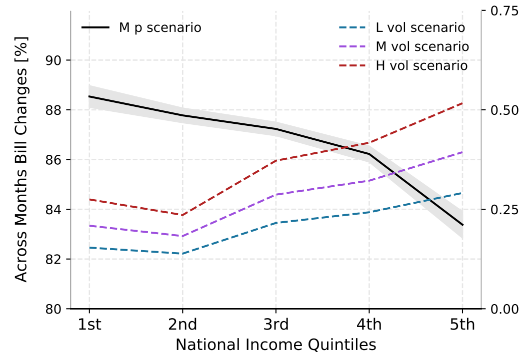
In Section 5 we emphasized the importance of taking into account the within income group heterogeneity, and concluded that location and appliance ownership are two main drivers behind the distributional impacts. We now explore this further by analyzing the heterogeneous impacts due to the various price shocks. Figure 14 shows the average bill impact of the switch to RTP across regions for households with and without electric heating, under the three price scenarios. As it can be seen, the difference between the light and dark red dots (representing the difference in the bill impact for households with and without electric heating under the high price scenario) is wider than the difference between the light and dark blue dots (representing the difference in the bill impact for households with and without electric heating under the low price scenario).

[29]Note that we do not have any December data. Therefore, we exclude December from our counterfactual simulation, which might lower our estimated impacts.

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, which would otherwise benefit the low income households). The volatility effect is nevertheless much weaker than the price effect. Panel (b) shows the distributional impact due to the across months channel (solid line; scale on the left axis) and due to the within month channel (dashed lines; scale on right axis) for the middle price scenario. The across months effect 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 of the policy change.

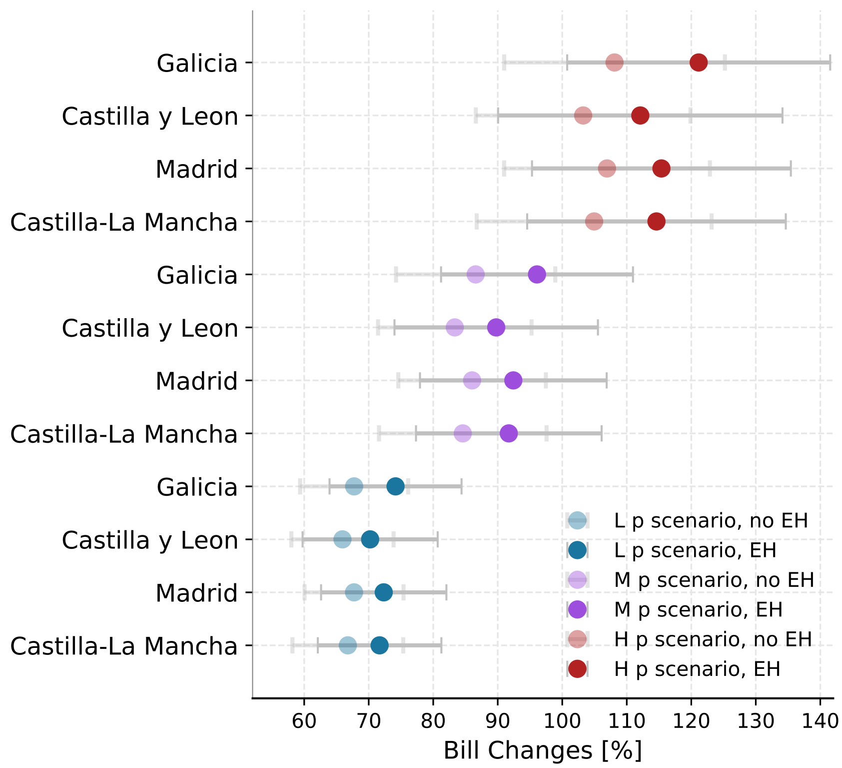
Also, for a given price shock, the difference between the most impacted region, which is also the lowest income region in our sample (Galicia), and the other regions is greater under the high price scenario. It follows that large price shocks increase the heterogeneity in the bill impacts across households through the two channels: location and appliance ownership.

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 these investments, and hence the scope for demand elasticity, will be positively

Figure 14: Heterogeneous impacts of the price shocks



Notes: This figure shows the impact of price shocks on real-time pricing bill changes in % under low, medium, and high price scenarios. Different colors represent different price scenarios, as shown in Figures 12 and 13. Price volatility is set at the medium level. The dots represent average impacts of each appliance ownership-location group and the gray lines represent the standard deviations. Regions are ordered from North to South in three blocks for the high, medium and low price scenarios. The darker dots represent households with electric heating, while the lighter dots are for the remaining households. As expected, the differences between the red dots become much larger than between the blue ones, i.e., stronger price shocks enlarge the differences in the bill impacts across the heating and non-heating households and across locations.

correlated with households' income. Furthermore, this equipment mostly provides insurance to short-run price changes, not to the seasonal price fluctuations that mostly affect the low income households.

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

$$kWh^e_{i,hd} = kWh_{i,hd} \times \left[1 + \frac{p_{hd} - \overline{p}_d}{A + \overline{p}_d} \times \tau_i\right], \tag{17}$$
$$\tau_i = -\alpha \hat{inc}_i,$$

where $kWh^e_{i,hdm}$ and $kWh_{i,hdm}$ denote household i's adjusted and actual consumption, respectively; p_{hd} is the real-time price, $\overline{p}_d$ is the daily average price, A is the bill's fixed fee,[30] τ_i is a negative parameter indicating the household's elasticity, α is a scale factor to adjust the elasticity to a reasonable magnitude, and $\hat{inc}_i$ is the household's estimated income. Because total consumption during the day tends to be relatively inelastic, we only allow households to adjust the timing of their consumption within the day by moving elastic activities to low price periods. In other words, households reduce (increase) their consumption at times when the real-time price is higher (lower) than the daily average. The magnitude of the change depends on the value of τ_i, which is positively correlated with income (in absolute terms).

Figure 15 shows the distributional impacts of a switch to RTP with elastic consumers under the assumption that income is positively correlated with the demand elasticity. As expected, as compared to our baseline results (solid red line in the figure), demand elasticity reduces the bills of the high income households under RTP relatively more given that they can adjust their consumption to the price changes. In the figure, we have considered two assumptions regarding the elasticity. The dashed line shows the results assuming an elasticity between 0.05-0.3. The dotted line shows the results assuming that the elasticity allows for maximum savings of 10% of the total bill.[31] Under the latter assumption, the within month impact (shown in panel b of Figure 15) also becomes slightly regressive as high income people adopt smart devices that allow them to also respond to the within day price changes.

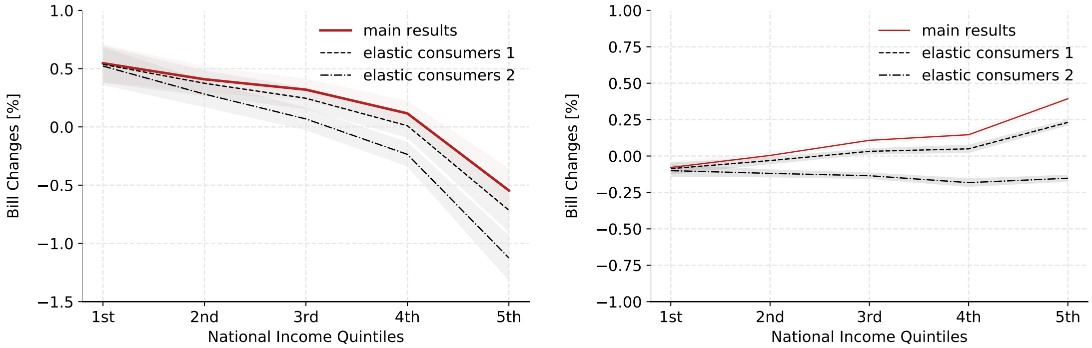
7 Conclusions

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) show that this regulatory change had little impact on households' electricity consumption, the question of

[30]As explained in the background section, the energy prices account for about 50% of the hourly prices, while the volumetric fixed fee accounts for the other half. This implies that hourly price fluctuations (the term before τ_i in equation (17)) generally do not exceed 50%. In our data, hourly price fluctuations usually vary from -5% to 5% around the daily average prices, and they can be as large as +/- 30%.

[31]This assumption is motivated by the fact that smart devices like Thermostat report average savings of 10%-15%.

Figure 15: Distributional implications of RTP under demand elasticity



(a) Aggregate impact

(b) Decomposition: within month impact

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

whether households were asymmetrically impacted by it remains 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 channel 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.

These findings are not a criticism to real-time pricing as a useful policy tool. Rather, they inform about its potential distributional effects in ways that should allow to design an equitable real-time pricing system. Our results highlight that the within day/month price signal can be preserved as it does not give rise to distributional concerns. This price signal is the relevant one for households to possibly adjust their electricity consumption within day/month, which is the most plausible source of demand elasticity. In contrast, the regressive effects stem from the across months price variation, as low income households lose the hedge against the high winter prices when they consume the most. While there are several options to achieve this,[32] understanding the channels by which RTP affects the various consumer groups is in any event necessary to design pricing schemes that retain the efficiency of RTP while making it socially feasible.

[32]One option would be to demean retail prices with the annual average, while still maintaining the hourly price signal.

References

Al-Wakeel, A., J. Wu, and N. Jenkins (2017). K-means based load estimation of domestic smart meter measurements. *Applied energy 194*, 333–342.

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

Bajari, P., J. T. Fox, and S. P. Ryan (2007). Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients. *American Economic Review 97*(2), 459–463.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy 112*(1), 68–105.

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.

Bonhomme, S., K. Jochmans, and J. Robin (2016). Non-parametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(1), 211–229.

Bonhomme, S., T. Lamadon, and E. Manresa (2021). Discretizing Unobserved Heterogeneity. *SSRN Electronic Journal*.

Bonhomme, S. and E. Manresa (2015). Grouped Patterns of Heterogeneity in Panel Data. *Econometrica 83*(3), 1147–1184.

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. Pérez-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.

Dyson, M. E., S. D. Borgeson, M. D. Tabone, and D. S. Callaway (2014). Using smart meter data to estimate demand response potential, with application to solar energy integration. *Energy Policy 73*, 607–619.

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.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics 2*(3), 381–418.

Haben, S., C. Singleton, and P. Grindrod (2015). Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE transactions on smart grid 7*(1), 136–144.

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, January). 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).

McLachlan, G. J., S. X. Lee, and S. I. Rathnayake (2019). Finite Mixture Models. *Annual Review of Statistics and Its Application 6*(1), 355–378.

Melzi, F. N., M. H. Zayani, A. B. Hamida, A. Same, and L. Oukhellou (2015). Identifying daily electric consumption patterns from smart meter data by means of clustering algorithms. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 1136–1141. IEEE.

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

Pébereau, C. and K. Remmy (2022, March). Barriers to Real-Time Electricity Pricing: Evidence From New Zealand. CRC TR 224 Discussion Paper Series crctr224$_2022_339, University of Bonn and University of Mannheim, Germany$.

Tureczek, A. M. and P. S. Nielsen (2017). Structured literature review of electricity consumption classification using smart meter data. *Energies 10*(5), 584.

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.

Westermann, P., C. Deb, A. Schlueter, and R. Evins (2020). Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. *Applied Energy 264*, 114715.

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

"The Distributional Impacts of Real-Time Pricing"

A Data Sources

In this appendix, we provide further details about the demographic data that we use in our analysis. These data are 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.[33] We only include places from which we have electricity consumption data. This limits our analysis to four regions: Galicia, Gastilla y Leon, Madrid, and Gastilla-La Mancha. Figure A.1 plots the location of these provinces.

We complement the data from the INE with data from MB Research at the postal code level.[34] 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.[35] On average, postal codes are matched to around 7 census districts. Once census districts and postal codes are matched, census district data are aggregated at 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").[36] A district section and a zip code are matched if the latitude and longitude of the address lays inside that section.

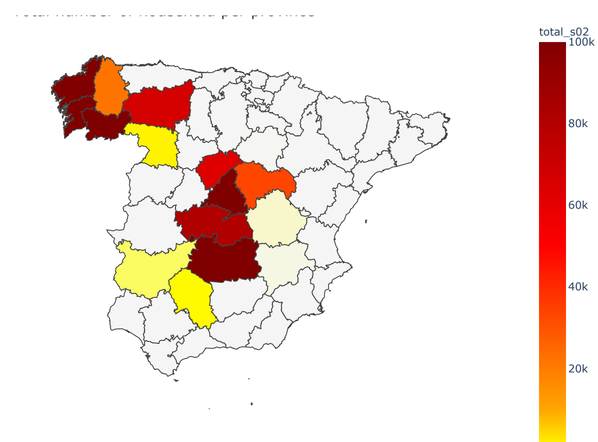
[33]For confidentiality reasons, sections are often not reported as they are a fairly small geographical units. For small to medium sized municipalities, data are often only available at the municipality level, which often coincides with the postal code. Very small municipalities might not have their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

[34]INE data reports median and mean income per household for each census. MB Research reports the distribution of household income, where the cutoffs are representative of the quintiles in the national distribution of income. Therefore, these two distributions of income complement each other at different parts of the support.

[35]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.

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

Figure A.1: Geographic distribution of households



B Alternative estimator

We consider an alternative estimator to our two-step approach, a more parametric approach to describe the relationship between income and electricity consumption. As explained in the main text, we assume that households' 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, their type θ_i, and some random shocks ϵ_{ih}. We reproduce equation (8) here:

$$kWh_{ih} = f(x_h, \epsilon_{ih}|\theta_i).$$

Allowing the household's type θ_i to correlate with its income helps us identify how income affects electricity consumption. This section uses a parametric model to describe the function f and to identify the relationship between household income and consumption, following the spirit of Berry et al. (1995) and Berry et al. (2004).

Just as we did in the main text, we estimate the demand system for each province separately because of the large heterogeneity across regions; see Section 5. All the parameters are province specific. To save notation, we suppress the province subscripts.

B.1 Alternative parametric approach

Assume that each household is optimizing daily consumption across 24 hours. Thus, we can think of each household-day as a "market." which we repeatedly observe every day. A household's consumption is correlated with his income $inc_i \in [1, 2, 3, 4, 5]$, indicating each quintile of the Spanish national income distribution. Let K denote the total number of income bins, with $K = 5$.

Our income data contains zip code level average income per household and the distribution of households' income within each zip code. We know the proportion of households that are in each national income quintile. Therefore, unlike most papers in the literature which assume inc_i follows normal distributions, we assume that inc_i follows discrete distributions $inc_i \in [1, 2, 3, 4, 5]$, representing the 1st to the 5th national income quintiles. We observe the PMF of income in each zip code.

In day d, household i from zip code z's consumption at hour h is:

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

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

We use $g(inc_i)$ to denote the daily household's scale factor and s_h denotes the share of kWh consumed in this hour. Note that we allow income to affect both of these processes. ξ_{zhd} is a zip code-hour-day level error term, and ε_{ihd} is a household-hour-day level error term, both of which follow a normal distribution with a zero-mean.

We assume that the factor $g(inc_i)$ follows a log-Normal distribution. η^{α} is a vector with length

K, where the kth element is the mean electricity consumption for income bin k, and σ^α is the scale parameter for this log-Normal distribution

We assume that the share s_h is a function of hourly prices p_{hd}, hour-zip code specific temperature $temp_{zhd}$, and household income inc_i. The detailed form is as follows:

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

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

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

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

where h is an index for an hour and t is an index for a time window, and $t = 1, ..., 5$ indicates the time intervals 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 a constant, respectively. We assume these random 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 $\theta^\beta = (\{\beta_t^0, , \eta_t^\beta, \sigma_t^\beta\}_{t=1}^{t=5})$ denote the parameter related to the consumption allocation within a day. Notice that the level of household consumption is independent from the kWh allocation parameters $\theta^\beta = (\beta_p, \beta_t, \beta_0, \eta^\beta)$. Therefore, we can identify consumption level parameters $\theta^\alpha = (\sigma^\alpha, \eta^\alpha)$ separately.

We use a simulated moment method to estimate θ^α and 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.

$$\sum_d \left(kWh_{zh,d} - kWh_{hd}(\theta^\alpha, \theta^\beta, inc_z) \right)^2 \quad \forall z, h, \quad \text{(B.7)}$$

$$kWh_{hd}(\theta, inc_z) = \int_{inc_i} \int_{\nu^i} kWh_{hd}(\theta, \nu, inc_i) dF_\nu dF_{inc_i}^z \quad \forall z, h, d. \quad \text{(B.8)}$$

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:

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

$$kWh_{hd}(inc_k) = \int_{\nu^i} kWh_{hd}(\theta^i) dF_\theta^k, \quad \forall k, h, d, \tag{B.10}$$

where F_θ^k is the distribution of demand parameters $\theta^i = (\beta_1^i, \beta_2^i, \beta_0^i)$.

In the parametric approach, the correlation between income and types is parameterized by η^β, and it is the same across all zip codes. Also the distribution of types $\theta^i = (\beta_1^i, \beta_2^i, \beta_0^i)$ is the same for all zip code regions conditionally on zip code demographics. We need these assumptions to give us identification power.

To make use of the repeated data for each household, we add a set of covariance moments to capture the relative attractiveness of different hours to the same consumer. These covariance moments allow households with a higher coefficient in the morning time window to have a lower (or higher) coefficient in the noon time window. They connect utility for different hours for the same household and help identify the distribution of the random coefficients, β_t^i. Because we assume that the β_t^i for different time windows can only be correlated through inc_i, these covariance moments help identify the income coefficient η^β.

Because data suggest considerable heterogeneity in household consumption patterns across months, we allow households to have month-specific utility, i.e., equation (B.1) and all the coefficients above are month-specific. Therefore, similar to the covariance moments above, we add one more set of across-month covariance moments to make use of the relationship between summer and winter coefficients, which gives us identification power from the seasonal patterns.

Because there are thousands of covariance moments by combining all the time windows and all the months, it is hard to use all the across time windows and across months variations (which is why we propose our two-step estimator). Due to the computational burden, we only use data from January and August 2016 and only use the following two sets of covariance moments:

$$\left[Cov\Big(kWh_{zh1,i}^m, kWh_{zh2,i}^m\Big) - Cov\Big(kWh_{h1,d}^m(\theta^\alpha, \theta^\beta, inc_z), kWh_{h2,d}^m(\theta^\alpha, \theta^\beta, inc_z)\Big)\right]^2 \quad \forall m, (h1, h2) \text{ pair}, \tag{B.11}$$

$$\left[Cov\Big(kWh_{zh,i}^{m1}, kWh_{zh,i}^{m2}\Big) - Cov\Big(kWh_{h,d}^{m1}(\theta^\alpha, \theta^\beta, inc_z), kWh_{h,d}^{m2}(\theta^\alpha, \theta^\beta, inc_z)\Big)\right]^2 \quad \forall h, (m1, m2) \text{ pair}, \tag{B.12}$$

where m1 indicates January 2016 and m2 indicates August 2016.

The estimated distributional impact of the parametric approach can be computed by predicting the bill impacts implied by different pricing schemes at different income levels. The results are presented in Table B.1. One can see that the results are closer to the naïve approach in the sense that they substantially underestimate differences across income bins. Several reasons may lead to the poor performance of these parametric approaches to electricity consumption, which we discuss

Table B.1: Estimated distributional impact of Madrid (Bill changes [%] by income)

	Naïve	Parametric approach	Two-step estimator
1st Quintile	0.54	0.93	1.73
	–	(0.82)	(0.12)
2nd Quintile	0.46	0.72	0.77
	–	(0.68)	(0.13)
3rd Quintile	0.38	0.42	0.04
	–	(0.11)	(0.11)
4th Quintile	0.35	0.13	0.09
	–	(0.05)	(0.09)
5th Quintile	0.49	-0.05	-1.47
	–	(0.06)	(0.19)

below.

B.2 Limitations of alternative estimator

Unobserved Heterogeneity within region The parametric approach uses demographics and the normal distributed error term $\sigma\mu^i$ to explain heterogeneity across types. It assumes that the distribution of types (β^is) is the same across zip codes, conditional on demographics. These parametric assumptions ignore important unobserved heterogeneity within a region.

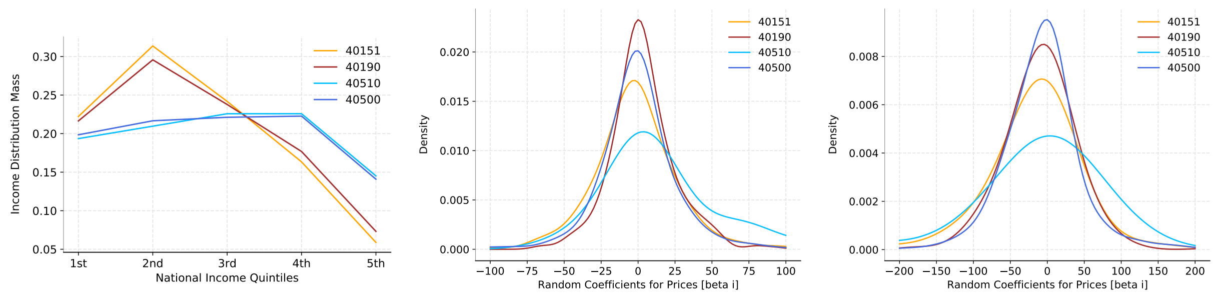
According to our data, different zip codes have different distributions of the random coefficients and different correlations of each household's consumption and income, as shown in Figure B.1. The figure shows the estimated distribution of individual price coefficients (Panels (b) and (c))[37] and the income distribution (Panel (a)) of four zip codes in the same region. One can see that the correlation between prices and electricity consumption can be substantially different among zip codes with similar distributions of income. The parametric approach has to use the σ^β coefficient to explain these differences. This suggests that including demographics as a way to classify households might be too limited to understand the heterogeneous impacts of real-time pricing. Because there is substantial unobserved heterogeneity within zip codes, our approach with discrete types attempts to allow for greater flexibility.

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

If we try to modify the moment conditions (B.7) and (B.8) in a more flexible way, a natural way

[37]Notice that, with individual-level data, we can estimate β^i for each individual using individual-level logit regressions. The estimated β^i helps reduce the distributions of the high dimensional outcome variable kWh into the distribution of β^i.

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



(a) Income distributions

(b) Distribution of individual price coefficients $\beta^i_{t=1,1}$, 4-7am

(c) Distribution of individual price coefficients $\beta^i_{t=3,1}$, , 12-3pm

Notes: This figure illustrates the distribution individual coefficients for different zip codes. The distributions are heterogeneous, even when conditioning on zip code income distributions. The two blue zip codes have a similar income distribution and the two red zip codes have a similar income distribution. However, the distributions of the β^i coefficients are different for each pair of zip codes, regardless of the time windows. This evidence violates the classic assumption in the parametric approach.

would be to consider the following two equations, which allow different zip codes to have different distributions of type $r \in [1, 2, .., R]$, denoted by $Pr_z(r)$, in a flexible manner. A type r can indicate a cluster of coefficient patterns (e.g. in the fixed grid nonparametric approach), or kWh patterns (e.g. in our main specification), or both. The relationship between income and household types is still characterized by η^k:

$$\sum_h \left(\overline{kWh}_{zh} - \sum Pr_z(r)kWh_{zh}(r)\right)^2, \quad \forall z, \tag{B.13}$$

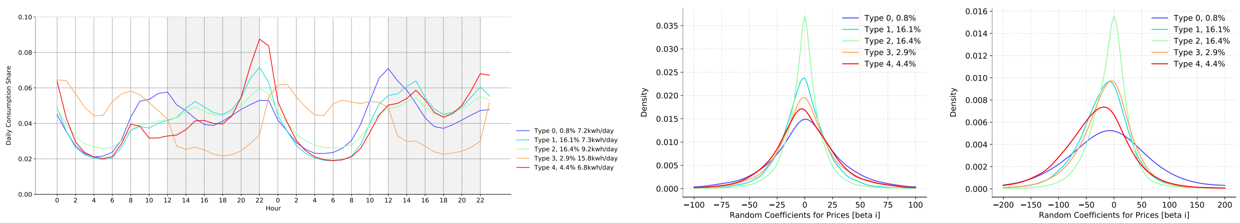
$$\sum_k \left(Pr_z(inc_k) - \sum_{i \in z}\sum_r \eta^k_{zr} Pr_z(r)\right)^2, \quad \forall z, \tag{B.14}$$

$$\text{s.t.} \quad \sum_k \eta^k_{zr} = 1, \quad \forall z, \forall r. \tag{B.15}$$

Allowing fully flexible income-type distributions for each zip code does not work, because when type and income are both unknown, the above system of equations (B.13) and (B.14) is greatly underidentified without imposing more structure. Allowing for zip-code-specific types, i.e., assigning only one type to a zip code with probability one, can perfectly match the aggregate moments. The result becomes equivalent to the "naïve" approach of assuming that all households within a zip code have the same income distribution. A natural modification to this is our two-step estimator, which allows for flexible types and can easily identify income-type distributions under the assumption that there is sufficient overlap in types across zip codes.

Limits to parameterization The second problem with using a parametric approach is that electricity consumption is a highly dimensional choice, and thus price or temperature variation

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



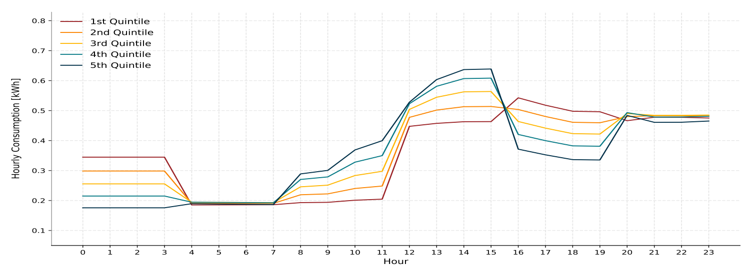
(a) Kmeans types for Madrid

(b) $\beta^i_{t=1,1}$, 4-7am

(c) $\beta^i_{t=3,1}$, 12-15am

Notes: This figure illustrates the agnostic consumption patterns by Kmeans types (Panel (a)) and the distribution of β^is for each Kmeans type (Panel (b) and (c)). The β^is are defined in the parametric model and are estimated through individual-level logit regressions. The Kmeans types are clearly distinct, while the distributions of β^is are very similar across Kmeans types. This indicates that the β^is, although capturing some aspects of the consumption patterns, are not informative enough for classifying consumer types.

Figure B.3: Estimated load curve from parametric estimation



Notes: This figure depicts the predicted load curve from the parametric model. Compared to Figure B.2 Panel (a), this predicted load curves are much less flexible.

is insufficient to explain it. The vector β might not be a proper primitive to model electricity consumption, and equation (B.3) might not be sufficiently detailed for our purposes.

Using an over-simplified model would ignore unobserved heterogeneity in consumption patterns. Two households of different consumption types might have very similar price correlations. If we ignored this, our income estimates would be biased. Figure B.2 provides an example showing that types with similar price coefficients may still have substantially distinct consumption patterns. The distributions of the individual price coefficients $\beta^i_{t=1,1}$ are homogeneous, as can be seen in Panel (b). However, as shown in Panel (a), the types are clearly heterogeneous even within the same time window.

Moreover, we can compare Figure B.2 Panel (b) and Figure B.1 Panel (b). These two figures plot the distributions of the same coefficients. The heterogeneity across types is even smaller than the heterogeneity across zip codes. Thirdly, both Panel (b) and (c) from Figure B.2 imply the blue type (type 1) and the orange type (type 3) are similar to each other. However, they might have different lifestyles (types) and probably have different income distribution. The parametric approach will ignore this heterogeneity and therefore give biased results.

The parametric model can capture some aspects of heterogeneity. Figure B.3 indicates that higher income households' consumption is more procyclical, consistent with our main model findings. However, the model does not allow for much further heterogeneity (other than via noise in the random coefficients). We miss the heterogeneity present in Panel (a) of Figure B.2, even though we have 45 parameters per month-province. Thus it may be hard to use any parametric functional form to describe the market share choices of different types. Comparing the consumption patterns and the price variation, we know that β in equation (B.6) is not a sufficient statistic for all household types.

Overall, we need some tools to simplify the high-dimensional heterogeneity in this setting and cluster households into a smaller number of types. In short, we need a dimension reduction tool. As explained in Bonhomme et al. (2021), allowing for discrete heterogeneity (clustered by *kmeans*) is an efficient dimension-reduction device to deal with an agnostic electricity consumption model. A parametric model, which describes the data variations using a small number of parameters, can serve a similar role. It reduces the distributions of the high-dimensional outcome variable kWh into the distributions of the β^i. However, as shown above, a parametric approach would be subject to limitations, making the more agnostic *kmeans* clustering approach preferable.

Repeated observations of each individual Panel data with repeated observations from each individual provide essential variations to identify the joint distribution of income and electricity consumption. The variation across individuals can be explained by income (or other individual features) variation, while other variables can explain the variation across time for the same individual, e.g., prices and temperature.

As mentioned above, the parametric approach uses repeated observations from the same individual through the covariance moments. These covariance moments are essential for identifying the income coefficients η because they connect household income with their across hour and month consumption patterns. We have hourly electricity consumption data from each household for more than one year, i.e., around 10,000 observations per individual, giving rise to thousands of candidate covariance moments. The large number of candidate moments and the complicated format of the income-β moments are due to the lack of an explicit dimension reduction device. Ideally, one would want to include many moments to remain agnostic about how to select them, but this is computationally not feasible.

In our approach, we simplify the highly-dimensional nature of smart meter consumption in our first step. We then connect households' income with their consumption patterns in our second step. The second step serves the same role as the covariance moments but it is computationally much simpler. We avoid the thousands of candidate covariance moments because we have reduced the dimensions non-parametrically in the first step. The *kmeans* clustering method has helped discretize the highly-dimensional household heterogeneity. Therefore, the relationship between income and types can be more agnostic and more explicit than the income-β relationship in the parametric approach, which is very much tied to the price coefficients.

Overall, these limitations highlight that it is difficult to summarize the heterogeneity in electricity consumption data using a parametric approach. These limitations also highlight that oftentimes the problem can become computationally expensive. Our approach tries to find a data-driven compromise to handle these two difficulties.

C Inferring appliance ownership

In this appendix, we infer households' appliance ownership by exploiting the richness of the smart meter data. The idea of using high-frequency data to infer appliance ownership has been applied to engineering papers like Westermann et al. (2020) and Dyson et al. (2014). For each household, we first run the following regression to obtain the correlation between its electricity consumption with temperature in winter and summer:

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

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 that time; α^i_{fe} are hour-month-weekends fixed effects, which we include to control for unobserved consumption heterogeneity across time. The coefficients of interest are β^i_w and β^i_s, which measure how much more a household consumes in response to a temperature increase in winter (summer) relative to other times of the year. We only account for daytime responses because according to the Household and Environment Survey 2008 carried out by the Spanish National Statistics Institute (INE), around 95% of households turn off their AC at night (and around 80% of households turn off their heating at night). We also include the term $\beta^i temp_{hdm}$ to control for the general trend of each household.

First, regarding the response to temperature, one would expect that a household with AC would consume more in the summer as temperature increases. Therefore, β^i_s should be positive. Similarly, households with electric heating are expected to increase their consumption as temperature decreases in winter. Therefore, β^i_w should be negative. Second, regarding the mean consumption level for different seasons across households, one would expect that households with AC (electric heating) would consume more in summer than on average, i.e., $\alpha^i_s > \alpha^i$ for households with AC and $\alpha^i_w > \alpha^i$ for households with electric heating.

After getting the seasonal consumption differences in slope and in levels for each household, we estimate the criteria for appliance ownership by matching the estimated state level market shares with the surveyed market share shown in Table C.1. To identify households' AC ownership, we perform the following optimization:

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

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

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

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

Table C.1: Statistics on the availability of heating systems

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: (1) Individual electric boiler (2) Electric radiators and accumulators (3) Radiant wire. Source: Spanish National Statistics Institute (INE), Household and Environment Survey 2008 (https://www.ine.es/dynt3/inebase/index.htm?type=pcaxis&path=/t25/p500/2008/p01/&file=pcaxis&L=0).

Table C.2: Estimated threshold values

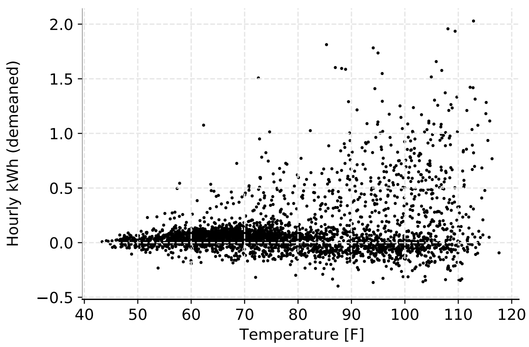
AC thresholds		EH thresholds	
$\underline{\beta}_s$	$\underline{\alpha}_s$	$\underline{\beta}_w$	$\underline{\alpha}_w$
0.50	0.01	-4.00	0.05

The first term in the objective function captures, for each state s, the difference between the surveyed AC market share, s_s^{AC}, and the estimated one, $\hat{s}_s^{AC}(\underline{\beta}_s, \underline{\alpha}_s)$, for given thresholds $(\underline{\beta}_s, \underline{\alpha}_s)$ for having or not having AC. The second term is a penalizing term that prevents the low contracted power group from having a higher share of AC ownership than 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 are expected to contract more power capacity. We follow a similar procedure to estimate electric heating ownership. Figure C.1 shows 4 example households and depicts the correlation of household consumption and temperature in our data. The patterns are similar the results from Dyson et al. (2014). Estimation threshold values are reported in Table C.2.

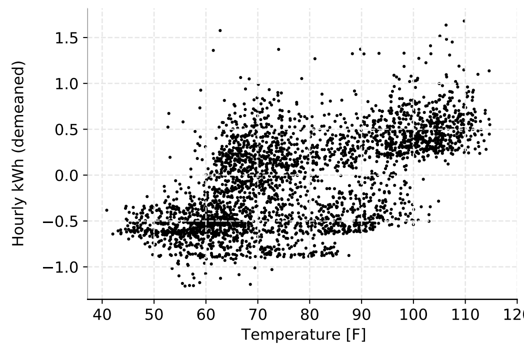
To show that our classification of household types is informative both about their consumption behavior, we plot consumer daily load curves by identified appliance ownership in Figure C.2. EH owners have relatively higher consumption during both day and night because electric heating devices are in general more energy consuming than AC, as show in Panel (a). We also observe that high consumption is particularly high during winter months for households that have electric heating, while it is peaking in the summer for those households with air conditioning, as shown in Panel (b).

We also use the EH ownership variable as an observable type in the GMM procedure to infer income. This allows us to infer the income distribution of households as a function of appliance ownership. Figure C.3 shows that we find that electric heating is particularly concentrated on the low income bins, while air conditioning is positively correlated with income, which is intuitive. We also show that the patterns of appliance ownership and income can change depending on the region. While air conditioning tends to be associated with high income (for regions with a meaningful share of air conditioning), electric heating is negatively correlated with income particularly in the most

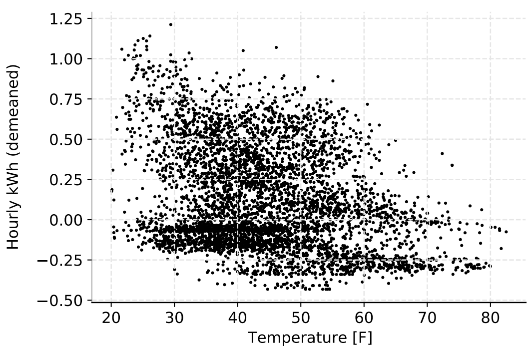
Figure C.1: Example: households' hourly consumption and temperature



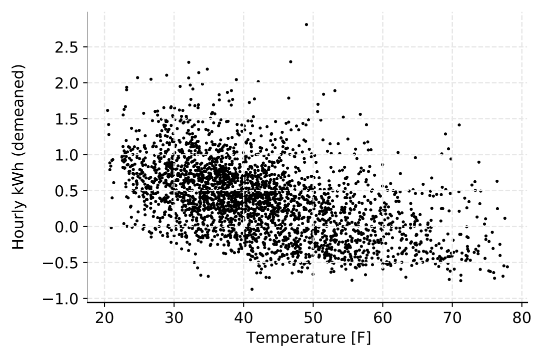
(a) Identified AC owner 1 hourly kWh (demeaned)



(b) Identified AC owner 2 hourly kWh (demeaned)



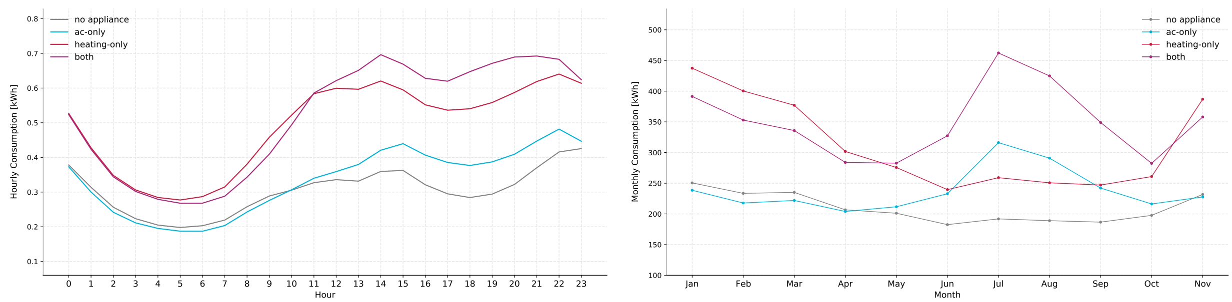
(c) Identified EH owner 1 hourly kWh (demeaned)



(d) Identified EH owner 2 hourly kWh (demeaned)

Notes: These figures show hourly consumption for one representative household. These figures show the correlation between household consumption and temperature and the variations that help us identify appliance ownership. All data points are demeaned at household-hour level. For identified AC owners (the upper panels), we plot data points during June, July, August, and September. For identified EH owners (the lower panels), we plot data points during the November, January, February, and March. The lower panels have more data points because we have January-March data for 2016 and 2017. The four panels are: (a) an AC owner that responses partially; (b) an AC owner that responses in all hours; (c) an EH owner that responses partially; (d) an EH owner that responses in all hours.

Figure C.2: Load curves by appliance ownership

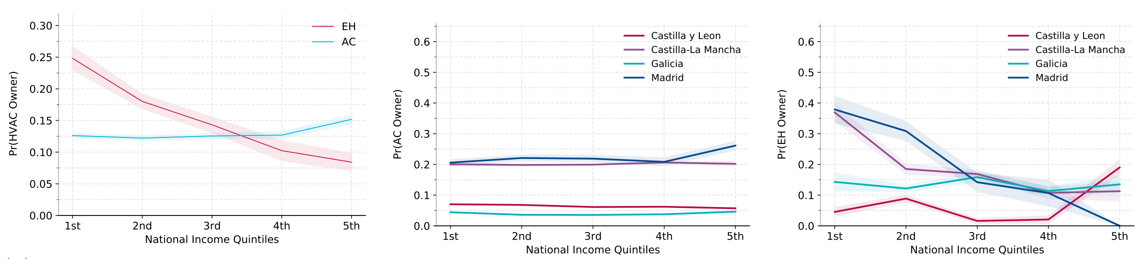


(a) Hourly consumption pattern [kWh]

(b) Annual consumption pattern [kWh]

Notes: These figures show consumption profiles over the day (the left panels) and the year (the right panels) for households with electric heating, AC, or both.

Figure C.3: Appliance Onwership and Income



(a) Share of electric heating owners and AC owners

(b) AC ownership by state

(c) EH ownership by state

urban regions (Madrid), as newer building tend to rely on city gas for heating.

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 use the smart-meter household-level consumption data from our sample and create a data generating process in which we know each individual's type. We assign types to individuals based on their consumption profiles, which we then use to assign them to a certain income bin, respecting an assumed distribution of income, and zip code. We then aggregate the randomly-assigned incomes to the zip code level, so that we can compute the distribution of income at the zip code level. At that point, we assume that types and household's income are not observed, and try to infer the distributional impacts of real-time pricing based on zip-code-level income data only.

The more detailed steps are as follows:

1. We summarize smart-meter household level data (one entry per household) to reduce the dimensionality of the data and assign them to a group.

2. We classify households into five types for each group using a kmeans algorithm based on their hourly market shares and total consumption.

3. We sort types based on their peak market share (hours 8 - 23). We assign the more "peaky" types to a higher income distribution, reflecting the within-month correlation of peak consumption and income.

4. This distribution is fixed conditional on type and is the same across all zip codes, but we introduce some noise to capture unmodelled randomness in the data.

5. Using the assigned types, we assign a zip code number to each household based on a pre-established probability that type θ belongs to zip code z, $Pr(z|\theta)$.

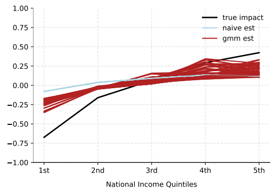
6. These zip codes and the zip-code-level income distribution, together with the household-level consumption patterns, is what is observed for the estimation.

These steps allow us to create an individual and zip-code-level distribution of income that is consistent with the underlying types and assumptions.

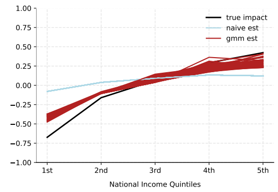
We then estimate the distributional impacts for each income bin given the calculated bills under RTP and flat tariffs, as in the main estimation. We sample 200 households from each zip code by default and increase the sample size to 1000 households per zip code along our discussion. We compare the results from our inferred income distribution vs. the one that we assumed to be the truth. Departures between our inferred impact and the true underlying generating process would suggest our model has some bias.

Figure D.1 shows our simulated distributional impacts compared to those of the "true" underlying data. In this Monte Carlo, bills are increasing with income, as the simulation is focused on peak vs. off-peak consumption. Therefore, the true impact is increasing in income. In Panel (a),

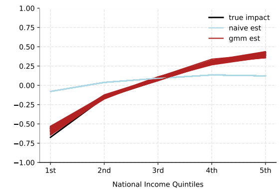
Figure D.1: Simulation results: different number of types



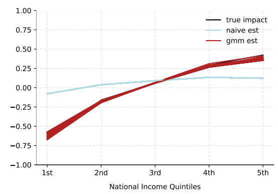
(a) 3 types, sample size 200



(b) 4 types, sample size 200



(c) 5 types, sample size 200



(d) 5 types, sample size 500

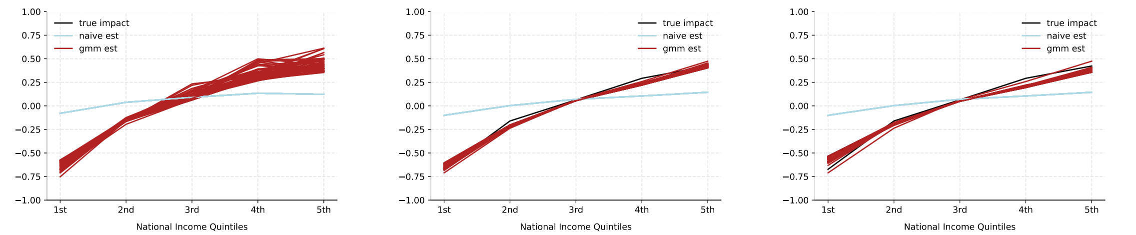
only three types are allowed, and one can observe that our estimator understates the distributional impacts of the policy. This is also true for four types (Panel b). It is only with five types that we recover the true impact (Panels c and d). This suggests that it is important to allow for flexibility in the number of types. It also suggests that it will be important to assess the sensitivity of our estimates to the number of types.

Another aspect that is unknown to us, as econometricians, is how to group zip codes in the data. The grouping of zip codes is crucial to our identification strategy, as the underlying assumption is that types are common across the zip codes that belong to one group. Figure D.2 explores this issue by assuming that we use the wrong groups of zip codes. Panel (a) shows that under the wrong grouping assumption, our estimates are noisier. Panel (b) shows that, with a larger sample (1000 households per zip code), the estimates can potentially be biased. Panel (c) shows the results when the number of zip codes per group is also off, and set to six instead of five, showing that it can also impact the results. The extra type allows the model to pick the new types of households from the mis-grouped zip code. This is another evidence that it will be important that we explore sensitivity of our estimates to the grouping choice.

It is important to note that the data generating process in the Monte Carlo assumes that the distributional impacts are similar across zip codes and groups of zip codes. In particular, the higher impacts are associated with higher income in all groups of zip codes, although the magnitude and exact shape varies. If the relationship between income and the impact of RTP were to be highly non-linear across zip code groups, then the bias in Figure D.2 could grow larger. This is why in our estimation we treat provinces as completely separate entities, to avoid mixing zip codes that have potentially very different characteristics and thus underlying distributional impacts.

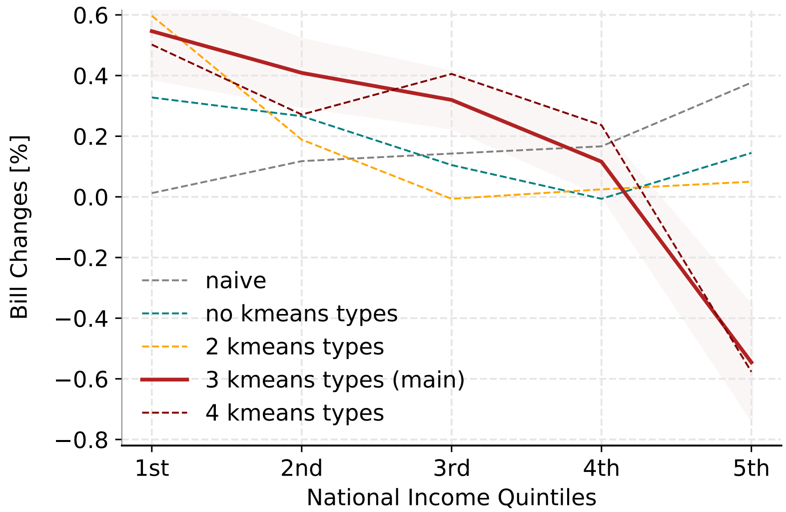
We conduct a sensitivity analysis of our estimates regarding the number of types to verify

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

Figure D.3: Estimated Bill Changes [%] from Alternative Specifications



Notes: This figure represents the estimated bill increase in % when moving from an annual time-invariant price to RTP under different specifications. Different lines represent different numbers of types set in the estimation. Only the main specification's standard error is included to keep the figure easy to read.

the Monte Carlo findings and the estimation results. Figure D.3 shows the results of setting different numbers of types. The gray line shows the naive approach without any within zip code heterogeneity. The green line shows results of including contracted power type and appliance ownership type, and no flexible type (Kmeans type) is allowed. The yellow dashed, red solid, and dark red dashed lines correspondingly show results from specifications with 2, 3, and 4 Kmeans types ($N^g = 4, 12$, and 16). All specifications include contracted power type and appliance ownership type. Only the main specification's standard error is included to keep the figure easy to read, the standard errors of other specifications are similar to the ones in the main specification.

As expected, Figure D.3 shows that the estimated distributional impacts are more pronounced with more types. Once we include contracted power type and appliance ownership type, the result is already opposite to the naive approach result, showing that households in the low income quintile lose from the switch to RTP. This is because contracted power is highly correlated with household income and is powerful in identifying the distributional impact. Adding more Kmeans types into the estimation makes the result closer to our main specification result. These results indicate that the actual estimation is aligned with the Monte Carlo results, which validates our estimator. Moreover, the result from the 4 Kmeans type ($N^g = 16$) specification falls in the confidence interval of the result from the main specification (3 Kmeans types, $N^g = 12$). This fact implies that the potential error generated from our zip code group choices is eliminated in the main specification.