The Distributional Impacts of Real-time Pricing

Mar Reguant
Northwestern and BSE

Sciences Po – 2021.11.15

Energy transition underway

- Need to reduce Green House Gas emissions (GHGs).
- Electricity sector ($\approx$35-40% of CO_2 emissions) has been **most active** and has the greatest potential in making the transition.
- Ambition to move towards **carbon-free electricity** by 2050.
- **Limits to decarbonization**:
  - **Renewables' intermittency** might lead to a potential mismatch between supply and demand, increasing need for flexibility.
  - **Extreme events** with adverse outcomes for households also intensify need for flexible demand.
  - **Important challenge** until better (cheap) storage solutions are found.

Energy transition's heterogeneous impacts

- The energy transition can have substantial impacts on households that can be highly heterogeneous.
- Net-metering of solar can leave poorer households stranded without policy action.
- **Uneven impacts combined with climate change impacts**:
  - Households most exposed to extreme events tend to have the lowest income (poor building construction and insulation, heat islands).
  - Also least able to adapt and upgrade with resilience equipment (solar + backup battery, solar + EV as battery).

Equity impacts recently in the news

After Days Of Mass Outages, Some Texas Residents Now Face Huge Electricity Bills

February 21, 2021 · 12:01 PM ET

REBECCA HERSHER

Equity impacts can be devastating

WINTER STORM 2021

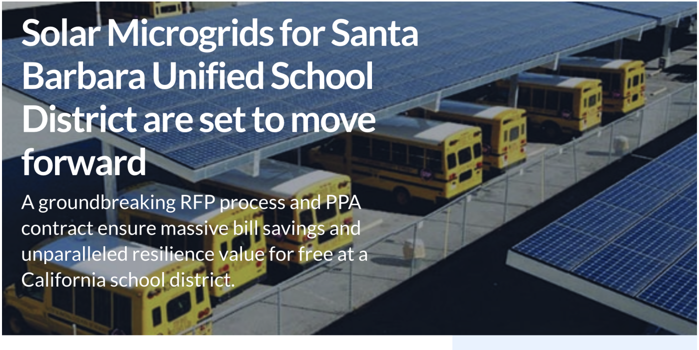
At least 111 people died in Texas during winter storm, most from hypothermia

> The newly revised number is nearly twice the 57 that state health officials estimated last week and will likely continue to grow.

BY SHAWN MULCAHY MARCH 25, 2021 4 PM CENTRAL

COPY LINK REPUBLISH

Resilience preparedness will not start where most needed

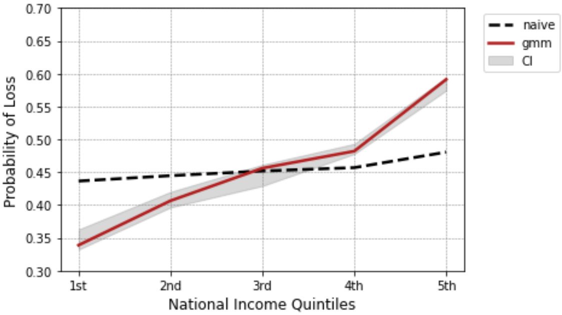


Our paper

- We study the distributional impact of real-time pricing in the Spanish electricity market, which rolled out **RTP as the default tariff** for a large share of residential customers.

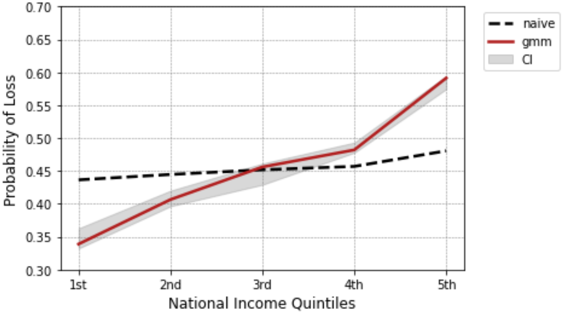
- Challenge: Our data **do not have** detailed income information, limiting the distributional analysis.
- Methodology: We complement aggregate patterns of distributional effects with a **method to infer individual income** using smart meter data and zip-code income distributions.

A preview of results



- RTP is mildly progressive in the aggregate.
- Result strengthened with household income heterogeneity.

A preview of results



- RTP is mildly progressive in the aggregate.
- Result strengthened with household income heterogeneity.
- Note: These results are in sample, future impacts need careful simulation if this was not already clear!

Related literature

- Papers on the role of RTP and efficiency:
  - Borenstein (2005) among related papers.
- Papers on the role of electricity pricing and equity:
  - Borenstein (2007) (industrial), Borenstein (2012) (nonlinear pricing), Borenstein (2013) (critical peak pricing), Faruqui et al. (2010), Horowitz and Lave (2017), Zethmayr and Kolata (2018), Burger et al. (2019).
- Papers on inferring income:
  - Pissarides and Weber (1989), Feldman and Slemrod (2007), Artavanis, Morse, and Tsoutsoura (2016), Dunbar and Fu (2015), etc.
- Papers unveiling household heterogeneity:
  - BLP (1995, 2004), Petrin (2002), Fox et al. (2011), etc.

Outline

1. Background and aggregate impacts
2. Inferring household income
3. Measuring impact of RTP

Dynamic electricity pricing in Spain

- April 2015: Spain becomes the only country in which RTP is the **default option for all households**.
  - *The case of Spain with a regulated default dynamic price contract is unique* (EC, 2019).

- Electricity **marginal price** composed of two parts:
  - **Energy price**: determined hourly as a function of the wholesale electricity market. (**RTP**), or time-invariant.
  - **Access price**: regulated costs charged as a function of consumption; peak/off-peak prices (**TOU**) or time-invariant.

- Customers defaulted into RTP and non-TOU.
- In this paper, focus on RTP vs. flat prices.

Data

- We obtain smart-meter data for over 4M households, from one large Spanish utility.
- For each household (January 2016-July 2017):
  - hourly electricity consumption
  - plan characteristics (pricing, contracted power (cap))
  - postal code
- We link the postal code with detailed Census data:
  - education, income and age distribution, avg number of rooms...

Data: electricity consumption area

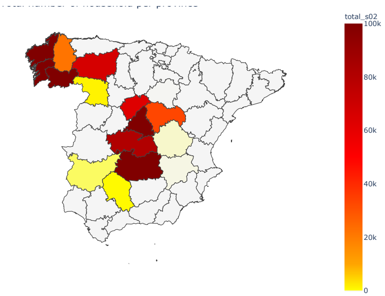


Figure: Naturgy area

A first look at the data: prices

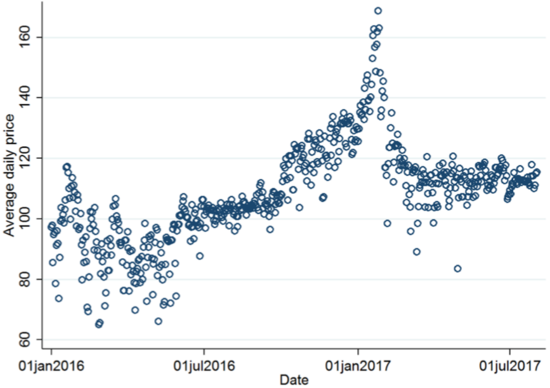


Figure: Average prices over the sample period

A first look at the data: price variation

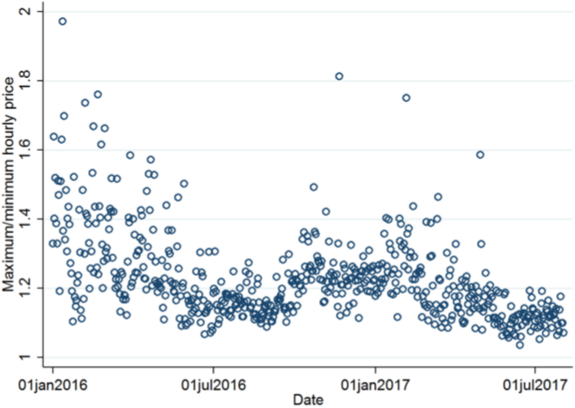


Figure: Ratio of the highest to lowest price every day

Policy implications

Goal: Analyze the **distributional effects of change to RTP**.

1. Describe RTP impacts assuming consumers are inelastic to prices in the short-run.
   → Justified by our previous project. ▸ Fabra, Rapson, Reguant, Wang
2. Assess relationship of RTP impacts with income.
   → Limited effects, not regressive.
3. Future work: consider impact of counterfactual experiments, such as responses to prices or extreme events.

Computing bills under RTP and flat tariffs

- We compute household bills with and without RTP pricing:

$$RTPBill_{im} = \sum_{h \in m} q^*_{ih} p^*_{ih}$$

$$FlatBill_{im} = \sum_{h \in m} q^*_{ih} \bar{p}_m$$

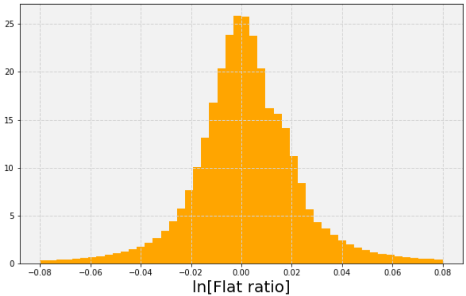
- Consider ratio of RTP bill to flat bill:

$$BillRatio_{im} = \frac{FlatBill_{im}}{RTPBill_{im}}$$

- Ratio lower than 1: consumers worse off under RTP.

Raw distribution: winners and losers

Figure: Ln[Bill Ratio] (individual, monthly)



- Centered around zero by construction (to focus on cross-subsidization due to consumption patterns).
- In practice, most consumers win under RTP (ongoing work).

Distributional impacts

	ln(kwh)	ln(kwh peak)	ln(bill ratio)
ln[IncPerHH]	0.095**	0.116***	-0.003***
	(0.036)	(0.034)	(0.001)
HHsize	0.385***	0.406***	0.001
	(0.029)	(0.027)	(0.001)
R-squared	0.617	0.701	0.508
N	685	685	685

- Income correlated with consumption.
- Modest negative correlation with RTP impacts, slightly negative.

Can we better infer household characteristics (income) exploiting the Census data?

The challenge: income data

- We observe the distribution of income at the zip code level.
- Zip codes can be substantially large.
- Inference of income common in other applications: tax fraud, subsidy fraud, refinements to coded income.
- Impacts of RTP depend on highly dimensional vector, so difficult to make intuitive bounding assumptions (e.g., Borenstein, 2012).

- **Research question**: how to better assign households' income exploiting richness of hourly consumption data?

Some notation and definitions

- Zip code as $z \in \{1, \ldots, Z\}$.
- Income bins as $inc_k \in \{inc_1, \ldots, inc_K\}$.
- Households in zip code z as $i \in \{1, \ldots, H_z\}$.

- Observed zip-code income distribution: $Pr_z(inc_k)$.
- Unknown household income distribution: $Pr_i(inc_k)$.

Inferring income in this setting

- We have detailed hourly consumption data for each household–1000's of observations per HH (panel).
- We have the distribution of income at the zip code (cross-section).
- We have the zip code of each household.
- Demand system approaches are a way to infer household income at the household level (e.g., seeing someone buy a Ferrari).
- Here we prefer to remain agnostic about the demand system (lots of heterogeneity), and **directly focus on inferring income** of households.

▸ Further discussion

Naïve approach

- Assign income distribution at the zip code level $Pr_z(inc_k)$ to all households in that zip code.
- Captures **across-zip-code heterogeneity**, but can miss important within-zip-code heterogeneity.
- One can get somewhat at within-income bin variance, but overstated if interpreted literally, as failing to sort households within each zip code.
  - Heterogeneity of policy impacts conditional on the same income can be large, e.g. Cronin, Fullerton and Sexton (2019).
  - Here, however, we would be attributing zip code heterogeneity to all quintiles.

Assigning a prob. income distribution to households

We introduce new additional objects:

- Zip code as $z \in \{1, \ldots, Z\}$.
- Income bins as $inc_k \in \{inc_1, \ldots, inc_K\}$.
- Households in zip code z as $i \in \{1, \ldots, H_z\}$.
- Discrete types as $\theta_n \in \{\theta_1, \ldots, \theta_N\}$.

- Observed zip-code income distribution: $Pr_z(inc_k)$.
- Unknown household income distribution: $Pr_i(inc_k)$.
- Unknown household type distribution: $Pr_i(\theta_n)$
- Unknown type-income distribution: η_n^k (probability that type n has income bin k).

Our approach: intuition

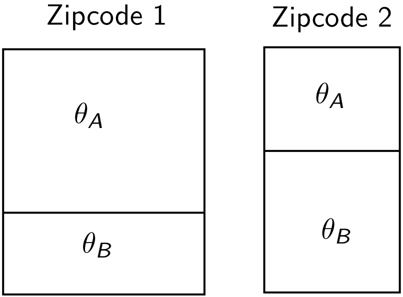
We propose an estimator in two steps:

1. Classify consumers into types (deterministic or mixtures).
2. Infer income distribution of the unobserved types based on zip code level distribution.

Key: Allow for sufficient unobserved heterogeneity to match income distribution at the zip code level.

Identifying assumption: Common types across (subsets of) zip codes.

Intuition follows similar settings (e.g., BLP, FKRB)



$$\eta_A^H Pr_1(\theta_A) + \eta_B^H Pr_1(\theta_B) = Pr_1(inc = H)$$

$$\eta_A^H Pr_2(\theta_A) + \eta_B^H Pr_2(\theta_B) = Pr_2(inc = H)$$

- Assume we have already inferred the distribution of unobserved types in each zip code.
- η_A^H represents the probability of income level H for unobserved type θ_A (similarly for θ_B), unknowns.
- Match zip code moments on the distribution of income, same underlying types across zip codes.

Identifying equations

Conditional on having identified the distribution of types for each zip code:

$$\min_{\eta} \sum_{z} \omega_z \sum_{k} \left(Pr_z(inc_k) - \sum_{i \in z} \sum_{n} \eta_n^k Pr_z(\theta_n) \right)^2$$

$$\text{s.t. } \sum_{k} \eta_n^k = 1, \forall n,$$

where ω_z is a sampling weight and

$$Pr_z(\theta_n) \equiv \sum_{i \in z} Pr_i(\theta_n) / H_z.$$

Further discussion

Step 1: Assigning households to types

- We break the approach in two steps to facilitate the computations: millions of households with individual hourly consumption data.
- Inefficient, but consistent under the proposed assumptions.

- We have explored several classification techniques:
  - Observable discrete characteristics (contracted power).
  - Inferred discrete characteristics based on smart-meter data (appliance ownership).
  - Deterministic classification based on summary stats from high-frequency data.
  - EM algorithm based on household-level regression outcomes.
  - **k-means clustering based on load profiles**

Step 1: k-means clustering of types

- We reduce dimensionality of data into market shares for daily consumption in weekdays and weekends for each individual household.
- We group nearby zip codes and cluster the population of consumers based on these market shares as well as the levels of production. Observable types based on contracted power.
- Our baseline has 5 zip codes with 5 types per observable types.
- We explore robustness of the method of choice under identifying assumptions via Monte Carlo simulation.

Step 2: GMM

- We have now a probabilistic assignment of types to each household.
  - Based on consumption patterns which correlate with price.
- We assume that types are shared across different zip codes.
  - What changes is the *proportion* of types.
- Assign **income probabilities to types** (η_n^k).
- Given the observed distribution of income for each zip code, we match these zip code-level moments.

GMM: income distribution

Income

1. The aggregated income distribution of each type should be consistent with the zip code level distribution.

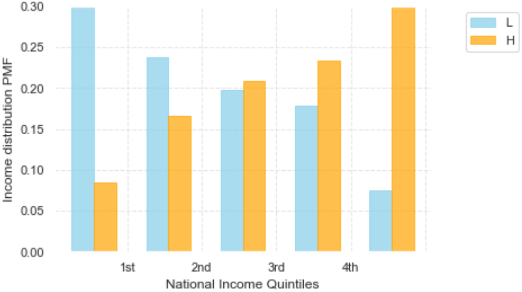
$$\hat{P}r_z(inc_k) = \sum_{n=1}^{N} \eta_n^k Pr_z(\theta_n) \qquad \forall k, z.$$

2. For each type, the probability of being in different income intervals sum up to 1:

$$\sum_{k=1}^{K} \eta_n^k = 1 \qquad \forall n$$

Results: Confirm importance of contracted power

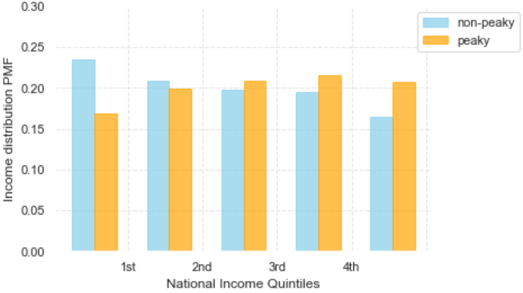
- Individual-level of contracted power predicts higher income distribution.



Results: Confirm relationship between income and peak consumption

- Individual-level monthly consumption by rich and poor

simulated income $>$ 17,000 Euro



Bringing it back to policy impacts

- We use the inferred distribution of income to derive implications about RTP pricing.

- *What is the distribution of income for winners and losers?*
- *What is the relevance of within-zip-code heterogeneity?*

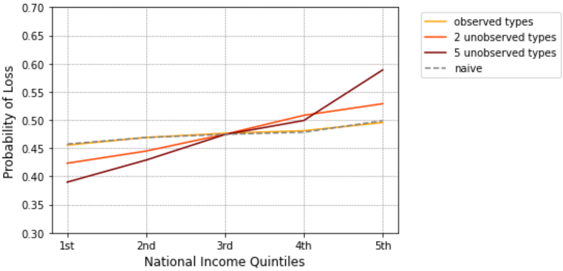
Predicting the probability of losing

- We focus on a summary statistic of the distributional income: the probability of losing under RTP.
- We assess the predicted probability of losing for each income bin under alternative income distributions.

$$Pr(lose|inc_k) = \frac{1}{H \times Pr(inc_k)} \sum_i \mathbb{1}(Loser)\ \hat{Pr}_i(inc_k).$$

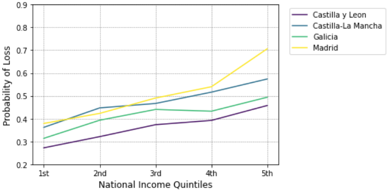
- Cases:
  - Naïve based on zip code distribution, $\widehat{Pr}_i(inc_k) \equiv Pr_{z_i}(inc_k)$;
  - Two unobservable types;
  - Five unobservable types.

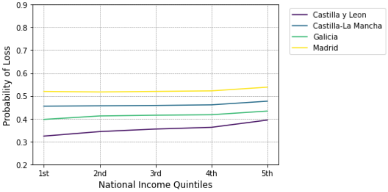
Losers by income bin



- Approach uncovers substantial within zip-code heterogeneity in impacts correlated with income.

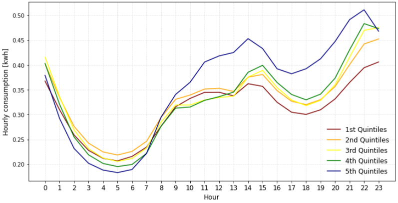
Substantial geographical heterogeneity

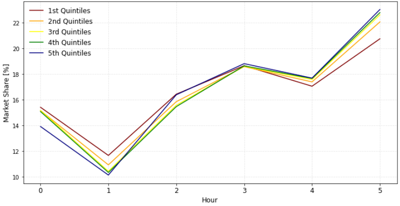




- But across-zip-code variation does not identify strong correlation of income with the policy.
- Due to geographical patterns in Spain, income vs geography cancel out in the aggregate.
- Heterogeneity within region via hidden types is driving the increase in heterogeneous impacts.

Mechanisms: consumption patterns



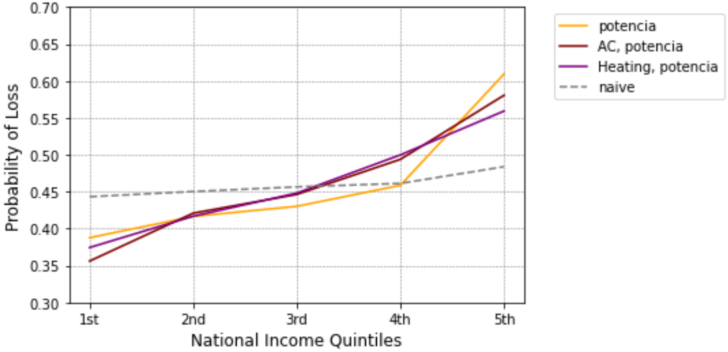


- Higher quintiles consume much more.
- They also consume proportionally more at peak.

Mechanisms: appliance ownership

- We use algorithm to infer appliance ownership by households.
- We then treat appliance ownership as an explanatory variable in heterogeneity.
- Appliance ownership is relevant to explain patterns.

Mechanisms: appliance ownership and income impacts

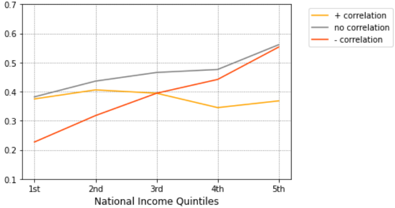


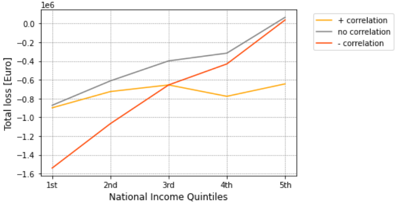
- Appliance ownership adds nuance to findings, e.g., poor households with electric heating disproportionally hurt.
- It does not affect the general average patterns if incorporated explicitly into types.
- In part due to minor role of electric heating in Spain.

Counterfactual experiments

- The distributional impacts in our sample are limited and bounded.
- However, patters could change going forward, with increasing extreme pricing and volatility.
- We plan to explore several counterfactuals:
  - Correlation of income and elasticity of demand.
  - Extreme prices under alternative assumption on their patterns and correlations of occurrence (e.g., temperature driven, peak/off-peak).
  - Natural gas price shocks.

Counterfactual experiments: preliminary elasticity results





- Elasticity (if positively correlated with income) can undo some of the patterns, but not revert them in this simple calibration.
- Important as high income households can better adapt to fluctuations in prices via smart technologies and batteries.
- As price fluctuations become large, income effects can be substantial (e.g., see Texas).

Counterfactual experiments: a big commodity shock

- Our results so far focus on monthly variation in prices.
- However, many flat tariffs (even when previously regulated) offered hedging during the year.
- Relevant for commodity price fluctuations.
- In the news lately due to the large cost shock in Europe.

- As prices increase, **income effects** and **heterogeneity** within quintiles matter for energy poverty.

Commodity shock and energy poverty in the news

Up to 1.5m more could struggle to pay energy bills next year

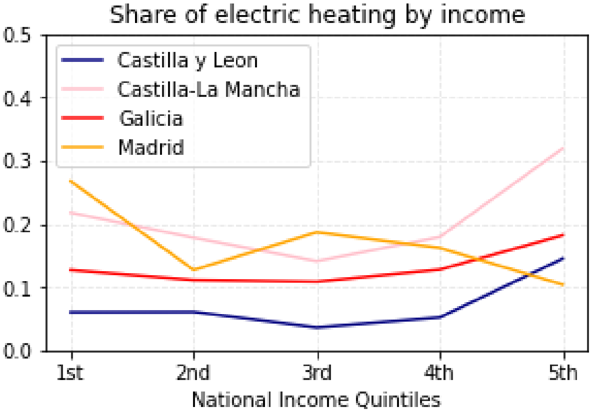
7 October | Comments

A big commodity shock: results

- Households with electric heating see the largest bill increases.
- Pass-through of costs (electricity, natural gas) and ownership of heating types important determinants of overall effect.

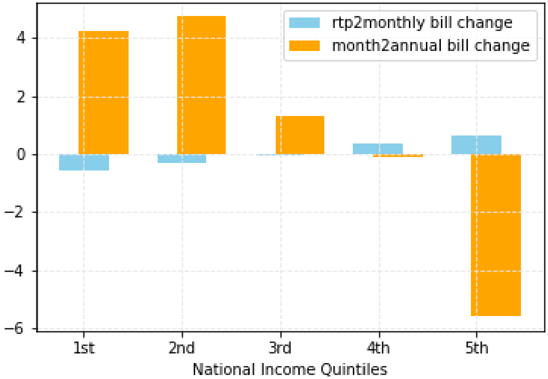
- For electricity, we find a **U-shape** in the monetary impacts:
  - High-income households consume more, and more in winter, and thus negatively affected by spikes.
  - Low-income households are least affected in *at the median* due to lower consumption.
  - A share of low-income households affected and at risk of energy poverty, *average impacts* are largest.

Heating ownership by region and quintile



U-shape in heating ownership, and large for low-income in Madrid

RTP can become regressive



Average impacts disproportionately fall on low-income

Conclusions and next steps

- We use a **two-step approach** to infer the distribution of individual income.
- The approach exploits detailed smart-meter data.
- We are exploring several aspects of the methodology:
  - What are the advantages/disadvantages of the different practical implementations?
  - Could we do a joint approach?
- We are also exploring broader counterfactual impacts of RTP.

Thank you.

Questions? Comments?

mar.reguant@northwestern.edu

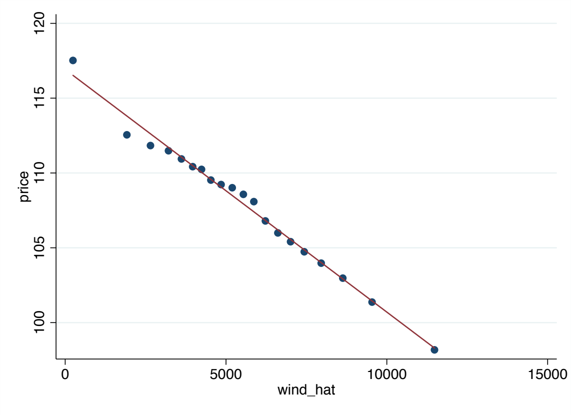
Measuring elasticity to RTP

- We estimate the short-run price elasticity of households.
- Main regression (individual by individual):

$$\ln q_{ith} = \beta_i \ln p_{ith} + \phi X_{ith} + \gamma_{ith} + \epsilon_{ith}$$

- In baseline specifications, we control for:
  - Temperature bins by hour
  - Fixed effects: hour x month, year x month, day of week
  - Wind power forecasts as an IV for short-run price changes

Instrumental Variable strategy



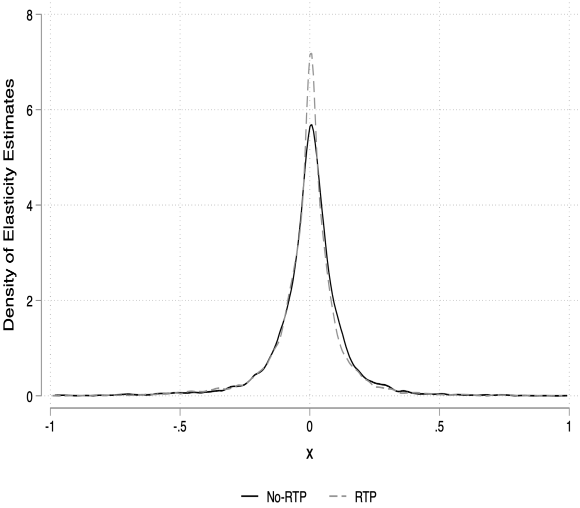
- Instrument shows strong first stage, also after conditioning
- Plausibly exogenous after controlling for local weather conditions

Instrumental Variable challenges

- Most consumers do not consume electricity explicitly based on wind patterns, so exclusion restriction plausibly valid.
- Yet, wind patterns are intertwined with weather.
- Weather can affect electricity consumption in many ways: temperature control, sunset/sunrise, type of activities, time at home, etc.
- Difficult to control for potentially all confounders.
- High-frequency data can easily lead to significant spurious patterns due to omitted variable bias.

We consider an array of fixed-effect individual specifications together with a lasso estimator.

We find similar distributions of price elasticities



- Distribution centered around zero, median of no response.

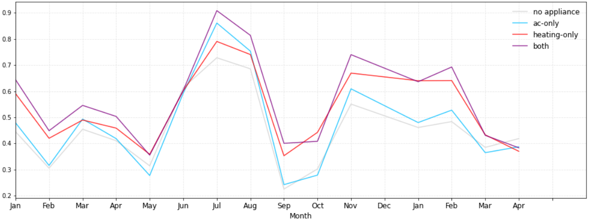
Average elasticities by group are close to zero

	(1) p_iv11	(2) p_iv21	(3) p_iv31	(4) p_lasso
rtp	-0.00513	-0.00430	-0.00374	-0.00468
	(0.00238)	(0.00237)	(0.00220)	(0.00217)
Constant	-0.00473	-0.00883	-0.0117	-0.0237
	(0.00244)	(0.00252)	(0.00182)	(0.00274)
Observations	14598	14598	14598	14598

Standard errors in parentheses

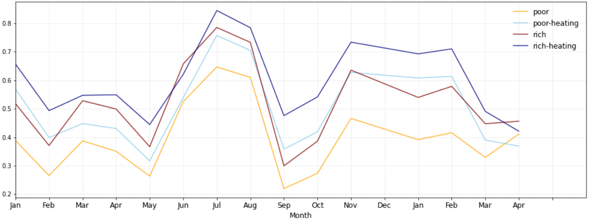
- Not much of an effect from RTP. ▸ Back distributional impacts

RTP shift affected by appliance ownership



- In spite of challenges with algorithm, we extract signal from inferred appliance ownership.

RTP shift modestly progressive: monthly



- AC explains the loss in summer
- More distributional effect in winter, poor have important share of electric heating but consumption smaller than richer households.

Back

An alternative approach

- Focus on zip-code level "aggregate" moments (cross-section).
- Make explicit parametric assumptions on the relationship between income and moments of the distribution of electricity consumption, e.g., integrating $\overline{kwh}_i(inc_i)$, $\overline{kwh}_{ih}(inc_i)$, etc. (drawing from zip-code income distribution)
- Estimate random coefficients that help explain the summarized aggregate data.
- Use Bayes rule to infer a households' income posterior.
- We did not follow this route to avoid simplifying the heterogeneity in the raw electricity consumption data for the policy analysis.

Identifying equations with aggregate moments

We could consider the zip-code level moments:

$$\sum_z \omega_z \sum_h \left(\overline{kwh}_{zh} - \sum Pr_z(\theta_n) kwh_h(\theta_n) \right)^2$$

$$\sum_z \omega_z \sum_k \left(Pr_z(inc_k) - \sum_{i \in z} \sum_n \eta_n^k Pr_z(\theta_n) \right)^2$$

$$\text{s.t.} \quad \sum_k \eta_n^k = 1, \forall n.$$

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

Our approach with micro data

$$\sum_z \omega_z \sum_k \left(Pr_z(inc_k) - \sum_{i \in z} \sum_n \eta_n^k Pr_z(\theta_n) \right)^2$$

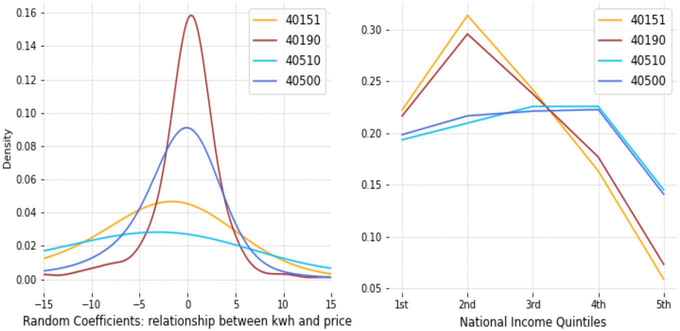
$$\text{s.t. } \sum_k \eta_n^k = 1, \forall n.$$

- Estimate $Pr_z(\theta_n)$ in a first step by classifying consumers into similar types with the micro data.
- Then allow up to Z types to fit income distribution system of equations.
- No longer underidentified subject to overlap in types (full rank).

Back

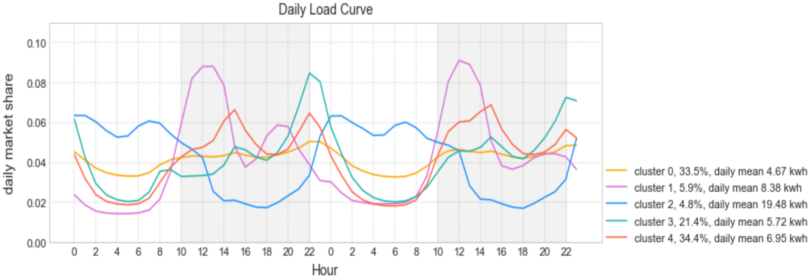
Why not BLP

- First, different zip code regions have different distribution of the random coefficients: the correlation of each household's consumption and income.



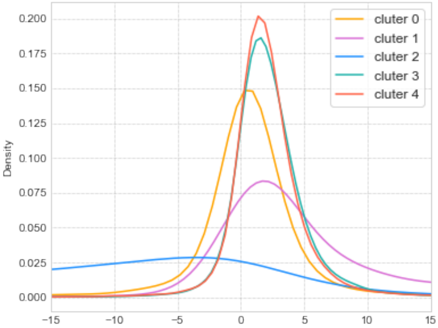
Why not BLP

- Second, electricity consumption is complicated and the price variation is too poor to explain it.



Why not BLP

- Second, electricity consumption is complicated and the price variation is too poor to explain it.



Why not BLP

- Considering these problems of BLP approach
  - We need a flexible model to understand consumption pattern and consumer heterogeneity
  - We need to reveal income from it.

Back