How to Use Python and Mathematical Modeling to Better Understand the Impact of Electricity Pricing on Consumption

Saba Nejad

November 2, 2023

PyData NYC 2023

Hi! I'm Saba. Thank you for coming to my talk.

My name is Saba. I'm a Data Engineer at Point72. Today, I'll be walking you through part of my master's thesis that I worked on and defended two years ago at MIT.

Goals for this talk

- Before starting to analyze your data, you should know details about your data:
  - how it was collected
  - its biases
  - dimenstionality, fill-rate, etc.

Goals for this talk

- Before starting to analyze your data, you should know details about your data:
    - how it was collected
    - its biases
    - dimenstionality, fill-rate, etc.
- Before trying to fit a model to your data, you should have a mathematical model rooted in the real world.

What are we going to talk about?

To that end, I'll be talking about,

- the data from a trial in London that implemented a particular pricing model on a subset of houses to quantify the impact of pricing on electricity consumption.

What are we going to talk about?

To that end, I'll be talking about,

- the data from a trial in London that implemented a particular pricing model on a subset of houses to quantify the impact of pricing on electricity consumption.
- the steps I took to understand this data, and methods I used to prep, clean, and process the data.

What are we going to talk about?

To that end, I'll be talking about,

- the data from a trial in London that implemented a particular pricing model on a subset of houses to quantify the impact of pricing on electricity consumption.
- the steps I took to understand this data, and methods I used to prep, clean, and process the data.
- the thinking behind the mathematical framing of the problem.

What are we going to talk about?

To that end, I'll be talking about,

- the data from a trial in London that implemented a particular pricing model on a subset of houses to quantify the impact of pricing on electricity consumption.
- the steps I took to understand this data, and methods I used to prep, clean, and process the data.
- the thinking behind the mathematical framing of the problem.
- the analysis based on the particular dataset and mathematical model.

What are we going to talk about?

To that end, I'll be talking about,

- the data from a trial in London that implemented a particular pricing model on a subset of houses to quantify the impact of pricing on electricity consumption.
- the steps I took to understand this data, and methods I used to prep, clean, and process the data.
- the thinking behind the mathematical framing of the problem.
- the analysis based on the particular dataset and mathematical model.
- results, takeaways, lessons learned.

Useful Details about the Problem Space

In electicity, there is a supply and demand constaint.

- Electricity is unique in that its **storage is prohibitively costly**.

In electicity, there is a supply and demand constaint.

- Electricity is unique in that its **storage is prohibitively costly**.
- As a result, **supply must meet demand** at all times.

In electicity, there is a supply and demand constaint.

- Electricity is unique in that its **storage is prohibitively costly**.
- As a result, **supply must meet demand** at all times.
- If demand exceeds supply, there will be a power outage.

In electicity, there is a supply and demand constaint.

- Electricity is unique in that its **storage is prohibitively costly**.
- As a result, **supply must meet demand** at all times.
- If demand exceeds supply, there will be a power outage.
- Power outages are extremely costly and we try to prevent them as much as possible.

Demand Response is a tool that helps lower demand.

- This unique feature of electricity, poses a **challenge**.

Demand Response is a tool that helps lower demand.

- This unique feature of electricity, poses a **challenge**.
- Imagine an unpredictably hot day where there is more demand than expected.

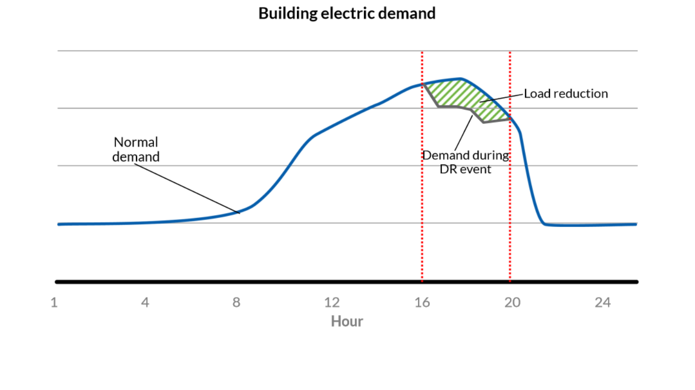
Demand Response is a tool that helps lower demand.

- This unique feature of electricity, poses a **challenge**.
- Imagine an unpredictably hot day where there is more demand than expected.
- To prevent a power outage, we need close to real-time methods to lower demand.

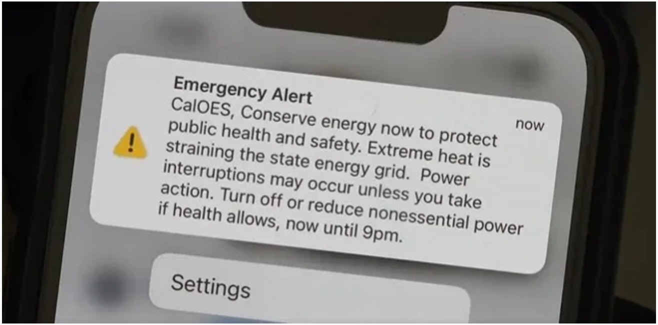
Demand Response is a tool that helps lower demand.

- This unique feature of electricity, poses a **challenge**.
- Imagine an unpredictably hot day where there is more demand than expected.
- To prevent a power outage, we need close to real-time methods to lower demand.
- **Demand Response** is one such method.

Demand Response visual example



Demand Response is a tragedy of the commons issue.



Demand Response also uses pricing models.

- Assumption 1: Demand can be shifted around.

Demand Response also uses pricing models.

- Assumption 1: Demand can be shifted around.
- Assumption 2: Consumers are price sensitive i.e. they will respond to price incentives.

Demand Response also uses pricing models.

- Assumption 1: Demand can be shifted around.
- Assumption 2: Consumers are price sensitive i.e. they will respond to price incentives.
- Two pricing models:

Demand Response also uses pricing models.

- Assumption 1: Demand can be shifted around.
- Assumption 2: Consumers are price sensitive i.e. they will respond to price incentives.
- Two pricing models:
    - Time of Use Pricing

Demand Response also uses pricing models.

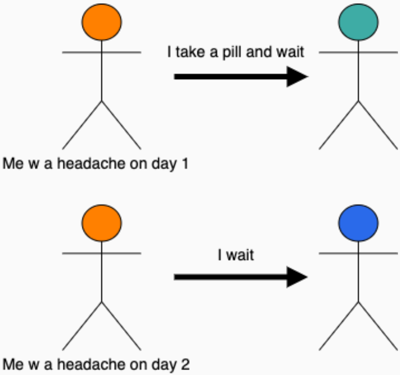
- Assumption 1: Demand can be shifted around.
- Assumption 2: Consumers are price sensitive i.e. they will respond to price incentives.
- Two pricing models:
    - Time of Use Pricing
    - Dynamic Time of Use Pricing

Repeat

There is a pricing model, called **Dynamic Time of Use Pricing**, that uses price incentives to lower demand on the grid and prevent a power outage.

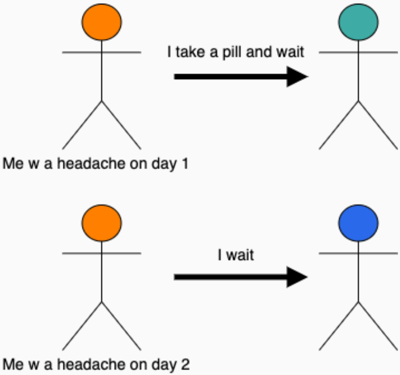
Understanding causality and some useful terminology

- What does it mean for something to have a causal effect on something else



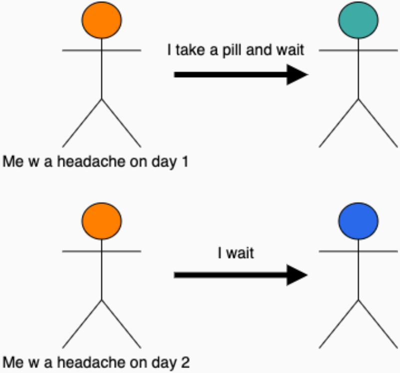
Understanding causality and some useful terminology

- What does it mean for something to have a causal effect on something else
- What is a treatment



Understanding causality and some useful terminology

- What does it mean for something to have a causal effect on something else
- What is a treatment
- What is the counterfactual



The Fundamental Problem of Causal Inference (FPCI)

- Assume $X \in \{0, 1\}$ is a binary causal variable and Y is a response variable (which may be continuous).

The Fundamental Problem of Causal Inference (FPCI)

- Assume $X \in \{0, 1\}$ is a binary causal variable and Y is a response variable (which may be continuous).
- Assume X has a causal effect on Y.
  - $Y_0 :=$ the value of Y when $X = 0$ (untreated unit)
  - $Y_1 :=$ the value of Y when $X = 1$ (treated unit)
  - $T = Y_1 - Y_0 :=$ the treatment effect

The Fundamental Problem of Causal Inference (FPCI)

- Assume $X \in \{0, 1\}$ is a binary causal variable and Y is a response variable (which may be continuous).
- Assume X has a causal effect on Y.
    - $Y_0 :=$ the value of Y when $X = 0$ (untreated unit)
    - $Y_1 :=$ the value of Y when $X = 1$ (treated unit)
    - $T = Y_1 - Y_0 :=$ the treatment effect
- Y_0 and Y_1 are **counterfactuals** of one another. We observe **only** Y_1 or Y_0, but not both at the same time.

Solutions Around FPCI

1. Temporal Stability & Causal Transience

Solutions Around FPCI

1. Temporal Stability & Causal Transience
2. Unit Homogeneity

Solutions Around FPCI

1. Temporal Stability & Causal Transience
2. Unit Homogeneity
3. Estimate causal effects for populations rather than units

Solutions Around FPCI for this Trial

1. Temporal Stability & Causal Transience X
2. Unit Homogeneity X
3. Estimate causal effects for populations rather than units ✓

Random samples of populations are studied to estimate treatment effects.

- It's difficult to estimate the treatment effect for a unit.
  $\implies$ analyses are done on a **population vs a unit**.

Random samples of populations are studied to estimate treatment effects.

- It's difficult to estimate the treatment effect for a unit.
  $\Longrightarrow$ analyses are done on a **population vs a unit**.
- When estimating treatment effect for populations, to remove bias terms, we want samples that behave similarly out of sample.
  $\Longrightarrow$ analyses are done on **random samples** of populations.

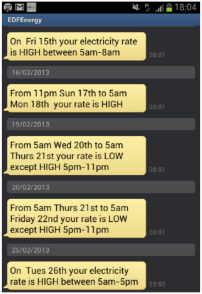
The Details of the Trial Studied

Low Carbon London Smart Meter Trial

- Motivation: The Climate Change Act of 2008 sets the target of reducing carbon emissions to 20% of 1990 levels by 2050.

Low Carbon London Smart Meter Trial

- Motivation: The Climate Change Act of 2008 sets the target of reducing carbon emissions to 20% of 1990 levels by 2050.
- "Dynamic": Prices were given a day ahead via the Smart Meter In Home Display or text message.



Low Carbon London Smart Meter Trial

- Data spans November 2011 and February 2014 (around 167M records).

Low Carbon London Smart Meter Trial

- Data spans November 2011 and February 2014 (around 167M records).
- Treatment took place in the calendar year 2013.

Low Carbon London Smart Meter Trial

- Data spans November 2011 and February 2014 (around 167M records).
- Treatment took place in the calendar year 2013.
- Readings were taken at half hourly intervals.

Low Carbon London Smart Meter Trial

- Data spans November 2011 and February 2014 (around 167M records).
- Treatment took place in the calendar year 2013.
- Readings were taken at half hourly intervals.
- dTou price bands vs static pricing model:

dToU Pricing Model	Static Pricing Model
High (67.20 p/kWh)	14.228 p/kWh
Normal (11.76 p/kWh)	14.228 p/kWh
Low (3.99 p/kWh)	14.228 p/kWh

Additional Features: Discrete Socio-Economic Categories

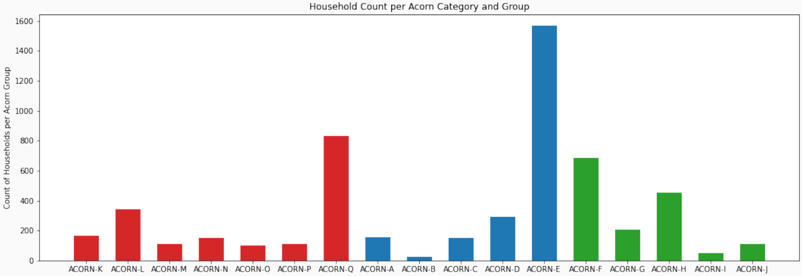
Additional Features: Discrete Socio-Economic Categories

Label	Acorn group	Acorn category
A	Wealthy executives	Wealthy achievers
B	Affluent greys	Wealthy achievers
C	Flourishing families	Wealthy achievers
D	Prosperous professionals	Urban prosperity
E	Educated urbanites	Urban prosperity
F	Aspiring singles	Urban prosperity
G	Starting out	Comfortably off
H	Secure families	Comfortably off
I	Settled suburbia	Comfortably off
J	Prudent pensioners	Comfortably off
K	Asian communities	Moderate means
L	Post industrial families	Moderate means
M	Blue collar roots	Moderate means
N	Struggling families	Hard pressed
O	Burdened singles	Hard pressed

Additional Features: Discrete Socio-Economic Categories

		house_count
acorn_category	acorn_group	
Adversity	ACORN-K	165
	ACORN-L	342
	ACORN-M	113
	ACORN-N	152
	ACORN-O	103
	ACORN-P	110

Additional Features: Discrete Socio-Economic Categories



Low Carbon London Smart Meter Trial: Biases?

The trial is double opt-in:

1. Households opt into sharing their data with the trial at all.

Low Carbon London Smart Meter Trial: Biases?

The trial is double opt-in:

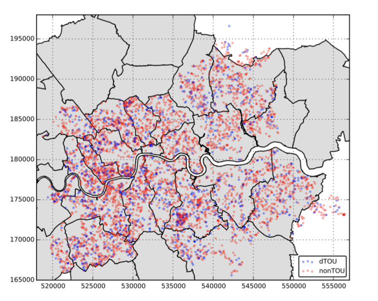
1. Households opt into sharing their data with the trial at all.
2. Some opted into the treatment group: subject to the dToU pricing model in 2013.

Low Carbon London Smart Meter Trial: Biases?

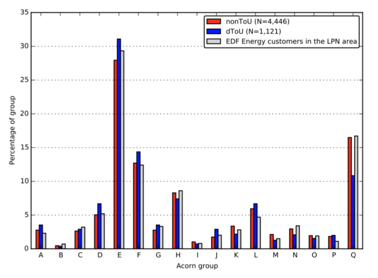
The treatment group was given some incentives:

- A guarantee that they will be reimbursed at the end of trial if they are worse off on the dToU tariff than they would have been on their previous tariff.
- Assurances regarding how many hours would be charged at the high price band.
- £100 for signing up to the dToU tariff.
- Another £50 for staying on the dToU tariff until the end of trial.
- Entry into a prize draw after completion of the post trial survey.

Low Carbon London Smart Meter Trial: Distribution?



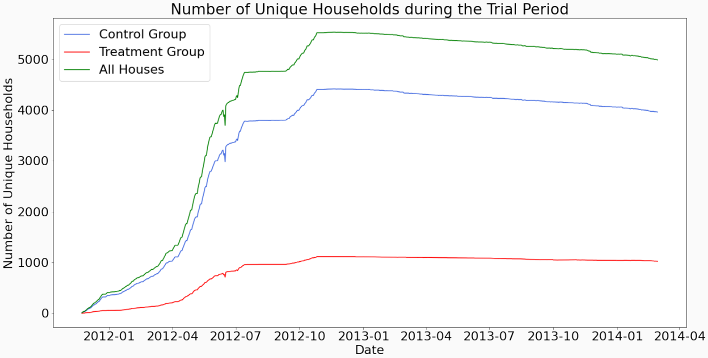
(a) Trial household sample locations overlaid on the borough boundary map of Greater London. This shows that the treatment and control group were representative samples of Greater London.



(b) The treatment and control groups were also representative socio-economic samples of Greater London.

Low Carbon London Smart Meter Trial: Missing Data

- 2012: Users were still onboarding.
- 2013: Some users dropped out of both the treatment and the control groups
- ~5,600 total households participated: ~4,500 were in the control group, ~1,100 were in the treatment group.



Treatment vs Control Group Baselines

- Though our groups are **geographically and socio-economically balanced**, one group opted into dToU pricing and was offered some incentives.

Treatment vs Control Group Baselines

- Though our groups are **geographically and socio-economically balanced**, one group opted into dToU pricing and was offered some incentives.
- What are some reasons the baselines might differ given what we know?

Treatment vs Control Group Baselines

- Though our groups are **geographically and socio-economically balanced**, one group opted into dToU pricing and was offered some incentives.
- What are some reasons the baselines might differ given what we know?
    - price sensitive group opted in

Treatment vs Control Group Baselines

- Though our groups are **geographically and socio-economically balanced**, one group opted into dToU pricing and was offered some incentives.
- What are some reasons the baselines might differ given what we know?
    - price sensitive group opted in
    - low consuming group opted in

Treatment vs Control Group Baselines

- Though our groups are **geographically and socio-economically balanced**, one group opted into dToU pricing and was offered some incentives.
- What are some reasons the baselines might differ given what we know?
    - price sensitive group opted in
    - low consuming group opted in
    - those households are flexible time-wise: particular jobs, hours, etc.

The Math

Mathematical Model

Consider the following segmentation of the data.

- $\alpha_y :=$ control group's consumption matrix during year y
- $\beta_y :=$ treatment group's consumption matrix during year y

Dimenstionality Analysis: Control Group Matrices

$$\alpha_y = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n_c} \\ a_{21} & a_{22} & \cdots & a_{2n_c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{t1} & a_{t2} & \cdots & a_{tn_c} \end{bmatrix}$$

α matrics are of size $\alpha_{t \times n_c}$.

Dimenstionality Analysis: Control Group Matrices

$$\alpha_y = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n_c} \\ a_{21} & a_{22} & \cdots & a_{2n_c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{t1} & a_{t2} & \cdots & a_{tn_c} \end{bmatrix}$$

α matrics are of size $\alpha_{t \times n_c}$.

t is the number of half-hour measurements in a year ($t = 365 \times 48 = 17,520$ for 2012 and 2013).

Dimentionality Analysis: Control Group Matrices

$$\alpha_y = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n_c} \\ a_{21} & a_{22} & \cdots & a_{2n_c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{t1} & a_{t2} & \cdots & a_{tn_c} \end{bmatrix}$$

α matrics are of size $\alpha_{t \times n_c}$.

t is the number of half-hour measurements in a year ($t = 365 \times 48 = 17,520$ for 2012 and 2013).

n_c is the number of households in the control group.

Dimenstionality Analysis: Treatment Group Matrices

$$\beta_y = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n_t} \\ b_{21} & b_{22} & \cdots & b_{2n_t} \\ \vdots & \vdots & \ddots & \vdots \\ b_{t1} & b_{t2} & \cdots & b_{tn_t} \end{bmatrix}$$

β matrics are of size $\alpha_{t \times n_t}$.

Dimenstionality Analysis: Treatment Group Matrices

$$\beta_y = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n_t} \\ b_{21} & b_{22} & \cdots & b_{2n_t} \\ \vdots & \vdots & \ddots & \vdots \\ b_{t1} & b_{t2} & \cdots & b_{tn_t} \end{bmatrix}$$

β matrics are of size $\alpha_{t \times n_t}$.

t is the number of half-hour measurements in a year ($t = 365 \times 48 = 17,520$ for 2012 and 2013).

Dimenstionality Analysis: Treatment Group Matrices

$$\beta_y = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n_t} \\ b_{21} & b_{22} & \cdots & b_{2n_t} \\ \vdots & \vdots & \ddots & \vdots \\ b_{t1} & b_{t2} & \cdots & b_{tn_t} \end{bmatrix}$$

β matrics are of size $\alpha_{t \times n_t}$.

t is the number of half-hour measurements in a year ($t = 365 \times 48 = 17,520$ for 2012 and 2013).

n_t is the number of households in the treatment group.

Mathematical Model

Year	Control Group	Treatment Group
2011	α_{2011}	β_{2011}
2012	α_{2012}	β_{2012}
2013	α_{2013}	$\hat{\beta}_{2013}$
2014	α_{2014}	β_{2014}

$\hat{\beta}_{2013}$ is the counterfactual consumption for β_{2013}.

Mathematical Model

Year	Control Group	Treatment Group
2011	α_{2011}	β_{2011}
2012	α_{2012}	β_{2012}
2013	α_{2013}	$\hat{\beta}_{2013}$
2014	α_{2014}	β_{2014}

$\hat{\beta}_{2013}$ is the counterfactual consumption for β_{2013}.

$T = \hat{\beta}_{2013} - \beta_{2013}$ is the treatment effect.

Mathematical Model

Year	Control Group	Treatment Group
2011	α_{2011}	β_{2011}
2012	α_{2012}	β_{2012}
2013	α_{2013}	$\hat{\beta}_{2013}$
2014	α_{2014}	β_{2014}

$\hat{\beta}_{2013}$ is the counterfactual consumption for β_{2013}.

$T = \hat{\beta}_{2013} - \beta_{2013}$ is the treatment effect.

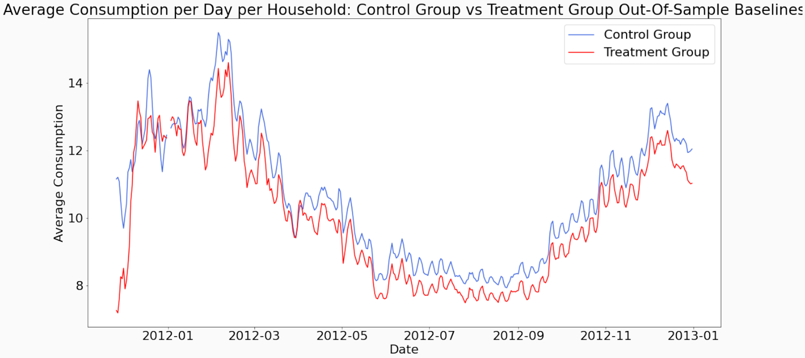
Goal is to **find a good estimate** for $\hat{\beta}_{2013}$.

Naive Model

If we didn't know about all these biases, we *could* use control group in 2013 as the counterfactual estimate for treatment group in 2013.

$$\begin{aligned} \hat{\beta}_{2013} &\approx \alpha_{2013} \\ T &= \beta_{2013} - \hat{\beta}_{2013} \\ &= \beta_{2013} - \alpha_{2013} \end{aligned} \tag{1}$$

Treatment vs Control Group Baselines



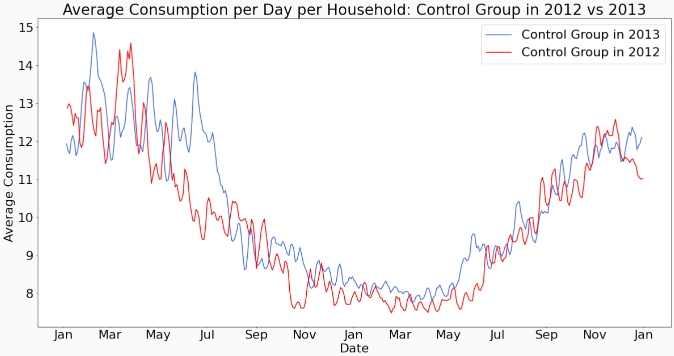
$\Longrightarrow$ the two groups don't have the same baseline electricity consumption patterns.

Sophisticated Naive Model

Let's instead use treatment group in 2012 as the counterfactual estimate for treatment group in 2013.

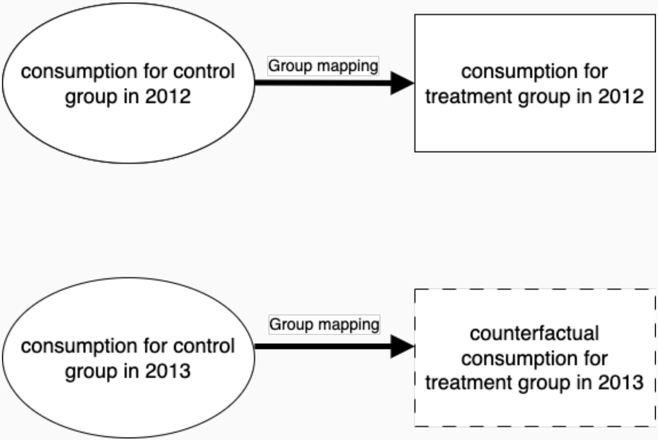
$$\begin{aligned} \hat{\beta}_{2013} &\approx \beta_{2012} \\ T &= \beta_{2013} - \hat{\beta}_{2013} \\ &= \beta_{2013} - \beta_{2012} \end{aligned} \tag{2}$$

Changes from Time: 2012 vs 2013

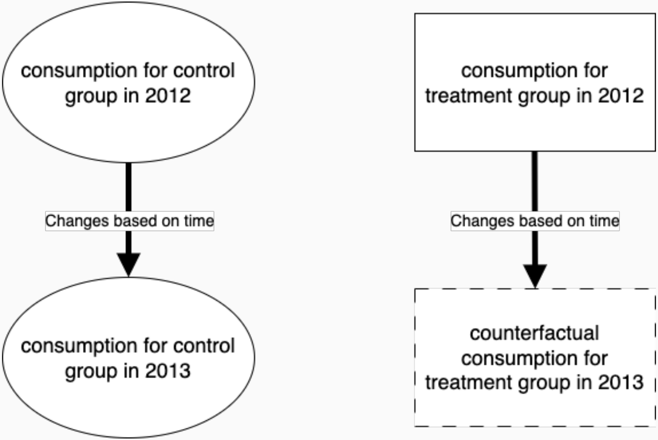


$\Longrightarrow$ Things changed between the two years even for the control group.

Treatment vs Control Group Baselines



Treatment vs Control Group Baselines



Mathematical Model: Using Group Mapping

$$
\begin{aligned}
\alpha_{2012} X &= \beta_{2012} \\
\alpha_{2013} X &= \hat{\beta}_{2013} \\
\Delta\text{treatment} &= \beta_{2013} - \hat{\beta}_{2013}
\end{aligned} \tag{3}
$$

The Analysis

Data Prep, Cleaning, Processing

UKPN-LCL-smartmeter-sample.csv	Nov 3, 2020 at 1:44 PM	1 MB	Comma...et (.csv)
Tariffs.xlsx	Jan 4, 2021 at 9:02 PM	235 KB	Microso...k (.xlsx)
Tariffs.csv	Jan 4, 2021 at 9:02 PM	371 KB	Comma...et (.csv)
tariffs_csv.csv	Jan 4, 2021 at 9:02 PM	371 KB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_168.csv	Aug 20, 2015 at 1:09 PM	64.7 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_167.csv	Aug 20, 2015 at 1:09 PM	68.5 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_166.csv	Aug 20, 2015 at 1:09 PM	68.7 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_165.csv	Aug 20, 2015 at 1:08 PM	68.8 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_164.csv	Aug 20, 2015 at 1:08 PM	68.5 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_163.csv	Aug 20, 2015 at 1:07 PM	68.5 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_162.csv	Aug 20, 2015 at 1:07 PM	68.9 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_161.csv	Aug 20, 2015 at 1:06 PM	69 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_160.csv	Aug 20, 2015 at 1:06 PM	69.1 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_159.csv	Aug 20, 2015 at 1:05 PM	69.4 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_158.csv	Aug 20, 2015 at 1:05 PM	69.4 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_157.csv	Aug 20, 2015 at 1:04 PM	69.5 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_156.csv	Aug 20, 2015 at 1:04 PM	69.5 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_155.csv	Aug 20, 2015 at 1:03 PM	69.7 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_154.csv	Aug 20, 2015 at 1:03 PM	69 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_153.csv	Aug 20, 2015 at 1:02 PM	69.1 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_152.csv	Aug 20, 2015 at 1:02 PM	68.8 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_151.csv	Aug 20, 2015 at 1:01 PM	68.7 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_150.csv	Aug 20, 2015 at 1:01 PM	68.7 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_149.csv	Aug 20, 2015 at 1:00 PM	68.8 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_148.csv	Aug 20, 2015 at 1:00 PM	68.8 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_147.csv	Aug 20, 2015 at 1:00 PM	68.9 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_146.csv	Aug 20, 2015 at 12:59 PM	68.5 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_145.csv	Aug 20, 2015 at 12:59 PM	69.4 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_144.csv	Aug 20, 2015 at 12:58 PM	68.6 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_143.csv	Aug 20, 2015 at 12:58 PM	69 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_142.csv	Aug 20, 2015 at 12:57 PM	68.8 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_141.csv	Aug 20, 2015 at 12:57 PM	68.8 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_140.csv	Aug 20, 2015 at 12:56 PM	68.8 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_139.csv	Aug 20, 2015 at 12:56 PM	68.7 MB	Comma...et (.csv)
Power-Networks-LCL-June2015(withAcornGps)v2_138.csv	Aug 20, 2015 at 12:55 PM	68.9 MB	Comma...et (.csv)

Data Prep, Cleaning, Processing

1. first set of files, segmented by year, contain the `house_id` (str), `treated` (bool), `date_time` (datetime), `KWH/hh` (float).

Data Prep, Cleaning, Processing

1. first set of files, segmented by year, contain the `house_id` (str), `treated` (bool), `date_time` (datetime), `KWH/hh` (float).
2. second contains `house_id` (str), `acorn` (str), `acorn_group` (str).

Data Prep, Cleaning, Processing

1. first set of files, segmented by year, contain the `house_id` (str), `treated` (bool), `date_time` (datetime), `KWH/hh` (float).
2. second contains `house_id` (str), `acorn` (str), `acorn_group` (str).
3. third is the tarrif file contains `date_time` (datetime), price per designation for that hh (float).

Data Prep, Cleaning, Processing

Name	Date Modified	Size	Kind
tariffs.gzip	Mar 29, 2021 at 9:51 PM	108 KB	gzip co...archive
total_acorn.gzip	May 16, 2021 at 4:30 PM	30 KB	gzip co...archive
total_usage_2011.gzip	Mar 2, 2021 at 12:18 AM	628 KB	gzip co...archive
total_usage_2012.gzip	Apr 26, 2022 at 7:30 PM	185.2 MB	gzip co...archive
total_usage_2013.gzip	Apr 26, 2022 at 10:28 AM	298.9 MB	gzip co...archive
total_usage_2014.gzip	Apr 26, 2022 at 10:28 AM	20.9 MB	gzip co...archive
total_usage.gzip	Mar 2, 2021 at 12:16 AM	549.6 MB	gzip co...archive

If we look at percent change without normalizing the baselines the treatment was effective.

Other Clustering?

I used different clustering methods to cluster houses in this data set based on the control group's consumption pre-intervention: k-means clustering, PCA3, TSNE4. I also tried clustering on the frequency responses which I found fast Fourier trans- form. PCA, TSNE, and Agglomerative clustering on the resulting data set was also inconclusive. This is reason for sticking to the pre-existing socio-economic clusters.

Assume matrix $\alpha^{t \times n_c}$ contains control group data, and $\beta^{t \times n_t}$ contains treatment group data. Where $t = 365 \times 48 = 17520$, n_c is the number of households in the control group and n_t is the number of households in the treatment group.

$$\begin{aligned} \overline{\beta^m_{2012}} &= a \times \overline{\alpha^m_{2012}} + b \\ \overline{\hat{\beta}^m_{2013}} &= a \times \overline{\alpha^m_{2013}} + b \\ \overline{\Delta \text{treatment}} &= \overline{\beta^m_{2013}} - \overline{\hat{\beta}^m_{2013}} \end{aligned} \tag{4}$$

The Conclusion

What Did We Learn?

- A little bit about electricity markets?

What Did We Learn?

- A little bit about electricity markets?
- Causal analysis.

What Did We Learn?

- A little bit about electricity markets?
- Causal analysis.
- How to break down your data to match the mathematical model at hand.

What Did We Learn?

- A little bit about electricity markets?
- Causal analysis.
- How to break down your data to match the mathematical model at hand.
- Questions to ask before analysis:

What Did We Learn?

- A little bit about electricity markets?
- Causal analysis.
- How to break down your data to match the mathematical model at hand.
- Questions to ask before analysis:
    - How was the data collected?

What Did We Learn?

- A little bit about electricity markets?
- Causal analysis.
- How to break down your data to match the mathematical model at hand.
- Questions to ask before analysis:
  - How was the data collected?
  - What biases are present?

What Did We Learn?

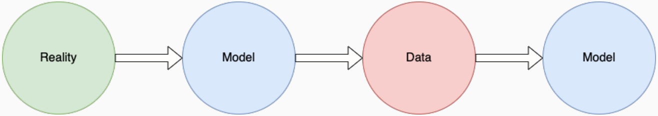
- A little bit about electricity markets?
- Causal analysis.
- How to break down your data to match the mathematical model at hand.
- Questions to ask before analysis:
    - How was the data collected?
    - What biases are present?
    - How does that impact your analysis and results?

Takeaways

- KYD: know your data: learn how it was collected, its biases, shortcomings, feature space, problem space.

Takeaways

- KYD: know your data: learn how it was collected, its biases, shortcomings, feature space, problem space.
- Before diving into fitting a model to your data, it's important to base your analysis on some mathematical model



Thank You! Questions?

- https://github.com/sabanejad
- https://www.linkedin.com/in/sabanejad/