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.

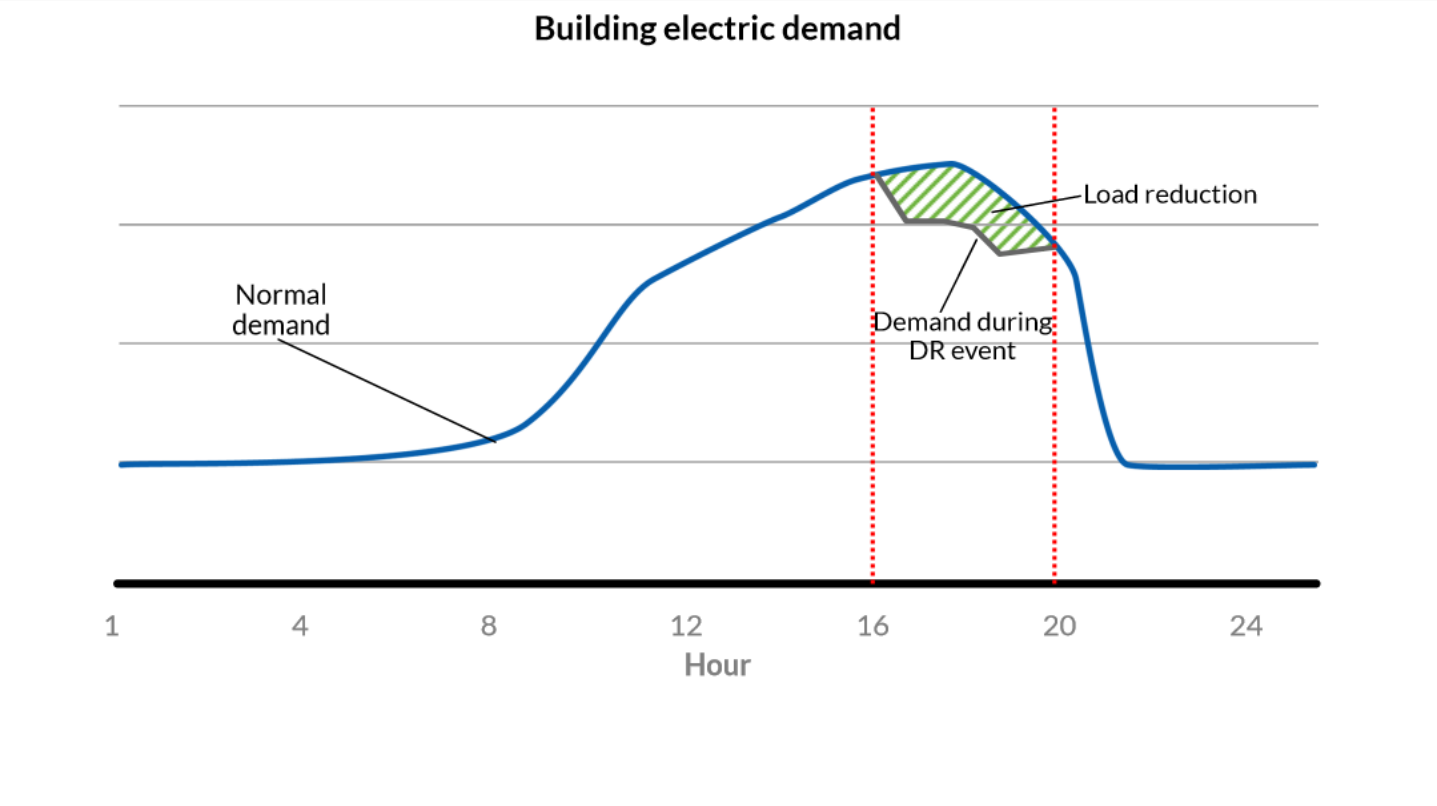
Demand Response is a tool that helps lower demand.

- 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.

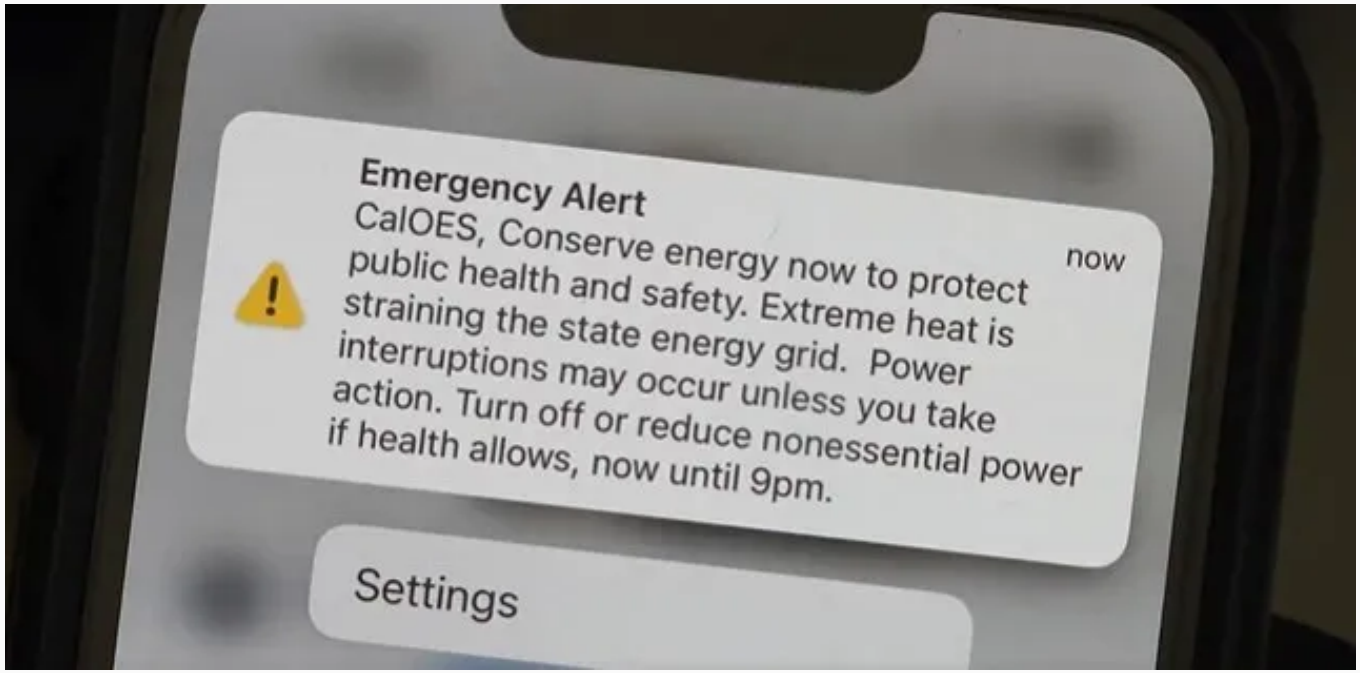
Demand Response is a tool that helps lower demand.

- 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



One demand response method are blast texts.



Another demand response method is using pricing models.

These pricing models **increase** price, to incentivize people to

- Assumption 1: Demand can be shifted around.

Another demand response method is using pricing models.

These pricing models **increase** price, to incentivize people to

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

Another demand response method is using pricing models.

These pricing models **increase** price, to incentivize people to

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

Another demand response method is using pricing models.

These pricing models **increase** price, to incentivize people to

- 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 (ToU)

Another demand response method is using pricing models.

These pricing models **increase** price, to incentivize people to

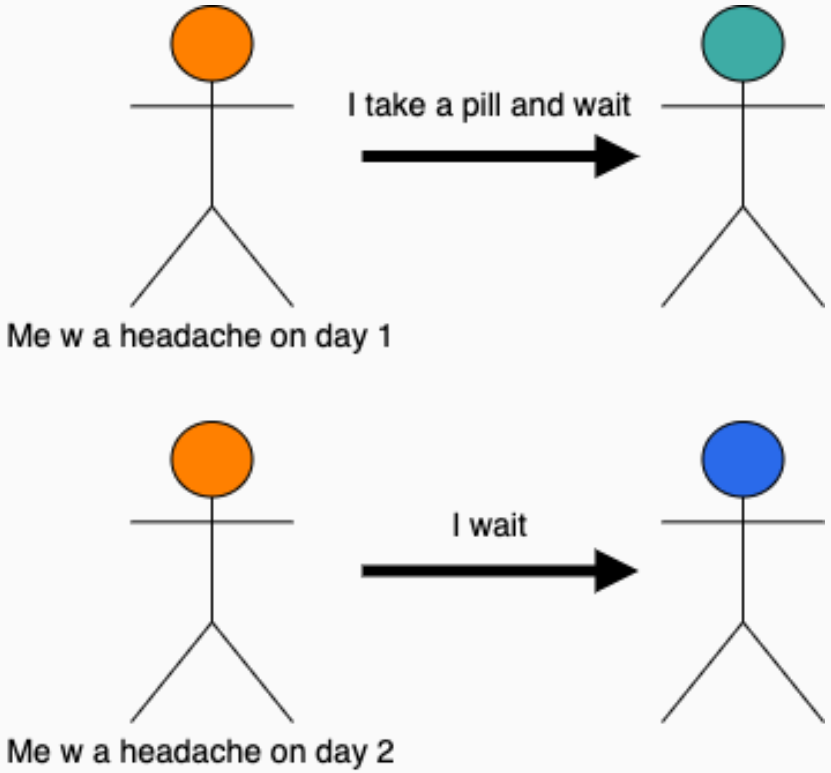
- Assumption 1: Demand can be shifted around.
- Assumption 2: Consumers are price sensitive i.e. they will respond to price incentives.
- Two pricing models:
  - Time of Use Pricing (ToU)
  - Dynamic Time of Use Pricing (dToU)

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.

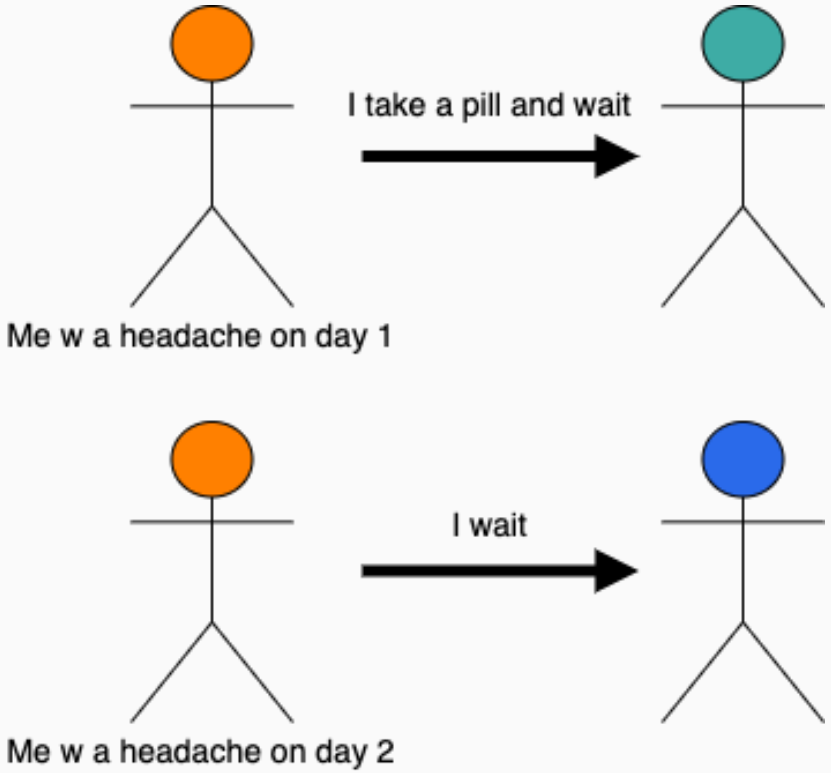
Understanding causality and some useful terminology

- What is a treatment? What is the treatment effect?



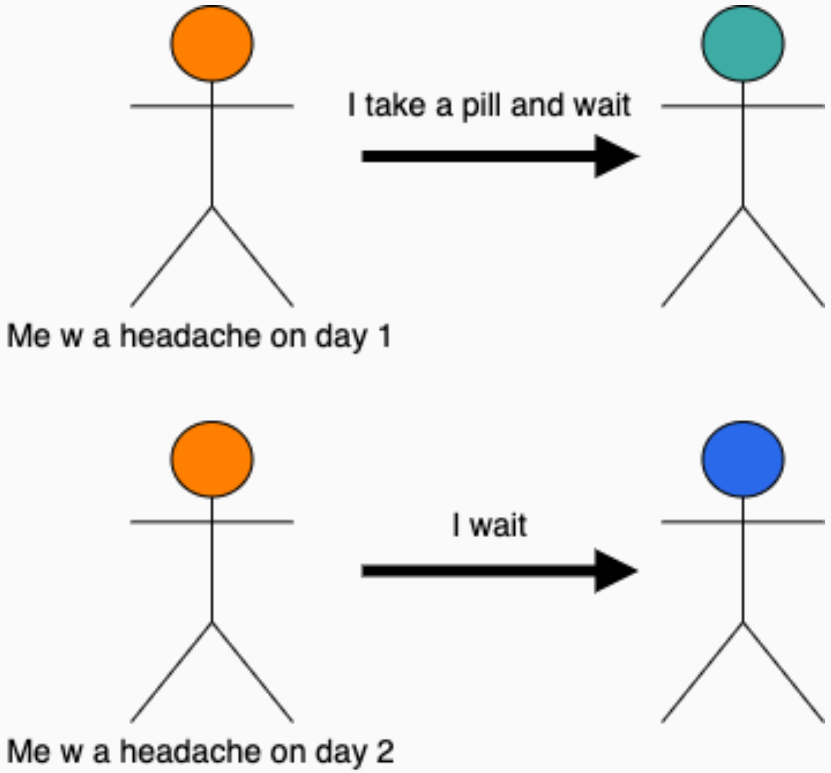
Understanding causality and some useful terminology

- What is a treatment? What is the treatment effect?
- What is a 'unit' undergoing treatment?



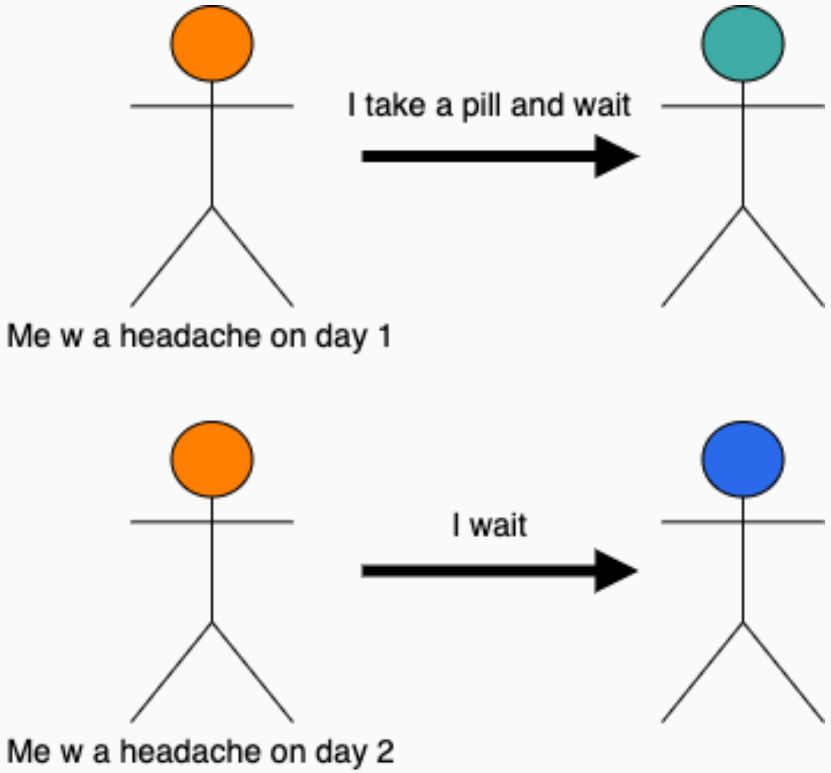
Understanding causality and some useful terminology

- What is a treatment? What is the treatment effect?
- What is a 'unit' undergoing treatment?
- What does it mean for something to have a causal effect on something else?



Understanding causality and some useful terminology

- What is a treatment? What is the treatment effect?
- What is a 'unit' undergoing treatment?
- What does it mean for something to have a causal effect on something else?
- 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.
  $\implies$ FPCI makes it **necessary** to study a **population vs a unit**. The **larger the population**, the easier to robustly estimate treatment effects.

Random samples of populations are studied to estimate treatment effects.

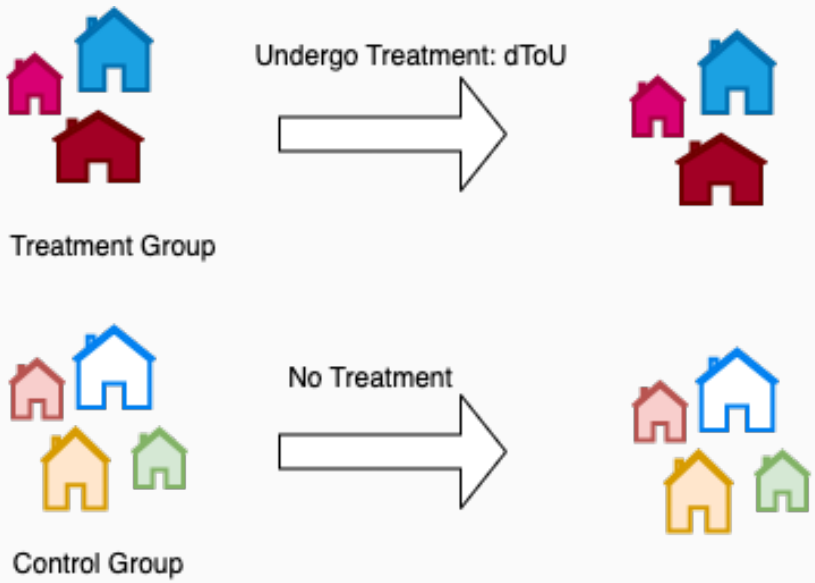
- It's difficult to estimate the treatment effect for a unit.
  $\implies$ FPCI makes it **necessary** to study a **population vs a unit**. The **larger the population**, the easier to robustly estimate treatment effects.
- When estimating treatment effect for populations, to remove bias terms, we want groups that behave similarly out of sample.
  $\implies$ analyses are done on **random samples** of populations.

What have we learned and where are we headed?

in summary, we want large, random samples of populations to estimate effect of a treatment. perhaps connecting this back to the second lesson of the talk, its important to base your analysis in some mathematical framework because it'll change what you are able to do with the data you have. now we have some mathematical tools in our toolbox to study causality and estimate treatment effects.we know a bit about the pricing model in this trial. now let's go learn more about the trial and the data

The Details of the Trial Studied

High Level Overview of the Trial



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.
- Data spans November 2011 and February 2014.

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.
- Data spans November 2011 and February 2014.
- Treatment took place in the calendar year 2013.

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.
- Data spans November 2011 and February 2014.
- Treatment took place in the calendar year 2013.
- Readings were taken at half hourly intervals, for the entire duration of the trial, and for 5600 houses in total (around 167M records).

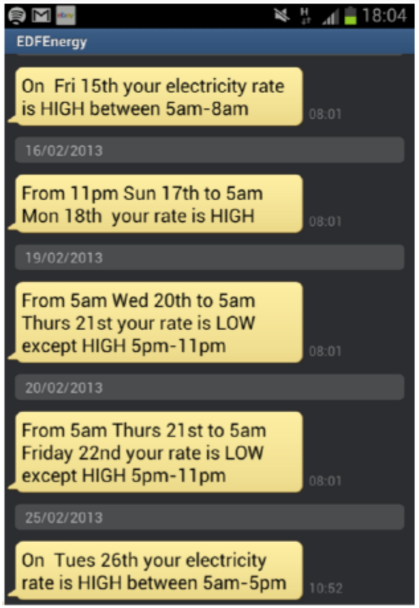
dToU: pre-set prices, implemented at different times of day

- dToU price bands vs static pricing model:

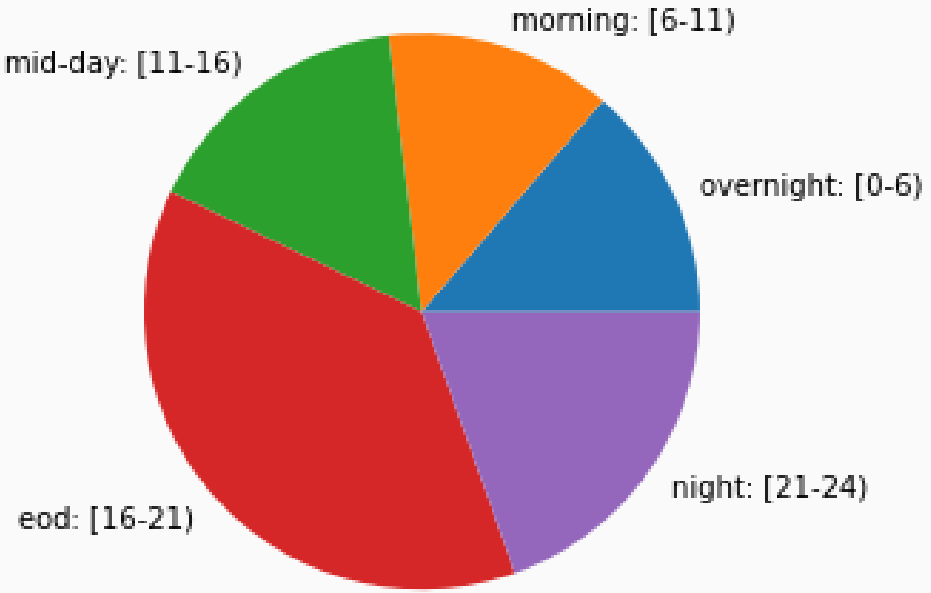
dToU Pricing Model	Static Pricing Model
High (67.20 p/kWh)	14.228 p/kWh
Normal (11.76 p/kWh)	14.228 p/kWh
Low (3.99 p/kWh)	14.228 p/kWh

dToU: pre-set prices, implemented at different times of day

- "Dynamic": Times of day when households would be subject to high price point was communicated a day ahead via the Smart Meter In Home Display or text message.



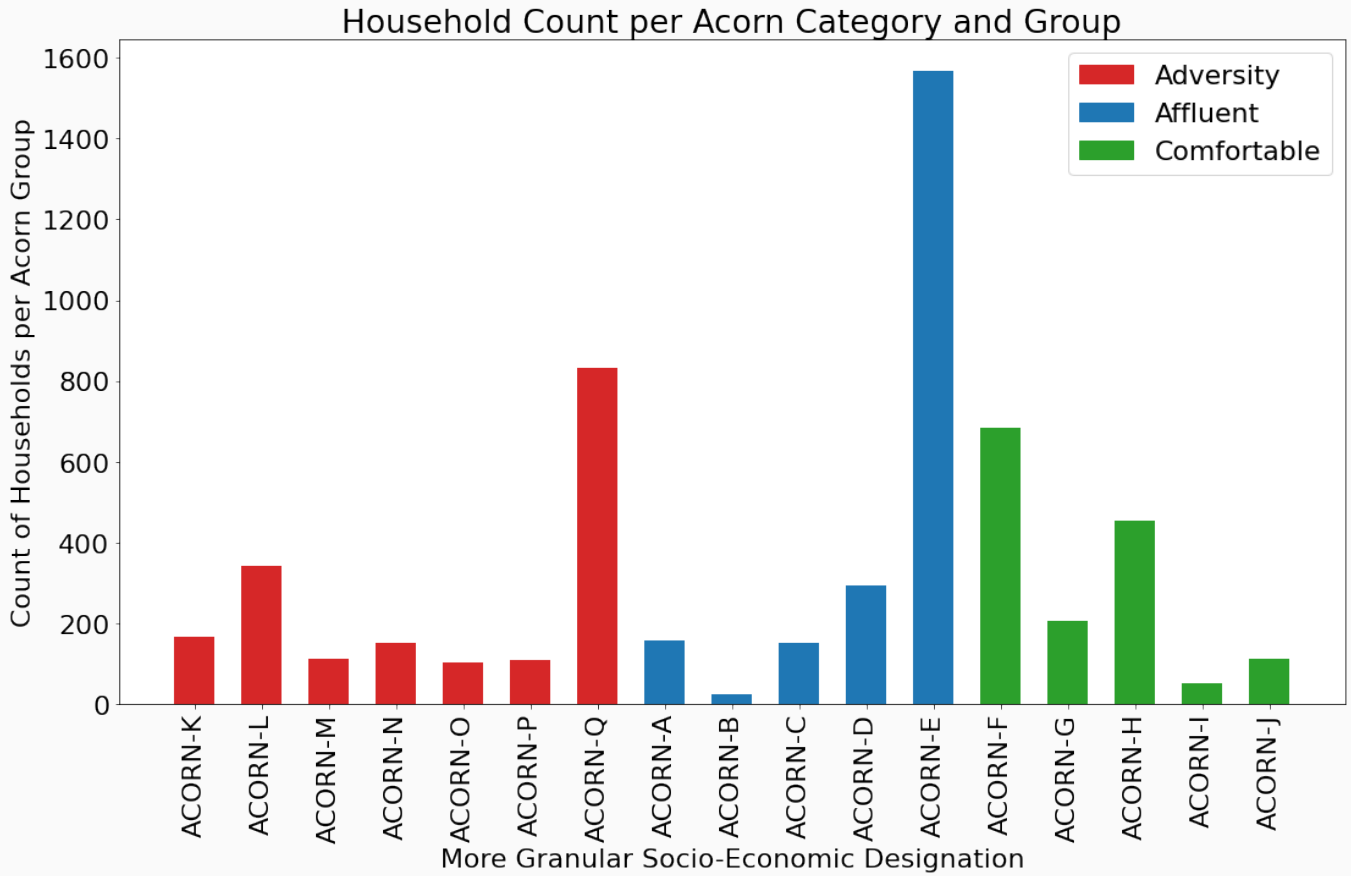
Distribution of High Price Periods



Features: Two Levels of 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
P	High rise hardship	Hard pressed
Q	Inner city adversity	Hard pressed

Features: Two Levels of Discrete Socio-Economic Categories



Biases I: Treatment group opted in.

The trial is double opt-in:

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

Biases I: Treatment group opted in.

The trial is double opt-in:

1. Households opt into sharing their data with the trial at all.
2. Some opted into undergoing dToU pricing in 2013.

Biases II: There were cash incentives for the treatment group.

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.

The two groups were not biased geographically

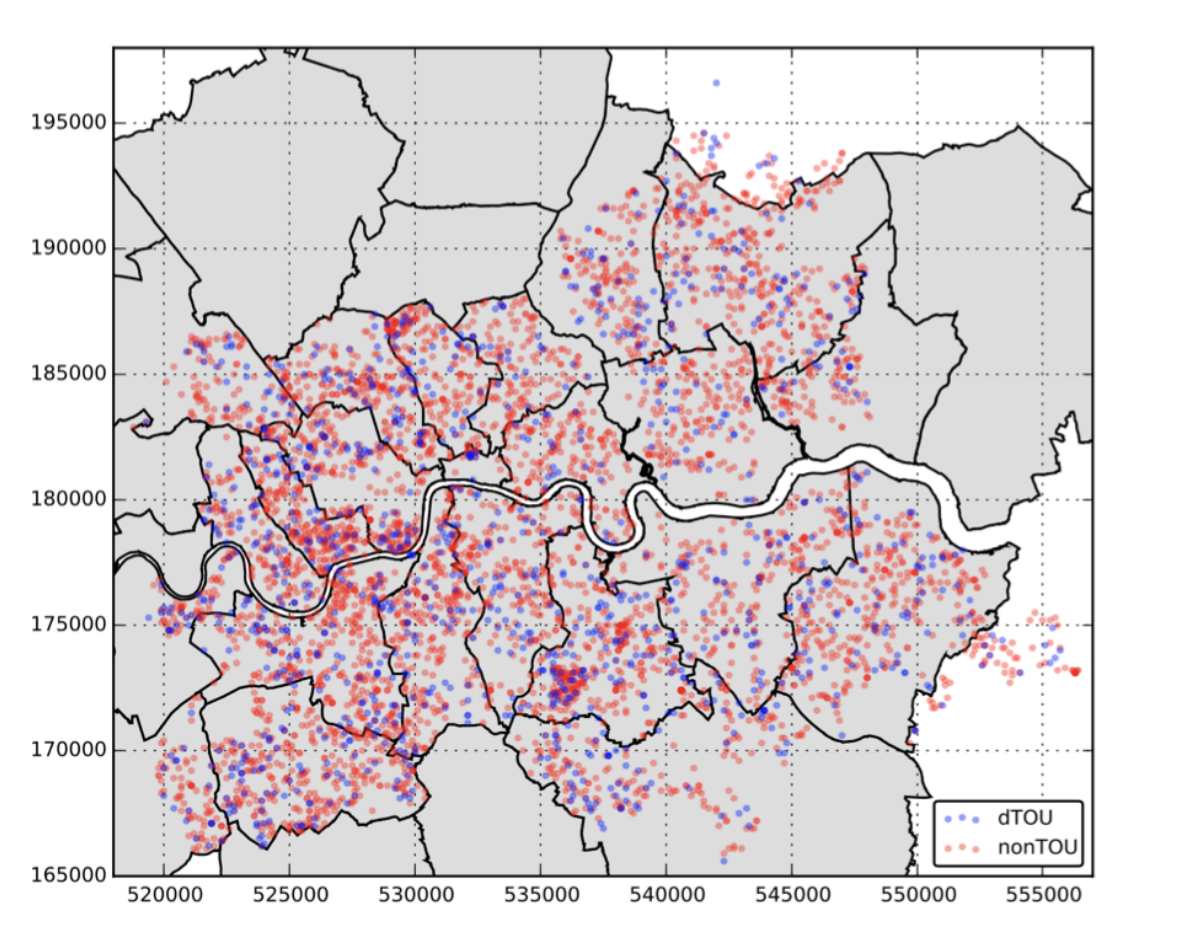


Figure 1: Trial household sample locations overlaid on the borough boundary map of Greater London. This shows that the treatment and control group were representative samples of Greater London.

The two groups were not biased socio-economically

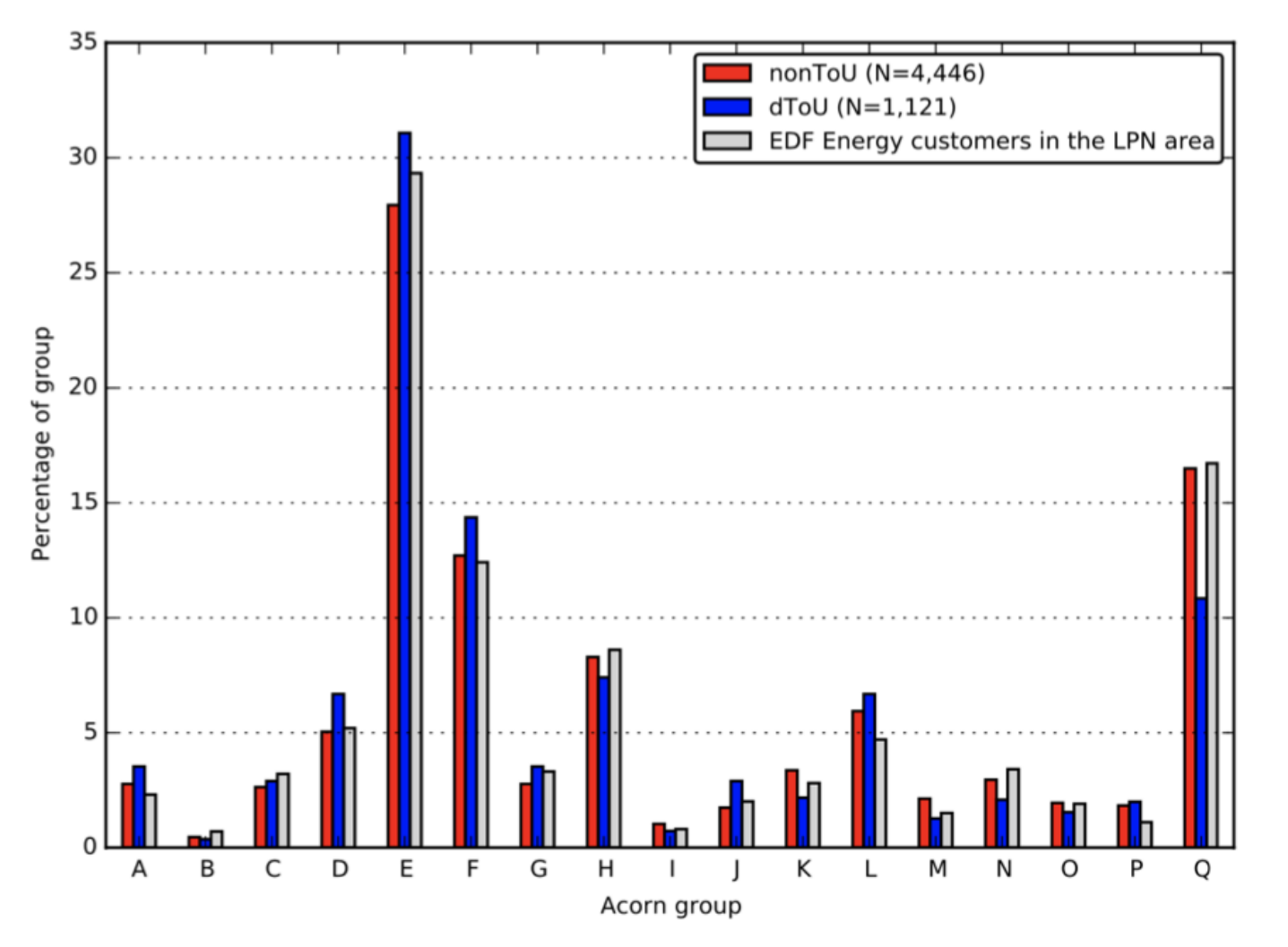
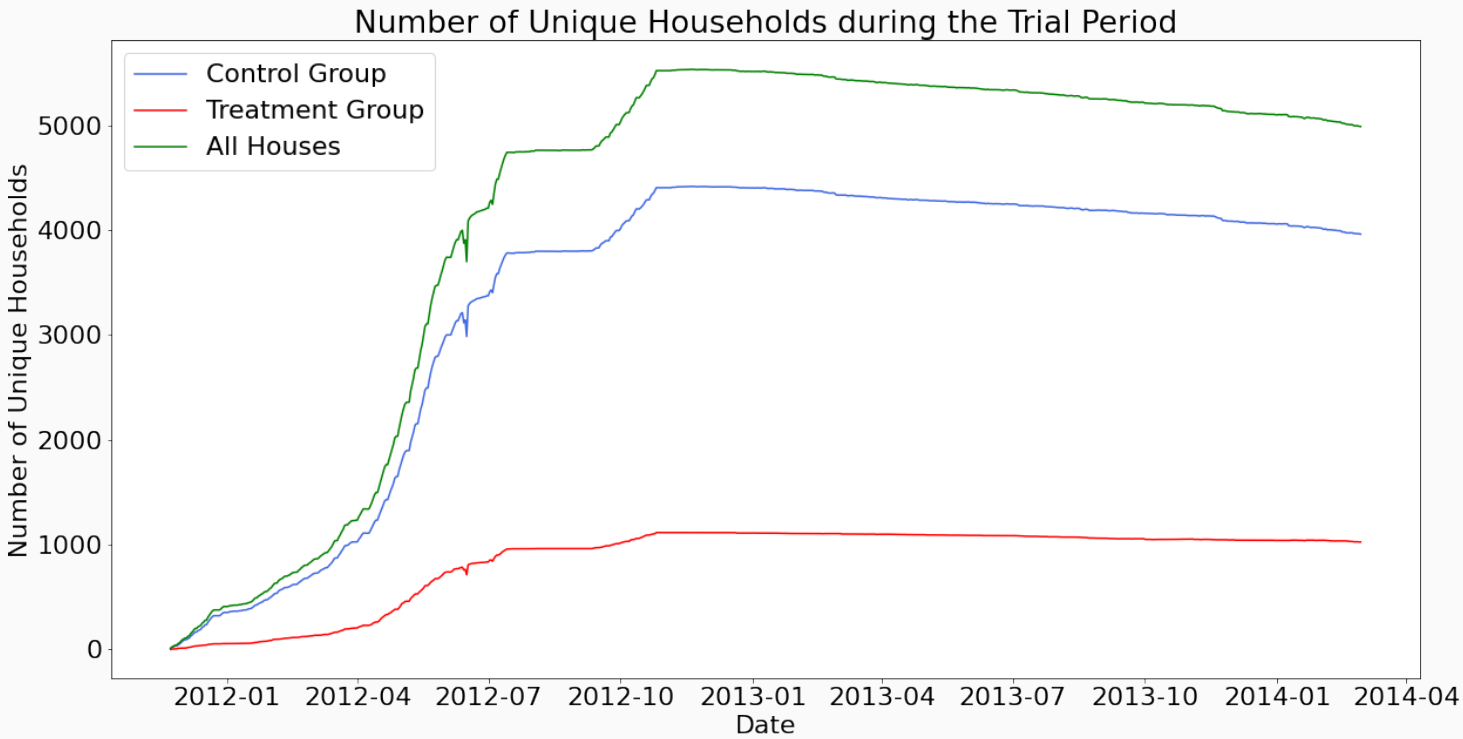


Figure 2: The treatment and control groups were also representative socio-economic samples of Greater London.

Onboarding households creates a missing data issue.

- 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.

Let's formulate the mathematical model that will inform our analysis.

Let's establish some symbology standards

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

What are these matrics and what is in them?

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

α matrics are of size $\alpha_{t \times n_c}$; β matrics are of size $\beta_{t \times n_t}$.

What are these matrics and what is in them?

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

α matrics are of size $\alpha_{t \times n_c}$; β matrics are of size $\beta_{t \times n_t}$.

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

What are these matrics and what is in them?

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

α matrics are of size $\alpha_{t \times n_c}$; β matrics are of size $\beta_{t \times n_t}$.

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

n_c is the number of households in the control group; n_t is the number of households in the treatment group.

Our goal is to find a good estimate for the counterfactual consumption.

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

Our goal is to find a good estimate for the counterfactual consumption.

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

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

Our goal is to find a good estimate for the counterfactual consumption.

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

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

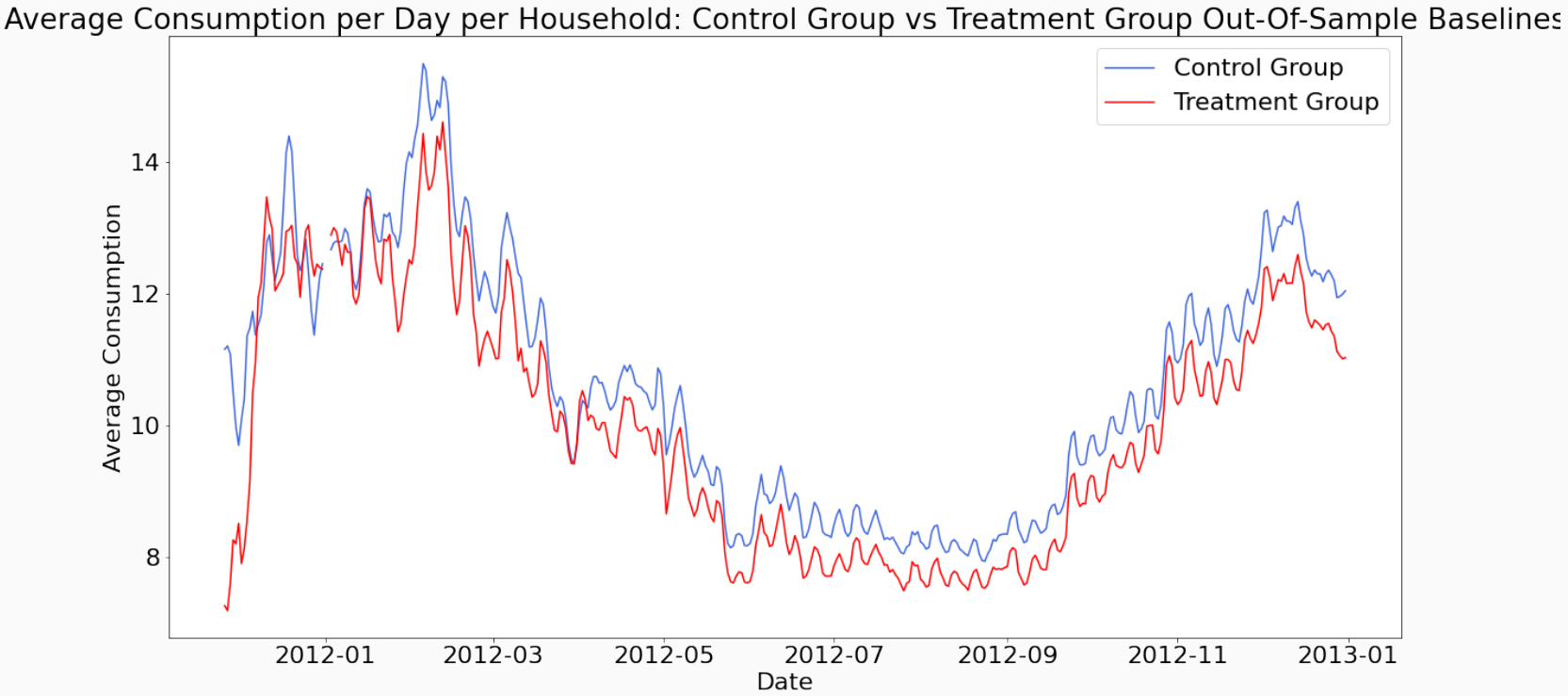
Goal is to **find a good estimate** for $\hat{\beta}_{2013}$.

Naive Model: Use Treatment Group in 2013 as Counterfactual

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



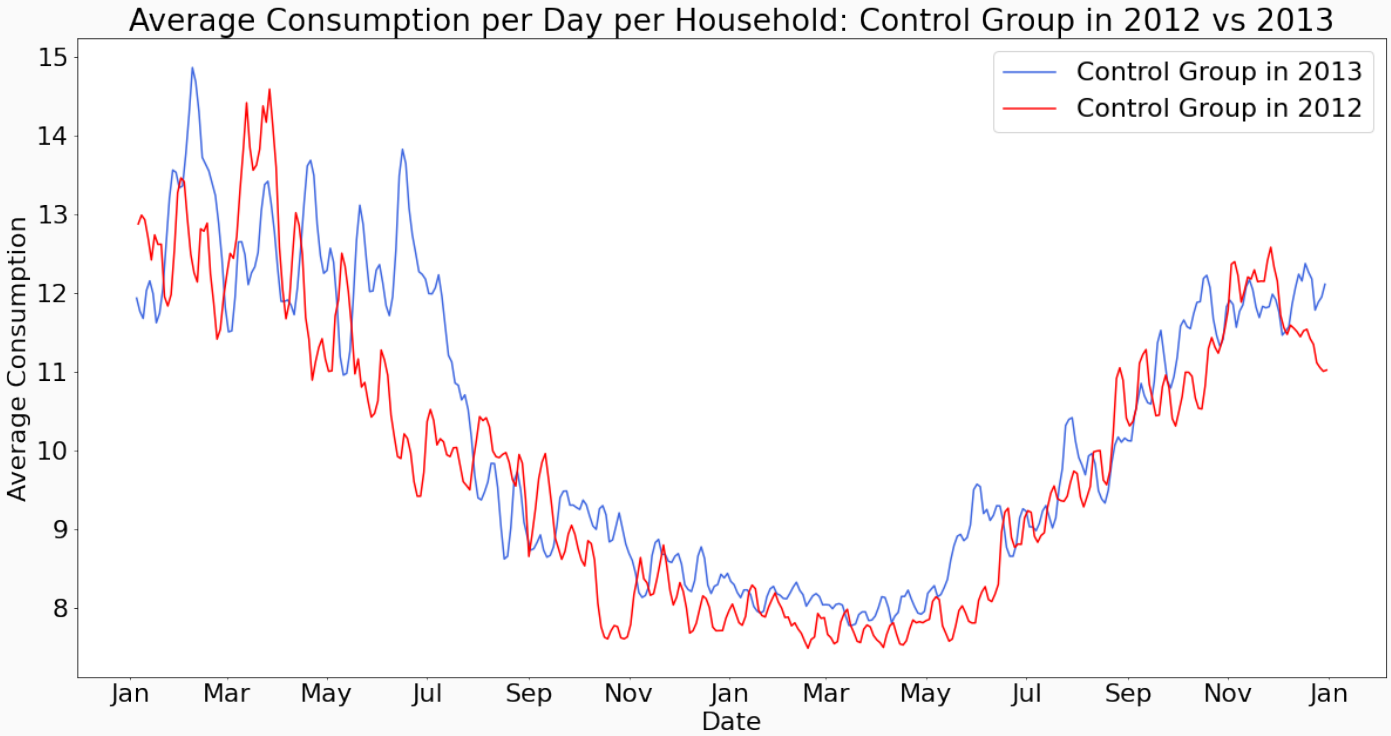
$\Longrightarrow$ the two groups don't have the same baseline electricity consumption patterns.

Sophisticated Naive Model

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

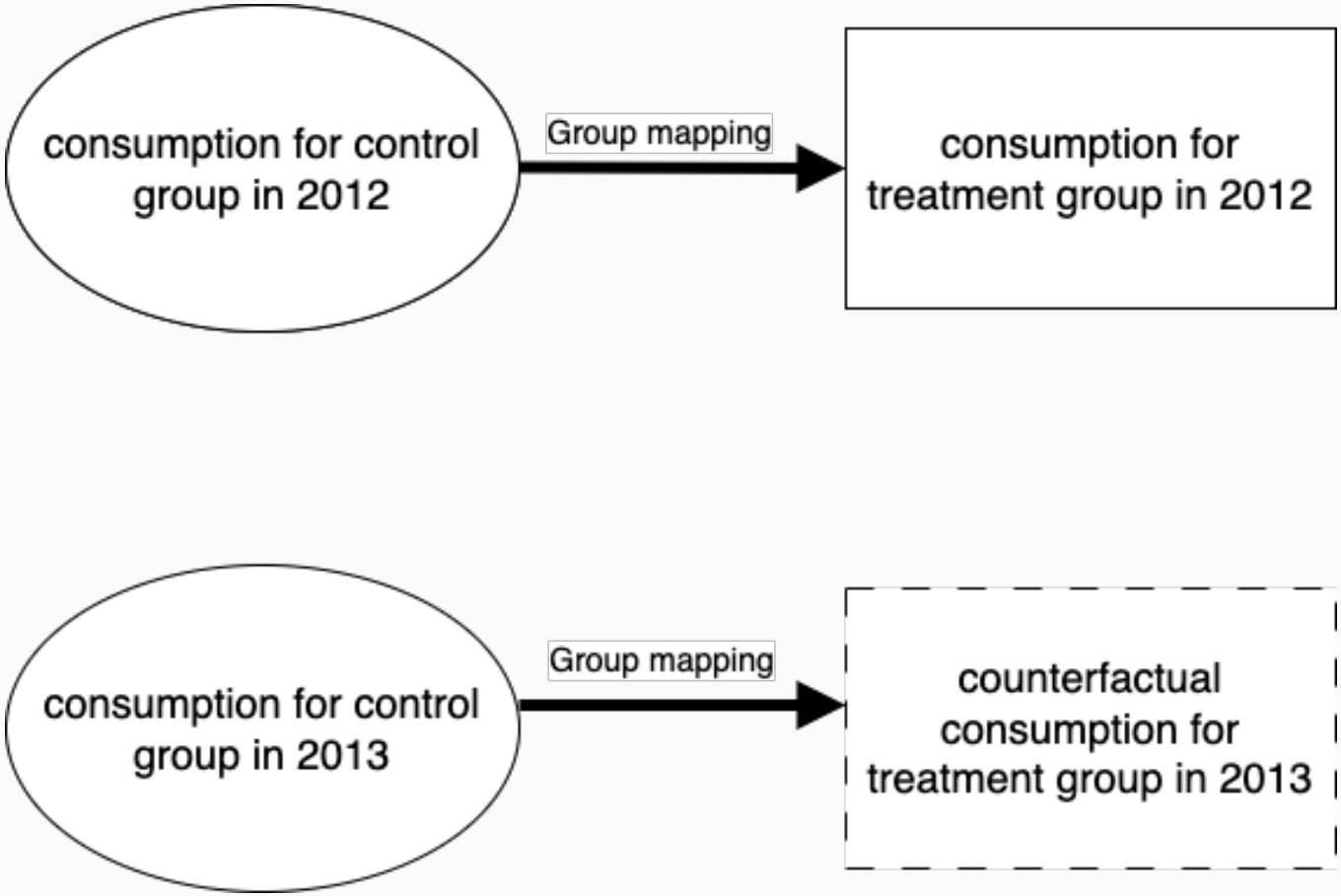
$$\begin{aligned} \hat{\beta}_{2013} &\approx \beta_{2012} \\ T &= \beta_{2013} - \hat{\beta}_{2013} \\ &= \beta_{2013} - \beta_{2012} \end{aligned} \tag{2}$$

Changes from Time: 2012 vs 2013

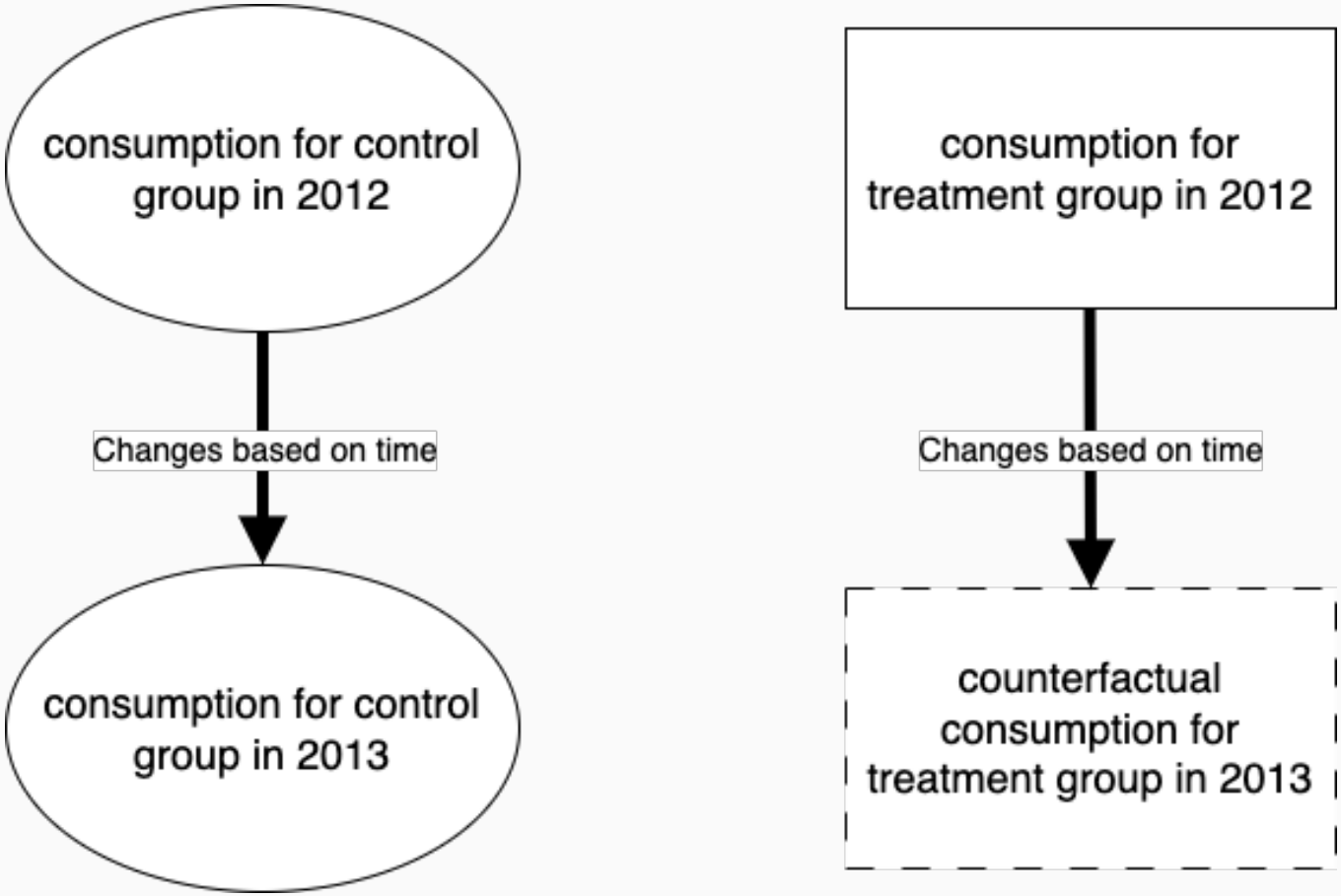


$\Longrightarrow$ Things changed between the two years even for the control group.

Treatment vs Control Group Baselines



Treatment vs Control Group Baselines

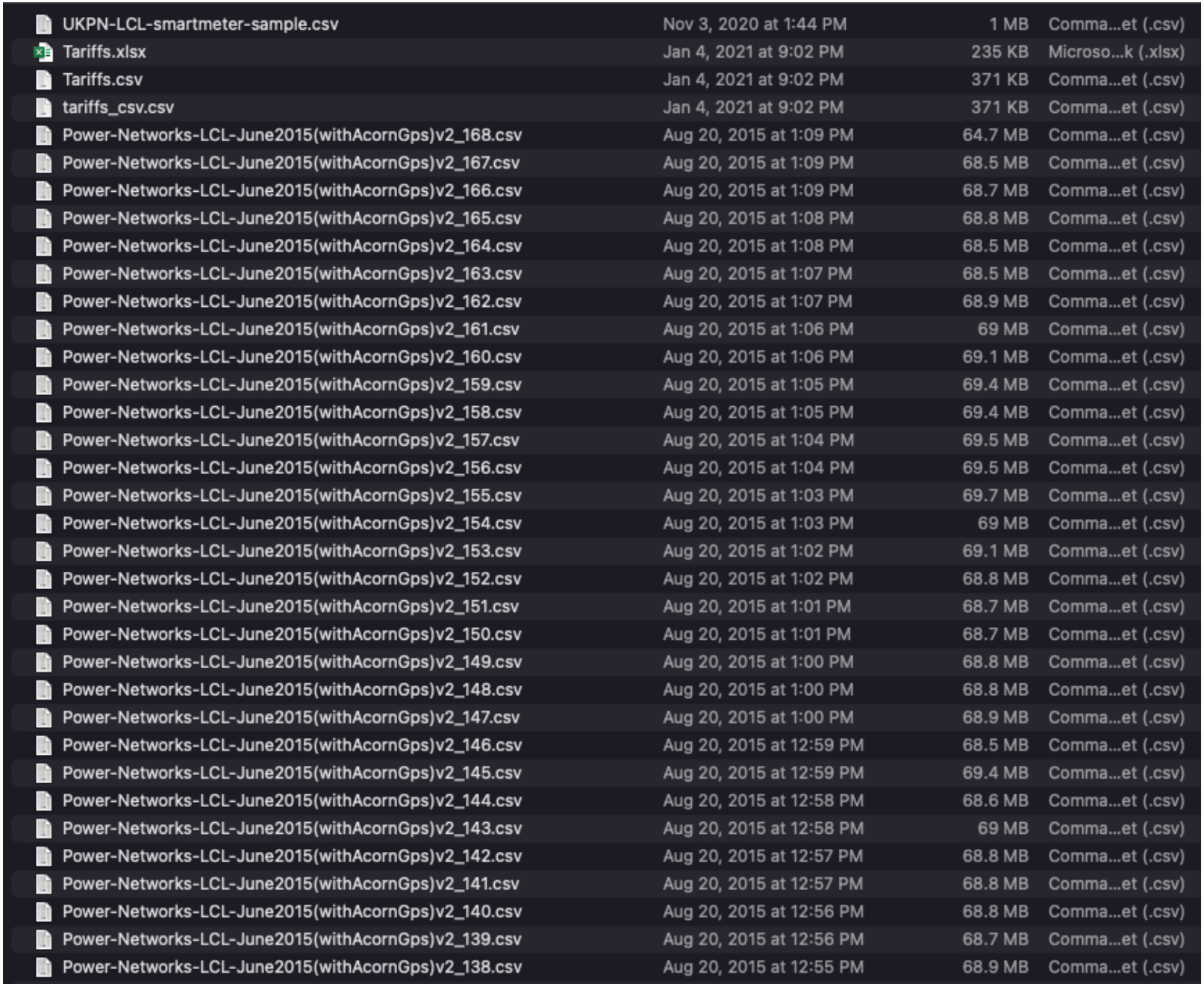


Mathematical Model: Using Group Mapping

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

Now we have a model, let's use run the analysis based on the model

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).

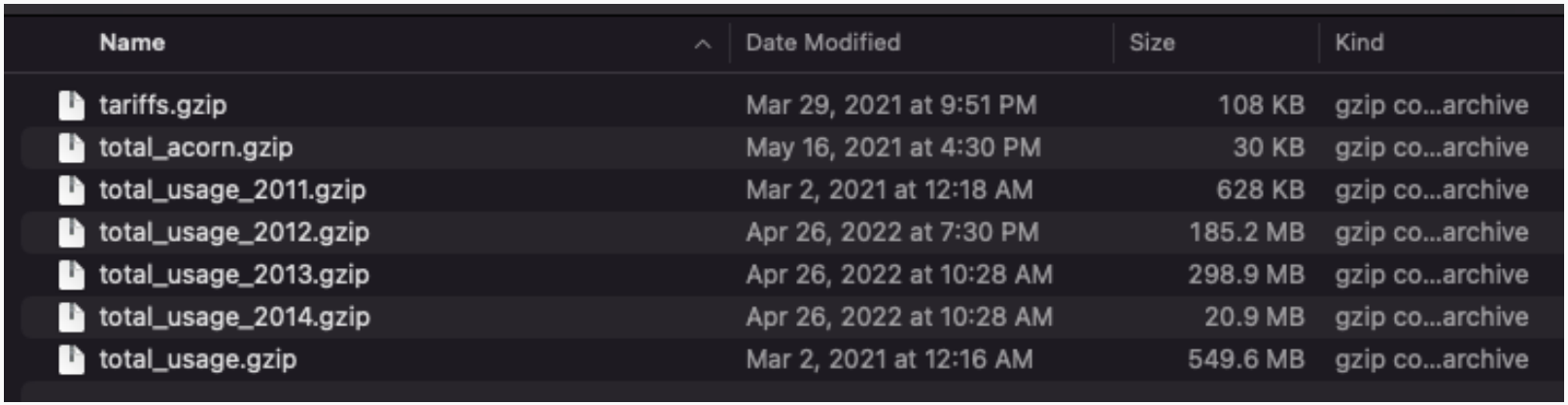
Data Prep, Cleaning, Processing

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).

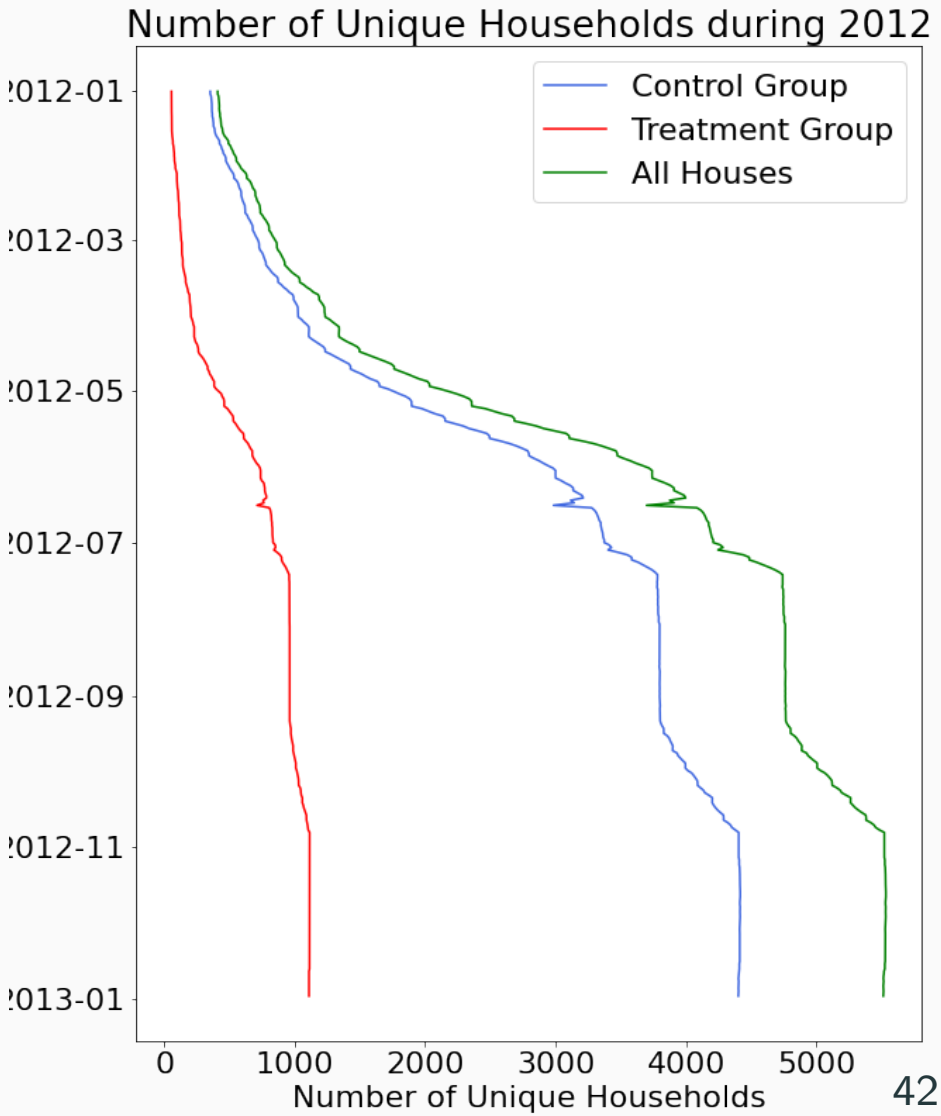
Data Prep, Cleaning, Processing



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

Our matrics are half full for 2012

$$\alpha_{2012} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n_c} \\ a_{21} & a_{22} & \cdots & a_{2n_c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{t1} & a_{t2} & \cdots & a_{tn_c} \end{bmatrix}$$



We have some paths forward

- impute values

We have some paths forward

- impute values
  - pro: you are keeping all of your data and you can capture the full trend.

We have some paths forward

- impute values
  - pro: you are keeping all of your data and you can capture the full trend.
  - how do you impute values?

We have some paths forward

- impute values
  - pro: you are keeping all of your data and you can capture the full trend.
  - how do you impute values?
  - are you introducing error?

We have some paths forward

- limit the timeframe to, for example, latter half of 2012

We have some paths forward

- limit the timeframe to, for example, latter half of 2012
    - con: can't capture annual seasonality

We have some paths forward

- use a fixed panel: only keep the houses that have values for all of 2012

We have some paths forward

- use a fixed panel: only keep the houses that have values for all of 2012
    - con: you are making your sample size small

We have some paths forward

- reduce dimenstionality

We have some paths forward

- reduce dimenstionality
  - go from matrix of $t \times n$ to vector of size t.

We have some paths forward

- reduce dimenstionality
  - go from matrix of $t \times n$ to vector of size t.
  - take the mean over all houses that you have at any given t to remove this issue.

We have some paths forward

- reduce dimenstionality
  - go from matrix of $t \times n$ to vector of size t.
  - take the mean over all houses that you have at any given t to remove this issue.
  - con: you are reducing dimenstionality and removing valuable datapoints that might teach you something

The path forward depends on your usecase.

There is no wrong answer here.

The path forward depends on your usecase.

There is no wrong answer here.

The choice of methodology depends on the assumptions you are comfortable making.

The path forward depends on your usecase.

There is no wrong answer here.

The choice of methodology depends on the assumptions you are comfortable making.

It's important to make informed decisions grounded in the type of analysis and with full knowledge of your data and its shortfalls.

Matrices $\rightarrow$ Vectors

$$\alpha_y = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n_c} \\ a_{21} & a_{22} & \cdots & a_{2n_c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{t1} & a_{t2} & \cdots & a_{tn_c} \end{bmatrix} \longrightarrow \overline{\alpha_y} = \begin{bmatrix} \overline{a_1} \\ \overline{a_2} \\ \vdots \\ \overline{a_t} \end{bmatrix}$$

where $\overline{a_i} = \frac{\Sigma a_{i,m}}{n}$ where $a_{i,m}$ has a value and n is the total number of elements with a value.

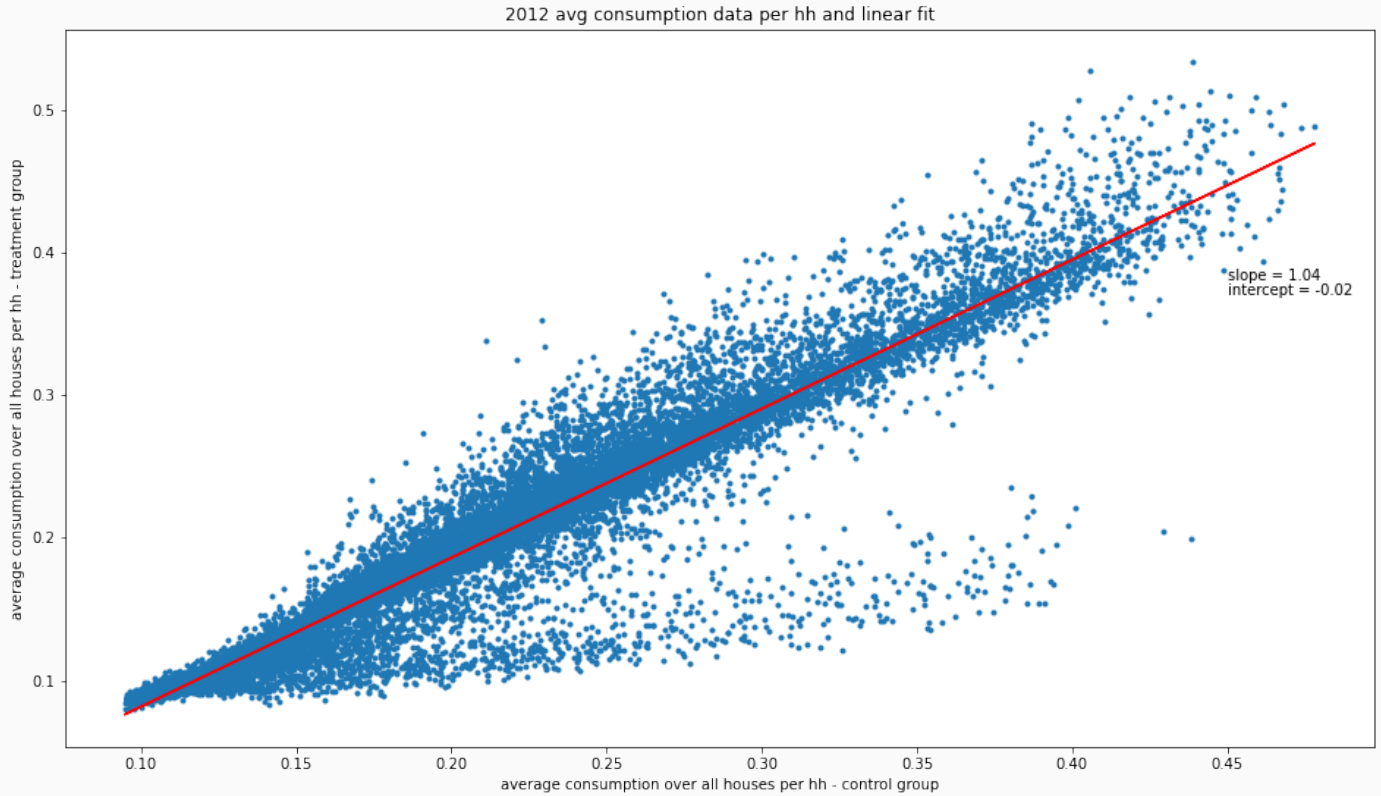
Reduce dimenstionality

$$\overline{\beta_{2012}} = a \times \overline{\alpha_{2012}} + b$$

$$\overline{\hat{\beta}_{2013}} = a \times \overline{\alpha_{2013}} + b$$

$$\overline{\Delta \text{treatment}} = \overline{\beta_{2013}} - \overline{\hat{\beta}_{2013}}$$

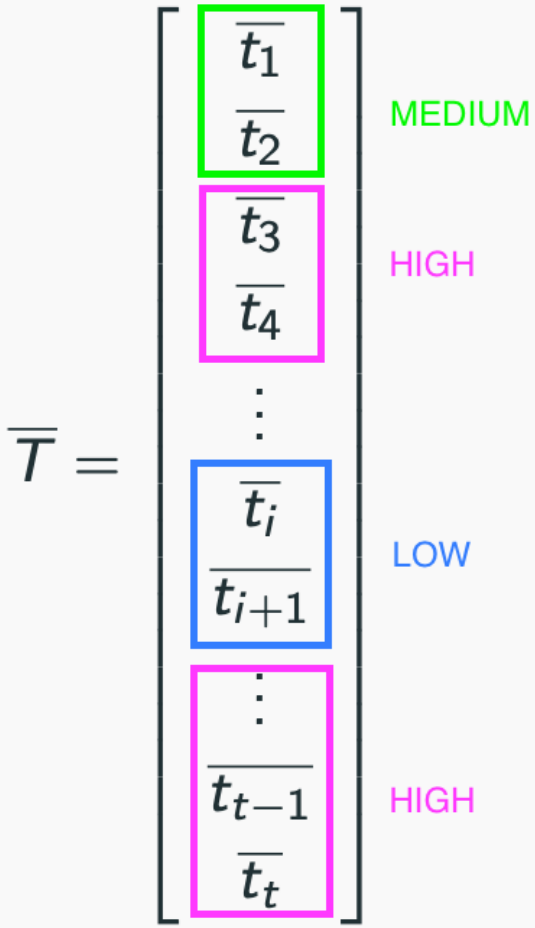
Learned linear relationship from 2012



Interpret the estimated treatment effect

$$\overline{T} = \begin{bmatrix} \overline{t_1} \\ \overline{t_2} \\ \overline{t_3} \\ \overline{t_4} \\ \vdots \\ \overline{t_i} \\ \overline{t_{i+1}} \\ \vdots \\ \overline{t_{t-1}} \\ \overline{t_t} \end{bmatrix}$$

Interpret the estimated treatment effect



Results

Mean Error Between Counterfactual and Real Consumption for 2013 (regression on socio-economic groups)				
	Low	Normal	High	Overall
Affluent	0.0085	0.0005	-0.0149	0.0005
Comfortable	0.0004	-0.0141	-0.0314	-0.0135
Adversity	-0.0022	-0.0129	-0.0281	-0.0126

Mean Percent Error Between Counterfactual and Real Consumption for 2013 (regression on socio-economic groups)				
	Low	Normal	High	Overall
Affluent	2.47%	0.49%	-5.40%	0.41%
Comfortable	-1.84%	-8.39%	-14.76%	-8.06%
Adversity	-2.63%	-7.83%	-14.71%	-7.65%

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?

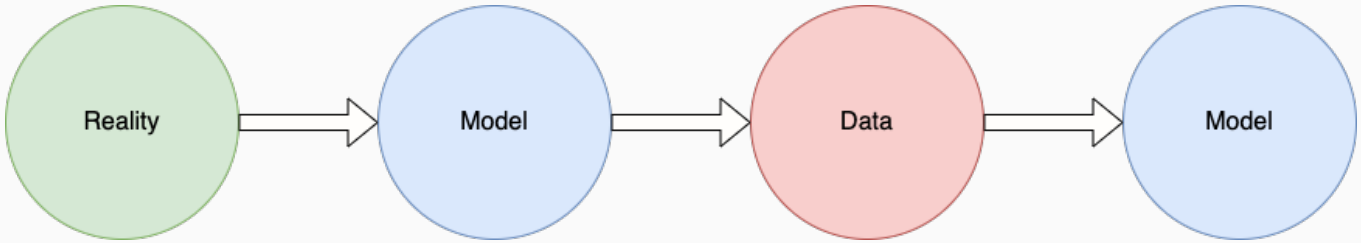
- A little bit about electricity markets.
- Causal analysis.
- How to break down your data to match the mathematical model at hand.
- Questions to ask before analysis:
  - How was the data collected?
  - What biases are present?
  - How does that impact your analysis and results?

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/
- https://dspace.mit.edu/handle/1721.1/144969