Assessing Unconfoundedness in Two Cases*

June 5, 2023

Contents

*This case is prepared solely for Causal Inference and Experimentation at Rotterdam School of Management. Please do NOT circulate.

Part I

Energy Conservation Field Experiments

Background

Following consultation with the industry, the Commission for Energy Regulation (CER) in Ireland established the Smart Metering Project Phase I in late 2007, powered by the installation of smart meters in all of Ireland. From the smart meters, CER has access to granular data of the electricity usage of households. As a matter of fact, CER record electricity of households at half-hour intervals for each household with a smart meter installed. Around March 2008, CER began to prepare a Customer Behavior Trial to ascertain the potential for smart metering technology, when combined with time of use tariffs and different DSM (Demand-Side Management) stimuli, to effect measurable change in consumer behavior in terms of reductions in peak demand and overall electricity use.

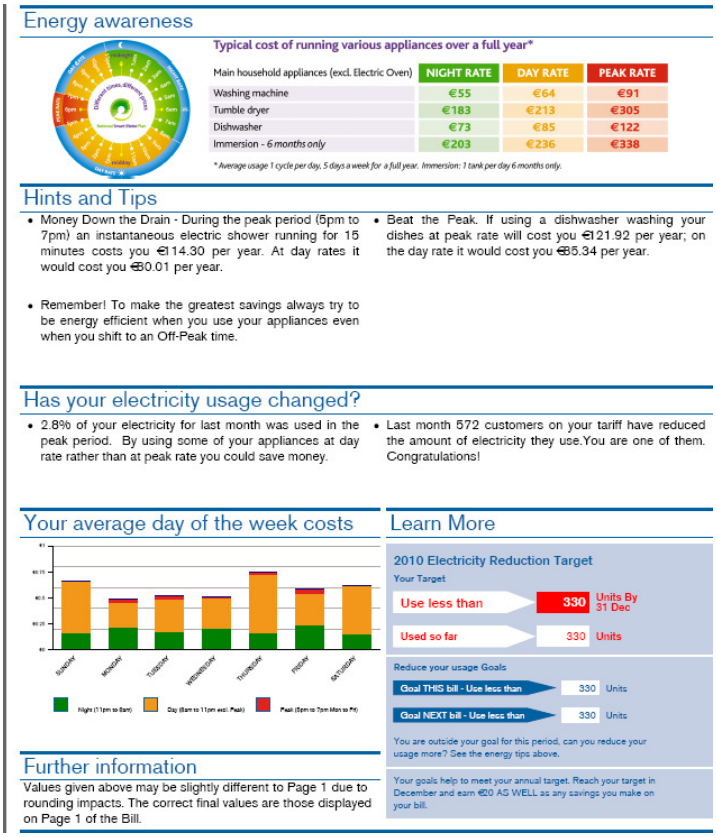
The field experiment design

The experiment focused on the time-of-use tariffs and DSM stimuli and possible interactions between them. Time of use tariffs and demand side management stimuli were specifically developed for use in the Customer Behavior Trial. These may be summarized as follows:

- Four specific time of use tariffs A, B, C and D offering different unit prices for the night time, day time and peak times.
- Specific DSM initiatives, which included: bi-monthly electricity bill with detailed energy statement, monthly electricity bill with detailed energy statement, electricity in-home display, and Overall Load Reduction (OLR) incentive.

The different tariffs are summarized in the table below. Two time-of-use tariffs were designed with two features:

1. Different tariffs were neutral in comparison with the standard regulated tariff to ensure that the "average" participant who did not alter their electricity consumption pattern was not penalized financially.



Energy Usage Statement Example

2. The relative differences of the low vs. normal vs. peak rate differ between tariffs. For example, in Tariff D, the peak vs. low rate are 38.00 vs. 9.00, but those for Tariff A 20.00 vs. 12.00. In expectation, a larger relative difference would reduce the peak usage of households.

Domestic Time of Use Tariffs				
Vs. Normal Rate = 14.1 € cents/ kWh		**Week Night** 23.00 – 8.00	**Week Day** 8.00 – 17.00 19.00 – 23.00	**Peak** 17.00 – 19.00 (Monday to Friday), ex. holidays
Tariff A	Cents per kWh	12.00	14.00	20.00
Tariff B	Cents per kWh	11.00	13.50	26.00
Tariff C	Cents per kWh	10.00	13.00	32.00
Tariff D	Cents per kWh	9.00	12.50	38.00

The DSM stimuli focused on two main features: 1) the frequency of information; and 2) different incentives. The design varied how frequent an energy usage statement was sent to consumers (monthly vs. bi-monthly). The design also included a condition with consumers having constant access to their energy usage via in-home display. Finally, an alternative monetary incentive was included to compare with sole information exposure.

In-home Display of Energy Usage

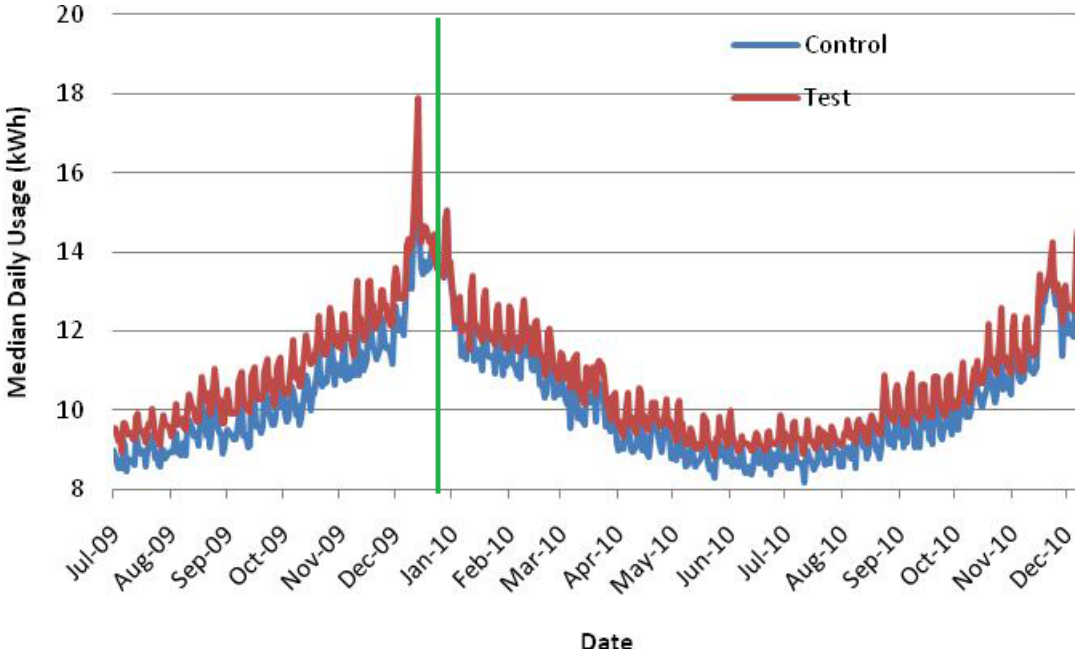
The recruitment and assignment of consumers

The main challenge of the experiment was the recruitment of consumers into treatment conditions. The CER must seek for the consent of consumers. In another word, consumer opted in the treatments. Out of 5,028 consumers contacted by CER, around 3,858 agreed to receive the treatments. And CER assigned the remaining 1,170 consumers to the control group. For the 3,858 consumers in the treatments, CER put them into "strata", with a variable synthesized from demographics and electricity usage, and then used a block randomization with these strata. In particular, a factor analysis was performed on a set of variables and factor loadings were used to calculate an index for each consumer, which the stratification was based on.

Composition of Eigen vectors	Eigen Vectors				
Factors	V1	V2	V3	V4	V5
Peak band	-0.178	0.577	-0.065	0.085	-0.087
Night band	0.103	-0.526	0.101	0.142	0.153
Overall Weekly Usage Variance	-0.039	0.1	0.133	0.164	0.485
Overall Peak Usage Variance	-0.152	0.53	0.01	0.193	0.207
Number in house	-0.395	-0.003	-0.11	-0.288	-0.276
# of bedrooms	-0.312	-0.119	-0.078	-0.247	-0.197
Internet access	-0.414	-0.119	0.006	0.095	-0.078
Income band	-0.463	-0.12	0.043	0.039	0.094
Education classification	-0.361	-0.178	0.071	0.196	0.288
Employment status	-0.401	-0.087	0.036	0.1	0.125
Wet	0.013	0.006	0.028	0.582	-0.304
Electronics	-0.004	-0.092	0.044	0.568	-0.398
Energy reduction engagement band	0.022	-0.047	-0.646	0.08	0.103

Factor Loadings of Different Variables

After the recruitment and assignment procedure, the distribution of consumers across different conditions are shown in the table below.



Median Daily Usage (kWh) of the Test and Control Group

Tariff	Bi-monthly bill and energy usage statement	Monthly Bill, and energy usage statement	Bi-monthly bill, energy usage statement and electricity Monitor	Bi-monthly bill, energy usage statement plus Overall Load Reduction	Total
Tariff A	342	342	342	342	1,368
Tariff B	127	129	127	128	511
Tariff C	342	342	343	343	1,370
Tariff D	127	129	126	127	509
Weekend					100
Control Group					1,170
	938	**942**	**938**	**940**	**5,028**

Challenges in data analysis

Overall, the final data include pre- and post-treatment period electricity usage (at half-hour intervals) for all consumers recruited into the experiment. The pre-treatment period lasts from July 2009 to Dec 2009, and the post-treatment period from Jan 2010 to Dec 2010. An overview of the median daily usage (in kWh) is in the figure below.

Give the data structure, analysts at CER decided to adopt a diff-in-diff design, which exploited the before-after and treated-control data structure to remove concerns about any confounders that 1) are stable over time but varying across consumers (i.e., individual fixed effects) and 2)

are common to all consumers but varying across time (i.e., time fixed effects). The credibility of the DID analysis rests on the parallel assumption, or the unconfoundedness of treatment assignment after controlling for individual fixed effects. Yet, this assumption is untestable. **Using the data and experiment design, try to construct tests for the implications of the unconfounded assumption.**

Part II

A Quasi-experiment of Online Reviews

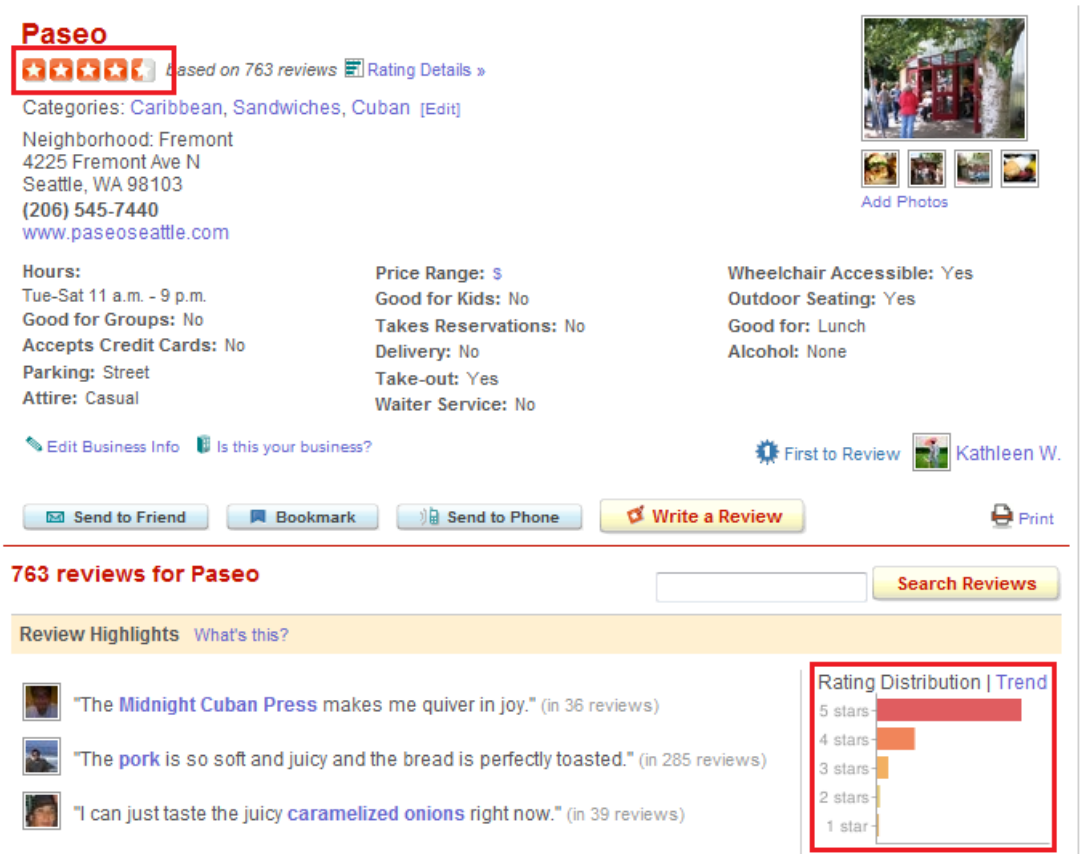
Background

Technological advances over the past decade have led to the proliferation of consumer review websites such as Yelp.com, where consumers can share experiences about product quality. These reviews provide consumers with information about experience goods, which have quality that is observed only after consumption. With the click of a button, one can now acquire information from countless other consumers about products ranging from restaurants to movies to physicians. But do online consumer reviews affect market outcomes?

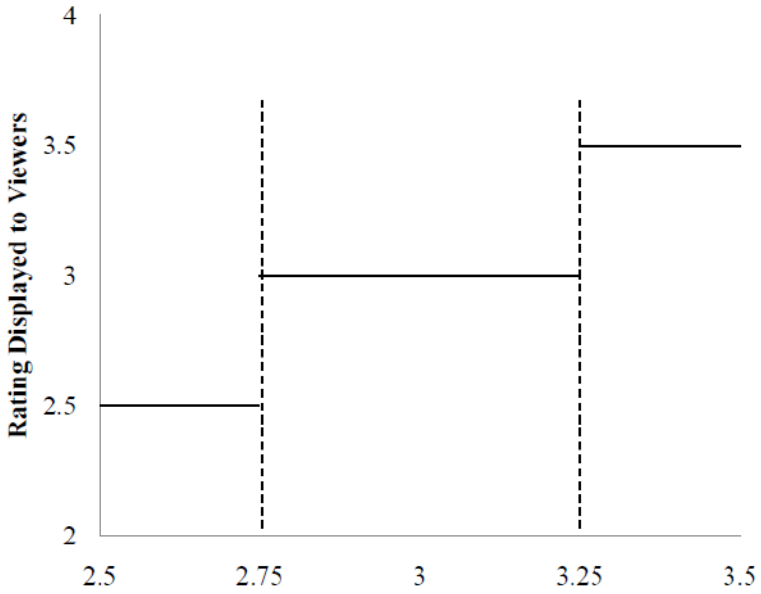
However, in non-experimental studies, identifying the causal effects of online reviews is a challenging task because of the potential endogeneity problem; the review rating is often correlated with unobserved heterogeneity that affects consumers' responses. For example, unobserved marketing expenditure is likely correlated with both the review ratings and consumers' social media endorsement.

Study design

To this end, data scientists from Yelp.com, using Yelp data, designed a quasi-experiment to study the effect of Yelp reviews on restaurant revenues. They worked with the state department of revenue and gathered revenues for all restaurants in a city from 2003 through 2009. To support the claim that Yelp reviews had a causal impact on revenue, the data scientists exploited the institutional features of Yelp to isolate variation in a restaurant's rating that was exogenous with respect to unobserved determinants of revenue. That is, in addition to specific reviews, **Yelp presents the average rating for each restaurant, rounded to the nearest half-star**. The following figure shows the review of a restaurant shown in the Yelp page at 4.5, but the actual average was different.

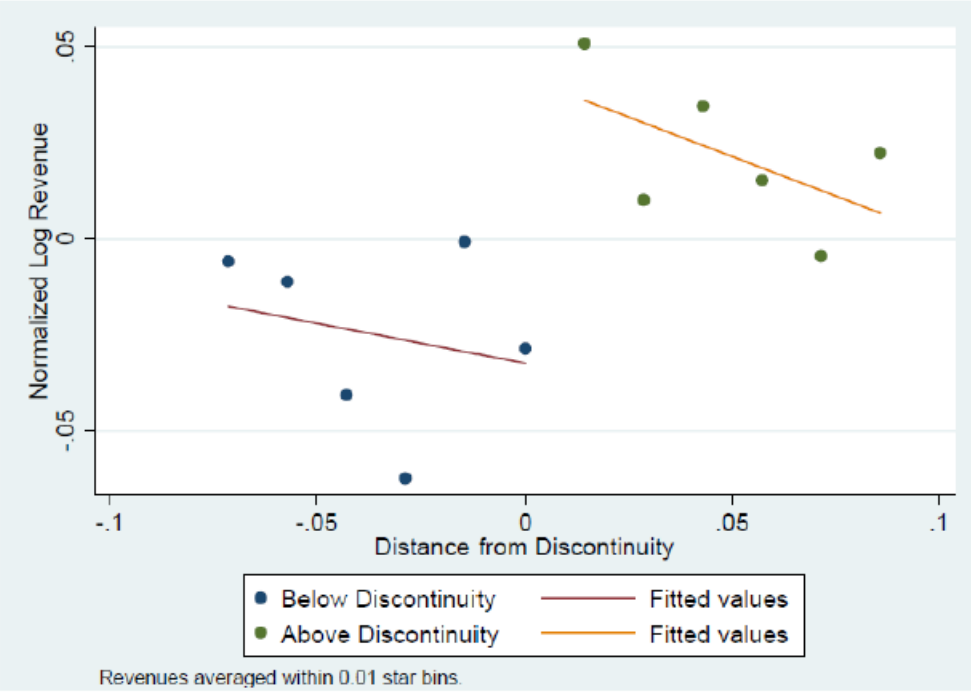


To further under stand how Yelp rounds the average ratings, see the following figure. In particular, Yelp prominently displays a restaurant's rounded average rating. Each time a restaurant's rating crosses a rounding threshold, the restaurant experiences a discontinuous increase in the displayed average rating.



Yelp Displays Each Restaurant's Rounded Average Rating

Exploiting the discontinuities from rounding the review ratings, the data scientists adopted a



Average Revenue around Discontinuous Changes in Rating

quasi-experiment design of regression discontinuity approach. They looked for discontinuous jumps in revenue that follow discontinuous changes in ratings. The following figure presents some initial evidence of the effect of ratings on revenue, with data on actual average ratings (i.e., the running variable), the rounded ratings (i.e., the treatment), and the revenue (i.e., the outcome).

Y-axis is the normalized log revenues. The normalized revenue are then averaged within bins based on how far the restaurant's rating is from a rounding threshold. The graph plots average log revenue as a function of how far the rating is from a rounding threshold. All points with a positive (negative) distance from a discontinuity are rounded up (down).

Evaluating the validity of the design

One challenge for identification in a regression discontinuity design is that any threshold that is seen by the investigators might also be known to the decision-makers of interest. This can cause concerns about gaming, as discussed in McCrary (2008). In the Yelp setting, the concern would be that certain types of restaurants submit their own reviews in order to increase their revenue.

How serious do you think the concern? Any idea how to test the implications of restaurant gaming the system? Suppose the data available to the data scientists include ratings, revenue and other restaurant features on Yelp.

References

[1] May 2011, Electricity Smart Metering Customer Behavior Trials (CBT) Findings Report, retrieved at: `https://www.cru.ie/wp-content/uploads/2011/07/cer11080ai.pdf`.

[2] McCrary, Justin, 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," Journal of Econometrics, Vol. 142, No. 2, 698-714.