Non-Parametric Estimation of Heterogeneous Treatment Effects

The Effectiveness of Electricity Demand Reduction Interventions

Bryan Bollinger and Wes Hartmann

Applying Field Experimentation to Behavior Research, Rady School of Management, March 2014
Research Questions

- **What is the incremental reduction in peak-load electricity usage that can be achieved by targeting consumers with the appropriate technology option to shift usage across time?**
  - Programmable Communicating Thermostats (PCTs) work on average, but:
    - They are costly to install
    - They are more/less effective for some people

- **How do we target treatments based on observable covariates when there might be thousands?**
  - Rossi, McCulloch and Allenby (1996) suggests targeting based on past usage
  - Estimating likelihood based models can be costly when we observe consumption every 15 min for 3+ months
    - Especially when the underlying model involves forward-looking dynamics
  - Can we develop a non-parametric estimator and avoid approaches that seek to throw away the data?
Peak-Pricing in Electricity

- Electricity prices are regulated
- There are environmental benefit of reducing usage in periods of peak demand to limit use of polluting technologies brought into use
- Aside from the environment, the firm would also rather move usage away from peak times
  - With current fixed pricing, costs are lower off-peak
  - The extent of variable pricing is limited due to price caps
  - A smoother distribution of usage can delay new plant investments
- The critical challenge involves providing the consumer:
  - Non-flat, real-time pricing
  - Information on real-time pricing
  - Mechanism to reduce switching costs in time of usage
Technology Options

- **Web portal**
  - Consumers can go online to learn about their usage and the current price of electricity if in a variable pricing plan

- **In-home display (IHD)**
  - Can solve the information problem by notifying consumers of current price or of a critical period

- **Programmable communicating thermostat (PCT)**
  - Provides the pricing information
  - Allows users to set thermostat based on time of day and current price

- **Latest technology not in study**
  - Firm learns habits and manages usage for you
  - Motion detector in the home
  - Big brother worries!
Empirical Setting

- We have experimental data from an electric utility.
- Consumers were randomly assigned to one of four technology treatment conditions or the control group.
  - Portal only, portal+IHD, portal+PCT, portal+IHD+PCT
- In addition to the technology treatments, households were also assigned to one of two price treatments.
- The price treatments are:
  - Time-of-use pricing (TOU): 4.2, 23, and 46 cents per kwh for off-peak, peak and critical
  - Variable peak pricing (VAR): four levels ranging from 4.5c to 46, where level for the day depends on demand
We have hourly electricity usage for 4,443 residential households in 2011.  
2,589 households appear in both the 2010 and 2011 data, with 5,491,443 hours of electricity usage data.
We exclude from the analysis households who have the low-income price rate (in the control group) and those without a treatment start data recorded, as well as accounts with multiple meterids.
We exclude 2011 consumption data for treated household before their treatment
The estimation sample includes 2,449 households with with 5,190,689 hours of electricity usage data.
The treatments for the household in both years are:

<table>
<thead>
<tr>
<th>Treatments</th>
<th>CON</th>
<th>TOU</th>
<th>VAR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>419</td>
<td>0</td>
<td>0</td>
<td>419</td>
</tr>
<tr>
<td>Portal</td>
<td>0</td>
<td>337</td>
<td>370</td>
<td>707</td>
</tr>
<tr>
<td>IHD</td>
<td>0</td>
<td>243</td>
<td>260</td>
<td>503</td>
</tr>
<tr>
<td>PCT</td>
<td>0</td>
<td>220</td>
<td>230</td>
<td>450</td>
</tr>
<tr>
<td>All three</td>
<td>0</td>
<td>191</td>
<td>179</td>
<td>370</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>419</td>
<td>991</td>
<td>1,039</td>
<td>2,449</td>
</tr>
</tbody>
</table>
Figure: Average Daily Electricity Consumption by Treatment
Average Hourly Usage by Treatment

Figure: Average Hourly Electricity Consumption by Treatment

Non-Parametric Estimation of Heterogeneous Treatment Effects
Kernal Density of On-Peak and Difference in On- vs. Off-Peak Usage

Figure: Average Hourly Electricity Consumption by Treatment, On-Peak Days
Targeting Heterogeneous Treatments

- Typically we measure average effect for a given population (ATE, ATT)
  - Obtain non-parametric estimate of the effect, conditional on observables
  - Integrate over desired distribution of observables

- Targeting skips the last step and optimizes based on the first
  - Non-parametrically estimate “treatment” effects, conditional on observables
  - Optimize the “treatment” for each set of observables
  - Apply to un-treated individuals based on observables
The Challenge

- Too many covariates
  - Can reduce statistical significance for effects estimated conditional covariates
  - May be more covariates than there are observations

- Manski (2004) considers the maximum number of covariates upon which to condition treatment

- Big data approaches seek relevant set of covariates
  - Lasso places zero effect on irrelevant variables
Treatment Effects Methods

\[ \hat{\alpha}_a = E[Y|A = a] - E[Y|A = 0] \quad (1) \]

- \( Y \) log usage
- \( A \) treatment

To estimate this we can use the following regression:

\[ y_{it} = \alpha \tilde{A}_i + \eta_t + \epsilon_{iat} \quad (2) \]

- \( \eta \) aggregate demand shock
- \( \tilde{A}_i \) treatment dummies
- \( \epsilon_{iat} \) Unobserved individual-level shock

To estimate heterogeneous effects, we could condition on state, \( x \):

\[ \hat{\alpha}_{ax} = E[Y|A = a, X = x] - E[Y|A = 0, X = x] \quad (3) \]
Targeting on Demographics?

Non-Parametric Estimation of Heterogeneous Treatment Effects
Targeting on Demographics?

No AC, Peak, Age

- family (N=45,899)
- old (N=68,505)
- young (N=60,903)
One solution is that we calculate the expected treatment outcome conditional on $i$’s vector of (pre-treatment) usage history, $h_i$, matching individuals on their entire history of usage:

$$\hat{\alpha}_{ai} = E [Y_j | A_j = a, H_j = H_i, X_j = X_i] - E [Y_j | A_j = 0, H_j = H_i, X_j = X_i]$$

In this true non-parametric approach, $X_i$ includes each hourly interval over the summer of 2010

- The expectation over $x_i$ is a simple sum over the hourly data.

We took a shot at this by defining:

$$D_{ij} = \sum_t |H_{jt} - H_{it}|$$

Issues:

- High-dimensional problem
- Not efficient: If two identical individuals roll dice each day to determine their usage, this distance will be quite large even though they are the same
What We Really Want

- Returning to the dice example, what we really want are people where the distribution of usage is the same/similar depending on their state i.e. people with the same policy function.
- A household with preference parameters $\theta$ makes choices $y$ dependent on the current state of the world.
- States observed by the econometrician can be segmented into a “treatment” $a$ faced by the household and all other observed states, $x$. The unobserved state is $\epsilon$.
- Choices are guided by a policy function $y = \sigma(x, \epsilon, a, \theta)$. One way to characterize this dynamic problem is as follows:

$$V(x_t, \epsilon_t | a, \theta) = \max_{\sigma(x, \epsilon, a, \theta)} \sum_{\tau = t}^{\infty} \beta^{t-\tau} u(y_t, x_t, e_t | a, \theta)$$

(4)

- We want a measure of the distances between individuals’ distributions of usage as a function of their state.
Our Approach

Formally, we consider:

\[ \hat{\alpha}_a = E \left[ Y_j | A_j = a, F_{j,t,x} (h|x) = F_{i,T_i,x} (h|x) \right] \]

\[ -E \left[ Y_j | A_j = 0, F_{j,t,x} (h|x) = F_{i,T_i,x} (h|x) \right] \]

where the expectation is taken across individuals indexed by \( j \neq i \) and \( F_{j,t,x} (h|x) \) is the empirical distribution function for \( T_{j,x} \) iid observations \( H_{j,t} \) occurring when \( j \) is in state \( x \).

- We want a distance in the distribution of the \( h|x \) that asymptotically approaches zero as our sample size increases
  - Kolmogorov-Smirnov statistic measures distances in distributions (0 to 1)
    - Note that this also allows us to utilize unbalanced pre-treatment panels, \( N_i \neq N_j \)
  - If we get this, we would asymptotically put all weight on only those individuals with the same policy function
    - Given the same policy function, a structural approach on the pre-treatment data would infer these people to have the same \( \theta \)
    - Recent work (e.g. Kasahara and Shimotsu) show that the number of unobserved types recoverable is related to the length of the panel.
Matching Method

- Weight each household with $A_j = a$ based on its distance from the focal household, $D_{ij} = D(X_j, x_i)$.

\[
D_{ij,X} = \sup_h |F_{j,T_{j,X}}(h|x) - F_{i,T_{i,X}}(h|x)|
\]  

(6)

- Derive the unconditional distance between $i$ and $j$ by taking a weighted average over state points:

\[
D_{ij} = \frac{\sum_x p_{i,X} D_{ij,X}}{\sum_x p_{i,X}}
\]  

(7)

where $p_{i,X}$ is the fraction of time $i$ is in state $x$. 

Non-Parametric Estimation of Heterogeneous Treatment Effects
Distribution of Distances

Non-Parametric Estimation of Heterogeneous Treatment Effects
We estimate the treatment effects using weighted OLS, where $K$ is the Epanechnikov kernel with optimal bandwidth $h_i$, and $A_{i-}$, $D_{ii-}$, and $Y_{i-}$ are the treatment dummy matrix, distance matrix to household $i$, and usage for all other households (indicated by $i-$)

$$\hat{\alpha}_{ai} = A' \left( A \ast \frac{1}{h_i} K \left( \frac{D_{ii-}}{h_i} \right) \right)^{-1} A' \left( Y_{i-} \ast \frac{1}{h_i} K \left( \frac{D_{ii-}}{h_i} \right) \right)$$
Non-Parametric Regression Results, Portal

Figure: Effect of Portal Treatment

Non-Parametric Estimation of Heterogeneous Treatment Effects
Non-Parametric Regression Results, IHD

Figure: Effect of IHD Treatment
Figure: Effect of PCT Treatment

Non-Parametric Regression Results, PCT

Non-Parametric Estimation of Heterogeneous Treatment Effects
Non-Parametric Regression Results, All Three

Figure: Effect of All Three treatment
Figure: CDFs Peak vs. Off-Peak, Portal and PCT, TOU
Figure: Scatterplot Peak vs. Off-Peak Coef., Portal and PCT, TOU
Effect by Age, PCT, TOU

Figure: Effect by Age, PCT, TOU
Implications

- Our methodology allows for the targeting of households based on individual-level treatment effects using electricity consumption data.
- Calculate the return to the firm for each technology: \( \pi (\alpha_{ai}) \):

\[
\pi (\alpha_{ai}) = (p_d - c - r) \alpha_{ai} - F_a
\]

- \( r \) is PV of a 1 kw reduction in peak usage for a household (\( \approx \) $700).
  - In our case based on delayed capacity investments.
- Assume marginal cost, \( c \), is zero because \( r \) has been calculated to account for all peak usage (actual and opportunity) costs.
- Assume regulatory pricing distortions prevent setting \( p_d \) to the PV-maximizing peak-price they could have obtained under treatment \( w \) if they had sold the kw of electricity.
- The utility chooses individual-specific profit maximizing technologies from the following set: \( \{ \pi (\alpha_{ai}) \} \), \( \forall a \).
In Conclusion

- We can explain the heterogeneity in treatment effects using past consumption data, specifically by matching individuals based on their pre-treatment distribution of usage as a function of state.
- With the exception of AC, demographic have little explanatory power in the treatment effect sizes.
- Targeting can be much more effective using individual-level treatment effects.
- This approach can be used in any setting with detailed, pre-treatment usage data (social media, frequent shopper purchase behavior, etc.)

“A substantial body of empirical evidence demonstrates that econometric models fit on individual-level data manifest heterogeneity in treatment effects that is present even after conditioning on observables.”
- James Heckman and Ed Vytlacil, 2001