Quasi Experimental Dating: You Matched Me with Who? Opportunities and Limitations of Matching Techniques

2019 IEPEC – Denver, CO

Jonathan Hoechst, Principal Consultant
Ethan Young, Principal Consultant
The Need for Accurate Matching Methods

• Control group needed to provide counterfactual argument
  ▪ What would have happened in lieu of the program’s treatment?

• Ideally, control group/members will mirror the treatment group/members in all observable and non-observable characteristics
  ▪ Allows evaluators to isolate effect of program

• Problem: evaluators often have only pre-intervention energy consumption data to use for matching control groups

Identical in all ways but one...
Does This Matter?

• Yes!
  ▪ Well maybe. Stay tuned.

• When evaluators have additional parameters to use for matching, the matched control group is more likely to accurately reflect the treatment group

• There are likely characteristics and demographics that are correlated with energy usage that are not directly captured by pre-intervention energy use

• Incorporating more of these characteristics should produce more accurate savings estimates
  ▪ Read: not necessarily higher savings estimates
Behavioral Program Analysis

• Client: utility in the Southern United States with 900,000 customers
• Program type: behavioral
  ▪ Email, bill inserts, or home energy reports in different waves
• This analysis focuses on customers that began receiving home energy reports via mail in April 2016
  ▪ Treatment group: approximately 40,000 customers
  ▪ Potential control group: approximately 300,000 customers
• Important:
  ▪ Since the program’s inception, utility obtained age and income data for all customers
• Using these data, could a better control group be built, and how would that affect savings?
Population of Control Group: Downwardly Biasing Energy Savings

• Unmatched control group v treatment
• Difference in average daily use statistically significant for each month
• Something systemically different (and not accounted for in energy consumption) between treatment and control group
• Result
  ▪ Control group showed lower average daily consumption during every month
Updated Control Groups: Adding Age & Income into Matching

- Analysis created four separate specifications for matching:
  - (1) pre-intervention energy use
  - (2) pre-intervention energy use and age of head of household
  - (3) pre-intervention energy use and income of head of household
  - (4) pre-intervention energy use, age, and income of head of household
- For each of the matched control groups produced from 1-4
  - Compare the difference in pre-intervention energy use, by month
    - Ideally this is as small as possible
  - Estimate energy savings using lagged dependent variable model
- How do savings estimates differ when the matched control group more accurately reflects the treatment group?
## Matching Results (Average Daily kWh)*

<table>
<thead>
<tr>
<th>Month</th>
<th>Spec 1: avg. daily energy use</th>
<th>Spec. 2: avg. daily energy use &amp; age</th>
<th>Spec. 3: avg. daily energy use &amp; income</th>
<th>Spec. 4: avg. daily energy use, age &amp; income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Treat</td>
<td>Control</td>
<td>Treat</td>
</tr>
<tr>
<td>June 2015</td>
<td>41.87</td>
<td>41.92</td>
<td>41.69</td>
<td>41.92</td>
</tr>
<tr>
<td>July 2015</td>
<td>59.21</td>
<td>59.48</td>
<td>59.07</td>
<td>59.48</td>
</tr>
<tr>
<td>Aug 2015</td>
<td>67.16</td>
<td>67.86</td>
<td>67.40</td>
<td>67.86</td>
</tr>
<tr>
<td>Sept 2015</td>
<td>53.39</td>
<td>54.01</td>
<td>53.62</td>
<td>54.01</td>
</tr>
<tr>
<td>Oct 2015</td>
<td>40.65</td>
<td>41.11</td>
<td>40.86</td>
<td>41.11</td>
</tr>
<tr>
<td>Nov 2015</td>
<td>31.08</td>
<td>30.98</td>
<td>30.90</td>
<td>30.98</td>
</tr>
<tr>
<td>Dec 2015</td>
<td>38.48</td>
<td>38.24</td>
<td>38.14</td>
<td>38.24</td>
</tr>
<tr>
<td>Jan 2016</td>
<td>47.39</td>
<td>47.00</td>
<td>46.71</td>
<td>47.00</td>
</tr>
<tr>
<td>Feb 2016</td>
<td>51.38</td>
<td>51.26</td>
<td>51.03</td>
<td>51.26</td>
</tr>
<tr>
<td>Mar 2016</td>
<td>39.35</td>
<td>39.34</td>
<td>39.21</td>
<td>39.34</td>
</tr>
</tbody>
</table>

*Cells presented in bold italic typeface denote significant differences between treatment and matched control at 95%.
Improved Control Group: Energy Usage & Age

- Adding age into propensity score matching method created accurate control group
- Matching Result:
  - No month has statistically significant difference in daily energy usage
  - Age differences between treatment and control no longer statistically significant
  - Income difference remains statistically significant between groups
Findings: Matching Specifications Effect on Savings

- The matching specification incorporating energy use and age produced highest savings.
- But only marginally more than the model using only energy use.
- Models using income (Specs. 3 & 4) produced lower savings than energy-only model (Spec. 1)
  - Likely due to collinearity of energy usage and income.

<table>
<thead>
<tr>
<th>Data used</th>
<th>Trtmnt effect (kWh)</th>
<th>Std. error</th>
<th>t-statistic</th>
<th>P-value</th>
<th>R²</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmatched</td>
<td>-0.207</td>
<td>0.028</td>
<td>-7.320</td>
<td>&lt; 0.001</td>
<td>0.948</td>
<td>0.948</td>
</tr>
<tr>
<td>Spec. 1</td>
<td>-0.306</td>
<td>0.035</td>
<td>-8.797</td>
<td>&lt; 0.001</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>Spec. 2</td>
<td>-0.308</td>
<td>0.035</td>
<td>-8.710</td>
<td>&lt; 0.001</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>Spec. 3</td>
<td>-0.248</td>
<td>0.035</td>
<td>-7.100</td>
<td>&lt; 0.001</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>Spec. 4</td>
<td>-0.243</td>
<td>0.035</td>
<td>-6.911</td>
<td>&lt; 0.001</td>
<td>0.949</td>
<td>0.949</td>
</tr>
</tbody>
</table>
Conclusions, Next Steps, and Continued Research

- Inclusion of demographic data for matching can produce improved control groups, but effect on savings is varied

- Expand demographic data by geography

- Conduct analysis with different program intervention

- Employ other matching techniques

- Ultimately, evaluators, implementers, and utilities must weigh high cost of obtaining data (both capital costs and cybersecurity risk) against gains in representativeness (and resulting affect on savings estimates)