Overview of the Matched Control Group Load Impact Estimation Methodology

Mike Clark
Christensen Associates Energy Consulting

DRMEC Spring Workshop

May 8, 2017

Presentation Outline

1. What is a matched control group?
2. Why would you use it?
3. How do you do it?
4. What do you do after the matching is done?
5. Are there any potential problems with the method?
6. Summary

1. What is a matched control group?

- It is a group of customers that is as similar as possible to the treatment customers being studied, but who are not subjected to the treatment (e.g., PG&E's Residential SmartRate program)
- The control group is **not** developed from an experimental design (e.g., a random draw of eligible customers where some are assigned the treatment and others are assigned to the control group)
- Rather, it is based on a matching process using **observable characteristics** (e.g., load data, location, industry, CARE status), where the most similar eligible control customer is matched to each treatment customer

2. Why would you use a matched control group?

- Matched control groups are useful when:
  - An experimentally designed control group is unavailable (which is typically the case for non-pilot programs)
  - There is a sizeable pool of eligible control customers (i.e., there are a lot of residential customers, but relatively few air products industrial customers)
  - The treatment is not event-based (e.g., TOU pricing)
  - The treatment is event based, but you are interested in potential non-event day treatment effects
  - The treatment is event-based, but most or all of the hottest days are called as events

3. How do you make a matched control group?

- Two commonly used methods for developing a matched control group:
  - **Propensity score matching (PSM)**
  - **Euclidean distance matching**

3. How do you make a matched control group? (2)

- **Propensity score matching**:
  - Discrete choice regression model with one observation for each treatment and eligible control customer
  - $(\quad = 1) = \Phi(\quad)$
    - Dependent variable = 1 if treatment, 0 if eligible control
    - Explanatory variables ( ) include observable characteristics (e.g., peak/off-peak usage ratio, other usage-based variables, LCA indicators, CARE status, industry group indicators)
  - Model predicts each customer's *propensity score*, which is the probability that the customer is a treatment customer conditional on the characteristics included in the model
    - Estimated coefficients, , indicate variable weights
  - Each treatment customer is matched to the eligible control customer with the closest propensity score ("nearest neighbor")
  - Regression can be segmented first by characteristics

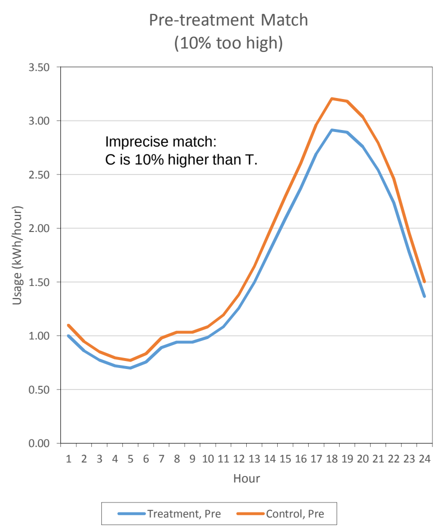
3. How do you make a matched control group? (3)

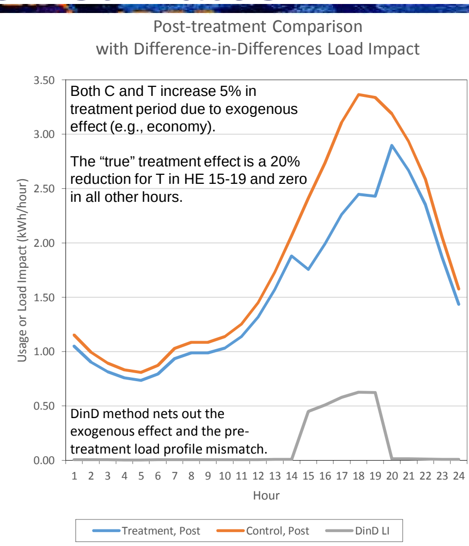
- **Euclidean distance matching:**
  - Calculate the “distance” between the characteristics of each treatment customer and every eligible control customer
  - Select the eligible control customer with the shortest distance
  - $_{,} = \sqrt{(\quad - \quad) + (\quad - \quad) \ldots + (\quad - \quad)}$
    - The *T* variables represent treatment customer characteristics and the *C* variables represent the corresponding eligible control customer characteristics
  - Places equal weights between the variables included. Consider standardization for variables with different units (see Mahalanobis matching)
  - Matches can also be segmented first by characteristics

4. What do you do with the matched control group?

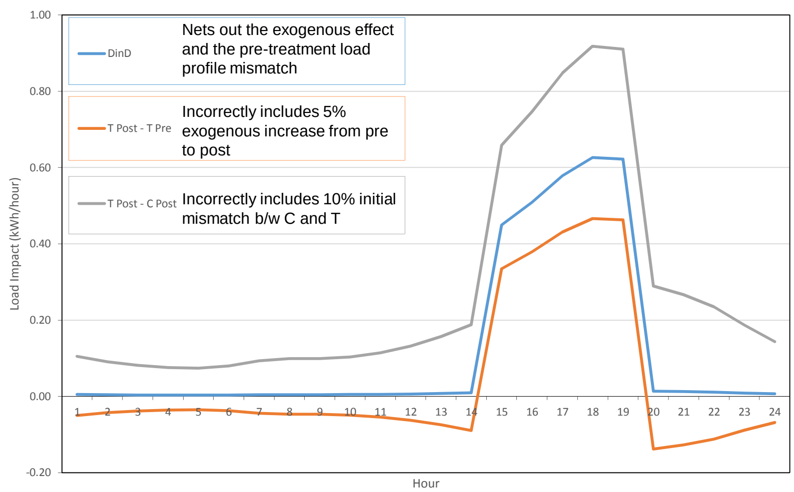
- Once you have a matched control group, load impacts can be estimated using a *difference-in-differences* approach
- Under this method, you compare treatment and control group usage during the treatment period (e.g., event days or when the customer faces TOU rates) adjusted for the pre-treatment difference between the groups
- The simple calculation is:
  - Load impact = $(kWh_{T,1} - kWh_{C,1}) - (kWh_{T,0} - kWh_{C,0})$
  - In this equation, T = treatment customer and C = control customer; 1 = treatment period and 0 = non- or pre-treatment period
- Can also estimate regressions that may or may not include other explanatory variables (e.g., weather)

4. Illustrating Difference-in-Differences Estimates





4. Illustrating Difference-in-Differences Estimates (2)



5. Any potential problems with matched control groups?

- Just because your match looks good doesn't guarantee it will produce a good load impact estimate
- For example, suppose:
  - Pre-treatment loads used in matching are from a mild weather year, the treatment period has hot weather, and the match doesn't account for thermostat set points or the willingness to endure days without turning on AC;
  - AND suppose the treatment group happens to contain a higher share of customers with a high thermostat set point or a high willingness to endure days without turning on AC (due to self selection into the program);
  - THEN the load impact estimate will be biased upward (i.e., even with no program-related load response, treatment loads will be less than control loads in the treatment period – and even DinD won't fix it)
- In practice, this problem seems unlikely to occur
  - Unobservable characteristics have to affect the change in loads (i.e., from pre-treatment to post-treatment or from non-events to events) differently for treatment and control groups
  - Unobservable effects on load that are constant across time are differenced out (by D-in-D method) and therefore do no affect load impact estimates

6. Summary

- Matched control groups are a way of mimicking an experimental design when one is not available
- Its effectiveness depends on the availability of a "good" pool of eligible control customers (large enough sample, adequate range of observed characteristics, sufficiently similar to treatment customers except for treatment)
- It is important to consider the potential effect of *unobservable characteristics* on the resulting estimates
  - Difference-in-differences method can compensate for many but not all biases that could arise due to a failure to include unobserved characteristics

Questions?

- Contact – Mike Clark,
  Christensen Associates Energy Consulting
  Madison, Wisconsin
  - mtclark@CAEnergy.com
  - 608-231-2266