MATCHING FOR EE AND DR IMPACTS

Seth Wayland, Opinion Dynamics

August 12, 2015
A Proposal

- Always use matching
  - Non-parametric preprocessing to reduce model dependence
  - Decrease bias and variance
  - Better understand your data
  - EE, DR
    - Quasi-Experiment
    - Randomized Experiment
Agenda

- Review current best practice for impact evaluation
- Review some matching methods
- Matching example
Impact Estimation

- Best Practice
  - RCT + Model (to reduce bias and variance)
  - Quasi-experiment + Matching + Model
Methods

- Model
  - Difference-in-Difference
  - Linear Fixed Effects
  - Lagged Dependent Variable

- Match
  - Propensity Score
  - Mahalanobis Distance
  - Coarsened Exact
  - Matching Frontier
Feeling Lucky?

- Randomized experiments are guaranteed to be unbiased over repeated experiments
  - There is only one actual experiment
  - How sure can we be that this one is unbiased?
    - Check the balance of treatment versus control
  - What can we do?
    - Match to reduce imbalance
    - Model to correct for dependence on known and (fixed) unknown covariates
Applying the Rubin Causal Model

For a particular unit, the causal effect of a treatment at time $t$ is the difference between what would have happened at time $t$ if the unit was exposed to the treatment and what would have happened at time $t$ if the unit was not exposed to the treatment.
Applying the Rubin Causal Model

- The customer cannot be simultaneously exposed to the treatment and not exposed to the treatment
- We need to make some assumptions
  - SUTVA
  - Ignorable treatment assignment
Ignorable treatment assignment

- Model
  - Parametrically adjust for the effect of covariates
- Match
  - Non-parametrically improve balance of all included covariates
- Both also usually reduce variance
- Matching yields insight into the data
Matching Procedure

1. Select a distance measure
2. Select and implement a matching method
3. Assess balance, return to 1 or 2 as necessary
4. Use the matched data to perform analysis
Matching Procedure - Considerations

- Choice of treatment effect (ITT, ATE, ATT, SATT, FSATT, etc.)
- Choice of variables to include in matching
- Choice of matching method
  - Choice of model in distance metric for Propensity Score matching
  - Choice of balance checks
Example

- Home energy report program with an RCT design
Matching Methods

- Exact
- K nearest neighbors
- Coarsened Exact
- Matching Frontier
- Many others
Balance Checks

- Difference in Means
  - Check all variables (don’t use statistical significance)
- Average Mahalanobis Imbalance
  - Mean Mahalanobis distance between all matched pairs
- Median L1 Distance
  - Distance between multivariate histograms
When Matching Doesn’t Help

- Coincident non-treatment changes
  - Some whole-house programs
- Missing information about treatment assignment
  - Opt-in bias?
- Modeling doesn’t help either
Coarsened Exact

- $N = 9,408$, $N_c = 9,355$
- Median L1 distance: 0.09
  - Much better
- Average mean distance: 0.37 kWh/day
  - Somewhat worse
Coarsened Exact

Feasible Group

Non-Feasible Group
Coarsened Exact

- FSATT ($N_f=9,408$, $N_c=9,355$) savings = 4.3%
- NFSATT ($N_{nf}=592$, $N_c=644$) savings = 9.6%
- Weighted SATT ($N=N_c=10,000$) savings = 4.6%
- Full Sample SATT ($N=N_c=10,000$) savings = 4.8%

\[
\text{weighted SATT} = \frac{\text{FSATT} \cdot N_f + \text{NFSATT} \cdot N_{nf}}{N}
\]
A Second Proposal

- How do we evaluate what are the best methods/approaches for impact evaluation?
- We need published data and well-defined metrics

**Common Task Method**

- Everyone works on the same problem
- Method
  - Publish data
  - Define evaluation metrics
  - Periodic public evaluation of methods
For More Information

Seth Wayland, Associate Director
Opinion Dynamics
swayland@opiniondynamics.com
Thank you

xkcd.com/925
Distance Metrics

- Exact
- Propensity Score
- Mahalanobis
  - Euclidian is a special case
Match Anyway

- Methods
  - K nearest neighbors (1:1) with SATT
    - Propensity score distance
    - Mahalanobis distance
  - Coarsened Exact with weighted SATT
    - L1 distance
Balance Metrics

- Treated group $N = 10,000$
- Comparison group $N_c = 10,000$
- Average mean difference for the 12 months of the pre-period: 0.03 kWh/day
- Median L1 distance: 0.56
K Nearest Neighbors

- Propensity score metric
  - Simple model with a variable for each month of pre-period usage
  - \( N = 10,000 \) and \( N_c = 5,580 \)
  - Average mean difference: \(-0.22 \text{ kWh/day}\)
  - Balance is a little worse
K Nearest Neighbors

- Mahalanobis distance
  - $N = 10,000$ and $N_c = 5,762$
  - Average mean difference: 2.9
  - Balance is much worse