UF UNIVERSITY of FLORIDA

Methods for Causal Inference with Observational Data

Walter L. Leite

walter.leite@coe.ufl.edu

http://www.coe.ufl.edu/Faculty/Leite/index.html

Structure of this presentation

- Introduce the problem of causal inference with observational data
  - The problem of selection bias.
  - The counterfactual framework.

- Solutions:
  - Propensity score matching.
  - Propensity score stratification
  - Inverse Probability weighting

The problem of estimating treatment effects in observational studies

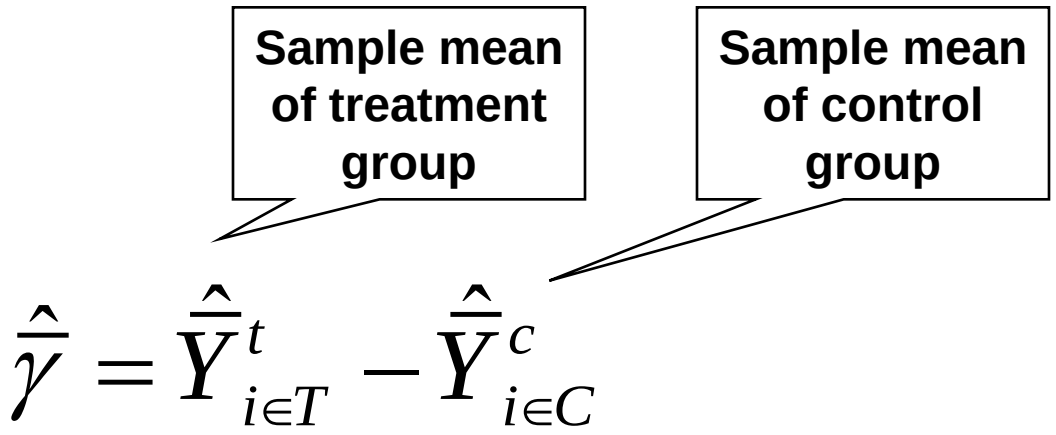
- In observational studies, estimates of treatment effects may be biased due to selection on observable and unobservable variables.

- Individuals that received the treatment may not comparable to individuals that do not receive the treatment.

The counterfactual framework (Rubin, 1974)

- All individuals have potential outcomes in both the presence and absence of treatment.
- Outcomes of the treatment group are only observed in the presence of treatment.
- Outcomes of the control group are only observed in the absence of treatment.
- Counterfactual outcomes: The potential outcomes that are *unobserved* for the treatment and control group.

Conventional estimation of treatment effects



- Strong ignorability of treatment assignment: The estimator $\hat{\bar{\gamma}}$ will only be unbiased if the treatment assignment is uncorrelated with the outcomes

Achieving conditional independence between treatment assignment and outcomes

- Conditional independence may be achieved by including covariates that correlate with treatment assignment in the analysis model

$$Y_i = \beta_0 + \beta_1 T + \beta_2 C_1 + \beta_3 C_1 + \beta_4 C_2 + \beta_5 C_3 + \beta_6 C_4 \ldots$$

Achieving conditional independence between treatment assignment and outcomes

- Problems:
  - Model can became very complex and difficult to estimate with small samples;
  - It is commonly assumed that the covariates are linearly related to the outcome;
  - Including covariates that have different distributions for treatment and control groups has been shown to lead to considerable bias in estimates of treatment effects (Winship & Morgan, 1999)

Matching

- A control group is selected that is equivalent to the treatment group if it had received no treatment;
- The matched group is the counterfactual outcome of the treatment group;
- Advantages:
  - Smaller models where fewer parameters are estimated;
  - Linearity assumptions are not made;
  - Problem of differences in distributions of covariates for treatment and control groups is eliminated.

Matching

- Matching can be performed using any number of covariates.
- Matching requires the existence of an area of common support, which is an area of the multivariate distribution of the covariates where values exist for both the treatment and control groups.
- As the number of covariates increase, it becomes more difficult to find an appropriate area of common support.

Propensity score matching

- The propensity score is defined as a predicted probability of treatment assignment, conditional on observed covariates;

- As compared to multivariate matching, propensity score matching eliminates the problem of finding an area of common support for all covariates;

- Several matching methods exist: Nearest neighbor one-one or many-one matching, Radius matching, Kernel matching, Mahalanobis metric matching, matching with a genetic algorithm.

Example of Application of Matching

- Research Question:

- Was the growth trajectory of mathematics and reading achievement of students in Lastinger Center schools different from the growth trajectory of students in comparable schools not in the Lastinger Center program?

Evaluating the Lastinger Model

- Complexities in estimating the effect of the Lastinger model:
  - Schools are not randomly assigned to participate in the Lastinger Center program.
  - Nonrandom differences between Lastinger schools and non-Lastinger schools may lead to biased estimates.
  - Solution: Matching

Matching

- Matching was performed with respect to the variables that correlated significantly with membership in the Lastinger Center program for each year.

- Matching was performed separately for the schools that joined the program in each year using covariates measured in the previous year 5 years.

Matching algorithm

- Best matches were identified using a genetic matching algorithm (Diamond & Sekhon, 2005).
- For each Lastinger school, the five best matches were identified.
- After matching, less than 5% of variables presented significant differences between Lastinger and matched schools.

School Demographics

Since we matched schools on a number of variables, non-Lastinger Schools are similar to Lastinger Schools in many ways, including

- Free & Reduced Lunch: 82-89%
- Percentage of non-white: 88-90%
- Percentage of retained 3rd graders: 11-33%

Propensity score stratification

- Selection bias can be controlled by grouping units into strata based on the propensity score, then comparing only treatment and control units who fall in the same strata.
- Research has shown that 5 strata are enough to control for 90% of the bias due to covariates.
- A common practice is to create strata based on the quintiles of the distribution of propensity scores.

Estimating treatment effects with propensity score stratification

- Before estimating treatment effects, the balance of the classes needs to be checked for each covariate.

- The balance of stratification can be checked with a 2 x 5 (treatments x strata) ANOVAS with the covariates as outcomes. If classes are balanced, all F ratios should be non-significant.

- Treatment effect can be estimated with any model that includes the strata as covariates.

$$Y_i = \beta_0 + \beta_1 T + \beta_2 S_1 + \beta_3 S_2 + \beta_4 S_3 + \beta_5 S_4 + \beta_6 S_5$$

The problem of estimating treatment effects in longitudinal observational studies

- In longitudinal studies, individuals that received the treatment at a later occasion are not comparable to individuals that receive the treatment at earlier occasions.
- The probability of receiving the treatment can be influenced by time-varying covariates, time-invariant covariates, and the previous treatment history.
- These three sources of bias can be controlled with inverse probability weighting.

Matching with time-varying treatments and time-varying outcomes

- The estimates of treatment effect will only be unbiased if there is conditional independence between treatment assignment and outcomes at each measurement occasion.

- Matching can be used to attain conditional independence at each measurement occasion by selecting time-varying control groups.

- The time-varying control groups provide counterfactual outcomes at each measurement occasion.

Inverse Probability Weighting

- If there is selection bias into the treatment condition at time *t*, unbiased estimates can be found by weighting the outcome of each person by the inverse of the conditional probability of receiving his or her own treatment:

$$w_i = \frac{1}{\prod_{k=0}^{t} p(T_i = 1 \mid L_0 = l_0)^{t_i} \; p(T_i = 0 \mid L_0 = l_0)^{1-t_i}}$$

- L_0 represents measured covariates. The weights are estimated from the data in a logistic regression of Ti on L0.

Logistic regression for estimating weights

$$\log it\left(T_i = 1\right) = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 C_3 + \beta_4 C_4 + \beta_5 C_5$$

$$P(T_i = 1) = \exp\left(\log it\left(T_i = 1\right)\right)$$

$$P(T_i = 0) = 1 - P(T_i = 1)$$

Stabilized Weights

vector of all time-invariant covariates

treatment history through time k-1

$$sw_i = \frac{\prod_{k=1}^{t} p(A(k) = a_i(k) \mid \overline{A}(k-1) = \overline{a}_i(k-1), V = v_i)}{\prod_{k=1}^{t} p(A(K) = a_i(k) \mid \overline{A}(k-1) = \overline{a}_i(k-1), \overline{L}_i = \overline{l}_k)}$$

vector of all time-varying covariates

- The numerator is the probability that the subject received his observed treatment conditional on his past treatment history and time-invariant covariates, but not adjusting for his past time-varying covariates.

- The denominator is the conditional probability that a subject had his or her own observed treatment, conditional on his past treatment history, time-invariant covariates, and time-varying covariates.

Interesting readings for beginners:

- Winship, C. & Morgan, S. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, *659–706.*

- Hahs-Vaughn, D. & Onwuegbuzie, A. J. (2006). Estimating and Using Propensity Score Analysis With Complex Samples. *The Journal of Experimental Education*, *75*(1), 31–65.

- Sampson, R. J., Laub, J. H., & Wimer, C. (2006). Does marriage reduce crime? A counterfactual approach to Within-individual causal effects. *Criminology*, 44(3), 465-510.