Constructing Counterfactuals in a Multisite Observational Study using Propensity Score Matching and Multilevel Modeling: An Empirical Example Looking at the Effect of 8th Grade Algebra Across Students and Schools

Jordan Rickles, MPP
University of California, Los Angeles

Abstract

This study presents a methodology for estimating causal effects from a multisite observational study that takes advantage of school-level variation in the assignment of students to schools to construct balanced treatment and control groups. The methodology extends Stuart and Rubin’s (2007) work on matching with multiple control groups and demonstrates how imputation of counterfactuals (Schafer & Kang, 2008) and multilevel modeling are useful tools for evaluating treatment effect heterogeneity. The methodological approach is demonstrated through an empirical example that seeks to determine whether students are better prepared for high school mathematics by taking a formal algebra course or a pre-algebra course in 8th grade.

Motivation

- Want to make causal inferences about the effects of an educational policy program, or practice when the following conditions apply:
  - Randomization is not possible or feasible.
  - Assignment to the treatment condition is highly selective.
  - The assignment mechanism can vary across sites.
  - The treatment effect can vary across students and sites.

- Example research questions where such conditions might apply:
  - Does an advanced placement course affect student learning?
  - Does a dropout prevention program keep students in school?
  - Does a school suspension affect student behavior?
  - Does grade retention affect long-term student outcomes?
  - Does taking algebra in 8th grade affect high school mathematics performance?

- Examples of standard methodological approaches to address questions like the above:
  - Regression model controlling for observed covariates
  - Non-parametrically estimated outcome, overlap, and propensity scores
  - Probit regression matching to construct treatment and control groups with similar covariate distributions
  - Requires reasonable overlap and sample size to find good matches
  - Other models: RD, interrupted Time Series, IV
  - Often not feasible given the assignment mechanism

- Other issues:
  - Inconsistently implemented or not at all to estimate effect heterogeneity

Data for Empirical Example

- 54 middle schools within a California school district that serve urban and suburban communities.

- Longitudinal data from 6th through 8th grade for 22,468 eligible students who are 8th graders during the 2006-2007 school year. 9,624 took algebra, 9,644 took pre-algebra.

- Assignment to 8th grade algebra is highly selective

- Previous research suggests a heterogeneous assignment mechanism across schools.

- Outcome: scale score on Algebra I California Standardized Test (CST) by end of 8th grade.

Proposed Methodology

- **Design Phase** Preprocess the data to construct comparable treatment (T=1) and control (T=0) groups:
  - Use matching methods to reduce dependence on modeling assumptions and extrapolation (reviews 2005).
  - Compute each site as a separate mini-study where we will approximate an experimental design:
    - Try to achieve within-site balance by matching across multiple control groups (Stuart & Rubin 2007).

- **Analysis Phase** Estimate average causal effects and describe heterogeneity:
  - Impose each student’s counterfactual potential outcome on the propensity score (Bickel & Myung, 2003) to facilitate analyses of treatment effect heterogeneity.
  - ATT = E(Y1 − Y0)
  - Use multilevel modeling techniques to explore effect heterogeneity across students, classrooms and schools.

- **Sensitivity Analysis Phase** Examine sensitivity of treatment effect estimates to deviation of assumptions:
  - How might omitted variable bias affect the estimates?
  - How do different specifications of the matching method affect the estimated effects?
  - Do the same results emerge in zero treatment effect for a non-equivalent outcome measure?

- **Design Phase: Preprocess the data via Matching**

  - Estimate propensity score for each student using a multilevel logistic regression that allows for a random intercept across schools (β0), a vector of student-level variables (ßs) where slopes vary across schools (βs), and a vector of student-level variables (ß0) with fixed slopes (β0s) logit (P) = β0 + β0s + βs + βs(school) + ε

  - Conduct propensity score matching following steps similar to Stuart & Rubin, 2007:
    - For each site, conduct a caliper 1-1 propensity score match without replacement using a caliper of 0.25 of the propensity score log odds. Call this the matched set M1 (see Figure 2, panel A)
    - For treatment students in schools that are not in M1, conduct a propensity score match with all control students not in school. Call this the matched set M2 (See Figure 2, panel B)
    - For control students in M1, conduct a propensity score match with all control students not in school. Call this the matched set M3 (See Figure 2, panel C)

  - Repeat matching steps for all schools.

  - Combine the M1 and M2 files for all schools into one data file (M) and combine the M3 files for all schools into one data file (M3).

- **Design Phase: Check Covariate Balance after Matching**

  - Matching created treatment and control groups with similar propensity score distributions (See Figure 3).

  - Matched groups also have similar 7th grade math achievement profiles (not shown).

  - Matching also created within-school treatment and control groups with similar propensity score distributions (See Figure 4).

- **Design Phase: Adjust for Outside School Matches**

  - Adjust for the treat that we had to go outside the school to find a match. What would be the control student’s outcome score if the control had been in school?

  - Using the MC file, estimate school bias based on a multilevel linear model where the outcome is a linear function of student propensity score log odds (PB) and an intercept that varies across schools:
   
   \[ Y_{ij} = \beta_0 + \beta_1 PB_{ij} + \beta_2 \text{school} + \epsilon_{ij} \]

  - Adjust the M2 control students’ observed outcome (Yij) for the difference between the estimated school effect (β2) for the student and an estimated school effect for the matched student (see Table 1).

- **Analysis Phase: Estimate Average Treatment Effects and Heterogeneity**

  - Estimate the counterfactual outcome for treatment students in M, conditional on a multilevel linear model based on control students in M:
   
   \[ \text{E}(Y_{ij}) = \beta_0 + \beta_1 PB_{ij} + \beta_2 \text{school} + \epsilon_{ij} \]

  - Compute the sample average treatment effect for the treated (ATT) and among-students (AOST) and school-level (al) variance in SAT (SAT_A) as:
   
   \[ \text{ATT} = \text{E}(Y_{ij}) - \text{E}(Y_{ij} \mid \text{PB} \text{= } 0) \]

  - Find average effect of 15.78 scale score points (see Table 2), or an effect size of about 0.25 sd.

  - The treatment effect varies significantly across classrooms and schools (see Figure 5).

  - 67% of treatment effect (46%) of classrooms with average positive effect (56%) of schools with average positive effect.

  - Examine whether certain student-level, classroom-level, and school-level characteristics mediate the SAT by including those characteristics in the 3-level model:

  - Student-level characteristics prior math achievement positively related to treatment effect, whereas student-level characteristics positively related to treatment effect explains about half the treatment effect.

  - Classroom-level diversity of student prior math achievement related to not treatment effect.

  - School-level characteristics achieve positive and diversity of achievement not related to treatment effect.

- **Sensitivity Analysis Phase**

  - Replicate methods for a non-equivalent outcome measure (matched student’s CST grade) and found no statistically significant SAT.

  - More sensitivity analyses planned

- **Next Steps**

  - Refine between-school matching strategy by restricting matches to “similar” schools.

  - Incorporate multiple imputation of school effect adjustment and student counterfactuals.

  - Expand analysis to look at additional outcomes and the treatment effect for the control students.

  - Compare treatment effect estimate to estimates from other methodological approaches.

  - Conduct simulation study to examine whether the proposed method can recover true treatment effect.

AUTHOR’S NOTES: References available upon request. The author has not been contacted or offered inducements. Part of this research is made possible by a pre-doctoral fellowship in educational policy and management (BS21024) awarded to the University of California, Los Angeles (UC Berkeley). The views expressed in this paper are those of the author and do not reflect the views of the project sponsors or funders or any organization.