Matching Across Schools with Multilevel Propensity Scores

Ben Kelcey
Wayne State University

Purposes of the Study
We examined the extent to which matching across groups using a multilevel propensity score (PS) produces a more effective estimator of the average treatment effect as compared to matching within groups using a single level PS. In particular, we structured our study to resemble nonexperimental multisite randomized trials and assessed the hypothesis that the availability of higher quality matches at different schools may remove more bias than when restricting matches to within schools. Because matching within schools potentially alleviates bias from unobserved variables, our study focuses on the extent to which the methods are sensitive to omitted covariates.

We assessed the performance of the respective treatment effect estimators in multiple scenarios using Monte Carlo simulation. We highlight four simple situations:

- No unmeasured covariates that influence the treatment assignment
- One unmeasured student level covariate that influences the treatment assignment
- One unmeasured school level covariate that influences the treatment assignment
- Both of which influence the treatment assignment

Selection Model
- Consider a nonexperimental study that resembles a multisite randomized trial such that individual students take up the treatment/control condition (e.g. retention Z=1) within each school.
- Because the avenues by which students take up the treatment may differ substantially by school (e.g. coefficient magnitude of student and school characteristics and their interactions), we allow each school to maintain different selection mechanisms (f).
- Further, because these avenues may not be observed, our true selection model includes unobserved school (U) and student (U) variables along with different predictive coefficients for each school.

Outcome Model
- Because matching across schools restricts the assessment of across school treatment effect variability, our causal estimand focuses solely on the average treatment effect. The model generating the outcome was:

Level 1: \( Y_{ij} = \beta_0 + \beta_j X_{ij} + \pi_j U_{ij} + \epsilon_{ij} \)

Level 2: \( \beta_j = \gamma_{00} + \gamma_j W_j + \delta U_{ij} + \epsilon_{ij} \)

\( \beta_j = \gamma_{00} + \gamma_j W_j + \delta U_{ij} + \epsilon_j \)

\( \pi_j = \gamma_{00} + \gamma_j W_j + \delta U_{ij} + \epsilon_{ij} \)

Propensity Scores & Matching
- Using several different single and multilevel PSs, students were matched within or across schools using full matching. We highlight the following scenarios:

**Single level:**
1. Matching within schools using a single level PS which constrained all random effects to zero
2. Matching across schools using a single level PS which constrained all random effects to zero

**Multilevel:**
3. Matching across schools using a multilevel PS which constrained random slopes to zero
4. Matching across schools using a multilevel PS with no constraints
5. Matching within schools using a multilevel PS with no constraints

Summary of Findings and Limitations
- Matching across school boundaries using multilevel PSs is potentially an effective strategy even when unobserved covariates influence treatment take-up.
- Bias attributed to unobserved student characteristics which influence the treatment selection was larger than that of unobserved school characteristics (see top right graph). This suggests that the benefit of matching students within schools to alleviate unobserved student level bias may be less central to bias reduction than unobserved student characteristics.
- Using only a random intercept may be insufficient in reducing bias if the predictive capacity of student level covariates also varies across schools (e.g. model (3)).
- High quality matches are central to bias reduction
- Limitations
  - Simulations limited.
  - Focused on a few covariates: Xs, Zs, and coefficients.
  - Focused on situations where treatment effect variability is zero.
  - Within school sample sizes were small (i.e. 10 students/school).
  - Neglects how one might establish covariate balance.