Weighting-Based Sensitivity Analysis for Evaluating the Average Treatment Effect: A Re-analysis of the Project STAR Data

Guanglei Hong
University of Chicago

Fan Yang
University of Colorado Denver

Xu Qin
University of Pittsburgh

Background

Identification of the average treatment effect (ATE) is subjected to selection bias when researchers analyze quasi-experimental data or experimental data with non-random attrition and/or noncompliance. In causal evaluations that rely on statistical adjustment for covariates, omitted confounding is always a major concern. These are covariates with different distributions between the treated group and the untreated group. In the case of non-random attrition, hidden bias may be associated with both pretreatment and posttreatment confounders. We further consider the case in which the treatment may last more than one time period. Covariates that are affected by the initial treatment assignment and predict the subsequent treatment and outcome are called “time-varying confounders” that require adjustment in evaluating the cumulative treatment effect on the final outcome. Weighting methods can flexibly adjust for observed pretreatment, posttreatment, and time-varying confounders (Hong & Raudenbush, 2008; Robins, 1987; Robins, Hernan, & Brumbeck, 2000; Rosenbaum, 1987). However, weighting cannot adjust for unobserved confounders. A sensitivity analysis (SA) quantifies the amount of hidden bias associated with the omitted confounders, observed or unobserved, and evaluates whether removing such a bias would qualitatively change the initial analytic conclusion. Conclusions that are harder to alter by scientifically plausible hidden bias are expected to add a higher value to knowledge about causality (Cornfield et al, 1959).

Purpose

We formalize a weighting-based SA approach (Hong, Qin, & Yang, 2018) for evaluations of ATE. It extends and supplements a number of existing propensity score weighting methods for identifying ATE. In its essence, the *discrepancy* between a new weight that adjusts for the omitted confounders and an initial weight that omits them captures the role of the confounders. We extend this innovative SA strategy to evaluations of ATE for assessing the consequences of not only omitted pretreatment covariates but also omitted posttreatment covariates associated with nonrandom attrition or omitted time-varying covariates associated with noncompliance. The broad utility of the weighting-based SA strategy is illustrated through a re-analysis of the well-known Student-Teacher Achievement Ratio experiment (Project STAR) data. The experiment was designed to evaluate the causal effect of class size reduction on student achievement.

Proposed Methods

Our derivation shows that, in ATE identification, the effect size of bias due to the omission of one or more pretreatment, posttreatment, or time-varying confounders denoted by *U* is $\rho_0\sigma_0 - \rho_1\sigma_1$. Here $\sigma_z = \sqrt{var(W_{zU} - W_z|Z = z)}$ is the standard deviation of the weight discrepancy in treatment group *z*; and $\rho_z = corr(W_{zU} - W_z, Y|Z = z)$ is the correlation between the weight discrepancy and the observed outcome in treatment group *z* for *z* = 0, 1. Generally speaking, the magnitude of σ_1 and σ_0 indicates the amount of *difficulty* that the omitted confounders *U* pose to the ATE identification; and the magnitude of ρ_1 and ρ_0 indicates the degree of *relevance* that the omitted confounders hold for the ATE identification. Together, $-\sigma_1\rho_1$ indicates the extent to which the weighted mean outcome of the experimental group deviates from the population average potential outcome associated with the experimental condition due to the omission of *U*; and $-\sigma_0\rho_0$ indicates the extent to which the weighted mean outcome of the control group deviates from the population average potential outcome associated with the control condition due to the omission.

Application

Starting from 1985, about 11,600 students who were enrolled in one of the 79 participating schools in Tennessee were assigned at random to classes of a small size (13~17) or classes of a regular size (22~25) with or without a teaching aide. Teachers were randomized to these different class types as well. Most students were randomly assigned in kindergarten and were expected to remain in the same class type for four years. The final outcomes are test scores measured at the end of the four-year study when most students were about to complete grade 3. Project STAR was initially designed to address causal questions including:

(1) What is the average effect of being assigned to a small class as opposed to a regular class in kindergarten on students' academic achievement four years later?
(2) What is the average effect of attending a small class as opposed to a regular class for four years on students' academic achievement?

Question (1) concerns the average effect of the treatment assigned (ATE-Assigned) while question (2) is about the average effect of the treatment received (ATE-Received). However, as pointed out by past researchers (Hanushek, 1999; Krueger, 1999; Nye, Hedges, & Konstantopoulos, 2002; Schanzenbach, 2006), non-random attrition and non-compliance complicate the causal analysis.

Unlike most of the past analyses of the Project STAR data, we employ weighting methods due to its flexibility in adjusting for observed pretreatment, posttreatment, and time-varying confounders that are associated with nonrandom attrition or noncompliance. In evaluating the ATE-Assigned, weighting transforms the subsample of students who were initially assigned to a certain class type and whose test scores were observed at the end of grade 3 to resemble the composition of the entire subsample of students who were assigned to that class type regardless of their attrition status. In evaluating the ATE-Received, the subsample of students who stayed in a small class type in all four years and the subsample of those who stayed in a regular class type for four years will each be transformed through weighting to resemble the composition of the entire sample of participants who entered the study in kindergarten. However, weighting does not remove selection bias associated with unobserved confounders. This limitation renders sensitivity analysis indispensable. We provide graphical and tabular representations of the SA results.

Conclusion

The weighting-based SA approach applies when the initial ATE analysis has been conducted through propensity score-based weighting. This new approach conveniently assesses bias associated with one or more omitted confounders. The number of sensitivity parameters does not increase with the complexity of the data generation forms. Hence this approach reduces the reliance on functional form assumptions and removes constraints on measurement scales. In a multi-year study such as Project STAR, the number of sensitivity parameters does not increase with the number of time intervals, which greatly simplifies sensitivity analysis and broadens its applicability.

References

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., & Wynder, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, *22*(4), 173-203.

Hanushek, E. A. (1999). “Some Findings from an Independent Investigation of the Tennessee STAR Experiment and fro Other Investigations of Class Size Effects,” *Educational Evaluation and Policy Analysis, 21*, 154-164.

Hong, G., Qin, X., & Yang, F. (2018). Weighting-based sensitivity analysis in causal mediation studies. *Journal of Educational and Behavioral Statistics*, *43*(1), 32-56.

Hong, G., & Raudenbush, S. W. (2008) Causal inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics*, *33*(3), 333-362.

Krueger, A. B. (1999). “Experimental Estimates of Education Production Functions,” *Quarterly Journal of Economics, 114*(2), 497-532.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2002). Do low achieving students benefit more from small classes? Evidence from the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, *24,* 201-217.

Robins, J. M. (1987). Addendum to “A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect.” *Computers and Mathematics with Applications*, *14*, 923-945.

Robins, J., Hernan, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*, 550–560.

Rosenbaum, P. R. (1987). Model based direct adjustment. *Journal of the American Statistical Association*, *82*(398), 387-394.

Schanzenbach, D. W. (2006). What have researchers learned from Project STAR? *Brookings Papers on Education Policy*, No. 9, pp. 205-228.