"Putting test scores on the right-hand side of your regression model: What works and what doesn't"

J.R. Lockwood

In observational studies of educational interventions using longitudinal student data, standardized test scores for students from year(s) prior to treatment are invariably used to adjust for pre-treatment differences among students exposed to different treatments, often through regression adjustment. However, observed test scores are noisy measures of true achievement due in part to the limited number of test items and their characteristics, and the resulting errors of measurement are typically both large and heteroskedastic. Measurement error erodes the ability of test scores to adequately control for pre-existing differences among students in different treatment groups, which typically will translate into bias in estimated treatment effects. In this presentation we discuss a variety of strategies to address test score measurement error in covariates that are tailored to the idiosyncrasies of test measurement error. The methods we discuss include using multiple prior test scores as covariates, augmenting models to include polynomial functions of test scores as covariates, method of moments corrections using test reliabilities, regression calibration methods, instrumental variable approaches, and latent variable regression methods. We discuss the pros and cons of the different methods and make recommendations about which approaches seem to provide the best balance of bias reduction and maintaining acceptable levels of precision. We present results from a case study of teacher value-added estimation using longitudinal data from a large suburban school district. We find that controlling for additional prior test scores is helpful, but is not as effective as explicit adjustments for test measurement error, and that including polynomial functions of observed test scores, a widespread practice in value-added estimation and related empirical work, is not generally effective.