

Comparing Covariate Prioritization via Matching to Machine Learning Methods for Causal Inference using Five Empirical Applications

Luke Keele, University of Pennsylvania

When investigators seek to estimate causal effects, they often assume that selection into treatment is based only on observed covariates. Under this identification strategy, analysts must adjust for observed confounders. While basic regression models have long been the dominant method of statistical adjustment, more robust methods based on matching or weighting have become more common. Of late, even more flexible methods based on machine learning methods have been developed for statistical adjustment. These machine learning methods are designed to be black box methods with little input from the researcher. Recent research used a data competition to evaluate various methods of statistical adjustment and found that black box methods out performed all other methods of statistical adjustment. Matching methods with covariate prioritization are designed for direct input from substantive investigators in direct contrast to black methods. In this article, we use a different research design to compare matching with covariate prioritization to black box methods. We use black box methods to replicate results from five studies where matching with covariate prioritization was used to customize the statistical adjustment in direct response to substantive expertise. We find little difference across the methods. We conclude with advice for investigators.