A simulation study of different approaches to mediation analysis in presence of unobserved heterogeneity and reverse causality

Dominik Becker

Federal Institute for Vocational Education and Training, Bonn, Germany

Whether an observed association between two social constructs is due to a causal effect is a fundamental methodological question in the social sciences. The additional question of how a causal effect is brought about is usually answered by mediation models investigating whether a significant parameter estimate from some type of regression of Y on X persists once mediator M is controlled for.

Concerning the analysis of panel data, unobserved heterogeneity and reverse causality are wellknown challenges that have yet been less frequently considered within statistical approaches to mediation analyses. This contribution compares the average bias of different approaches to mediation analysis – i.e., simple mediation within pooled OLS regressions (POLS), fixed-effects (FE) regressions, generalized-method-of-moments (GMM) regressions, causal mediation analysis without (CM) and with fixed effects (CMFE), and fixed-effects cross-lagged panel models (FECLPMs) – in presence of unobserved heterogeneity and reverse causality. To do so, I conduct simulation analyses of generated panel data within which intercorrelations between predictor, mediator and outcome are varied across different scenarios of causal order. Special emphasis will be laid on the sensitivity of each approach in case of an unobserved confounder affecting the mediation effect.

Preliminary results suggest that POLS estimates are generally upwardly biased, FE and CMFE estimates by trend downwardly biased, while estimates of CM models (without FEs) can be biased in both directions. In contrast, coefficients and confidence intervals estimated by both GMM regressions and FE-CLPMs are most accurate – particularly if the estimated structure of lags meets the consecutive order implied by the data-generating process. Unlike GMMs, FECLPMs are hardly sensitive to whether the first lag of the outcome variable is included as an additional predictor. Next steps involve to explore the average bias of each approach in estimating the mediation effect when an unobserved confounder affects the mediator with variable simulated regression weights.