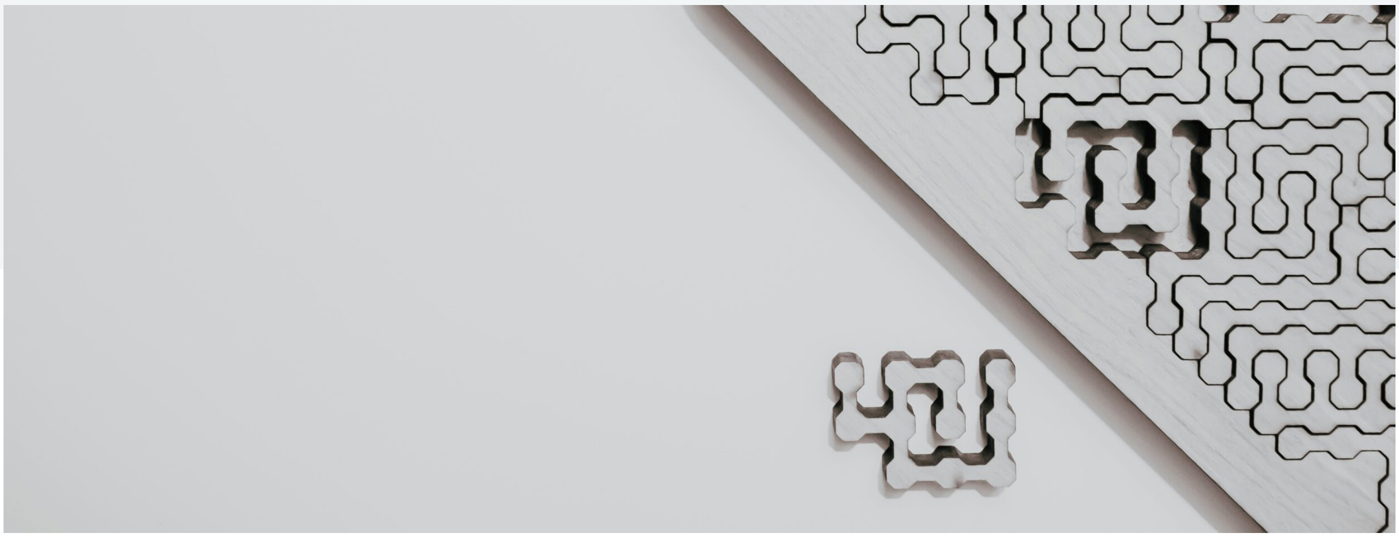
The Experimental Selection Correction Estimator: Using Experiments to Remove Biases in Observational Estimates

SUSAN ATHEY, RAJ CHETTY, GUIDO W. IMBENS



Identifying the effects of interventions on long-term outcomes is a core challenge for research on economic mobility and other questions. This study proposes a new technique to identify such impacts by combining observational and experimental data, weakening assumptions made by existing methods.

In the age of "big data," it is increasingly common for analysts to have access to two types of data: *observational* data (large datasets where treatments are not randomly assigned, but many outcomes are observed) and *experimental* data (smaller datasets where treatments are randomly assigned, but only a subset of outcomes are observed). Both types of data have strengths and limitations: observational data allow us to study a wide range of outcomes, but the inferences we draw from them can be biased because of selection bias in treatment assignment, while experimental data makes it straightforward to identify causal effects, but often lack information on outcomes of interest.

These problems are particularly important in the study of economic mobility, which typically focuses on analyzing the impacts of interventions on outcomes observed years later. For example, in the context of education, there is much interest in identifying the causal effects of classroom sizes and teacher quality in elementary school on high school graduation rates. Observational data with information on class sizes, teachers, and graduation rates are now widely available from school districts' administrative records. But causal inference using these data is challenging because of selection biases arising from non-random assignment to classrooms. Causal inference is more straightforward in experimental data – such as the widely studied Project STAR class size experiment – but experimental datasets often do not contain information on outcomes such as graduation rates because they are observed with long delays.

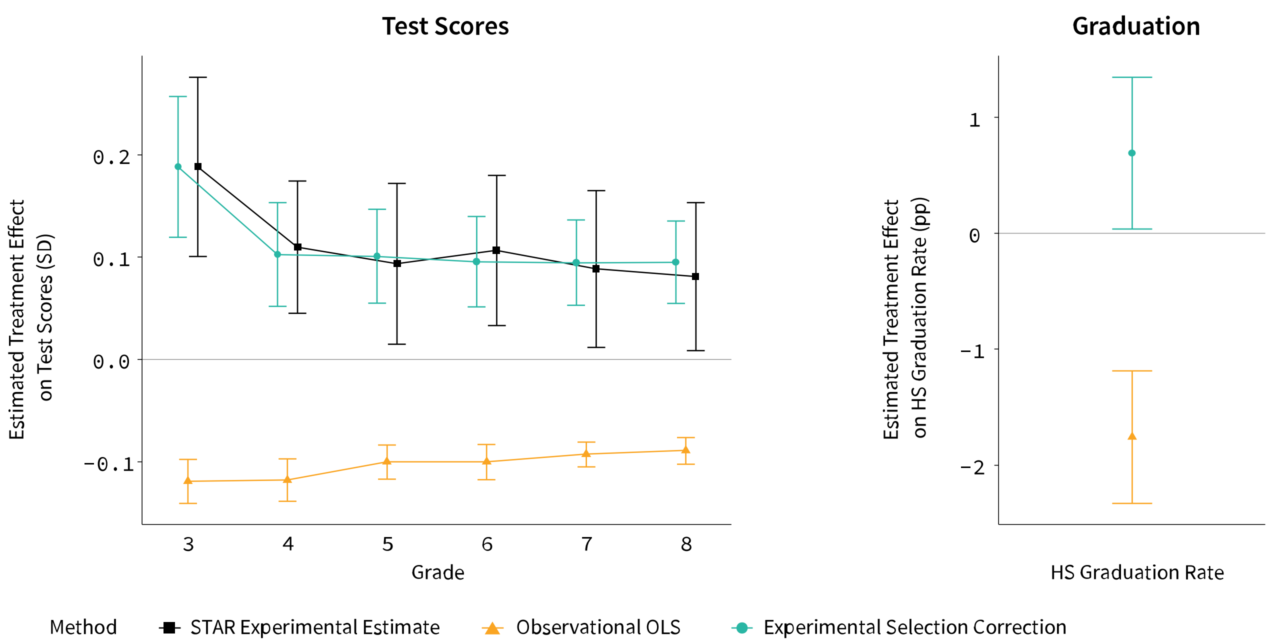
The most common approach currently used to address this problem is to use intermediate outcomes as *surrogates* for the primary outcomes of interest (Athey, Chetty, Imbens, Kang 2025). For example, test scores are widely used as a surrogate for later outcomes such as graduation rates and earnings. The surrogate approach is intuitive: we estimate an intervention's effect on test scores in experimental data, and then multiply that estimated impact by the relationship between the long-term outcomes (earnings or graduation rates) and test scores estimated in observational data. But even though it is widely applied, there is often concern that the key assumptions underlying the surrogate approach - namely that the only causal pathways through which interventions affect long-term outcomes are via the surrogate – may not be valid in practice. For example, education researchers have identified many other pathways through which childhood interventions affect long-term outcomes outside test scores, such as non-cognitive skills.

In this study, we propose a new method of estimating treatment effects on long-term outcomes that relaxes the assumptions required for the surrogate estimator in settings where both the treatment and the long-term outcome of interest are observed in the observational dataset. The approach we propose here, which we term the Experimental Selection Correction (ESC) estimator, uses the difference between observed outcomes and predicted outcomes (based on experimental data) to correct for biases in observational data. It relies on a new assumption called *latent unconfoundness*, which requires that the same unobserved factors affect both primary and secondary outcomes. This assumption is strictly weaker than the assumptions underlying existing surrogate estimators.

The estimator is straightforward to implement in three steps: First, we estimate the effect of the treatment (e.g., class size) on the intermediate outcome (e.g., test scores). Second, for all observations in the observational sample, we calculate an experimental "selection correction" term as the difference between the secondary outcome (test score) and the predicted test score based on the student's assigned class size (with the prediction based on the experimental estimate). Finally, we regress the primary outcome (e.g., graduation rates) on treatment (class size) in the observational data, controlling for the experimental selection correction term.

We apply this ESC estimator to identify the effect of third grade class size on students' outcomes. Estimated impacts on test scores using OLS regressions in observational school district data have the opposite sign of estimates from the Tennessee STAR experiment. In contrast, selection-corrected estimates in the observational data replicate the experimental estimates. Our estimator reveals that reducing class sizes by 25% increases high school graduation rates by 0.7 percentage points. Controlling for observables does not change conventional regression estimates, demonstrating that experimental selection correction can remove biases that cannot be addressed with standard controls.

FIGURE 1: Treatment Effects of Assignment to Small Class in 3rd Grade



In summary, the approach of using experimental data for selection correction weakens the assumptions underlying the popular approach of using intermediate outcomes as surrogates and yields more credible estimates of causal effects. The ESC estimator thus provides a new tool for researchers to understand the determinants of economic mobility as well as interventions of interest in other applications.

[Explore more of Opportunity Insights' methodological research in econometrics.]

Want to learn more?

Read the Paper ↗

See Presentation Slides ↗

All materials are freely available for use with citation