Abstract Title Page
Not included in page count.

Title: Compliance-Effect Correlation Bias in Instrumental Variables Estimators

Author(s): sean f. reardon

Abstract Body

Background/context:

Instrumental variables estimators hold the promise of enabling researchers to estimate the effects of educational treatments that are not (or cannot be) randomly assigned but that may be affected by randomly assigned interventions. Examples of the use of instrumental variables in such cases are increasingly common in educational and social science research. For example, although we cannot randomly assign students to attend private school, we can use the variation in private school enrollment induced by the (randomly-assigned) offer of a tuition voucher to identify and estimate the effect of private school enrollment (Krueger and Zhu 2004). Although we cannot randomly assign teachers to use specific instructional practices, we can use the variation in instructional practices induced by a (randomly-assigned) professional development program to identify and estimate the effect of specific instructional practices. The most commonly used instrumental variables estimator is two-stage least squares (2SLS). Many of the properties of the 2SLS estimator are well-understood, including its identifying assumptions (Angrist, Imbens and Rubin 1996) and the issue of finite sample bias (Bound, Jaeger and Baker 1995). Less well understood, however, is the behavior of the 2SLS estimator when individuals are heterogeneous in their compliance with the instrument (that is, heterogeneous in the extent to which the instrument affects their behavior). Heckman, Urzua, and Vytlacil (2006), for example, note a peculiar asymmetry in the assumptions of many instrumental variables models: they typically allow heterogeneity of compliance, but assume homogeneity of treatment effects. This paper investigates the effects of heterogeneous compliance on 2SLS estimates and the conditions under which such heterogeneous compliance results in bias.

A limitation of instrumental variables models is that they are able to identify the effects of only (at most) as many mediators as there are ignorably assigned instruments. Thus, if we think that a specific intervention may affect student outcomes through several different mediators, we require at least as many instruments as mediators. Multi-site randomized trials can be used to define a set of instruments, each defined as the interaction of treatment status with a site indicator variable. If there are more sites than mediators, then such multisite trials hold the promise of enabling researchers to estimate the effects of multiple mediators at once.

The issue of heterogeneous compliance, however, is particularly important in the case where site-by-treatment interactions are included as instruments, for in this case, it is precisely the heterogeneity of the compliance across sites that is being leveraged to provide identification of effects. In fact, if sites are homogeneous with respect to their compliance, the inclusion of site-by-treatment interactions as instruments leads to finite sample bias because the combined set of instruments is much weaker in this case than a single instrument would be.

Purpose / objective / research question / focus of study:

The above discussion suggests that the use of instrumental variables estimators to identify the effects of mediators in multi-site randomized trials requires a thorough understanding of the effects of compliance heterogeneity and the conditions under which it may lead to bias. In this paper, I investigate the properties of the 2SLS IV estimator under conditions of heterogeneous compliance. I do so in both the simple case where a 2SLS model is used to estimate the effect of single mediator in a single site, and the the more complex case where multiple site-by-treatment interactions are used to estimate the effects of multiple mediators.

Setting:
N/A: methodological paper

Population / Participants / Subjects:
N/A: methodological paper

Intervention / Program / Practice:
N/A: methodological paper

A Stylized Example:
To fix ideas, suppose we are interested in investigating how a teacher professional development (PD) program (call it T) affects student achievement (call it Y). Further, suppose the theory of action of the PD program is that it will affect teachers' use of three instructional practices (call them A, B, and C), which in turn will affect student achievement. We say that T acts on Y through the *mediators* A, B, and C. We therefore have three research questions: 1) what is the average effect of the PD program on student achievement? 2) what is the average effect of the PD program on teachers' instructional practices? and 3) what are the average effects of each of the instructional practices A, B, and C on student achievement?

Suppose we design a study in which we sample a moderate number of school districts, and in each district we randomize schools to receive T or not. In each school, we observe teachers' instructional practices and measure student achievement at the end of some time period.

At the end of the time period, we can easily obtain an unbiased estimate the average effect of the PD program T on student achievement (because T is randomly assigned). Moreover, we can likewise obtain an unbiased estimate the average effect of the PD program T on teachers' instructional practices (again, because T is randomly assigned). Obtaining an unbiased estimate of the effect of instructional practices A, B, and C on student achievement, however, is more complex, because teachers' instructional practices are not randomly assigned in this design.

In cases like this, researchers sometimes rely on instrumental variables (IV) methods to obtain estimates of the effects of mediators on outcomes. Under the assumption that the treatment T can only affect the outcome Y through its effects on mediators A, B, and C (as well as some additional assumptions), we may be able to identify the effects of the mediators themselves by relying on the associations among the treatment-induced variation in the mediators and the treatment-induced variation in the outcome. If districts where the PD program had larger average effects on instructional practice A are also those where the PD program had larger average effects on student achievement, controlling for the average effects of the PD program on instructional practices B and C, then we may be able to conclude that A affects Y.

Research Design:
N/A: methodological paper

Data Collection and Analysis:
N/A: methodological paper

Notation and Preliminaries:
Let there be N individuals, indexed by i, who are located in S sites, indexed by s. Let there be P mediators $M^1, M^2, \dots M^P$, indexed by p (and sometimes indexed by q). Each individual is assigned a treatment T_{is} (we assume T is ignorably assigned within each site) and we observe mediator values m_{is}^p (which may be affected by T) and outcome Y_{is} for each individual. In what follows, we will assume each of the five IV assumptions described by Angrist, Imbens, and Rubin hold (that is, we assume SUTVA; we assume a non-zero effect of T on each M^p; we assume monotonicity; ignorable treatment assignment; and the exclusion restriction) (Angrist, Imbens and Rubin 1996). Further, we will assume that the samples are large enough, and the treatment-induced variation in the mediators is strong enough, that finite sample bias is inconsequential (Bound, Jaeger and Baker 1995).

We assume that the structural model describing the relationship between the outcome Y and the mediators M^p is of the form:

$$Y_{is} = \Delta_s + \sum_{p=1}^{P} \delta^p m_{is}^p + u_{is}$$

where Δ_s is a site-specific fixed effect (we need this because we have assumed ignorable treatment assignment conditional on sites), u_{is} is an *iid* mean-zero, normally distributed error, and $\bar{\delta}^p$ is the average effect of M^p on Y in the sampled population. By the exclusion restriction, T does not appear in the structural model. We are interested in estimating each of the δ^p's.

The goal of this paper is to specify the conditions under which 2SLS will yield an unbiased estimate of each of the δ^p's. We are particularly interested in knowing whether the $\hat{\delta}^p$ estimated via 2SLS is equal to the average of the δ_i's.

The paper proceeds by deriving the estimand of the 2SLS estimator in several cases. I begin with the simple case when there is one site and one mediator. I then consider cases where treatment is randomly assigned within multiple sites but there is only a single mediator. Finally, I consider the most general case, where there are multiple sites and multiple mediators.

Findings / Results:
Case 1: S=1, P=1: Consider the system of equations

$$m_i^1 = \gamma^1 T_i + \epsilon_i^1$$

$$Y_i = \delta^1 m_i^1 + u_i$$

Here, γ^1 is the average effect of T on mediator m^1 (that is, it is the average of the person-specific *compliance*, γ_i^1), and δ^1 is the average effect of mediator m^1 on outcome Y (that is, it is the average of the person-specific effect, δ_i^1). In the paper, I show that, in infinite samples, 2SLS estimates converge to the estimand

$$\begin{aligned} plim\left[\hat{\delta}^{1(2SLS)}\right] &= \sum_i \frac{\gamma_i^1}{N\gamma^1}\delta_i^1 \\ &= \delta^1 + \frac{Cov\,(\gamma_i^1, \delta_i^1)}{\gamma^1} \end{aligned}$$

That is, the 2SLS estimand is the *compliance-weighted average treatment effect* (as shown in the first line), not the simple average treatment effect we desire. The second line writes this same estimand as the average treatment effect plus a bias term, which will be non-zero if the person-specific compliance (the effect of T on M^1) is correlated with the person-specific effect of M^1 on Y among individuals (I refer to this as the *compliance-effect correlation*). The compliance-effect correlation bias will be exacerbated when T is a weak instrument (when γ^1 is small).

Case 2: S>1, P=1: In this case, we have multiple sites, but only a single mediator. A standard approach is to use the multiple site-by-treatment interactions to generate S instruments (see, for example, Katz, Kling and Liebman 2007). We then have the following pair of equations:

$$\begin{aligned} m_{is}^1 &= \Gamma_s^1 + \sum_{r=1}^{S} \gamma_r^1 (I_{is}^r \cdot T_{is}) + \epsilon_{is}^1 \\ Y_{is} &= \Delta_s^1 + \delta^1 m_{is}^1 + u_{is} \end{aligned}$$

In the paper, I show that estimating this model via 2SLS yields

$$plim\left[\hat{\delta}^{1(2SLS)}\right] = \sum_s \frac{n_s \sigma_s^2 (\gamma_s^1)^2}{\sum_s n_s \sigma_s^2 (\gamma_s^1)^2}\left(\delta_s^1 + \frac{Cov_s(\gamma_{is}^1, \delta_{is}^1)}{\gamma_s^1}\right)$$

where n_s is the number of sampled individuals in site s; σ_s^2 is the variance of the treatment in site s; γ_s^1 is the average compliance within site s; δ_s^1 is the average effect of M^1in site s, and $Cov_s(\gamma_{is}^1, \delta_{is}^1)$ is the compliance-effect covariance within site s.

This result indicates that there are two sources of bias in the 2SLS estimator: 1) one due to the correlation between the square of the site-average compliance ($(\gamma_s^1)^2$) and the site-average effect (δ_s^1) (I call this the *between-site compliance-effect bias*;); and 2) one due to the average within-site correlations between person-specific compliance and person-specific effects (I call this the *within-site compliance-effect bias*).

Case 3: S>P>1: In this case, we have multiple sites, and multiple mediators (but more sites than mediators in order that the model be identified). A standard approach is to use the multiple site-by-treatment interactions to generate S instruments (see, for example, Morris, Duncan and Rodrigues 2006). We then have the following system of equations:

This implies the system of equations:

$$
\begin{aligned}
m_{is}^{1} &= \Gamma_{s}^{1} + \sum_{r=1}^{S} \gamma_{r}^{1}(I_{is}^{r} \cdot T_{is}) + \epsilon_{is}^{1} \\
m_{is}^{2} &= \Gamma_{s}^{2} + \sum_{r=1}^{S} \gamma_{r}^{2}(I_{is}^{r} \cdot T_{is}) + \epsilon_{is}^{2} \\
&\vdots \\
m_{is}^{P} &= \Gamma_{s}^{P} + \sum_{r=1}^{S} \gamma_{r}^{P}(I_{is}^{r} \cdot T_{is}) + \epsilon_{is}^{P} \\
Y_{is} &= \Delta_{s} + \sum_{p} \delta^{p} m_{is}^{p} + u_{is}
\end{aligned}
$$

In the paper, I show that fitting this set of equations via 2SLS yields the following:

$$
plim\left[\hat{\delta}^{p(2SLS)}\right] = \sum_{s}\left[\left(n_{s}\sigma_{s}^{2}\gamma_{s}^{p}\sum_{q=1}^{P}\alpha_{pq}\gamma_{s}^{q}\right)\left(\delta_{s}^{p} + \frac{Cov_{s}\left(\gamma_{is}^{p},\delta_{is}^{p}\right)}{\gamma_{s}^{p}}\right)\right]
$$

where α_{pq} is a weight term (it is the element in the p^{th} row and q^{th} column of the inverse of the weighted variance-covariance average compliance matrix). This shows that the 2SLS estimator converges to an estimand that is a complex weighted average of the person-specific effects in the case of a multi-site multiple-mediator estimator. There are two kinds of bias here: within-site compliance-effect correlation bias; and bias due to a correlation between the site-specific weights and the site average treatment effects. The weights are complex and relatively uninterpretable, suggesting that the 2SLS estimator in this case does not in general yield a meaningful estimand.

Conclusions:

The paper shows that the 2SLS estimator may be severely biased by correlations between instrument compliance and mediator effects, even if all the standard IV assumptions are met. This bias is in addition to finite sample bias, and is, in fact, exacerbated by weak instruments. Because there are many conditions that might yield compliance-effect correlations, this source of bias should be considered in designing and interpreting the results of studies that rely on instrumental variables estimation.

Appendices

Appendix A. References

References are to be in APA version 6 format.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91:444-455.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90:443-450.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. "Understanding instrumental variables in models with essential heterogeneity." *Review of Economics and Statistics* 88:389-432.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2007. "Experimental estimates of neighborhood effects." *Econometrica* 75:83-119.

Krueger, Alan B., and Pei Zhu. 2004. "Another Look at the New York City Voucher Experiment." *American Behavioral Scientist* 47:658-698.

Morris, Pamela, Greg J. Duncan, and Christopher Rodrigues. 2006. "Does Money Really Matter? Estimating Impacts of Family Income on Young Children's Achievement with Data from Random-Assignment Experiments." *Unpublished manuscript*.

Appendix B. Tables and Figures

Not included in page count.