Paper 2

Abstract Title Page

Not included in page count.

Title: **Assessing Compliance-Effect Bias in the Two Stage Least Squares Estimator**

Author(s): Sean Reardon (Stanford University) sreardon@suse.stanford.edu
Fatih Unlu (Abt Associates, Inc., contact author) Fatih_Unlu@abtassoc.com
Pei Zhu (MDRC) Pei.Zhu@mdrc.org
Howard Bloom (MDRC) Howard.Bloom2@mdrc.org

Abstract Body

Background / Context:

In the past eight years, education research has taken a quantum leap forward based on a large and growing number of high-quality randomized field trials and regression discontinuity studies of the effects of educational interventions.[2] Most of this new research and existing methodologies for conducting it focus on the response of student academic outcomes to specific educational interventions.[3] Such information is invaluable and can provide a solid foundation for accumulating much-needed knowledge. However, this information only indicates how well specific interventions (which comprise complex bundles of features) work for specific students in specific settings. Therefore by itself, the information is not sufficient to ascertain "What works best for whom, when and why?" And it is this more comprehensive understanding of educational interventions that is needed to guide future policy and practice.

In other words, it is necessary to "unpack the black boxes" being tested by randomized experiments or high-quality quasi-experiments in order to learn how best to improve the education—and thus life chances—of students in the U.S., especially those who are economically disadvantaged. This unpacking job comprises learning more about the relative effectiveness of the active ingredients of educational interventions (their mediators) and learning more about factors that influence the effectiveness of these interventions (their moderators).[4] Now that multi-site randomized experiments and rigorous quasi-experiments have been shown to be feasible for educational research[5] it is an opportune time to begin to explore these subtler and more complex questions.

Of particular relevance for the present paper is the use of instrumental variables analysis in the context of multi-site randomized experiments or quasi-experiments to study the effects of mediating variables on final outcomes. In particular, recent applications of the approach have begun to use it to explore causal effects of one or more mediating factors. For example, data from a randomized trial of subsidies for public housing residents to stimulate movement to lower-poverty neighborhoods were used to study the effects of neighborhood poverty on child outcomes (Kling, Liebman, and Katz, 2007). Using a similar strategy, Morris, Duncan, and Rodrigues (2010) used data from 16 implementations of welfare-to-work experiments to identify the impact of family income, average hours worked, and receipt of welfare as mediators.

Even though this strategy for generating multiple instruments has potentially great appeal in research on causal effects through multiple mediators in education policy, the conditions under which this strategy can be used to identify the average treatment effect (ATE) has not been addressed until recently. Specifically, Reardon and Raudenbush (2010) fills this vacuum by

[2] Spybrook and Raudenbush (2009) identified 75 randomized studies of a broad range of interventions and Gamse et al. (2008) and Jackson et al. (2007) report on regression discontinuity studies of the federal Reading First and Early Reading First programs.
[3] An important exception involves a series of randomized tests of interventions for improving students' social and emotional outcomes (Jones, Brown, and Aber, 2008 and Haegerick and Metz, under review).
[4] Cook (2001) speculates about why, until recently, the education research community strenuously resisted randomized experiments.
[5] Greenburg, Meyer, Michalopoulos and Wiseman (2003) argue for using multi-site experiments to study moderators of program effectiveness.

demonstrating that a number of assumptions above and beyond the canonical instrumental variable analysis assumptions (Angrist, Imbens,and Rubin, 1996) are needed to identify the average treatment effect in the case of a multi-site study in which an instrument may affect the outcome through multiple mediators. One of their key assumptions is that the effect of the treatment (instrument) on a mediator ("compliance") should not be correlated with the effect of that mediator on the outcome of interest ("effect"), i.e., no compliance-effect covariance. The proposed paper zooms in on this assumption and assesses the properties of the most common instrumental variable estimator (the two-stage least squares, or 2SLS, estimator) when there is compliance-effect covariance.

Purpose / Objective / Research Question / Focus of Study:

The proposed paper studies the bias in the 2SLS estimator that is caused by the compliance-effect covariance (hereafter, the compliance-effect bias). It starts by deriving the formula for the bias in an infinite sample (i.e., in the absence of finite sample bias) under different circumstances. Specifically, it considers the following cases:

a) A single site study with one mediator;
b) A multiple site study with one mediator; and
c) A multiple site study with multiple mediators.

The formulas demonstrate how the magnitude of the compliance-effect bias varies with different parameters (e.g., compliance-effect correlation, mean and variance of the compliance and effect) in infinite samples. However, as the situation under consideration gets more complicated, the bias formula quickly becomes intractable. The second part of the paper, therefore, uses simulations to demonstrate the relationship between the compliance-effect bias and various parameters, as well as the behavior of the estimated 2SLS standard errors. Furthermore, the simulation exercise assesses how the compliance-effect bias interacts with the finite sample bias when the analysis sample is small or when the instrument is weak. The paper also uses simulations to compare the properties of the 2SLS estimator with those of the ordinary least squares (OLS) estimator in the presence of the compliance-effect bias, the finite sample bias, and the omitted variable bias.

Significance / Novelty of study:

This is the first paper that systematically studies the form and behavior of the compliance-effect bias. It provides valuable insights to under what circumstances the compliance-effect correlation is likely to be problematic. It also compares the performances of the 2SLS estimator and the OLS estimator when various combinations of bias sources exist, thereby providing guidance to researchers as to which estimation method is more suitable for a given situation.

Statistical, Measurement, or Econometric Model:

The paper starts with the derivation of the formulas for the compliance-effect bias in the absence of the finite sample bias. In particular, it studies the following cases:

a. <u>A single-site study with one mediator and one instrument</u>

In this case, the following set of models are used to estimate the effect of the instrument, T, on the mediator, m^1, and the effect of T on the outcome, Y:

$$m_i^1 = \gamma^1 T_i + \epsilon_i^1 \quad (1)$$

$$Y_i = \beta^1 T_i + u_i \quad (2)$$

The model in Equation 1 is sometimes called the first-stage equation and the one in Equation 2 is referred to as the second-stage equation. It has been shown that the 2SLS estimator of the mediator effect on the outcome can be expressed as $\hat{\delta}^{1(2SLS)} = \frac{\hat{\beta}^1}{\hat{\gamma}^1}$ (Wald estimator). In the absence of finite sample bias, it can be further shown that:

$$E\left[\hat{\delta}^{1(2SLS)}\right] = \delta^1 + \frac{Cov\left(\gamma_i^1, \delta_i^1\right)}{\gamma^1} \quad (3)$$

That is, the 2SLS estimator will be biased if the effect of the T on m^1 is correlated with the effect of m^1 on Y (the *compliance-effect correlation*). The compliance-effect correlation bias will be exacerbated when T is a weak instrument (or when γ^1 is small).

b. <u>A multiple-site study with one mediator</u>

Since there are more sites than mediators in this case, there are at least three options to estimate the effect of mediator on the outcome.

<u>*Option 1: Average of the within-site 2SLS estimates*</u>

A separate 2SLS model like the one described in Equations 1 and 2 can be fitted within each site and the resulting within-site 2SLS estimates can be averaged across sites, weighting by sample sizes within sites. It can be shown that:

$$E\left[\hat{\delta}^{1(2SLS)}\right] = \delta^1 + \sum_s \frac{n_s}{N}\left[\frac{Cov_s\left(\gamma_{is}^1, \delta_{is}^1\right)}{\gamma_s^1}\right] \quad (4)$$

where $n_s$is the sample size for site s, N is the total sample size across all sites, and $Cov_s(\gamma_{is}^1, \delta_{is}^1)$ is the compliance-effect correlation for site s. Equation 4 shows that the average of the within-site 2SLS estimates will be a biased estimate of the average effect of M^1 on Y in the sampled population unless the second term above is zero.[6]

<u>*Option 2: 2SLS with site fixed effects and a single instrument*</u>

Here the following model with a single instrument (the treatment indicator) and site fixed effects will be fitted to the pooled dataset:

$$m_{is}^1 = \Gamma_s^1 + \gamma^1 T_{is} + \epsilon_{is}^1 \quad (5)$$

$$Y_{is} = \mathrm{B}_s^1 + \beta^1 T_{is} + u_{is} \quad (6)$$

[6] The pooled 2SLS estimate is unbiased if the within-site compliance-effect correlation is either zero in all sites or positive in some sites and negative in others in such a way that the weighted average is zero, which is unlikely.

where Γ_s^1 and B_s^1 are the site fixed effects for the first and second-stage regression. As before, $\hat{\delta}^{1(2SLS)} = \frac{\hat{\beta}^1}{\hat{\gamma}^1}$. Assuming very large samples, it can be shown that:

$$E\left[\hat{\delta}^{1(2SLS)}\right] = \sum_s \frac{w_s \gamma_s^1}{\sum_s w_s \gamma_s^1} \delta_s^1 + \frac{\sum_s w_s Cov_s\left(\gamma_{is}^1, \delta_{is}^1\right)}{\sum_s w_s \gamma_s^1} \qquad (7)$$

where $w_s = \frac{n_s \sigma_s^2}{\sum_s n_s \sigma_s^2}$. That is, this approach yields a weighted average of the within-site average effects of M^1 on Y and a bias-term, where the weights are proportional to the product of the sample size, the variance of the treatment, and the compliance within each site and the bias-term is a function of the within-site compliance-effect covariance.

If the variance of the treatment effect and the sample size is the same for all sites, (that is, if $w_s = 1$ for all s):

$$\begin{aligned} E\left[\hat{\delta}^{1(2SLS)}\right] &= \sum_s \frac{\gamma_s^1}{\sum_s \gamma_s^1} \delta_s^1 + \frac{\sum_s Cov_s\left(\gamma_{is}^1, \delta_{is}^1\right)}{\sum_s \gamma_s^1} \\ &= \delta^1 + \frac{Cov(\gamma_s^1, \delta_s^1)}{\gamma^1} + \frac{\sum_s Cov_s\left(\gamma_{is}^1, \delta_{is}^1\right)}{S\gamma^1} \qquad (8) \end{aligned}$$

So the site fixed effects 2SLS estimator will be an unbiased estimate of the average effect of M^1on Y if the between-site compliance-effect correlation is zero and the average within-site compliance-effect correlation is zero. The extent of these biases will be exacerbated when the average compliance is low (weak instrument).

Option 3: 2SLS with site fixed effects and site-by-treatment interactions as instruments

A third option utilizes multiple site-by-treatment interactions to generate as many instruments as the number of sites (S):

$$m_{is}^1 = \Gamma_s^1 + \sum_{r=1}^{S} \gamma_r^1 (I_{is}^r \cdot T_{is}) + \epsilon_{is}^1 \qquad (9)$$

$$Y_{is} = \Delta_s^1 + \delta^1 m_{is}^1 + u_{is} \qquad (10)$$

where I_{is}^r is a site indicator that equals one if individual *i* is in site r and zero otherwise. We then show that:

$$E\left[\hat{\delta}^{1(2SLS)}\right] = \sum_s \frac{w_s \gamma_s^1 \gamma_s^1}{\sum_s w_s \gamma_s^1 \gamma_s^1} \left[\delta_s^1 + \frac{Cov_s\left(\gamma_{is}^1, \delta_{is}^1\right)}{\gamma_s^1}\right] \qquad (11)$$

where w_s is defined as in Equation 7. Equation 11 shows that this option yields a weighted average of the within-site average effects of M^1 on Y, where the weights are proportional to the product of the sample size, the variance of the treatment, and the *square* of the compliance within each site.

c. A multiple-site study with multiple mediators

Here we have multiple mediators and more sites than mediators (multiple site multiple instruments, or MSMM). Suppose that there are S sites and P mediators and that S > P > 1. This implies that there are P first-stage and one second-stage equations:

$$m_{is}^p = \Gamma_s^p + \sum_{r=1}^{S} \gamma_r^p (I_{is}^r \cdot T_{is}) + \epsilon_{is}^p \quad \text{(p=1,2,..,P)} \qquad (12)$$

$$Y_{is} = \Delta_s + \sum_p \delta^p m_{is}^p + u_{is} \tag{13}$$

It can be shown that in this case,

$$\delta^p = \sum_s \left[\left(w_s \gamma_s^p \sum_r \alpha_{pr} \gamma_s^r \right) \left(\delta_s^p + \frac{Cov_s(\gamma_{is}^p, \delta_{is}^p)}{\gamma_s^p} \right) \right] \tag{14}$$

In other words, the MSMM 2SLS estimand is a weighted average of the site-specific δ_s^p's, where the weights include the sample size, the treatment variance, the site-specific compliance with mediator M^p, and a weighted average of the site-specific compliances with each of the mediators. Unless the correlation of the site-specific effects with each of these weight terms is zero, the estimator will be biased. In addition, there is another source of bias in the MSMM 2SLS estimator – the ratio of the within-site mediator M^p compliance-effect covariance to the site-specific compliance (which is also weighted by the same factors as above).

These analyses show that the 2SLS may yield biased estimates of the effects of mediator(s) on the outcome. There are several sources of bias: within-site compliance-effect correlation (i.e., individuals whose mediator values are most strongly affected by the treatment (instrument) respond more, on average, to the mediators); between-site compliance-effect correlations; and unequal treatment variance across sites.

Usefulness / Applicability of Method:

The paper utilizes simulated data corresponding to the situations described above to facilitate the understanding of the derived expressions for the compliance-effect bias in the 2SLS estimator. It further demonstrates how substantial the compliance-effect bias can be in different situations and the relationship between the bias and the various parameters that affect the bias.. For example, it shows that for the multiple-sites two mediators case, for certain parameter values, holding other parameters fixed, the 2SLS bias for the first mediator tends to increase as:

- the strength of the instrument decreases,
- the variance of the effect of treatment on mediator increases,
- the effect of mediator on outcome increases, or
- the compliance-effect correlation (either within a mediator or cross-mediators) increases.

The paper also relies on simulations to study the standard error of the 2SLS estimator in the presence of just compliance-effect bias, just finite sample bias, or both. The simulated 2SLS estimates are also compared to the corresponding OLS estimates to assess which estimation method produces less bias or is more efficient under given conditions.

Conclusions:

This paper derives the expressions for the bias in the 2SLS estimator when the effects of treatment on mediator(s) are correlated with the effects of mediator(s) on the outcome in various situations and uses simulated data to demonstrate the behavior of the compliance-effect bias. It shows that the compliance-effect bias can be substantial under certain conditions. Therefore it is important for researchers to assess the potential magnitude of this bias before selecting the method to conduct mediational analyses.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Angrist, J. D., G. Imbens, & D. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*, 444-55.

Cook, T. D. (2001). Sciencephobia: Why Education Researchers Reject Randomized Experiments. *Education Next*, *Fall*, 63-68.

Gamse, B. C., H. S. Bloom, J. J. Kemple, & R. T. Jacob (2008). *Reading First Impact Study: Interim Report (NCEE 2008-4016).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Greenberg, D. H., R. Meyer, C. Michalopoulos, & M. Wiseman (2003). Explaining Variation in the Effects of Welfare-to-Work Programs. *Evaluation Review, 27(4),* 359-94.

Haegerich, T.M., & E. Metz (under review). The Social and Character Development Research Program: Development, Goals and Opportunities. *Journal of Research in Character Education*.

Jackson, R., A. McCoy, C. Pistorino, A. Wilkinson, J. Burghardt, M. Clark, C. Ross, P. Schochet, & P. Swank (2007). *National Evaluation of Early Reading First: Final Report*, U.S. Department of Education, Institute of Education Sciences, Washington, DC: U.S. Government Printing Office.

Jones, S. M., J. L. Brown, & J. L. Aber (2008). Classroom Settings as Targets of Intervention and Research. In M. Shinn and H. Yoshikawa (editors) *Changing Schools and Community Organizations to Foster Positive Youth Development.* New York: Oxford University Press.

Kling, J, J. B. Liebman, & L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica* 75, 1: 83-119.

Morris, P., Duncan, G. J., & Rodrigues, C. (2010). Does Money Really Matter? Estimating Impacts of Family Income on Young Children’s Achievement with Data from Random-Assignment Experiments. *Unpublished manuscript*.

Reardon, S. F. & S. W. Raudenbush (2010). Under What Assumptions Do Site-by-Treatment Instruments Identify the Average Causal Effects? *Unpublished manuscript*.

Spybrook, J. & S. W. Raudenbush (2009). An Examination of the Precision and Technical Accuracy of the First Wave of Group Randomized Trials Funded by the Institute of Education Sciences. *Educational Evaluation and Policy Analysis*, 31(3): 298–318.