

NBER WORKING PAPER SERIES

THE PERILS OF PEER EFFECTS

Joshua Angrist

Working Paper 19774

<http://www.nber.org/papers/w19774>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

December 2013

Presented at the European Association of Labor Economists annual meeting, September 2013, in Torino. This research was partially funded by the Institute for Education Sciences. Gaston Illanes and Gabriel Kreindler provided expert research assistance. Seminar participants at EALE, Maryland, Warwick, and Queens provided helpful comments. Special thanks go to Bruce Sacerdote, who patiently walked me through his earlier analyses and graciously supplied new results, and to Steve Pischke, for extensive discussions and feedback repeatedly along the way. Thanks also go to many of my other peers for helpful discussions and comments, especially Daron Acemoglu, Andrea Ichino, Guido Imbens, Patrick Kline, Guido Kuersteiner, Steven Lehrer, Victor Lavy, Parag Pathak, and Rob Townsend. The effects of their interventions were mostly modest, but that's entirely my fault. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Joshua Angrist. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Perils of Peer Effects  
Joshua Angrist  
NBER Working Paper No. 19774  
December 2013, Revised January 2014  
JEL No. C18,C31,C36,I21,I31

### **ABSTRACT**

Individual outcomes are highly correlated with group average outcomes, a fact often interpreted as a causal peer effect. Without covariates, however, outcome-on-outcome peer effects are vacuous, either unity or, if the average is defined as leave-out, determined by a generic intraclass correlation coefficient. When pre-determined peer characteristics are introduced as covariates in a model linking individual outcomes with group averages, the question of whether peer effects or social spillovers exist is econometrically identical to that of whether a 2SLS estimator using group dummies to instrument individual characteristics differs from OLS estimates of the effect of these characteristics. The interpretation of results from models that rely solely on chance variation in peer groups is therefore complicated by bias from weak instruments. With systematic variation in group composition, the weak IV issue falls away, but the resulting 2SLS estimates can be expected to exceed the corresponding OLS estimates as a result of measurement error and other reasons unrelated to social effects. Randomized and quasi-experimental research designs that manipulate peer characteristics in a manner unrelated to individual characteristics provide the strongest evidence on the nature of social spillovers. As an empirical matter, designs of this sort have uncovered little in the way of socially significant causal effects.

Joshua Angrist  
Department of Economics, E17-226  
MIT  
77 Massachusetts Avenue  
Cambridge, MA 02139  
and NBER  
angrist@mit.edu

# 1 Introduction

In a regression rite of passage, social scientists around the world link student achievement to the average ability of their schoolmates. A typical regression in this context puts individual test scores on the left side, with some measure of peer achievement on the right. These regressions reveal a strong association between the performance of students and their peers, a fact documented in Sacerdote's (2011) recent survey of education peer effects. Peer effects are not limited to education and schools; evidence abounds for associations between citizens and neighbors in every domain, including health, body weight, work, and consumption, to name a few (a volume edited by Durlauf and Young (2001) points to some of the literature.)

Most people have a powerful intuition that "peers matter," so behavioral interpretations of the strong positive association between the achievement of students and their classmates or the labor force status of citizens and their neighbors ring true. Correlation among peers is a reliable descriptive fact, but the scope for spurious correlation in peer analysis is wide. Others have made this point (see, especially, Deaton, 1990; Manski, 1993; Boozer and Cacciola, 2001; Moffitt, 2001; and Hanushek, Kain, Markman, and Rivkin, 2003). Nevertheless, I believe there's value in a restatement and synthesis of the many perils of econometrically estimated peer effects. I find it especially useful to link econometric models of peer effects to the behavior of instrumental variables (IV) estimators.

The link with IV shows that models which assign a role to group averages in the prediction of individual outcomes should be expected to produce findings that look like a peer effect, even in a world where behavioral influences between peers are absent. The vacuous nature of many econometric peer effects is not an identification problem; the parameters of the models I discuss are identified. More often than not, however, these parameters teach us little about human behavior or what we should expect from changes in group composition. If the group average in question involves the dependent variable, the estimated peer effect is a mechanical phenomenon, either affirming an identify in the algebra of expectations or providing a measure of group clustering devoid of behavioral content. If the model in question draws in individual covariates, the putative peer effect is a test for the equality of two-stage least squares (2SLS) and OLS estimates of the effect of these covariates on outcomes. There are many reasons why 2SLS estimates might differ from OLS; peer effects are on the list, but should not be at the top of it.

## 2 Peer Theory

Like many in my cohort, I smoked a lot of dope in high school. Most of my friends smoked a lot of dope too. Ten years later, my youngest brother went to the same high school, but he didn't smoke nearly as much dope as I did, something that worried me at the time. My brother's friends also smoked little. In fact, by the time my brother made it to our high school, nobody smoked as much dope as we did in 1975. That must be why my brother smoked so much less than me.

This story calls for some research. Let  $\bar{s}_j$  be the smoke-alotta-dope rate among students attending high school  $j$ , the school average of  $s_{ij}$ , a dummy for whether student  $i$  smokes. Am I more likely to smoke when rates are high? We can explore this by estimating the following regression,

$$s_{ij} = \alpha + \beta \bar{s}_j + \xi_{ij} \tag{1}$$

Estimation of (1) is superfluous, of course. Any regression of  $s_{ij}$  on  $\bar{s}_j$  produces a coefficient of unity:

$$\frac{\sum_j \sum_i s_{ij} (\bar{s}_j - \bar{s})}{\sum_j n_j (\bar{s}_j - \bar{s})^2} = \frac{\sum_j (\bar{s}_j - \bar{s}) (n_j \bar{s}_j)}{\sum_j n_j (\bar{s}_j - \bar{s})^2} = 1$$

In fact, the properties of equation (1) emerge without algebra: The group average on the right hand side is a fitted value from a regression of the left hand side on dummies indicating groups (high schools, in this case). The covariance between any dependent variable and a corresponding set of fitted values for this variable is equal to the variance of the fits.

The tautological nature of the relationship between individual data and group averages is not a story about samples. Let  $\beta$  denote the population regression coefficient from a regression of (mean zero)  $y$  on  $\mu_{y|z} = E[y|z]$ , for any random variables,  $y$  and  $z$ . The scenario I have in mind is that  $z$  indexes peer-referent groups (like high schools). For any  $z$ , we can be sure that

$$\beta \equiv \frac{E[y\mu_{y|z}]}{V[\mu_{y|z}]} = 1, \tag{2}$$

a relation that follows by iterating expectations:

$$\begin{aligned} E[y\mu_{y|z}] &= E\{E[y|z, \mu_{y|z}] \times \mu_{y|z}\} = E\{E[y|z] \times \mu_{y|z}\} \\ &= E[\mu_{y|z}^2] = V[\mu_{y|z}]. \end{aligned}$$

Others have commented on the vacuous nature of regressions of individual outcomes on group mean outcomes. Manski (1993) described the problem this way: "... observed behavior is *always* consistent with the hypothesis that individual behavior reflects mean reference-group behavior" (ital-

ics mine). Implicit in Manski’s extended discussion, however, is the suggestion that the tautological nature of (2) is a kind of troubling special case, perhaps a bad equilibrium that can in principle be avoided. In the same spirit, Brock and Durlauf (2001) and Jackson (2010), among others, describe regressions like (2) as posing an identification problem, suggesting we might, with suitable econometric magic, find a solution. Yet, the coefficient in my simple regression of individual outcomes on high school mean outcomes is identified in a technical sense, by which I mean, *Stata* or even *SAS* should have no trouble finding it!

Econometric models of endogenous peer effects are typically more sophisticated than the one I’ve used to describe the Angrist brothers’ smoking habits. Discussing peer effects in the Tennessee STAR class size experiment, Boozer and Cacciola (2001, p.46) observed: “Of course, since the setup just discussed delivers a coefficient of exactly 1, it is improbable a researcher would not realize his error, and opt for a different estimation strategy.” Sophistication, however, need not produce a sound causal framework. In another analysis of the STAR data, for example, Graham (2008) models achievement in STAR classrooms as satisfying this equation:

$$y_{ci} = \alpha_c + (\gamma - 1)\bar{\varepsilon}_c + \varepsilon_{ci}, \tag{3}$$

where  $\alpha_c$  is a class or teacher effect and  $\gamma > 1$  captures social interactions. The residual  $\varepsilon_{ci}$  is a kind of placeholder for individual heterogeneity, but not otherwise specified. Graham (2008)’s narrative imbues (3) with a causal interpretation: “Consider the effect of replacing a low- $\varepsilon$  with high- $\varepsilon$  ... mean achievement increases for purely compositional reasons and ... because ... a high- $\varepsilon$  raises peer quality” (p. 646). I can fit equation (3) as follows: set  $\alpha_c = \bar{y}_c$  and  $\varepsilon_{ci} = y_{ci} - \bar{y}_c$ ; since  $\bar{\varepsilon}_c = 0$ , any  $\gamma$  will do. This works for any sample or data generating process, including random assignment to groups, so the parameters in (3) seem no more useful than my tautological slope,  $\beta$ , in (2).

## 2.1 Control Yourself

Many econometric models of peer effects build on a theoretical framework that includes both individual and group regressors. Townsend (1994), for example, hypothesized that, controlling for household demographic structure, individual household consumption responds to village average consumption in a theoretical relationship generated by risk sharing. Bertrand, Luttmer, and Mullainathan (2000) described spillovers in welfare use that emerge as a result of ethnic networks - these are parameterized as acting through neighborhood and ethnicity group averages, controlling for individual characteristics. With individual covariates included as controls, a regression of  $y$  on group average  $y$  need not produce a coefficient of unity. This methodological improvement notwith-

standing, I'm skeptical that the coefficient on group averages in a multivariate model of endogenous peer effects reveals the action of social forces.

I interpret covariate-controlled endogenous peer relationships here using a model for the population expectation of outcomes conditional on individual and group characteristics. I focus on a specification from Manski (1993), who notes that the following conditional expectation function (CEF) is typical of econometric research on peer effects:

$$E[y|x, z] = \beta\mu_{y|z} + \gamma x. \quad (4)$$

In this model,  $z$  defines groups,  $x$  is an individual covariate, and all variables are mean zero.

A natural first step in the study of (4) is to iterate over  $x$ , and then solve for  $E[y|z]$ . This generates a reduced form relation which can be written,

$$E[y|z] = \frac{\gamma}{1-\beta} E[x|z]. \quad (5)$$

Because  $\beta$  is thought to lie between 0 and 1, and  $\frac{1}{1-\beta}$  scales the effect of individual covariates in (4), the term  $\frac{\gamma}{1-\beta}$  is said to reflect a social multiplier that magnifies the impact of covariate changes. Becker and Murphy (2001, p.14), for example, argued that social multipliers make the effects of changes in group composition large even when "there is only a small response to idiosyncratic (individual) variation." In a recent study of cheating behavior at service academies, Carrell, Malmstrom, and West (2008, p. 193) estimated a version of the endogenous peer effects model where peer cheating in college has a multiplier effect, controlling for whether students cheated in high school (an individual covariate). They describe the multiplier idea as follows: "Hence, in full equilibrium, our models estimate the addition of one college cheater 'creates' roughly three new college cheaters."

I'll return to the social multiplier interpretation of (5) shortly. For now, I note that the regression of average outcomes on average covariates suggested by (5) is algebraically two-stage least squares (2SLS) using group dummies to instrument  $x$ , an estimand I label  $\psi_1$ . Specifically, we have,

$$\psi_1 = \frac{E[y\mu_{x|z}]}{V[\mu_{x|z}]} = \frac{E(E[y|z]E[x|z])}{V[\mu_{x|z}]}, \quad (6)$$

where  $\mu_{x|z}$  is shorthand for  $E[x|z]$ . The first equals sign in (6) comes from the fact that the first stage in this case is  $E[x|z]$ , while the second follows by iterating expectations. Because 2SLS is the same as OLS on group means, we also have that

$$\psi_1 = \frac{\gamma}{1-\beta}. \quad (7)$$

With or without the interpretation of  $\psi_1$  derived from (4), the econometric behavior of the sample

analog of  $\psi_1$  is that of a 2SLS estimator. Evidence for social effects should be evaluated in light of this fact.

Suppose the CEF is indeed as described by (4). This implies that we can write

$$E[xy] = \beta E[x\mu_{y|z}] + \gamma\sigma_x^2. \quad (8)$$

The combination of (8) and (7) facilitate a link between  $\beta$  and  $\gamma$  in (4) and more familiar econometric parameters, specifically,  $\psi_1$  and its OLS counterpart, defined as:

$$\psi_0 = \frac{E[xy]}{\sigma_x^2}. \quad (9)$$

Dividing (8) by  $\sigma_x^2$ , we have

$$\psi_0 = \beta\tau^2\psi_1 + \gamma,$$

where  $\tau^2 = \frac{V[\mu_{x|z}]}{\sigma_x^2}$  denotes the (population) first stage R-squared associated with  $\psi_1$ . Using this and (7), we find

$$\beta = \frac{\psi_1 - \psi_0}{\psi_1} \times \frac{1}{(1 - \tau^2)}. \quad (10)$$

Since  $\tau^2$  is likely to be small, this analysis shows that

$$\frac{1}{1 - \beta} \cong \frac{\psi_1}{\psi_0}. \quad (11)$$

In other words, the social multiplier implied by (4) is approximately the ratio of the 2SLS to OLS estimands for the effect of individual covariates on outcomes. Consequently, any excess of IV over OLS looks like a social multiplier.<sup>1</sup>

In an influential recent discussion of peer effects in social networks, Bramoullé, Djebbari, and Fortin (2009) described models like (4) as posing an identification problem. Again, I see the problem here differently. Just as in the context of the tautological bivariate regression of individual outcomes on group mean outcomes,  $\beta$  in (4) and (11) is identified. My concern is that this parameter captures a mechanical relationship, divorced from any social significance that you might wish for the underlying CEF.

## 2.2 Greek Peers

I illustrate the value of the 2SLS interpretation of econometric peer models by re-examining the Dartmouth College roommates research design pioneered by Sacerdote (2001). This design exploits the fact that, conditional on a few preference variables, Dartmouth College matches freshman room-

---

<sup>1</sup>A similar observation appears in Boozer and Cacciola (2001), who wrote (p. 47): “As long as the Between coefficient ... lies above this [OLS coefficient] ... the estimated peer effect will be non-zero.” In the Boozer-Cacciola setup, the “between coefficient” is the regression of average  $y$  on average  $x$ , which I have characterized as the 2SLS estimand,  $\psi_1$ .

mates randomly. Sacerdote (2001) used this to look at peer effects in academic achievement. In a follow-up analysis, Glaeser, Sacerdote, and Scheinkman (GGS, 2003) used random assignment of roommates to ask whether the propensity of Dartmouth freshman to join fraternities reflects a social multiplier.

In the GSS application, the dependent variable,  $y$ , is an indicator of fraternity (or sorority) membership (about half of Dartmouth College undergraduates go Greek). High school drinking is a strong predictor of pledge behavior; a dummy variable indicating high-school beer drinking is my  $x$ . Finally, peer reference groups, indexed by  $z$ , consist of dormitory rooms, dormitory floors, and dormitory buildings. Each of these grouping schemes creates an increasingly coarse partition of a fixed sample consisting of 1,579 freshmen.

The OLS estimand here consists of a regression of fraternity participation on a dummy for whether students drank in high school. The resulting estimate of  $\psi_0$ , computed in a model that controls for own SAT scores, own high school GPA, and own and family income, appears in column 1 of Table 1 (taken from GSS). This estimate is about 0.10 with a standard error of 0.03, showing that (self-reported) high school drinking is a strong and statistically significant predictor of fraternity participation. The remaining columns of Table 1 report results from regressions that put  $E[y|z]$  on the left hand side and  $E[x|z]$  on the right. These are estimates of  $\psi_1$  using room, floor, and building dummies as an instrument for  $x$  (The regression of individual  $y$  on  $E[x|z]$  is the same as the regression of  $E[y|z]$  on  $E[x|z]$  since the grouping transformation is idempotent.) Because these regressions use grouped data, the resulting standard errors are similar to those that would be generated by 2SLS after clustering individual data on  $z$ .<sup>2</sup>

As can be seen in column 2 of Table 1, the estimate of  $\psi_1$  with data grouped at the room level is 0.098, remarkably similar to the corresponding OLS estimate of  $\psi_0$ . Coarser grouping schemes generate larger estimates: 0.15 with data grouped by floor and 0.23 with data grouped by building. Using (11), the implied social multiplier is about one for dorm rooms, 1.4 for dorm floors, and 2.2 for dorm buildings. GSS interpret these findings as showing that social forces multiply the impact of individual causal effects in large groups.

I believe that the estimates in Table 1 are explained by the finite sample behavior of 2SLS using many or not so many weak instruments. The forces determining the behavior of 2SLS estimates as the number of instruments change are divorced from those determining human behavior. Note first that the instruments driving 2SLS estimates of the parameter I've labelled  $\psi_1$  are - by construction - both many and weak. The instruments are weak because group membership is randomly assigned.

---

<sup>2</sup>A detail here is that the grouped data estimates in Table 1 are unweighted, while 2SLS implicitly weights groups by their size (see, for example, Angrist and Pischke, 2009).



Asymptotically on group size,  $E[x|z] = E[x]$ , and the first stage relationship supporting  $\psi_1$  disappears. The instruments are many because there are many groups: 700 dorm rooms for the estimates in column 2, in particular. This extreme version of a many-weak IV scenario seems likely to produce an estimate close to the corresponding OLS estimate.

GSS observed that estimates of  $\psi_1$  increase as the level of aggregation increases. More important from my point of view, however, is the fact that the standard errors increase sharply as aggregation coarsens: the estimated standard errors in column 4 are five times larger than those in column 2. Moving from dorm rooms to dorm floors and then from dorm floors to dorm buildings increases group size with a fixed overall sample size. The resulting increase in imprecision is what I expect from 2SLS estimates with a collapsing first stage, as are increasingly extreme magnitudes. Given this simple, mechanical explanation for the pattern of estimates reported in Table 1, I'm reluctant to acknowledge a role for elaborate social forces.

### 3 Leave Me Outta This!

In an influential study of risk sharing in Indian villages, Townsend (1994) regressed individual household consumption on the leave-out mean of village average consumption (as one of a number of empirical strategies meant to capture risk sharing). The tautological nature of “y on y-bar” regressions would appear to be mitigated by replacing full group means with leave-out means. In my notation, the model of endogenous peer effects with leave-out means can be written,

$$s_{ij} = \alpha + \beta \bar{s}_{(i)j} + \xi_{ij}, \tag{12}$$

where the leave-out mean is constructed using,

$$\bar{s}_{(i)j} = \frac{N_j \bar{s}_j - s_{ij}}{N_j - 1},$$

for individuals in a group of size  $N_j$ .

In contrast with estimates of (1), estimates of equation (12) are not preordained. In my view, however, these results are also bereft of information about human behavior. Students in the same school and households from the same village are similar in many ways, almost certainly including aspects of their behavior captured by the variable  $s_{ij}$ , be this drug use, achievement, or consumption. A simple model of this correlation allows for a group random effect,  $u_j$ , defined as  $u_j = E[s_{ij}]$  in group  $j$ . Random effects are shorthand for the fact that individuals in the same group are likely to be more similar than individuals in different groups, just by virtue of the fact that they're grouped together. If we live in the same village, for example, we're subject to the same weather.

The random effects notation allows us to model  $s_{ij}$  as,

$$s_{ij} = u_j + \eta_{ij}, \tag{13}$$

where  $E[\eta_{ij}u_j] = 0$ . To see the implication of this for estimates of (12), suppose that group size is fixed at 2 and that  $\eta_{ij}$  is homoskedastic and uncorrelated within groups. Then  $\beta$  is the regression of  $s_{1j}$  on  $s_{2j}$  and vice versa, a coefficient that can be written,

$$\frac{C(s_{1j}, s_{2j})}{V[s_{ij}]} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\eta^2}, \tag{14}$$

where  $\sigma_u^2$  is the variance of the group effects and  $\sigma_\eta^2$  is the variance of what's left over. In a discussion of Townsend's (1994) empirical strategies, Deaton (1990) observed that in a regression of individual consumption on a leave-out mean, any group-level variance component such as described by (13) generates the correlation captured by (14). Risk sharing and other sorts of behavior might contribute to this, but generic clustering makes models like (12) scientifically uninformative.

### Dartmouth Do-Over

Sacerdote (2001) estimated a version of (12) for the freshman grades of Dartmouth College roommates. My version of the roommate achievement analysis appears here in Table 2. The first column shows the coefficient on roommate GPA from a model for 1,589 Dartmouth roommates in 705 rooms. These models include 41 block (preference-group) effects to control for the fact that roommates are matched randomly within blocks. The resulting precisely estimated coefficient of about 0.11 shows that roommates' GPAs are highly correlated.

A useful summary statistic for roommate ability is the SAT reasoning score, computed here as the sum of SAT math and SAT verbal scores (divided by 100). SAT tests are taken in high school, before roommates are matched. As can be seen in column 2 Table 2, roommates' own SAT reasoning score is also a strong predictor of own GPA, with an effect of about the same magnitude as the roommate GPA coefficient, and estimated more precisely. At the same time, roommate's SAT score is unrelated to a student's own GPA, as can be seen in column 3 of Table 2, which reports estimates from a model that predicts each student's GPA using his roommate's as well as his own SAT scores.

A social planner interested in boosting achievement among college freshman can work only with the information he or she has, information like SAT scores that's necessarily collected before freshman year. Because SAT scores strongly predict college grades, aspiring social planners might be tempted to mix and match new students using information on their SAT scores. The estimates

in Table 2 suggest any such manipulation is likely to be of no consequence. Estimates showing a strong correlation in roommate GPAs would seem to be driven by common variance components in outcomes. These are the sort of variance components that motivate empiricists to report clustered standard errors, but not themselves usually seen as a causal force subject to external manipulation.

### Shocking Peer Effects

Causal interpretations of common shocks appear frequently in scientific publications. In a widely discussed study of social networks in the Framingham Heart Study, for example, Christakis and Fowler (2007) report strong correlations in obesity across friends and family, with the strongest correlations for mutual friends. This finding is offered as evidence of social transmission, described in the study as a causal force. In particular, the within-network correlation this study reveals is said to have predictive value for policy (p. 376-377): “Our study suggests that obesity may spread in social networks in a quantifiable and discernible pattern that depends on the nature of social ties ... Consequently, medical and public health interventions might be more cost-effective than initially supposed, since health improvements in one person might spread to others.” In an investigation motivated by the Christakis and Fowler (2007) study, however, Cohen-Cole and Fletcher (2008) find strong within-friend correlations in acne, height, and headaches. The fact that correlation in outcomes like height cannot be explained by transmission across social networks casts doubt on the predictive value of social correlations in health outcomes and health-related behaviors.

You might hope that endogenous network effects can be uncovered by IV. Suppose, for example, that students have better peers when assigned to the honors floor, indicated by  $h_j$ . We might therefore instrument  $\bar{s}_{(i)j}$  with this peer-changing group-level instrument, which is correlated with  $\bar{s}_{(i)j}$  and, I’ll assume, nothing else. This is not a very realistic scenario, as it requires social engineers to know the future. In any case, such awesome power is econometrically misplaced. Booser and Cacciola (2001) show that IV estimation of an equation like (12) produces a coefficient of unity, much like the tautological model I started with.<sup>3</sup> Straightforward regression algebra reveals why this must be so: in this IV setup, where each observation of  $s_{ij}$  provides both an outcome and a treatment, the first stage (regression of roommates’ GPA on  $h_j$ ) and reduced form (regression of own GPA on  $h_j$ ) are the same, since everybody in this data set is somebody’s roommate. Recognizing this difficulty, however, opens the door to more informative strategies that separate research subjects from the peers whose characteristics might influence them. I return to this point in Section 5.

---

<sup>3</sup>Kelejian, Prucha, and Yuzefovich (2006) derive related results.

## 4 Socially Awkward

The theory of human capital externalities suggests that a more educated workforce makes everyone more productive, whether educated or not. Acemoglu and Angrist (2001) therefore asked whether a man’s earnings are affected by the average schooling in his state. Human capital externalities illustrate a class of peer effects where the group average of one variable is presumed to influence individual outcomes that come later. Motivated by the schooling example, I call the effect of an average predetermined variable,  $x$ , on an outcome variable,  $y$ , a social return. Social returns are sometimes said to be contextual. Manski (1993) also calls such effects exogenous peer effects, as opposed to the model of endogenous outcome-on-outcome peer effects meant to be captured by (4).

The typical population social returns CEF looks like this,

$$y = \pi_1 \mu_{x|z} + \pi_0 x + \varepsilon, \tag{15}$$

where  $\pi_1$  is meant to capture the causal effect of changes in average  $x$ . This differs from (4) by swapping  $\mu_{x|z}$  for  $\mu_{y|z}$ . As with (4),  $\pi_1$  and  $\pi_0$  are determined by more fundamental parameters. Specifically, Acemoglu and Angrist (2001) showed that,

$$\pi_0 = \frac{\psi_0 - \psi_1 \tau^2}{1 - \tau^2} = \phi \psi_0 + (1 - \phi) \psi_1 = \psi_1 - \phi(\psi_1 - \psi_0) \tag{16}$$

$$\pi_1 = \frac{\psi_1 - \psi_0}{1 - \tau^2} = \phi(\psi_1 - \psi_0) \tag{17}$$

where  $\psi_0$  and  $\psi_1$  are as defined in (9) and (6),  $\phi = \frac{1}{1 - \tau^2}$ , and  $\tau^2$  is again the first stage R-squared associated with the use of group dummies to instrument the individually-varying covariate,  $x$ . It’s easy to see where (17) comes from: Equation (15) is the regression version of the Hausman (1978) specification test comparing OLS and 2SLS estimates of the effect of  $x$  on  $y$ .

The social returns parameter in a contextual effects model is proportional to the difference between 2SLS and OLS, while in the endogenous effects model, the social multiplier is proportional to the ratio of these two. Either way, however, IV might exceed OLS due to measurement error. As an empirical matter, Ashenfelter and Krueger (1994) find that adjustment for measurement error produces a substantial increase in schooling coefficients. Many other regressors are measured accurately, of course. But “measurement error” here is a metaphor for anything that gets averaged out in grouped data. Perhaps schooling, though accurately measured on its own terms, has group-specific variance components that affect earnings especially strongly.<sup>4</sup>

IV might exceed OLS for other reasons as well. For one thing, selection bias can push IV

---

<sup>4</sup>Moffitt 2001 noted that measurement error complicates the interpretation of estimates of equations like (15).

estimates above or below the corresponding OLS estimates. Card (1995, 2001) and others note the common finding that IV estimates of the returns to schooling tend to exceed the corresponding OLS estimates. Here, the omitted variables bias seems to go the wrong way (though the theory of optimal schooling choice is ambiguous on this point). This finding might also reflect discount rate bias, a scenario first described by Lang (1993), in which those affected by compulsory schooling laws and similar instruments tend to have unusually high returns, leading IV estimates to exceed OLS estimates even when they are uncompromised by selection bias. Nonlinearity may also drive IV estimates away from OLS. Suppose, for example, that the returns to college are below the returns to secondary schooling, as seems true for middle-aged men in the 2000 Census (see Angrist and Chen, 2011). Grouping by state - implicitly instrumenting by state - might produce estimates closer to the average secondary school return than to the average college return.

## 4.1 Social Returns Details

### Models With Controls

Empirical social returns models typically allow for additional controls beyond the individual covariate,  $x$ . Acemoglu and Angrist (2001), for example, control for state and year effects. A version of equation (15) with controls can be written

$$y = \pi_1 \mu_{x|z} + \pi_0 x + \delta' w + \varepsilon \tag{18}$$

where  $w$  is a vector of controls other than  $x$ . At first blush, the introduction of additional controls complicates the interpretation of  $\pi_1$  and  $\pi_0$  since  $\mu_{x|z}$  is no longer the first stage fitted value for a 2SLS model with covariates (as always, the relevant first stage must include the covariates). In Acemoglu and Angrist, however, and probably not untypically, the key covariates are linear combinations of the grouping dummies or instruments,  $z$ . In such cases, my interpretation of the parameters in (18) stands with only minor modification.

To see this, let  $P_w$  and  $P_z$  denote the projection matrices associated with  $w$  and  $z$  and let  $M_w = I - P_w$  be the residual-maker matrix for  $w$ . The scenario I have in mind has  $P_z P_w = P_w$ , in which case it's straightforward to show that

$$M_w P_z x = P_z M_w x.$$

In other words, the order of instrumenting (with  $z$ ) and covariate adjustment (for  $w$ ) can be swapped. From here it's straightforward to show that (16) and (17) apply after dropping  $w$  from (18) and replacing  $x$  by  $\tilde{x} \equiv M_w x$  throughout.

Table 3 reports estimates of a version of equation (18) using the 1950-1990 census extracts used in the Acemoglu and Angrist (2001) study. The average schooling variable in this case is constructed using the same sample of white men in their forties that I used to construct the regression estimates (The Acemoglu and Angrist study used an hours-weighted average for all workers). The covariates here consist of a full set of state and census year effects, so the social returns formulas apply after partialing them out. The estimate of  $\psi_0$  in column 1 of Table 3 comes in at 0.076, while the estimate of  $\psi_1$  in column 2 is larger at 0.105. Because the first-stage R-squared in this case is close to zero, the estimate of  $\pi_1$  in column 3 is the difference between  $\psi_1$  and  $\psi_0$ , at 0.029, a seemingly reasonable magnitude for human capital externalities. Regardless of interpretation, however, we learn from these estimates that 2SLS using state and year dummies as instruments for schooling are (marginally) significantly larger than the corresponding OLS estimates. This result can arise for a variety of reasons. For example, any omitted variables bias (OVB) associated with this 2SLS procedure seems very likely to be positive since states with high average schooling probably have high average wages for other reasons as well. If so, the fact that 2SLS exceeds OLS may be unrelated to human capital externalities.

Equally important, I can tune the findings in Table 3 as I wish: Columns 5-7 report estimates of the social returns CEF after adding noise to the individual highest grade completed variable. The reliability ratio relative to unadulterated schooling is 0.7. The addition of measurement error leaves the estimate of  $\psi_1$  in column 6 largely unchanged, but the estimate of  $\psi_0$  in column 5 is attenuated. Consequently, social returns come in larger, at almost 5 percent, a result with no predictive value for the effects of social policy.<sup>5</sup>

## Back to School Again

Building on Sacerdote's (2001) seminal analysis, columns 4-7 of Table 2 sketch a social returns scenario for Dartmouth roommates. To make sure the social returns algebra applies in detail, I've limited the sample to the 804 roommates living in doubles. My estimates also omit roommate preference block effects, which turn out to matter little in the doubles subsample. In my social returns analysis, freshman GPA plays the role of  $y$ , while the role of  $x$  is played by SAT scores. Just as in the full sample, SAT achievement is a strong predictor of freshman GPA in the doubles sample: every 100 point score gain (about two-thirds of a standard deviation) again boosts GPA by almost 0.11 points. This can be seen in the estimate of  $\psi_0$  shown in column 4 of Table 2.

---

<sup>5</sup>See also Ammermueller and Pischke (2009), who discuss models in which measurement error in peer group composition makes evidence of peer effects harder to uncover.

The regression of individual GPA on room average SAT, the parameter  $\psi_1$  in this context, is 0.09, just under the corresponding estimate of  $\psi_0$ . Because  $\psi_1 < \psi_0$ , estimates of the social returns equation, (18), show negative peer effects. The first-stage R-squared associated with column 5 is surprisingly large, at 0.52, a consequence of the fact that there are half as many instruments in the form of room dummies as there are observations. Using the formula in (17) produces the estimate of  $\pi_1$  found in column 6, in this case,  $-0.042$ .

It's worth asking why 2SLS doesn't exceed OLS in this case, thereby producing an apparent positive peer effect. I believe the answer lies in the many-weak nature of the roommate grouping instruments, much as for the GSS table discussed earlier. Although the first stage R-squared in this case is large, the joint F for 401 room dummies in the first stage is small. With so many small groups - equivalently, many weak instruments - a world without peer effects generates 2SLS estimates with a sampling distribution centered near that of the corresponding OLS estimate. By contrast, the state and year dummy instruments used to construct the estimates of  $\psi_1$  and  $\pi_1$  reported in Table 3 have real predictive value for schooling. As I've noted, however, the strong first stage in the schooling example isn't necessarily an asset, since 2SLS estimates with strong instruments may diverge from the corresponding OLS estimates for reasons unrelated to social returns.

### I've Got Issues

The juxtaposition of peer effects estimates using samples of states and roommates raises two further issues. The first is the importance of using leave-out means in place of full means in social returns models. The sample analog of (15) for roommates can be written

$$g_{ij} = \mu + \pi_1 \bar{s}_j + \pi_0 s_{ij} + \nu_{ij}, \quad (19)$$

where  $g_{ij}$  is the GPA of roommate  $i$  in room  $j$ ,  $s_{ij}$  is his SAT score, and  $\bar{s}_j$  is the room average. Suppose that instead of the full room average, we use the leave-out mean,  $\bar{s}_{(i)j}$ . In a room with two occupants, this is my roommate's score, while with three, this is the average SAT for the other two. The estimating equation becomes

$$g_{ij} = \lambda + \kappa_1 \bar{s}_{(i)j} + \kappa_0 s_{ij} + u_{ij}. \quad (20)$$

Equation (20) seems to resonate more than equation (19) in the context of social spillovers. Perhaps use of the leave-out mean ameliorates social awkwardness of the sort described by (16) and (17).

Substitution of the leave-out mean for the full mean typically matters little, and less and less as

group size increases. Here's the algebra showing this for fixed group sizes:

$$\begin{aligned}
g_{ij} &= \lambda + \kappa_1 \bar{s}_{(i)j} + \kappa_0 s_{ij} + u_{ij} \\
&= \lambda + \kappa_1 \left[ \frac{N \bar{s}_j - s_{ij}}{N-1} \right] + \kappa_0 s_{ij} + u_{ij} \\
&= \lambda + \underbrace{\frac{\kappa_1 N}{N-1}}_{\pi_1} \bar{s}_j + \underbrace{\left[ \kappa_0 - \frac{\kappa_1}{N-1} \right]}_{\pi_0} s_{ij} + u_{ij}
\end{aligned} \tag{21}$$

Estimated social returns differ by a factor of  $\frac{N}{N-1}$  according to whether or not the peer mean is full or leave out. This rescaling is as large as 2 for roommates, but the econometric behavior of social returns equations is similar regardless of group size. Column 7 of Table 2 substantiates this with estimates of (20) for Dartmouth roommates. At  $-0.021$ , the estimate of  $\kappa_1$  is half that of  $\pi_1$ .

A second issue here is the role of the individual control variable in equations like (20). Perhaps the mechanical link between estimates of social returns and the underlying estimates of  $\psi_0$  and  $\psi_1$  can be eliminated by dropping the individual  $s_{ij}$  control in equation (20). After all, when peer groups are formed randomly, we might reasonably expect a bivariate regression linking outcomes with peer means to produce an unbiased estimate of causal peer effects. Setting  $\kappa_0 = 0$  in equation (20) generates a bivariate model that can be written like this,

$$g_{ij} = \alpha + \beta \bar{s}_{(i)j} + v_{ij}. \tag{22}$$

How should we expect estimates of this equation to behave?

Here too, a link with IV is helpful. As noted by Kolesár, Chetty, Friedman, Glaeser, and Imbens (2011), OLS estimates of equation (22) can also be interpreted as a jackknife IV estimator (JIVE; Angrist, Imbens, and Krueger, 1999). JIVE estimates in this case are from regression of  $g_{ij}$  on  $s_{ij}$  using group dummy instruments. If there is an underlying first stage, that is, if groups are formed systematically, we can expect JIVE estimates to behave much like 2SLS estimates when groups are large. The resulting estimates of (22) therefore provide misleading estimates of peer effects, since 2SLS estimates in this case surely reflects the effect of individual  $s_{ij}$  on outcomes in a setting with or without peer effects.

The interpretation of (22) in a no-first-stage or random groups scenario is more subtle. In data with a group structure, the leave-out mean,  $\bar{s}_{(i)j}$ , is likely to be negatively correlated with individual  $s_{ij}$ , regardless of how groups are formed. This correlation strengthens as between-group variation falls, that is, as the first stage implicit in grouping grows weaker. More generally, the regression of



individual data on leave-out means can be written as

$$\theta_{01} = \frac{E[s_{ij}\bar{s}_{(i)j}]}{V[\bar{s}_{(i)j}]} = \frac{\tau^2 - \frac{(1-\tau^2)}{N-1}}{\tau^2 + \frac{(1-\tau^2)}{(N-1)^2}}, \quad (23)$$

where  $\tau^2$  again is the first-stage R-squared associated with grouping, that is,  $\frac{V[\mu_x|z]}{\sigma_x^2}$ . I derive this formula in the appendix.<sup>6</sup> Note that when  $\tau^2 = 0$ ,  $\theta_{01} = -(N-1)$ , in which case individual data and leave-out means are highly negatively correlated. On the other hand, with large groups and a strong first stage,  $\theta_{01} \approx 1$ .

Equations (22) and (20) describe short and long regression models that can be used in conjunction with (23) to understand the behavior of the short. The OVB formula tells us that

$$\beta = \kappa_1 + \kappa_0\theta_{01}, \quad (24)$$

that is, short equals long plus the effect of omitted in long times the regression of omitted on included. Using (24) in combination with the social returns formulas, (16) and (17), we have:

$$\beta = \theta_{01}\psi_1 + (1 - \theta_{01}) \left[ \frac{N-1}{N} \right] \phi(\psi_1 - \psi_0). \quad (25)$$

This confirms that with large groups and a strong first stage,  $\beta \approx \psi_1$  since  $\theta_{01} \approx 1$ . By contrast, in the absence of peer effects, a many-weak IV scenario produces  $\psi_1 \approx \psi_0$ , in which case,

$$\beta \approx \theta_{01}\psi_0 \approx -(N-1)\psi_0,$$

a strong negative effect (assuming  $\psi_0 > 0$ ). To see why the bivariate regression on leave-out means is potentially misleading, consider (22) with only one group, say a single classroom. It would seem there's little to be learned about peer effects from a single classroom, yet the slope coefficient  $\beta$  in (22) is identified and may be estimated precisely if the class is large. In the one-group case, however,  $\tau^2 = 0$  is a constraint of the data, and negative estimates of  $\beta$  a foregone conclusion.

I document the correlation between individual data and leave-out means using the sample of Kenyan first-graders studied by Duflo, Dupas, and Kremer (2011). This study reports on a randomized evaluation of tracking by ability in Kenyan primary schools: in the control group, students were randomly assigned to one of two classes, while in the treatment group, students were grouped by ability using a baseline test score. Along the way, Duflo, Dupas, and Kremer (2011) also looked at classroom peer effects in the control group. My re-analysis of their data is similarly limited to the control sample, which consists of 2,190 students from 61 schools, randomly split into two classes. Outcome data come from a sample of up to 30 students drawn from each class, though many classes

---

<sup>6</sup>See also Boozer and Cacciola (2001) and Guryan, Kroft, and Notowidigdo (2009) for closely related discussions.

are smaller, and 18% of those originally assigned were lost to follow-up.

As a benchmark, I estimated a version of (18) with peer means computed using students in the analysis sample only. The covariates here consist of school effects, which are absorbed by grouping into classes (so my analysis of (18) applies). When group means are constructed using the follow-up sample, the grouping first stage has an R-squared under 0.02. The results, reported in columns 1-4 of Table 4, show  $\psi_0 = 0.496$ ,  $\psi_1 = 0.785$ , and a marginally significant estimate of  $\pi_1$  equal to 0.294. Swapping leave-out means for full class means changes this little, as can be seen in the estimate of  $\kappa_1$  reported in column 4.<sup>7</sup> The original Duflo, Dupas, and Kremer (2011) study computes peer means including students for whom follow-up data is unavailable; the resulting estimate of  $\kappa_1$ , reported in column 5 of Table 4, is 0.359. This differs little from the corresponding estimate in column 4.

As can be seen in column 6 of Table 4, the omission of own-baseline controls reduces the estimated peer mean coefficient to 0.092. Consistent with a low value of  $\tau^2$  and the moderately large  $N$  for peer groups, the regression of own on leave-out means in this case is strongly negative, on the order of  $-0.53$  in a model with school effects. This estimate of  $\theta_{01}$  is reported in column 7 of the table, with an estimated standard error of 0.18. The peer effect necessarily falls here as a result: applying (24), we have that  $0.092 = 0.359 + (0.499 \times -0.534)$ .

The mechanical forces generating a small estimate of  $\beta$  for the Kenya study bring us back to equation (20), with controls for own baseline scores. The principle threat to validity here is divergence between OLS and 2SLS for reasons unrelated to social returns. With a weak grouping first stage such as produced by the Kenya tracking study, we can expect  $\psi_1 \approx \psi_0$  in the absence of peer effects. The fact that  $\psi_1 > \psi_0$  and the consequent large positive estimate of  $\pi_1$  and  $\kappa_1$  in columns 1-4 of Table 4 may therefore signal positive peer effects, though ambiguities remain.

These ambiguities are documented in columns 8-10 of Table 4, which report estimates of (20) in samples stratifying by the quantiles of baseline scores (the original Duflo, Dupas, and Kremer study reported estimates using the same stratification scheme). Positive estimates of  $\kappa_1$  are driven by students in the upper and lower baseline quartiles; there's no apparent peer effect for students with baseline scores in the middle of the distribution. Duflo, Dupas, and Kremer (2011) give a structural interpretation of this result, which they see as generated by complex interactions between students and teachers. Weighing against this causal interpretation, in my view, is the fact that the estimated effect of classmates' baseline scores on outcome scores is much larger than the effect of a student's own baseline score. In column 10, for example, peer means raise achievement twice as much as students' own baseline scores. This suggests some kind of measurement error may be at

---

<sup>7</sup>The scale factor linking  $\pi_1$  and  $\kappa_1$  differs from  $\frac{N}{N-1}$  because group size varies in this application.

work after all, perhaps related to the fact that baseline scores in the study aren't comparable across schools.

## 5 A Little Help for My Friends

In an ambitious and original study of peer effects among freshmen at the United States Air Force Academy (USAFA), Carrell, Sacerdote, and West (2013) explored the consequences of peer group manipulation. They began by estimating econometric peer effects using a version of (18). The outcome here is freshman GPA at USAFA, while peer characteristics include SAT scores and other pre-treatment variables. The results from this initial investigation suggested that groups of students who are predicted to do poorly in their first year at USAFA benefited from exposure to classmates who have high SAT verbal scores. Motivated by this finding, the authors randomly assigned incoming cadets to peer groups whose composition was informed by these estimates. As it turns out, this manipulation had no overall effect, with marginally significant negative estimates for the group the intervention was meant to help. Carrell, Sacerdote, and West (2013) attributed these unexpected results to social stratification within squadrons.

I read these findings as illustrating the proposition that estimates of equations like (18) are unlikely to have predictive value for interventions that change peer groups. My diagnosis identifies the problem as originating in the spurious nature of peer effects estimated using equations like (18), as opposed to endogenous stratification. I'll therefore conclude with a brief discussion of estimation strategies that seem to me most likely to generate evidence on social interactions that has predictive value. Two features strike me as especially important. The first is a clear separation between the *subjects* of a peer effects investigation and the peers who provide the mechanism for causal effects on these subjects. The second is a set-up where fundamental OLS and 2SLS parameters ( $\psi_0$  and  $\psi_1$ , in my notation) can be expected to produce the same result in the absence of peer effects.

Imagine a peer experiment that takes a sample of  $J \times N$  individuals and randomly allocates  $J$  groups of size  $N$  to different peer environments, say neighborhoods. The analyst focuses on the original  $J \times N$  subjects; the peers are a mechanism for causal effects but not themselves subjects for study. By construction, peer characteristics in this design are orthogonal to individual characteristics. As a result we needn't control for the latter, avoiding the mechanical forces at work in estimates of models like (18) and (22), where peers and subjects are treated symmetrically. The design I'm describing fails to capture outcome-on-outcome causal effects of the sort that are sometimes said to reflect social multipliers, but this design captures the causal effects of peer group manipulation

nevertheless.

An important experimental implementation of this design is the randomized evaluation of Moving to Opportunity housing vouchers, analyzed in Kling, Liebman, and Katz (2007). Members of the MTO treatment groups were randomly offered housing vouchers to cover rent for units located in low poverty neighborhoods. Randomized voucher offers were orthogonal to subjects' baseline characteristics. The neighbors' data plays no role in the statistical analysis of MTO, other than to provide descriptive statistics that help to characterize the treatment in terms of average peer characteristics for treatment and control groups. Although social scientists have long documented correlation in the labor market outcomes of citizens and their neighbors, the well-designed MTO intervention uncovered no evidence of causal peer effects.

Observational studies with similar design features include the Angrist and Lang (2004) exploration of the consequences of busing low-income students into suburban schools through a program known as Metco. The analysis sample here is limited to children found in classrooms receiving bused-in peers, omitting the Metco students who produce the change in peer composition. The Angrist and Lang (2004) research design attempts to isolate exogenous variation in the number bused, variation unrelated to Metco-receiving students' characteristics. The Abdulkadiroglu, Angrist, and Pathak (2014) analysis of selective public schools likewise focuses on the effect of exam school offers on quasi-experimental subjects (in this case, exam school applicants). The Duflo, Dupas, and Kremer (2011) tracking study also implements an RD analysis of the tracking treatment group, comparing those who cross the high-ability threshold in tracked schools to those just below.

The MTO, Metco, exam school, and Kenya treatment group analysis can be understood as constructing IV estimates of equations like (22), where constant-within-group manipulation becomes an instrument for ex ante peer characteristics summarized by  $\bar{x}_{(i)j}$ . The instruments are meant to be orthogonal to individual baseline variables, so that own-baseline controls such as found in equation (15) are needless, or at least irrelevant. When successful, these designs eliminate OVB in estimates linking peer characteristics with individual outcomes, including the own/leave-out bias described by equation (24), and the spurious social returns generated by equation (15). Not coincidentally, in my view, these studies also uncover little evidence of peer effects.

In designs that fail to separate subjects from their peers or produce an orthogonal-to-baseline peer group manipulation, we'd like the 2SLS estimates we'd get using group dummies as instruments for ex ante characteristics in a world without peer effects to be close to the corresponding OLS estimates of the effects of these characteristics. As I've noted, random group formation implies a many-weak IV scenario that has this feature. Yet, some amount of group-to-group variation in peer characteristics

is required for any peer effects design to be informative. This raises the question of just how weak is weak enough to avoid bias from divergent 2SLS and OLS estimates under the no-peer-effects null hypothesis. My reanalysis of the Kenya control sample illustrates the ambiguity here, yielding what would seem to be implausibly large peer effects even under random assignment to groups.

A second robust research design for peer effects creates a strong first stage, while ensuring  $\pi_1 = 0$  under the no-peer-effects null. A recent job training study by Crepon, Duflo, Gurgand, Rathelot, and Zamora (2013) uses this approach to study job search assistance in French labor markets. The Crepon experiment randomly assigned treatment proportions,  $p_c$ , from the set  $\{0, 25, 50, 75, 100\}$  to each of 235 local labor markets (cities). Within cities, treatment was randomly assigned at rate  $p_c$  to the population of eligible job seekers. The social returns equation motivated by this design can be written,

$$y_{ic} = \mu + \pi_1 p_c + \pi_0 t_{ic} + v_{ic}, \quad (26)$$

where  $y_{ic}$  is an employment outcomes for individual  $i$  in city  $c$  and  $t_{ic}$  is his treatment status (an offer of job search assistance). Equation (26) is meant to uncover externalities or spillovers engendered by living in a city with many treated workers. If treated workers displace others, these spillovers are negative. As an empirical matter, estimates of (26) indicate substantial negative spillovers for some groups of workers.

As always, the parameters of a social returns model like (26) are determined by the corresponding OLS and 2SLS fundamentals,  $\psi_0$  and  $\psi_1$ . In this case,  $\psi_0$  is the slope coefficient from a regression of  $y_{ic}$  on  $t_{ic}$ , a simple treatment-control contrast, while  $\psi_1$  is the slope coefficient from a regression of  $y_{ic}$  on  $p_c$ . This is what we'd get using dummies for cities to instrument  $t_{ic}$ . Note that  $E[t_{ic}|c] = p_c$ , implying a strong first stage since  $p_c$  has been set to vary across cities, while within cities samples are large. As this is not a many-weak IV scenario, we might expect  $\psi_0 \neq \psi_1$  in a world without peer effects. In this case, however, there's no measurement error, omitted variables bias, nonlinearity, or LATE-type heterogeneity to drive a wedge between 2SLS and OLS estimates for reasons other than peer effects.

To see why this is a robust peer effects research design, let  $Y_{1ic}$  and  $Y_{0ic}$  denote individual potential outcomes indexed against treatment status,  $t_{ic}$ . The observed outcome,  $y_{ic}$ , is

$$y_{ic} = t_{ic}Y_{1ic} + (1 - t_{ic})Y_{0ic}.$$

By virtue of random assignment within cities, we have,

$$\{Y_{1ic}, Y_{0ic}\} \perp\!\!\!\perp t_{ic} | p_c.$$

In other words, potential outcomes are independent of individual treatment status conditional on treatment rates. Consequently, treatment-control comparisons within cities capture the average causal effect of treatment when treatment is at rate  $p_c$ :

$$E[y_{ic}|t_{ic} = 1, p_c] - E[y_{ic}|t_{ic} = 0, p_c] = E[Y_{1ic} - Y_{0ic}|p_c].$$

This comparison is a misleading guide to overall program impact, however, if externalities make  $E[Y_{0ic}|p_c]$  a decreasing function of  $p_c$ . On the other hand, in the absence of externalities, the probability of treatment is also ignorable:

$$\{Y_{1ic}, Y_{0ic}\} \perp\!\!\!\perp t_{ic}, p_c,$$

in which case, we have,

$$\begin{aligned} \psi_0 &= E[y_{ic}|t_{ic} = 1, p_c > 0] - E[y_{ic}|t_{ic} = 0] \\ &= E\{E[Y_{1ic}] - E[Y_{0ic}]\} \\ &= E[Y_{1ic} - Y_{0ic}]. \end{aligned}$$

To evaluate  $\psi_1$ , I begin by noting that 2SLS estimation using dummy instruments produces a weighted average of estimates using the dummies one at a time (see, e.g., Angrist and Pischke (2009)). It's therefore enough to look at a single just-identified dummy-IV estimate, comparing, say, cities with  $p_c = p > 0$  to cities with  $p_c = 0$ . Let  $T_{ic}(p)$  indicate  $i$ 's treatment status when  $p_c$  in his or her city is set to  $p$ . Note that  $T_{ic}(p)$  is defined for all  $p$  for each  $i$  and not just for the realized  $p_c$ . In the Crepon, Duflo, Gurgand, Rathelot, and Zamora (2013) design,  $T_{ic}(p) = t_{ic}$  for all  $p > 0$  and is zero otherwise. The additional notation for latent treatment status is useful nonetheless.

With spillovers, use of a dummy for  $p_c = p$  to instrument for  $t_{ic}$  violates the exclusion restriction. Without spillovers, however, this procedure estimates the local average treatment effect,

$$E[Y_{1ic} - Y_{0ic}|T_{ic}(p) = 1, T_{ic}(0) = 0].$$

Because  $T_{ic}(0) = 0$  for everyone, this is the average treatment effect on the treated in cities with  $p_c = p$ . Formally, we have,

$$\begin{aligned} E[Y_{1ic} - Y_{0ic}|T_{ic}(p) = 1, T_{ic}(0) = 0] \\ = E[Y_{1ic} - Y_{0ic}|t_{ic} = 1, p_c = p]. \end{aligned}$$

Without spillovers, random assignment of  $t_{ic}$  and  $p_c$  makes this the population average treatment effect. Consequently,  $\psi_1$  here is the population average treatment effect. This implies in turn that

$\psi_1 = \psi_0$  under the no-peer-effects null hypothesis.

## 6 Summary

Powerful mechanical and statistical forces link data on individuals with the characteristics of the groups to which they belong. The relationships these forces generate have no behavioral implications and no predictive value for the consequences of peer group manipulation. Because mechanical and statistical artifacts make spurious correlation among individuals and their peers likely, I set a high bar for a causal interpretation of econometrically estimated peer effects. My reading of the body of recent empirical work implementing robust peer effects research designs is that this research has uncovered little in the way of causal effects.

## References

- ABDULKADIROGLU, A., J. ANGRIST, AND P. A. PATHAK (2014): “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools,” .
- ACEMOGLU, D., AND J. ANGRIST (2001): “How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws,” in *NBER Macroeconomics Annual*, ed. by B. S. Bernanke, and K. Rogoff, vol. 15, pp. 9 – 74. MIT Press.
- AMMERMUELLER, A., AND J.-S. PISCHKE (2009): “Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Study,” *The Journal of Labor Economics*, 27 (3), 315–348.
- ANGRIST, J., AND S. H. CHEN (2011): “Schooling and the Vietnam-Era GI Bill: Evidence from the Draft Lottery,” *American Economic Journal: Applied Economics*, 3(2), 96–118.
- ANGRIST, J. D., G. W. IMBENS, AND A. B. KRUEGER (1999): “Jackknife Instrumental Variables Estimation,” *Journal of Applied Econometrics*, 14(1), 57–67.
- ANGRIST, J. D., AND K. LANG (2004): “Does School Integration Generate Peer Effects? Evidence from Boston’s Metco Program,” *American Economic Review*, 94, 1613–1634.
- ANGRIST, J. D., AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- ASHENFELTER, O., AND A. B. KRUEGER (1994): “Estimates of the Economic Returns to Schooling from a New Sample of Twins,” *American Economic Review*, 84(5), 1157–1173.
- BECKER, G. S., AND K. M. MURPHY (2001): *Social Economics: Market Behavior in a Social Environment*. Harvard University Press.
- BERTRAND, M., E. F. LUTTMER, AND S. MULLAINATHAN (2000): “Network Effects and Welfare Cultures,” *The Quarterly Journal of Economics*, 115(3), 1019–1055.
- BOOZER, M., AND S. E. CACCIOLA (2001): “Inside the ‘Black Box’ of Project Star: Estimation of Peer Effects Using Experimental Data,” *Yale Economic Growth Center Discussion Paper No. 832*.
- BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): “Identification of peer effects through social networks,” *Journal of Econometrics*, 150, 41–55.



- BROCK, W. A., AND S. N. DURLAUF (2001): “Discrete Choice with Social Interactions,” *The Review of Economic Studies*, 68(2), 235–260.
- CARD, D. (1995): “Earnings, Schooling, and Ability Revisited,” in *Research in Labor Economics*, ed. by S. Polachek, vol. 14. JAI Press.
- (2001): “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems,” *Econometrica*, 69(5), 1127–1160.
- CARRELL, S. E., F. V. MALMSTROM, AND J. E. WEST (2008): “Peer Effects in Academic Cheating,” *The Journal of Human Resources*, 43(1), 173–207.
- CARRELL, S. E., B. I. SACERDOTE, AND J. E. WEST (2013): “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation,” *Econometrica*, 81(3), 855–882.
- CHRISTAKIS, N. A., AND J. H. FOWLER (2007): “The Spread of Obesity in a Large Social Network over 32 Years,” *New England journal of medicine*, 357(4), 370–379.
- COHEN-COLE, E., AND J. M. FLETCHER (2008): “Detecting Implausible Social Network Effects in Acne, Height, and Headaches: Longitudinal Analysis,” *BMJ: British Medical Journal*, 337.
- CREPON, B., E. DUFLO, M. GURGAND, R. RATHELOT, AND P. ZAMORA (2013): “Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment,” *Quarterly Journal of Economics*, 128(2), 531–580.
- DEATON, A. (1990): “On Risk, Insurance, and Intra-Village Consumption Smoothing,” *Princeton University Working Paper*.
- DUFLO, E., P. DUPAS, AND M. KREMER (2011): “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, 101(5), 1739–1774.
- DURLAUF, S. N., AND H. P. YOUNG (2001): *Social Dynamics*. Brookings Institution Press.
- GLAESER, E. L., B. I. SACERDOTE, AND J. A. SCHEINKMAN (2003): “The Social Multiplier,” *Journal of the European Economic Association*, 1(2-3), 345–353.
- GRAHAM, B. S. (2008): “Identifying Social Interactions Through Conditional Variance Restrictions,” *Econometrica*, 76(3), 643–660.

- GURYAN, J., K. KROFT, AND M. J. NOTOWIDIGDO (2009): “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments,” *American Economic Journal: Applied Economics*, 1(4), 34–68.
- HANUSHEK, E. A., J. F. KAIN, J. M. MARKMAN, AND S. G. RIVKIN (2003): “Does Peer Ability Affect Student Achievement?,” *Journal of Applied Econometrics*, 18(5), 527–544.
- HAUSMAN, J. A. (1978): “Specification Tests in Econometrics,” *Econometrica*, 46(6), 1251–1271.
- IMBENS, G. W., AND J. D. ANGRIST (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62(2)(2), 467–475.
- JACKSON, M. O. (2010): *Social and Economic Networks*. Princeton University Press.
- KELEJIAN, H. H., I. R. PRUCHA, AND Y. YUZEFOVICH (2006): “Estimation Problems in Models with Spatial Weighting Matrices Which Have Blocks of Equal Elements,” *Journal of Regional Science*, 46 (3), 507–515.
- KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 75(1), 83–119.
- KOLEŠÁR, M., R. CHETTY, J. FRIEDMAN, E. GLAESER, AND G. W. IMBENS (2011): “Identification and Inference with Many Invalid Instruments,” *NBER Working Paper No. 17519*.
- LANG, K. (1993): “Ability Bias, Discount Rate Bias and the Return to Education,” MPRA Paper, University Library of Munich, Germany.
- MANSKI, C. F. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *The Review of Economic Studies*, 60(3)(3), 531–542.
- MANSKI, C. F. (2000): “Economic Analysis of Social Interactions,” *Journal of Economic Perspectives*, 14(3), 115–136.
- MOFFITT, R. A. (2001): “Policy Interventions, Low-level Equilibria, and Social Interactions,” in *Social dynamics*, ed. by S. N. Durlauf, and P. H. Young, pp. 45–82. MIT Press.
- SACERDOTE, B. (2001): “Peer Effects With Random Assignment: Results For Dartmouth Roommates,” *The Quarterly Journal of Economics*, 116(2), 681–704.

SACERDOTE, B. (2011): “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, vol. 3. Elsevier, first edn.

SCHEINKMAN, J. A. (2008): “Social Interactions,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, second edn.

TOWNSEND, R. M. (1994): “Risk and Insurance in Village India,” *Econometrica*, 62(3), 539–591.

## Appendix: The Regression of Own on Leave-Out

We're interested in the regression of  $x_{ij}$  on

$$\bar{x}_{(i)j} = \frac{N\bar{x}_j - x_{ij}}{N-1}$$

in  $J$  groups of size  $N$ . In what follows, the total mean of  $x_{ij}$  is set to zero.

To simplify, we first write

$$\bar{x}_{(i)j} = \frac{N\bar{x}_j - x_{ij}}{N-1} = \bar{x}_j - \frac{x_{ij} - \bar{x}_j}{N-1},$$

the difference in two orthogonal pieces. The variance in the denominator is therefore

$$V[\bar{x}_{(i)j}] = E[\bar{x}_j^2] + \frac{E[V_j(x_{ij})]}{(N-1)^2},$$

where  $V_j(x_{ij}) = \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2$ , and  $E[\bar{x}_j^2] = \frac{1}{J} \sum_{j=1}^J \bar{x}_j^2$ . As always, total variance,  $V[x_{ij}]$ , can be written as the sum of between-group variance,  $E[\bar{x}_j^2]$ , and average within-group variance,  $E[V_j(x_{ij})]$ .

That is,

$$V(x_{ij}) = \sum_{j=1}^J \sum_{i=1}^N x_{ij}^2 = E[\bar{x}_j^2] + E[V_j(x_{ij})].$$

With this notation in hand, the numerator simplifies as follows:

$$\begin{aligned} E[x_{ij}\bar{x}_{(i)j}] &= E\left[x_{ij}\left(\bar{x}_j - \frac{x_{ij} - \bar{x}_j}{N-1}\right)\right] \\ &= E[\bar{x}_j^2] - \frac{E[V_j(x_{ij})]}{N-1} \end{aligned}$$

The regression of own on leave-out is therefore

$$\begin{aligned} \theta_{01} &= \frac{1}{V[\bar{x}_{(i)j}]} \times \left\{ E[\bar{x}_j^2] - \frac{E[V_j(x_{ij})]}{N-1} \right\} \\ &= \frac{E[\bar{x}_j^2] - \frac{E[V_j(x_{ij})]}{N-1}}{E[\bar{x}_j^2] + \frac{E[V_j(x_{ij})]}{(N-1)^2}} \end{aligned}$$

Relabeling between and within variance components  $E[\bar{x}_j^2] = \sigma_b^2$ ;  $E[V_j(x_{ij})] = \sigma_w^2$ , and defining  $\tau^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$ , we can write

$$\theta_{01} = \frac{E[x_{ij}\bar{x}_{(i)j}]}{V[\bar{x}_{(i)j}]} = \frac{\sigma_b^2 - \frac{\sigma_w^2}{N-1}}{\sigma_b^2 + \frac{\sigma_w^2}{(N-1)^2}} = \frac{\tau^2 - \frac{(1-\tau^2)}{N-1}}{\tau^2 + \frac{(1-\tau^2)}{(N-1)^2}}.$$

The reverse regression produces,

$$\theta_{10} = \frac{E[x_{ij}\bar{x}_{(i)j}]}{V[x_{ij}]} = \frac{\sigma_b^2 - \frac{\sigma_w^2}{N-1}}{\sigma_b^2 + \sigma_w^2} = \tau^2 - \frac{(1-\tau^2)}{N-1}.$$

Finally, note that  $\tau^2 = \frac{V[\mu_{x|z}]}{\sigma_x^2}$ , the first stage R-squared from a regression of  $x_{ij}$  on a full set of group dummy instruments.

Table 1. Social Multipliers in Fraternity Participation

	(1)	(2)	(3)	(4)
	OLS	Room average	Floor average	Dorm average
Drank beer in high school	0.104 (0.03)	0.098 (0.04)	0.145 (0.08)	0.232 (0.19)
Observations	1579	700	197	57
Average group size	1	2.3	8.0	28

Notes: Adapted from Glaeser, Sacerdote, and Scheinkman (2003). Data are for Dartmouth Freshmen. Roommates and dormmates are randomly assigned as described in Sacerdote (2001). Regressions include math and verbal SAT scores, dummy for male, family income, and high school GPA. SAT scores are from Dartmouth Admissions data. Family income, use of beer, and high school GPA are self-reported on the UCLA Higher Education Research Institute's Survey of Incoming Freshmen. Standard errors in parentheses. Column (1) shows the OLS regression of individual fraternity participation on own use of beer in high school. Columns (2-4) show the results of grouped data regressions at various levels of aggregation. All regressors are averaged.

Table 2. Dartmouth Roommates Redux

	All Rooms			Doubles Only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Roommate GPA	0.111 (0.037)	0.111 (0.036)					
Own SAT Reasoning		0.109 (0.010)	0.109 (0.010)	.110 (.013)		.132 (.011)	.109 (.010)
Room Average SAT Reasoning					.090 (.020)	-.042 (.025)	
Roommate SAT Reasoning			-0.003 (0.010)				-.021 (.012)
First Stage R2					0.52		
Block Effects	x	x	x				

Notes: The sample used to construct the estimates in columns 1-3 includes 1589 Dartmouth roommates in 705 rooms. The sample used to construct the estimates in columns 4-7 includes 804 Dartmouth roommates in 402 rooms. The dependent variable is freshman GPA. Standard errors, clustered on room, are reported in parentheses.

Table 3. Human Capital Externalities

	Reported Schooling			With Reliability 0.7		
	(1)	(2)	(3)	(5)	(6)	(7)
Own Schooling	.076 (.001)		.076 (.001)	.052 (.001)		.052 (.001)
State Average Schooling		.105 (.016)	.029 (.016)		.098 (.016)	.046 (.016)
First Stage R2		.0022			.0015	

Notes: Based on Angrist and Acemoglu (2001). The dependent variable is the log weekly wage. The sample includes 729,695 white men aged 40-49 in the 1950-1990 IPUMS files. Standard errors, clustered on state, are reported in parentheses. All models include state of residence and census year effects.



Table 4. Kenya Leave-Me-Out

	Peer Means Computed in Estimation Sample				Peer Means Computed in Full Sample			By Baseline Percentile		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	25-75 (8)	< 25 (9)	> 75 (10)
Own Baseline	0.496 (0.024)		0.492 (0.025)	0.505 (0.024)	0.499 (0.024)			0.531 (0.057)	0.370 (0.098)	0.480 (0.089)
Class Mean Baseline		0.785 (0.152)	0.294 (0.158)							
Classmates' Baseline (Leave-out) Mean				0.292 (0.151)	0.359 (0.161)	0.092 (0.157)	-0.534 (0.179)	-0.050 (0.246)	0.573 (0.207)	0.966 (0.313)
N	2188	2188	2188	2188	2190	2190	2190	1092	525	573
Dependent Var			Outcome Scores				Baseline Scores		Outcome Scores	

Notes: Estimates computed using the DDK (2011) control sample. The sample includes first graders in 61 schools, with two classes each. The dependent variable is an outcome test score. All models control for school effects. Standard errors, clustered on class, are reported in parentheses. The first stage R2 for column 2 is 0.016. The peer means used for columns 8-10 were computed in the full sample.