Labour Economics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

The perils of peer effects☆

Joshua D. Angrist *

MIT Department of Economics, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

ARTICLE INFO

Article history:
Received 8 January 2014
Received in revised form 22 May 2014
Accepted 22 May 2014
Available online xxxx

JEL classification:
C18
C31
C36
I21
31

Keywords:
Causality
Social returns
Instrumental variables

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 a leave-out mean, 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 for other reasons unrelated to social effects. Research designs that manipulate peer characteristics in a manner unrelated to individual characteristics provide the most compelling evidence on the nature of social spillovers. As an empirical matter, designs of this sort have mostly uncovered little in the way of socially significant causal effects.

1. Introduction

In a regression rite of passage, social scientists around the world link students' 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 positive association between the achievement of students and their classmates or the labor force status of citizens and their neighbors ring true.

I argue here that although correlation among peers is a reliable descriptive fact, the scope for incorrect or misleading attributions of causality in peer analysis is extraordinarily wide. Many others have made this point (see, especially, Deaton, 1990; Manski, 1993; Boozer and Cacciola, 2001; Moffitt, 2001; Hanushek et al., 2003). Nevertheless, I believe there's value in a restatement and synthesis of the many perils of econometrically estimated peer effects. Because both peer analysis and instrumental variables (IV) estimates involve statistical correlations between group means, I find it especially useful to link econometric models of peer effects with the behavior of IV estimators.

The link with IV shows that models which assign a role to group averages in the prediction of individual outcomes should often 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 reveal little about human behavior or what we should expect from policy-induced 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 includes individual covariates, putative peer effects are a test for the

☆ Presented as the Adam Smith Lecture 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. I thank seminar participants at EALE, Maryland, Warwick, and Queens for 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 an anonymous referee, the editor, and many of my peers for helpful discussions and comments, especially Daron Acemoglu, Bryan Graham, Andrea Ichino, Guido Imbens, Patrick Kline, Guido Kuersteiner, Steven Lehrer, Robert Moffitt, Victor Lavy, Parag Pathak, and Rob Townsend. The effects of their interventions were modest, but that's entirely my fault.

* Tel.: +1 857 225 0879.
E-mail address: angrist@mit.edu.

http://dx.doi.org/10.1016/j.labeco.2014.05.008
0927-5371/© 2014 Elsevier B.V. All rights reserved.

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 the corresponding OLS estimates. While peer effects are on the list of causes behind such divergence, they should not usually be at the top of it.

2. Peer theory

Like many in my cohort, I smoked marijuana repeatedly throughout the day 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 my friends and I did, something that worried me at the time. My brother's friends also smoked little. In fact, by the time my brother went 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 youthful story bears econometric investigation. Let $\bar{s}_j$ be the smoke-alotta-dope rate among students attending high school j, that is, the school average of s_{ij}, a dummy for whether student i at school j smokes. Is there a school-level dope-smoking peer effect? It's tempting to explore the peer effects hypothesis by estimating this regression:

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

a model that seems to quantify the essence of my story.

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

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

In fact, the properties of Eq. (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 variable and a corresponding set of regression fitted values for this variable is equal to the variance of the fits, producing the result that covariance over variance equals one.

The tautological nature of the relationship between individual data and group averages is not a story about samples. Let β 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

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

a relation that follows by iterating expectations:

$$\begin{aligned} E\left[y\mu_{y|z}\right] &= E\left(E\left[y|z,\mu_{y|z}\right]\times\mu_{y|z}\right) = E\left(E[y|z]\times\mu_{y|z}\right) \\ &= E\left[\mu^2_{y|z}\right] = V\left[\mu_{y|z}\right]. \end{aligned}$$

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" (italics mine). Manski's extended discussion, however, suggests that the tautological nature of Eq. (2) is a kind of troubling special case, one that can in principle be avoided given sufficient ex ante information on how individuals choose their peer reference groups. In the same spirit, Brock and Durlauf (2001) and Jackson (2010), among others, describe regressions like Eq. (2) as posing an identification problem, one for which we might, with suitable econometric ingenuity, 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 elaborate 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." Elaboration, however, need not produce a coherent causal framework. In a more recent 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 α_c is a class or teacher effect and $\gamma > 1$ captures social interactions. The residual ε_{ci} is a kind of placeholder for individual heterogeneity, but not otherwise specified.

As in many discussions of peer effects, Graham (2008)'s narrative imbues Eq. (3) with a causal interpretation: "Consider the effect of replacing a low-ε with high-ε … mean achievement increases for purely compositional reasons and … because … a high-ε raises peer quality" (p. 646). Graham (2008)'s subsequent discussion introduces covariates that might be causally linked with changes in α_c. On it's own, however, Eq. (3) is a weak foundation for causal inference. I can fit this model perfectly as follows: set α_c equal to the group average, $\bar{y}_c$, and $\varepsilon_{ci} = y_{ci} - \bar{y}_c$. Since $\bar{\varepsilon}_c = 0$ in this specification, any γ will do. My proposal, which identifies α_c with the only conditional mean function that can be constructed given information on individuals and groups and nothing else, satisfies Eq. (3) under any sample design or data generating process, including those with random assignment to groups and groups of differing or even infinite size. Eq. (3) therefore seems no more useful than the tautological relation described by Eq. (2).

2.1. Control yourself

Many econometric models of peer effects build on a theoretical framework that explains behavior as a function of both individual and group characteristic. 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 et al. (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 typically does not produce a coefficient of unity. This feature notwithstanding, I don't believe 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 characteristics and peer group membership. My discussion focuses on a specification from Manski (1993), who notes that the following conditional expectation function (CEF) is typical of econometric research on peer effects:

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

Please cite this article as: Angrist, J.D., The perils of peer effects, Labour Econ. (2014), http://dx.doi.org/10.1016/j.labeco.2014.05.008

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 Eq. (4) is to iterate over x, and then solve for $E[y|z]$. This generates a reduced form relation that can be written,

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

Because β is thought to lie between 0 and 1, so multiplication by $\frac{1}{1-\beta}$ inflates the effect of individual covariates in Eq. (4), the term $\frac{\gamma}{1-\beta}$ is said to reflect a social multiplier that magnifies the impact of changes in individual covariates. 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 et al. (2008, p. 193) estimated a 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 Eq. (5) shortly. For now, I note that the regression of average outcomes on average covariates suggested by Eq. (5) is the two-stage least squares (2SLS) estimand defined by using dummies for all possible groups (values of z) to instrument x. I label this 2SLS estimand ψ_1, which can be written as follows,

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

where $\mu_{x|z}$ is shorthand for $E[x|z]$. The first equals sign in Eq. (6) comes from the fact that first-stage fitted values with dummy instruments in a saturated model are given by the first-stage conditional mean function, $E[x|z]$, while the second follows by iterating expectations. The reduced form conditional mean function, Eq. (5), implies that ψ_1 also satisfies

$$\psi_1 = \frac{\gamma}{1-\beta}, \tag{7}$$

with parameters defined by Eq. (4). With or without the interpretation of ψ_1 derived from Eq. (4), however, the econometric behavior of the sample analog of ψ_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 Eq. (4). This implies that we can write

$$E[xy] = \beta E\left[x\mu_{y|z}\right] + \gamma\sigma_x^2. \tag{8}$$

The combination of Eqs. (8) and (7) produces a link between β and γ in Eq. (4) and more familiar econometric parameters, specifically, ψ_1 and its OLS counterpart, with the latter defined as:

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

Dividing Eq. (8) by σ_x^2, we now 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 ψ_1. Using this and Eq. (7), we find

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

Since τ^2 is likely to be small, this analysis shows that

$$\frac{1}{1-\beta} \cong \frac{\psi_1}{\psi_0}. \tag{11}$$

In other words, the social multiplier implied by Eq. (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.[1]

In an influential recent discussion of peer effects in social networks, Bramoullé et al. (2009) described models like Eq. (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, the parameter β in Eqs. (4) and (11) is identified. But because this parameter describes the difference between OLS and IV, which can be expected to diverge for quotidian reasons, it's value is unlikely to have any social significance.

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 roommates randomly. Sacerdote (2001) used this to look at peer effects in academic achievement. In a follow-up analysis, Glaeser, Sacerdote, and Scheinkman, (GSS; 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 1579 Dartmouth freshmen.

The OLS estimand here come from a regression of fraternity participation on a dummy for whether students drank in high school. The resulting estimate of ψ_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 ψ_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. The population version of this fact is my Eq. (6).) 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.[2]

As can be seen in column 2 of Table 1, the estimate of ψ_1 with data grouped at the room level is 0.098, close to the corresponding OLS

[1] 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, ψ_1.

[2] 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).

Please cite this article as: Angrist, J.D., The perils of peer effects, Labour Econ. (2014), http://dx.doi.org/10.1016/j.labeco.2014.05.008

Table 1
Social multipliers in fraternity participation.

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

Notes: Adapted from Glaeser et al. (2003). Data are for Dartmouth freshmen. Roommates and dormmates are randomly assigned as described in Sacerdote (2001). Regressions control for math and verbal SAT scores, a dummy for males, family income, and high school GPA. Standard errors are reported 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 to produce the estimates in columns 2-4.

estimate, ψ_0, in column 1. Coarser grouping schemes generate larger estimates: 0.15 with data grouped by floor and 0.23 with data grouped by building. Using Eq. (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 and fewer 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 labeled ψ_1 are – by construction – weak. The instruments are weak because group membership is randomly assigned. Asymptotically on group size, the first stage disappears since $E[x|z] = E$. The instruments are many because there are many groups: an extreme of 700 instruments (dorm rooms) for the estimates in column 2, in particular. This version of a many-weak IV scenario seems likely to produce an IV estimate close to the corresponding OLS estimate, a conjecture supported by the results in column 2.

GSS observed that estimates of ψ_1 increase as the level of aggregation increases, arguing that this shows more powerful social forces at work in larger groups. Importantly 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 almost 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 we should expect from 2SLS estimates with a collapsing first stage, as are increasingly extreme magnitudes. This simple, mechanical explanation for the pattern in Table 1 leaves little room for causal peer effects.

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,

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

where the leave-out mean is constructed as,

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

for individuals in a group of size N_j.

In contrast with Eq. (1), estimates of Eq. (12) are not preordained. In my view, however, estimates of Eq. (12) are similarly bereft of information about human behavior. Like students in the same school, households from the same village are similar in many ways, almost certainly including aspects of their behavior captured by the variable s_{ij}, whatever this may be. 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, because they're close in space or time, individuals in the same group are likely to be more similar than individuals in different groups. If we live in the same village at the same time, for example, we're subject to the same weather. Such shared influences are sometimes called "common shocks".

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

$$s_{ij} = u_j + \eta_{ij}, \qquad (13)$$

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

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

where σ_u^2 is the variance of the group effects and σ_η^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 Eq. (13) reflects the intraclass correlation summarized by Eq. (14). Risk sharing and other sorts of behavior might contribute to this, but the likelihood of generic clustering makes models like Eq. (12) scientifically uninformative.

3.1. Dartmouth do-over

Sacerdote (2001) estimated a version of Eq. (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 1589 Dartmouth roommates in 705 rooms. Theses models include 41 block

Table 2
Dartmouth roommates redux.

	All rooms			Doubles only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Roommate GPA	0.111	0.111					
	(0.037)	(0.036)					
Own SAT		0.109	0.109	.110		.132	.109
Reasoning		(0.010)	(0.010)	(.013)		(.011)	(.010)
Room average					.090	−.042	
SAT reasoning					(.020)	(.025)	
Roommate SAT			−0.003				−.021
Reasoning			(0.010)				(.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.

Please cite this article as: Angrist, J.D., The perils of peer effects, Labour Econ. (2014), http://dx.doi.org/10.1016/j.labeco.2014.05.008

(preference-group) effects to control for the fact that roommates are matched randomly only within blocks. The results, a precisely estimated coefficient of about 0.11, shows that roommate GPAs are highly correlated.

A useful summary statistic for roommate ability is the SAT reasoning score, defined here as the sum of SAT math and SAT verbal scores (divided by 100). Importantly, SAT tests are taken in high school, before roommates are matched. As can be seen in column 2 of Table 2, own SAT reasoning scores are a strong predictor of own GPAs, with an effect of about the same magnitude as the roommate GPA coefficient, and estimated more precisely. At the same time, roommates' SAT scores are unrelated to students' own GPAs, a result shown in column 3 of the table, 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 that any such manipulation is likely to be of no consequence. Estimates showing a strong correlation in roommate GPAs would seem to be driven solely by common variance components in outcomes. These variance components motivate empiricists to report clustered standard errors for regression estimates that come from samples with a group structure, but they are not a causal force subject to external manipulation.[3]

3.2. Shocking peer effects

In a widely cited *New England Journal of Medicine* study investigating social networks, one of many related publications on the same topic, Christakis and Fowler (2007) report strong correlations in obesity across friends and family, with the strongest correlations seen for mutual friends. This finding is offered as evidence of social transmission of obesity-related behavior, described in the study as a causal force. In particular, the within-network correlation this study documents 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 variables like height cannot be explained by transmission across social networks casts doubt on the predictive value of social correlations in health and health-related behavior.

Many economists draw causal conclusions from intragroup correlations in dependent variables as well, especially in analyses of teachers and schools. For example, in a recent reexamination of STAR data, augmented with adult earnings outcomes, Chetty et al. (2011) document strong intraclass correlation in achievement and earnings. The classes in this case are the kindergarten classrooms randomly assigned in STAR. The Chetty et al. (2011) study interprets this finding as indicative of a lasting causal kindergarten teacher effect, a characteristic that those in the same kindergarten classroom share. Intraclass correlations emerge with markedly more precision than do the effects of either randomly assigned class size or of observable teacher characteristics like experience. This leads me to believe that the intraclass correlation seen in this context is largely spurious, reflective of the intraclass correlation in outcomes that we should expect to find in most data with a group structure.[4]

You might hope that the common shocks problem can be ameliorated by treating the leave-out mean as endogenous in an IV setup. Returning to Dartmouth for illustration, suppose that some students are (randomly) 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. Boozer and Cacciola (2001) show that IV estimation of an equation like Eq. (12) produces a coefficient of unity, much like the tautological model I started with.[5] It's easy to see why this is so: in this IV setup, where every 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 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.

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 human capital example, I call the effect of an average predetermined variable, x, on an outcome variable, y, a social return. Manski (1993) calls such effects exogenous peer effects, contrasting them with the endogenous outcome-on-outcome peer effects meant to be captured by Eq. (4).

The typical econometric social returns model describes $E[y|x, z]$ like this:

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

where π_1 is meant to capture the causal effect of changes in average x. This CEF differs from Eq. (4) in that it swaps $\mu_{x|z}$ for $\mu_{y|z}$. As with Eq. (4), π_1 and π_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 ψ_0 and ψ_1 are as defined in Eqs. (9) and (6), $\phi = \frac{1}{1-\tau^2}$, and τ^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

[3] Sacerdote (2001) noted but dismissed the absence of a relationship between roommate high school background and college GPA (p. 697): "The effects on GPA from randomly assigned roommate background are modest in size and statistical significance … The correlation in own and roommate outcomes for GPA delivers larger t-statistics and is highly robust to changes in specification. I interpret both findings as supporting the existence of peer effects."

[4] Chetty et al. (2011) find that randomly assigned class size has no detectable effect on earnings (See their Table V). Models for the effect of observable teacher characteristics on earnings show no significant effect of experience measured linearly (57 with a standard error of 38 from Table 6 in a working paper version), and a marginally significant effect of having a teacher with more than 10 years experience (1093 with an estimated standard error of 546, from their Table VI). Classmate's *predicted* scores constructed using demographic characteristics (things like race and free lunch status) are unrelated to earnings (See their Table VI). These marginal-to-insignificant findings for observable teacher and class characteristics contrast with those from, a regression of individual kindergarten scores on classmate's average scores (a peer mean), which generate estimates with a t-statistic close to 30 (.662 with a standard error of .024, from their Table VIII). Regressions of earnings on kindergarten peer mean scores generate a relatively precise effect of 61, with a standard error of 20 (also from Table VIII).

[5] Kelejian et al. (2006) derive related results.

Please cite this article as: Angrist, J.D., The perils of peer effects, Labour Econ. (2014), http://dx.doi.org/10.1016/j.labeco.2014.05.008

to see where Eq. (17) comes from: Eq. (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 coeffficient in (15) is proportional to the difference between 2SLS and OLS estimands, while in the endogenous effects model, the social multiplier is proportional to the ratio of these two. Either way, however, measurement error can cause IV estimates to exceed OLS estimates. As an empirical matter, Ashenfelter and Krueger (1994) find that adjustment for measurement error produces a substantial increase in schooling coefficients. Many regressors are measured accurately, of course, so this finding need not be relevant for the interpretation of all social returns estimates. But "measurement error" here is shorthand for any source of variation that is averaged away in grouped data. Perhaps schooling, though accurately measured on its own terms, has group-specific variance components that affect earnings especially strongly.[6]

IV estimates can exceed OLS estimates for other reasons as well. For one thing, selection bias may push IV 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 variable 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 to schooling, leading IV estimates to exceed OLS estimates even when the latter are not compromised 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 to be 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

4.1.1. 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 Eq. (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. At first blush, the introduction of additional controls complicates the interpretation of π_1 and π_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 includes the covariates). In Acemoglu and Angrist, however, and probably not untypically, the key covariates can be expressed as linear combinations of the grouping dummies or instruments, z. In such cases, my interpretation of the parameters in Eq. (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 $M_w = \mathrm{I} - P_w$ be the residual-maker matrix for w. The scenario I have in mind has $P_z P_w = P_w$ (since $P_z w = w$), 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 Eqs. (16) and (17) apply to the CEF defined by Eq. (18) after dropping w and replacing x by $\tilde{x} \equiv M_w x$ throughout.

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 Acemoglu and Angrist (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 3 reports estimates of a version of Eq. (18) using the 1950–1990 census extracts used in the Acemoglu and Angrist (2001) study. The average schooling variable in this case (an average by state and year) is constructed using the same sample of white men in their forties 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 ψ_0 in column 1 of Table 3 comes in at 0.076, while the estimate of ψ_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 π_1 in column 3 is the difference between ψ_1 and ψ_0. This is 0.029, a seemingly reasonable magnitude for human capital externalities. Yet these estimates merely reveal that 2SLS estimation using state and year dummies as instruments for schooling are (marginally) significantly larger than the corresponding OLS estimates, a finding that can arise for many reasons. States with high average schooling may have high average wages for other reasons as well, in which case state and year instruments fail to satisfy the exclusion restriction required for a causal interpretation of the 2SLS estimates using these variables as instruments. The fact that 2SLS exceeds OLS then reflects a form of omitted variables bias (OVB) in the 2SLS procedure.

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 ψ_1 in column 6 largely unchanged, but the estimate of ψ_0 in column 5 is attenuated. Consequently, the estimated social returns are larger, at almost 5%, a result with no predictive value for the effects of social policy.[7]

4.2. Back to school again

Columns 4–7 of Table 2 sketch a social returns scenario for Dartmouth roommates. To ensure that the social returns algebra discussed below applies in every 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 ψ_0 shown in column 4 of Table 2.

A regression of individual GPA on room average SAT, that is, an estimate of ψ_1 using room dummy instruments, generates a coefficient of 0.09, just under the corresponding estimate of ψ_0. Because $\psi_1 < \psi_0$, estimates of the social returns Eq. (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

[6] Moffitt (2001) was among the first to observe that measurement error complicates the interpretation of estimates of equations like Eq. (15).

[7] 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.

in the form of room dummies as there are observations. Using the this value in Eq. (17) produces the estimate of π_1 found in column 6, in this case, −0.042.

It's worth asking why 2SLS estimates don't exceed OLS estimates in the roommates analysis, thereby producing an apparent positive peer effect as with the investigation of human capital externalities in Table 3. I believe the answer lies in the many-weak nature of roommate grouping instruments, much as for the GSS analysis of social multipliers 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 many small groups – equivalently, many weak instruments – a world without peer effects generates 2SLS estimates that have 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 ψ_1 and π_1 reported in Table 3 have real predictive value for schooling, so that many-weak IV bias is less relevant. As I've noted, however, the strong first stage in the schooling example is not an asset in this case: Table 3 shows how 2SLS estimates with strong instruments can diverge from the corresponding OLS estimates for reasons unrelated to social returns.

4.3. I've got issues

The juxtaposition of peer effect estimates using research designs based on states and roommates raises two further issues. The first is the importance of replacing full means with leave-out means on the right hand side of social returns models. The sample analog of Eq. (15) for roommates can be written

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

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 SAT score. 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 score for the other two. The estimating equation becomes

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

Eq. (20) resonates more than Eq. (19) in the context of social spillovers. Perhaps use of the leave-out mean ameliorates social awkwardness of the sort described by Eqs. (16) and (17).

Substitution of a leave-out mean for the corresponding full mean typically matters little, and less and less as group size increases. For fixed group sizes, we have:

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

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 in Table 2 substantiates this with estimates of Eq. (20) for Dartmouth roommates. At −0.021, the estimate of κ_1 is half that of π_1.

A second issue here is the role of the individual control variable that appears in social return models like Eq. (20). Perhaps the mechanical link between estimates of social returns and the underlying estimates of ψ_0 and ψ_1 can be eliminated by omitting s_{ij} altogether. 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 Eq. (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 et al. (2011), OLS estimates of Eq. (22) can be interpreted as a jackknife IV estimator (JIVE; Angrist et al., 1999). The JIVE estimates in this case use group dummy instruments to capture the effect of g_{ij} on s_{ij}. 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 Eq. (22) will then provide misleading estimates of peer effects, since 2SLS estimates in this case surely reflect the effect of individual s_{ij} on outcomes in a setting with or without social returns.

The interpretation of Eq. (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 coefficient produced by regressing individual data on leave-out means can be written as

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

where τ^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 Appendix A.[8]) 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, $\theta_{01} \approx 1$.

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

$$\beta = \kappa_1 + \kappa_0 \theta_{01}, \tag{24}$$

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

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

This confirms that with large groups and a strong first stage, $\beta \approx \psi_1$ since $\theta_{01} \approx 1$. On the other hand, a many-weak IV scenario with no peer effects produces $\psi_1 \approx \psi_0$, in which case,

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

a substantially negative coefficient (assuming $\psi_0 > 0$).

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

I illustrate the correlation between individual data and leave-out means using the sample of Kenyan first-graders studied by Duflo et al. (2011). This study reports on a randomized evaluation of ability tracking in Kenyan primary schools: in the control group, students were randomly assigned to one of two classes, while in the treatment group,

[8] See Boozer and Cacciola (2001) and Guryan et al. (2009) for similar derivations.

Table 4
Kenya leave-me-out.

	Peer means in estimation sample				Peer means in full sample			By baseline percentile		
								25–75	<25	>75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Own baseline score	0.496		0.492	0.505	0.499			0.53	10.370	0.480
	(0.024)		(0.025)	(0.024)	(0.024)			(0.057)	(0.098)	(0.089)
Class mean baseline score		0.785	0.294							
		(0.152)	(0.158)							
Classmates'score baseline score (The leave-out mean)				0.292	0.359	0.092	−0.534	−0.050	0.573	0.966
				(0.151)	(0.161)	(0.157)	(0.179)	(0.246)	(0.207)	(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 Duflo, Dupas, and Kremer (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.

students were grouped by ability using a baseline test score. The Duflo et al. (2011) study also includes an investigation of classroom peer effects in the control group. My re-analysis of their data is similarly limited to the control sample, which consists of 2190 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 are smaller, and 18% of those originally assigned were lost to follow-up.

As a benchmark, I estimated a version of Eq. (18) with peer means computed using students in the follow-up sample only. The covariates here consist of school effects, which can be expressed as linear combinations of dummies for classes, so my analysis of Eq. (18) applies. When class means are constructed using the follow-up sample, the grouping first stage has an R-squared under 0.02.

The results of estimating Eq. (18) with these data, reported in columns 1–4 of Table 4, show $\psi_0 = 0.496$, $\psi_1 = 0.785$, with a marginally significant estimate of π_1 equal to 0.294. Swapping leave-out means for full class means changes this little, as can be seen in the estimate of κ_1 reported in column 4.[9] The original Duflo et al. (2011) study computes peer means including students for whom follow-up data is unavailable; the resulting estimate of κ_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 τ^2 and the moderately large N for classroom peer groups, the coefficient from a 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 θ_{01} is reported in column 7 of the table, with an estimated standard error around 0.18. The peer effect necessarily falls here as a result: applying Eq. (24), we have that $0.092 = 0.359 + (0.499 \times -0.534)$.

The mechanical forces generating a small estimate of β in Eq. (22) for the Kenya study bring us back to Eq. (20), with controls for students' own baseline scores, may be more useful than models without individual controls. 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 might be generated by random assignment if groups are large enough, we can also 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 π_1 and κ_1 in columns 3 and 4 of Table 4 may therefore signal positive peer effects.

Such a conclusion nevertheless strikes me as premature. My doubts arise from columns 8–10 of Table 4, which report estimates of Eq. (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 κ_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 et al. (2011) offer a structural interpretation of this result, which they see as generated by complex interactions between students and teachers. Weighing against this causal interpretation, however, 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 implausibly large peer coefficient suggests some kind of measurement error may be behind the divergence between OLS and 2SLS estimates in this case, perhaps reflecting the fact that baseline scores in the study aren't comparable across schools.

5. A little help for my friends

In a creative study of peer effects among freshmen at the United States Air Force Academy (USAFA), Carrell et al. (2013) explored the consequences of peer group manipulation. This study begins with econometric peer effects estimated using a version of Eq. (20) that includes demographic covariates. 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 predicted to do poorly in their first year at USAFA benefit from exposure to classmates with high SAT verbal scores. Motivated by these results, the authors randomly assigned incoming cadets to peer groups (squadrons) whose composition was informed by these estimates. As it turns out, random peer group manipulation had no overall effect on grades, with marginally significant negative estimates for the group of (predicted) low achievers that the intervention was meant to help. Carrell et al. (2013) attributed these unexpected results to social stratification within squadrons.

I read these findings as illustrating the proposition that estimates of equations like (20) are unlikely to have predictive value for interventions that change peer groups. The disappointing Carrell et al. (2013) results seem to me more likely to originate in the spurious nature of econometric estimates of peer effects than in endogenous social stratification. Because USAFA peer groups are formed randomly, 2SLS estimates aren't systematically biased in favor of peer effects. However, differences between OLS and 2SLS estimates that generate putative peer effects may be chance findings, driven in part by the imprecision of 2SLS without a real first stage. The non-experimental results that motivate the Carrel, Sacerdote, and West experiment indeed have this flavor: the study emphasizes a large significant estimate for a single group (high SAT verbal scores in the group with low predicted GPA), estimated in a model with 6 peer interactions (see their Table 2). Yet a model with three interactions generates much smaller effects, significant only in the high predicted GPA group and not the low (see the article supplement, Table A.VII). A model with 18 peer interactions

[9] The scale factor linking π_1 and κ_1 differs from $\frac{N}{N-1}$ because group size varies in this study.

(see supplement Table A.II) generate a single statistically significant effect, with the other estimated peer effects negative as often as positive.

My position here naturally provokes the question of how we might generate evidence on social interactions that is likely to have predictive value. To this end, two design features strike me as especially important. The first is clear distinction between the *subjects* of a peer effects investigation on the one hand and the peers who potentially provide the mechanism for causal effects on these subjects on the other. This distinction eliminates mechanical links between own and peer characteristics, making it easier to create or to isolate variation in peer characteristics that is independent of subjects' own characteristics. The second is a set-up where fundamental OLS and 2SLS parameters (ψ_0 and ψ_1, in my notation) can be expected to produce the same result in the absence of peer effects.

Imagine a peer experiment that randomly allocates J groups of size N to varying 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. 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 Eqs. (20) and (22), where peers and subjects are treated symmetrically. This design fails to capture "endogenous" outcome-on-outcome causal effects but it captures the causal effects of peer group manipulation nevertheless.

The randomized evaluation of Moving to Opportunity housing vouchers, analyzed in Kling et al. (2007), fits this mold. Members of the MTO treatment groups were randomly offered housing vouchers to cover rent for units located in low poverty neighborhoods. Randomized voucher offers are unrelated to subjects' baseline characteristics, obviating the need for any controls other than those use to stratify random assignment. Moreover, the neighbors neighbors' data plays no role in the statistical analysis of MTO, other than to provide descriptive statistics that characterize the treatment delivered in terms of average peer characteristics for those who were and were not offered vouchers. Although social scientists have long documented correlation in the labor market outcomes of citizens and their neighbors, the well-designed MTO intervention uncovered little evidence of causal effects on these outcomes (treated subjects reported improved mental health, perhaps a consequence of the opportunity to live in lower-crime neighborhoods).

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 Massachusetts 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 changes in peer composition. The Angrist and Lang (2004) research design attempts to isolate exogenous variation in the number bused that is, variation unrelated to Metco-receiving student characteristics. The Abdulkadiroglu et al. (2014) analysis of selective public schools likewise focuses on the effect of exam school offers on subjects (in this case, exam school applicants), under a manipulation that balances subject characteristics in quasi-experimentally formed treatment and control groups. The Duflo et al. (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 analyses can be understood as constructing IV estimates of equations like Eq. (22), where a constant-within-group treatment variable becomes an instrument for ex ante peer characteristics summarized by $\bar{x}_{(i)j}$. The instruments are orthogonal to individual baseline variables (or at least meant to be), so that own-baseline controls such as found in Eq. (15) are irrelevant. These designs eliminate mechanical sources of bias in estimates linking peer characteristics with individual outcomes, including the own/leave-out bias described by Eq. (24), and the spurious social returns reflected in estimates of equations like (15) or (18). Not coincidentally, in my view, these studies also uncover little evidence of peer effects.

Designs that fail to produce orthogonal-to-baseline peer group variation can reduce threats to validity by ensuring that 2SLS estimates generated using group dummies as instruments for peer composition are likely to be close to the corresponding OLS estimates of the effects of these characteristics in the absence of peer effects. In other words, I look for credible claims that $\psi_0 = \psi_1$ under the no-peer-effects null hypothesis. As I've noted, random group formation with many small groups generates 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 research design to be informative. This raises the question of just how weak is weak enough to avoid bias from divergent 2SLS and OLS estimates for reasons unrelated to peer effects. My reanalysis of the Kenya control sample illustrates the tension here, yielding what would seem to be implausibly large peer effects even under random assignment to groups.

An alternative robust peer effects research design uses random assignment to create a strong first stage for peer characteristics, while still ensuring OLS and IV estimates of own-effects are the same under the no-peer-effects null. A recent job training study by Crepon et al. (2013) exemplifies this approach. This experiment randomly assigned treatment proportions p_c, from the set {0, 25, 50, 75, 100} to each of 235 local labor markets in (French cities). Within cities, treatment was randomly assigned at this rate to the population of eligible job seekers. The social returns equation motivated by this design can be written,

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

where y_{ic} is an employment outcome for individual i in city c and t_{ic} is his treatment status (an offer of job search assistance). Eq. (26) is meant to uncover externalities in cities with many treated workers. If treated workers displace others, these spillovers are negative. As an empirical matter, estimates of Eq. (26) indicate substantial negative spillovers, especially for men.

As always, the parameters of a social returns model like Eq. (26) are determined by the corresponding OLS and 2SLS fundamentals, ψ_0 and ψ_1. In this case, ψ_0 is the slope coefficient from a regression of y_{ic} on t_{ic}, a simple treatment-control contrast, while ψ_1 is the slope coefficient from a regression of y_{ic} on P_c. The latter is what we'd get from 2SLS estimation using dummies for cities to instrument t_{ic}. Since $E[t_{ic}|c] = p_c$ and samples within cities are large, this design has a strong first stage. We might therefore expect $\psi_0 \neq \psi_1$ in a world without peer effects. In this case, however, there's no measurement error, omitted variable bias, nonlinearity, or LATE-type heterogeneity to drive a wedge between 2SLS and OLS estimates for reasons other than peer effects. This point is detailed further in the Appendix A.

6. The social network

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 spurious correlation among individuals and their peers arises so easily, I set a high bar for any causal interpretation of econometrically estimated peer effects.

The likelihood of spurious correlation notwithstanding, growing interest in social networks has lifted the tide of credulous peer research to a new high-water mark. In an elaboration on the simple models discussed here, work by Lee (2007) and a host of more recent studies consider peer effects in social networks, discussing both identification and empirical applications. In my view, this new literature follows the old in confusing technical questions of identification with the more fundamental question of whether any of the effects that might be identified should be seen as causal.

Please cite this article as: Angrist, J.D., The perils of peer effects, Labour Econ. (2014), http://dx.doi.org/10.1016/j.labeco.2014.05.008

To illustrate the risk of inappropriate attributions of causality in research on network effects, I'll borrow notation and a simple example from an influential study by Bramoullé et al. (2009), which models high school friendship networks and their affect on athletic activities. Assume the data are arrayed in column vectors Y and X. The goal of this work is identification of the (scalar) parameters β, γ and δ in equations like this,

$$Y = \beta GY + \gamma X + \delta GX + \varepsilon, \tag{27}$$

where G is an $N \times N$ matrix of known constants that defines the reference population or social network and γ captures own effects of X. For example, G might be defined so that the rows of GY contain leave-out means. The moment restriction that identifies the parameters in Eq. (27) is

$$E[\epsilon|X] = 0. \tag{28}$$

This restriction implies that the CEF, E[Y|X], satisfies

$$E[Y|X] = \beta GE[Y|X] + \gamma X + \delta GX,$$

a model reminiscent of my Eq. (15).

Bramoullé et al. (2009) offer an illustrative simplification of Eq. (27), generated by assuming that individuals are ordered from left to right, standing with friends perhaps in the school gym. Everyone has one friend standing to the left, but alas friendship is not transitive, so that Eq. (27) becomes

$$y_i = \beta y_{i-1} + \gamma x_i + \delta x_{i-1} + \varepsilon_i. \tag{29}$$

I'll motivate this simplified version of the network model with a final story from my high school years. Let y_i be the enthusiasm expressed for basketball played in the school gym, measured in stanines. Let x_i be the height of the basketball hoop in child i's driveway. My father helpfully mounted our home hoop below regulation height in the interest of boosting his boys' self esteem. Of course, in practice, the relationship between basketball hoop height at home (HHH) and enthusiasm for high school basketball (EHSB) need not be causal. We can ensure, however, that

$$E[y_i|x_i] = \gamma_0 + \gamma_1 x_i$$

for some γ_0 and γ_1 by using dummy variables for all possible values of HHH (My father experimented with one foot increments; too low isn't good, either). This saturated model therefore satisfies $E[u_i|x_i] = 0$ for $u_i \equiv y_i - \gamma_0 - \gamma x_i$, regardless of how y_i is generated. Adding to this the assumption that my friend's HHH is unrelated to my EHSB, we also have that $E[u_i|X] = 0$.

Conditional on HHH, my friend and I share a low EHSB. To be concrete, I'll describe our similarity with an AR(1) model,

$$u_i = \alpha u_{i-1} + \varepsilon_i,$$

where ε_i is assumed to be an i.i.d. residual from the regression of u_i on u_{i-1}. Now, we can write:

$$y_i = \gamma_0(1-\alpha) + \alpha y_{i-1} + \gamma x - \alpha\gamma x_{i-1} + \varepsilon_i, \tag{30}$$

where $E[\varepsilon_i|X] = 0$, since ε_i is a linear combination of my own and my friends u_i, and $E[u_i|X] = 0$.

Eq. (30) fits the Bramoullé et al. (2009) template in the form of Eq. (29), yet the parameters here reflect tautological relationships and quotidian correlation in unobservables, in a world otherwise characterized by social indifference. As with the naive regression of outcomes on outcomes in Section 3 and the social return models described in Section 4, here too, I'm provoked to ask why we should attend to the question of whether such models are identified. As evidence belying the predictive value of spurious peer effects continues to mount, I hope that other scholars will increasingly ask this question as well.

Appendix A

A.1. The regression of own on leave-out

Consider 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, I 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 of this regression coefficient is therefore

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

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

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

The numerator of the coefficient of interest simplifies as follows:

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

The regression of own on leave-out is therefore

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

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

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

The reverse regression coefficient changes the denominator to total variance:

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

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.

Please cite this article as: Angrist, J.D., The perils of peer effects, Labour Econ. (2014), http://dx.doi.org/10.1016/j.labeco.2014.05.008

A.2. More on Crepon et al. (2013): Robust peer effects with a strong first stage

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,

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

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 if externalities make $E[Y_{0ic}|p_c]$ a function of p_c. On the other hand, in the absence of externalities, the probability of treatment is also ignorable:

$$\{Y_{1ic}, Y_{0ic}\} \amalg t_{ic}, p_c,$$

in which case, we have,

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

To evaluate ψ_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 $p_c = \mathrm{p} > 0$ to cities with $p_c = 0$. Let $T_{ic}(\mathrm{p})$ indicate $i's$ treatment status when P_c in his or her city is set to p. Note that $T_{ic}(\mathrm{p})$ is defined for all p for each i and not just for the realized p_c. In the Crepon et al. (2013) design, $T_{ic}(\mathrm{p}) = t_{ic}$ 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 since those who live in cities where many are treated are affected even if they are not treated. Without spillovers, however, this 2SLS 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, $\psi_1 = \psi_0$ under the no-peer-effects null hypothesis.

References

Abdulkadiroglu, A., Angrist, J., Pathak, P.A., 2014. The Elite Illusion: Achievement Effects at Boston and New York Exam Schools,.

Acemoglu, D., Angrist, J., 2001. How large are human-capital externalities? Evidence from compulsory-schooling laws. In: Bernanke, B.S., Rogoff, K. (Eds.), NBER Macroeconomics Annual, vol. 15. MIT Press, pp. 9–74.

Ammermueller, A., Pischke, J.-S., 2009. Peer effects in European primary schools: evidence from the progress in international reading study. J. Labor Econ. 27 (3), 315–348.

Angrist, J., Chen, S.H., 2011. Schooling and the Vietnam-era GI bill: Evidence from the draft lottery. Am. Econ. J. Appl. Econ. 3((2), 96–118.

Angrist, J.D., Lang, K., 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. Am. Econ. Rev. 94, 1613–1634.

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press,.

Angrist, J.D., Imbens, G.W., Krueger, A.B.., 1999. Jackknife instrumental variables estimation. J. Appl. Econ. 14 (1), 57–67.

Ashenfelter, O., Krueger, A.B.., 1994. Estimates of the economic returns to schooling from a new sample of twins. Am. Econ. Rev. 84 (5), 1157–1173.

Becker, G.S., Murphy, K.M., 2001. Social Economics: Market Behavior in a Social Environment. Harvard University Press,.

Bertrand, M., Luttmer, E.F., Mullainathan, S., 2000. Network effects and welfare cultures. Q. J. Econ. 115 (3), 1019–1055.

Boozer, M., Cacciola, S.E., 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., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. J. Econ. 150, 41–55.

Brock, W.A., Durlauf, S.N., 2001. Discrete choice with social interactions. Rev. Econ. Stud. 68 (2), 235–260.

Card, D., 1995. Earnings, schooling, and ability revisited. In: Polachek, S. (Ed.), Research in Labor Economics. , 14. JAI Press.

Card, D., 2001. Estimating the return to schooling: progress on some persistent econometric problems. Econometrica 69 (5), 1127–1160.

Carrell, S.E., Malmstrom, F.V., West, J.E., 2008. Peer effects in academic cheating. J. Hum. Resour. 43 (1), 173–207.

Carrell, S.E., Sacerdote, B.I., West, J.E., 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. Econometrica 81 (3), 855–882.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D., Yagan, D., 2011. How does your kindergarten classroom affect your earnings? Evidence from project STAR. Q. J. Econ. 126 (4), 1593–1660.

Christakis, N.A., Fowler, J.H., 2007. The spread of obesity in a large social network over 32 years. N. Engl. J. Med. 357 (4), 370–379.

Cohen-Cole, E., Fletcher, J.M., 2008. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. BMJ 337.

Crepon, B., Duflo, E., Gurgand, M., Rathelot, R., Zamora, P., 2013. Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. Q. J. Econ. 128 (2), 531–580.

Deaton, A., 1990. On Risk, Insurance, and Intra-Village Consumption Smoothing. Princeton University Working Paper,.

Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. Am. Econ. Rev. 101 (5), 1739–1774.

Durlauf, S.N., Young, H.P., 2001. Social Dynamics. Brookings Institution Press,.

Glaeser, E.L., Sacerdote, B.I., Scheinkman, J.A., 2003. The social multiplier. J. Eur. Econ. Assoc. 1 (2-3), 345–353.

Graham, B.S., 2008. Identifying social interactions through conditional variance restrictions. Econometrica 76 (3), 643–660.

Guryan, J., Kroft, K., Notowidigdo, M.J., 2009. Peer effects in the workplace: evidence from random groupings in professional golf tournaments. Am. Econ. J. Appl. Econ. 1 (4), 34–68.

Hanushek, E.A., Kain, J.F., Markman, J.M., Rivkin, S.G., 2003. Does peer ability affect student achievement? J. Appl. Econ. 18 (5), 527–544.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46 (6), 1251–1271.

Jackson, M.O., 2010. Social and Economic Networks. Princeton University Press,.

Kelejian, H.H., Prucha, I.R., Yuzefovich, Y., 2006. Estimation problems in models with spatial weighting matrices which have blocks of equal elements. J. Reg. Sci. 46 (3), 507–515.

Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica 75 (1), 83–119.

Kolesár, M., Chetty, R., Friedman, J., Glaeser, E., Imbens, G.W., 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.

Lee, L.-F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. J. Econ. 140, 333–374.

Manski, C.F., 1993. Identification of Endogenous Social Effects: The reflection Problem. The review of Economic Studies. 60 (3), 531–542.

Moffitt, R.A., 2001. Policy interventions, low-level equilibria, and social interactions. In: Durlauf, S.N., Young, P.H. (Eds.), Social Dynamics. MIT Press, pp. 45–82.

Sacerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. Q. J. Econ. 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: Hanushek, E., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education, First edn. , vol. 3. Elsevier.

Townsend, R.M., 1994. Risk and insurance in village India. Econometrica 62 (3), 539–591.

Please cite this article as: Angrist, J.D., The perils of peer effects, Labour Econ. (2014), http://dx.doi.org/10.1016/j.labeco.2014.05.008