NBER WORKING PAPER SERIES

EXCLUSION BIAS IN THE ESTIMATION OF PEER EFFECTS

Bet Caeyers
Marcel Fafchamps

Working Paper 22565
http://www.nber.org/papers/w22565

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2016

We benefitted from useful comments from Matt Jackson, Arun Chandrasekhar, Steve Durlauf, Emily Breza, Giacomo De Giorgi, Xu Tan, Adam Szeidl, Michael Gechter, Bernard Fortin and Elena Manresa as well as from participants to seminars at Gothenburg University, the Institute for Fiscal Studies, LICOS, UC Davis and the International Food Policy Research Institute, as well as participants to the November 2016 NEUDC conference, the March 2016 RES conference, the May 2016 Stanford conference on Social Networks, the March 2016 CSAE conference on African development, and the April 2016 Laval University conference on Social Networks, and from students at Stanford University, the 2016 PODER Summer School, and the 2016 Nova University Summer School. Finally, we are grateful to the ESRC Centre for Microeconomic Analysis of public policy (CPP) at the Institute for Fiscal Studies for funding parts of this research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Exclusion Bias in the Estimation of Peer Effects
Bet Caeyers and Marcel Fafchamps
NBER Working Paper No. 22565
August 2016
JEL No. C31

ABSTRACT

We formalize a noted [Guryan et al., 2009] but unexplored source of bias in peer effect estimation, arising because people cannot be their own peer. We derive, for linear-in-means models with non-overlapping peer groups, an exact formula of the bias in a test of random peer assignment. We demonstrate that, when estimating endogenous peer effects, the negative exclusion bias dominates the positive reflection bias when the true peer effect is small. We discuss conditions under which exclusion bias is aggravated by adding cluster fixed effects. By imposing restrictions on the error term, we show how to consistently estimate, without the need for instruments, all the structural parameters of an endogenous peer effect model with an arbitrary peer-group or network structure. We show that, under certain conditions, 2SLS do not suffer from exclusion bias. This may explain the counter-intuitive observation that OLS estimates of peer effects are often larger than their 2SLS counterpart.

Bet Caeyers
Institute for Fiscal Studies
London
United Kingdom
bet_c@ifs.org.uk

Marcel Fafchamps
Freeman Spogli Institute
Stanford University
616 Serra Street
Stanford, CA 94305
and NBER
fafchamp@stanford.edu

1 Introduction

It is widely assumed that randomized peer assignment ensures no correlation between the ex ante characteristics of individuals and their peers. As a result, any correlation in outcomes after peer assignment, conditional on common shocks, is attributed to peer effects (e.g. Sacerdote, 2001). However, as Guryan et al. [2009] first pointed out and a few others have since re-emphasized (Wang, 2009; Angrist, 2014; Stevenson, 2015a; Stevenson, 2015b), even with random peer assignment a mechanical negative relationship exists between people's ex ante characteristics and those of their peers. This is because peers are drawn without replacement: individuals cannot be their own peers and therefore each individual is excluded from the pool of its potential peers. It follows that the expected value of the pool of someone's potential peers - and thus any random draw from that pool - is negatively correlated with the characteristics of that individual. Guryan et al. [2009] - through Monte Carlo simulatons - and Angrist [2014] - through basic algebra - illustrate how this correlation yields a downward bias in the ordinary least squares (OLS) estimate of peer effects.The purpose of this paper is to move beyond the basics on the bias and to gain important insights into its properties, causes, consequences and solutions, all of which have largely been ignored to date.

We demonstrate that this bias - henceforth referred to as 'exclusion bias' - can seriously bias point estimates downwards and thus affect inference when estimating peer effects. This negative bias is on top of other well-known sources of (positive) bias such as reflection bias and correlated effects (Manski, 1993; Brock and Durlauf, 2001; Moffitt, 2001). We show: that the negative exclusion bias can be substantial; that it is stronger when peer groups are large relative to the peer selection pool; that it does not disappear asymptotically in large samples; that it can dominate the positive reflection bias when the true peer effect is small; and that it is larger when cluster fixed effects are included in settings where peer group formation is correlated with cluster formation. We offer a number of simple statistical solutions to obtain consistent peer effect estimates and to avoid incorrect inference.

Although exclusion bias is also present in models for which peer selection is non-random, the focus of this paper is on random peer assignment. This is a deliberate choice since the random assignment of peers is typically assumed to yield consistent estimates of peer effects. We show that this is, in general, not true. We evaluate the magnitude of the exclusion bias without conflating it with sources of bias caused by endogenous peer group formation.

This paper contributes to the literature in a number of ways. First, we formalize the simulation results presented in Guryan et al. [2009] and derive a simple, exact formula for the magnitude of the exclusion bias in standard linear-in-means tests of random peer assignment. Unlike top-level expressions of the bias provided for instance in Angrist [2014] - which main objective merely is to introduce the basic idea behind the bias - all formulas we present in this paper are a function of the core parameters driving the bias, namely the size of the peer group and the size of the peer selection pool. As such, these formulas allow the reader to assess the magnitude of the bias in different settings.

Second, we derive, for groups of size two, an exact formula for the combined exclusion and reflection bias in standard peer effect estimation models (note that Guryan et al. [2009] focused on the test of random peer assignment and did not consider the presence of exclusion bias in the actual peer effect estimation models). We show that, while reflection bias tends to inflate peer effects estimates, exclusion bias operates in the opposite direction. We identify conditions under which the exclusion bias dominates the reflection bias and changes the sign of peer effect estimates. Using simulations, we generalize these findings to peer groups of size greater than two.

Third, we identify conditions under which the inclusion of cluster fixed effects can magnify exclusion bias. We show that exclusion bias becomes drastically more severe in models that include selection pool fixed effects. The same holds for cluster fixed effects at levels other than the selection pool, whenever peer group formation is correlated with cluster formation. In such cluster fixed effects models, the bias does not disappear as the sample size tends to infinity.

Fourth, we offer several possible solutions to the exclusion bias. The applicability of each solution depends on the objective of the researcher and the type of data available. Our main focus is on the case where one is interested in the correlation between characteristics of individuals and the equivalent characteristics of their peers. We first discuss how to obtain a correct test of the null hypothesis of zero peer effect or, equivalently, random peer assignment. We start by illustrating the various limitations of the methods suggested by Guryan et al. [2009], Stevenson (Stevenson, 2015b; Stevenson, 2015a), and Wang [2009]. We propose instead to rely either on a method that combines our formula with the clustering of standard errors at the level of the peer selection pool; or, more generally, on a particular type of bootstrapping alternatively called permutation method or randomization inference. We show that these simple methods provide a way of conducting correct inference in spite of the bias. They do not, however, allow to correct point estimates of the peer effect. For this purpose, we show how to consistently estimate all the structural parameters of a general model allowing for a large variety of network structures

– including partly overlapping peer groups or arbitrary network data. This method does not require instruments but relies on the assumption of zero correlated shocks between peers.[1]

Since the assumption of zero correlated effects does not suit all estimation situations, we show the conditions under which two-stage least squares (2SLS) estimation procedures do not suffer from exclusion bias. We note that 2SLS requires the availability of suitable strong instruments of a particular type. Furthermore the method is consistent only in large samples (Bound et al., 1995). The various correction methods that we propose do not require instruments and yield valid inference even in small finite samples. Nonetheless, if the data allow consistent 2SLS estimation, we show that it has the added advantage, under certain conditions, of eliminating the exclusion bias. This property can account for a counter-intuitive yet common finding. Many studies on social interactions obtain 2SLS estimates of endogenous peer effects that are significantly larger than OLS estimates. This is counter-intuitive: OLS estimates ought to be biased upwards due to reflection bias (Manski, 2000), endogenous peer selection, or unobserved correlated effects (Brock and Durlauf, 2001; Moffitt [2001]). This paper provides a new explanation for this finding: the negative exclusion bias that affects OLS disappears when a valid 2SLS estimation is used.

The paper is organized as follows. In Section 2 we start off with exclusion bias in a standard test of random peer assignment. We provide the intuition for the bias, derive an exact formula, discuss its properties, and suggest various methods for correct inference. We conclude Section 2 by discussion the conditions under which the addition of cluster fixed effects aggravate the bias relative to pooled OLS.

Section 3 moves on to a treatment of the exclusion bias in the estimation of endogenous peer effects. In a simple model with peer group size equal to two, we derive exact formulas for the exclusion bias and the reflection bias. These formulas can be used to recover correct peer effect estimates from naive OLS estimates under the assumption of zero correlated effects between peers. We then illustrate how the permutation method can be used to correct p-values, thus allowing for correct inference about the null hypothesis of zero peer effects. Next, we generalize the treatment of the exclusion bias in the estimation of endogenous peer effects, first to peer groups of size greater than two and then to network data. We also discuss possible extensions to more complex settings, such as peer groups or peer selection pools that differ in size. Section 4 illustrates the practical relevance of our

[1]But it allows for correlated shocks within selection pools – e.g., classroom fixed effects in models where all peers are selected within the same classroom.

findings with a simple empirical application derived from the seminal work of Sacerdote [2001]. We also discuss which type of empirical studies are expected to be affected. Not all peer effect studies in the literature are affected by the bias, however. This is discussed in Section 5, which among other things explains the conditions under which 2SLS estimation strategies do not suffer from the bias. Section 6 concludes.

2 Testing random peer assignment

2.1 Intuition

We are interested in the properties of a data generating process in which individual units of observations – which we refer to as 'individuals' – are assigned a number of peers. In this Section we focus on non-overlapping, mutually exclusive peer groups because this is the most relevant case of random peer assignment in practice (e.g., assignment to a room, a class, a neighborhood). It is nonetheless possible to extend our results to networks in general. We revisit this point in Section 3.3 when we generalize our findings to a generalized network model.

Here we imagine that a researcher has data on peer assignment and wishes to test whether assignment is random based on an observable pre-treatment characteristic x_{ikl}, where i indexes individuals, k indexes (peer) groups, and l indexes the pool or cluster from which i's peers are selected. One such example is the study of Dartmouth college freshmen by Sacerdote [2001], who exploits the random allocation of students to roommates to study peer effects. In that study, i denotes an individual student, l is the pool to which the student is assigned based on her stated housing preferences ('block'), and k is the room within pool l to which the student is randomly assigned.

Peer effect studies that rely on random peer assignment typically start off by testing whether peer assignment is random. The purpose of this test is akin to testing the 'balancedness' of random assignment to treatment: it verifies that baseline characteristic x_{ikl} of individual i is not correlated with the average characteristic of its peers (excluding individual i), which we denote $\bar{x}_{-i,k,l}$. Specifically, the researcher estimates:

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \epsilon_{ikl} \tag{1}$$

This is, for example, the test for random assignment reported in Sacerdote [2001]. In that application, model (1) regresses freshman i's pre-treatment test score (e.g. SAT math) on the average pre-treatment test score of his/her roommates. In the case of stratified randomization, cluster dummies δ_l are typically added to control for the sub-level at which randomization is carried out – e.g., block dummies in the case of Sacerdote [2001].

Researchers typically proceed as if random assignment of peers implies that the estimate of the coefficient β_1 in regression (1) should be 0. As initially argued by Guryan et al. [2009], this is incorrect: a mechanical *negative* relationship exists between i's characteristics and those of i's peers prior to treatment. The intuition is as follows. Randomly allocating people to peer groups of size K is like randomly drawing $K-1$ peers for each individual. Given that individuals cannot be their own peers, they are excluded from the urn from which peers are drawn. This implies that the mean characteristic of an individual's peers selection pool is negatively correlated with the characteristic of the individual herself. As an example, consider the context of Sacerdote [2001]: If a student has a higher than average ability relative to other students in the sample, then excluding her from her pool of potential peers lowers the average ability of the remaining pool. Vice versa, if the student has a lower than average ability, then excluding her from the urn from which her peers are drawn yields a pool with an average ability that is higher than the overall sample mean. This mechanism leads to a negative correlation between an individual's characteristic and the average characteristic of their peers. Using Monte Carlo simulations, Guryan et al. [2009] illustrate how this mechanical correlation yields a downward bias in the OLS estimate of β_1. They also show that this bias is decreasing in the size of the pool from which peers are drawn.

In this section, we formalize the bias in regression (1). We call this bias the 'exclusion bias' given that it is driven by a systematic exclusion of an individual from her peer group. We derive an exact formula for this bias and we discuss the features of the data generating process behind it. We then propose a general method for testing random peer assignment.

2.2 Formulas

Let us assume that we have a population Ω of N individuals. Each individual $i \in \Omega$ is randomly assigned to a group of K individuals. Let $P_i \subseteq \Omega$ be the pool of people from which individual i's $(K-1)$ peers are randomly drawn. The pool P_i can be the entire sampled network (e.g., the entire grade population in the school), i.e. $P_i = \Omega$. Alternatively, each pool can be a subset of the network, i.e. $P_i = l \subset \Omega$ (e.g., a classroom). The latter is the case, for instance, in Sacerdote [2001], Glaeser et al. [2003], Zimmerman [2003] and Duflo and Saez [2011], where the population of interest consists of students of multiple schools and students are randomly assigned to a dormitory or to a work group *within* each school.

Throughout this paper, we make the natural assumption that individual i is excluded from her own pool, that is, $i \notin P_i$. This is equivalent to assuming that individuals are drawn from the pool without replacement. This implies that i is excluded from being her own peer. This feature is what causes a bias.

In the first sub-section below, we consider the case where $P_i = \Omega$, i.e., when peers are drawn from the entire sampled population, with no stratification. In the sub-section that follows, we introduce stratification into selection pools. In the latter case, we follow best practice and add pool fixed effects to regression model (1). To keep the exposition simple, we assume that the size of each pool P_i takes the same value N_P for all $i \in \Omega$. Similarly, we assume that all peer groups are of the same size K.[2]

2.2.1 Without stratification

We first consider the simple case when peers are randomized at the level of the sample population, i.e. $P_i = \Omega$ and $N_P = N$. Regression (1) simplifies to:

$$x_{ik} = \beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik} \tag{2}$$

We formally show that, even with random peer assignment, $\bar{x}_{-i,k}$ is correlated with the error term ϵ_{ik}. To see this, let us expand $\bar{x}_{-i,k}$ in equation (2). Since each individual i is randomly assigned to a peer group k of size K, i's peer set consists of a random sub-set of $(K-1)$ individuals selected from the population. Let $\bar{x}_{-i}$ be the average characteristic of the pool of $N-1$ individuals, i.e., omitting i. Given random peer assignment, $E(\bar{x}_{-i,k}) = \bar{x}_{-i}$. But the actual draw $\bar{x}_{-i,k}$ deviates from $\bar{x}_{-i}$ by a random component u_{ik}:

$$\bar{x}_{-i,k} = \bar{x}_{-i} + u_{ik} \tag{3}$$

with $E(u_{ik}) = 0$. Inserting equation (3) into equation (2), we obtain:

$$x_{ik} = \beta_0 + \beta_1 (\bar{x}_{-i} + u_{ik}) + \epsilon_{ik} \tag{4}$$

where, under random peer assignment, we have $E(u) = E(\epsilon) = 0$, $var(u) = \sigma_u^2$, $var(\epsilon) = \sigma_\epsilon^2$ and $cov(u, \epsilon) = 0$.

There is a close relationship between σ_u^2 and σ_ϵ^2 since i's peers are drawn from the same population as i. As demonstrated in Appendix A, for a model without stratification

[2] Section 3.4 discusses how the results extend to the more general case where selection pools or peer groups differ in size.

(i.e., $N_P = N$) this relationship is:

$$\sigma_u^2 = \frac{(N-K)}{(N-1)(K-1)}\sigma_\epsilon^2 \tag{5}$$

which indicates that σ_u^2 is decreasing in K – the size of the peer group – and it is increasing in N – the size of the pool from which peers are drawn in the case of non-stratified randomization.

We note that $\bar{x}_{-i}$ is nothing but:

$$\bar{x}_{-i} = \frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} x_{js}\right] - x_{ik}}{N-1} \tag{6}$$

Equation (4) can thus be rewritten as:

$$x_{ik} = \beta_0 + \beta_1 \left(\frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} x_{js}\right] - x_{ik}}{N-1} + u_{ik} \right) + \epsilon_{ik} \tag{7}$$

The presence of dependent variable x_{ik} on the right-hand side of equation (7) is what leads the OLS estimate of β_1 to be biased downwards. To derive this correlation more formally, insert equation (4) into equation (6) to obtain a reduced form for $\bar{x}_{-i}$:

$$\bar{x}_{-i} = \frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} x_{js}\right] - \beta_0}{N-1+\beta_1} - \frac{\beta_1 u_{ik}}{N-1+\beta_1} - \frac{\epsilon_{ik}}{N-1+\beta_1} \tag{8}$$

Under random peer assignment (i.e. $\beta_1 = 0$), this equation reduces to:

$$\bar{x}_{-i} = \frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} x_{js}\right]}{N-1} - \frac{\epsilon_{ik}}{N-1} \tag{9}$$

Comparing equation (4) to equation (9) it is immediately apparent that $cov(\bar{x}_{-i,k}, \epsilon_{ik}) \neq 0$ in equation (2) – even though, under random peer assignment, $cov(u, \epsilon) = 0$. It follows that OLS estimation of equation (2) leads to a biased estimate of β_1:

$$cov(\bar{x}_{-i,k}, \epsilon_{ik}) = cov(\bar{x}_{-i} + u_{ik}, \epsilon_{ik}) = cov(\bar{x}_{-i}, \epsilon_{ik}) = \frac{-\sigma_\epsilon^2}{N-1} < 0 \tag{10}$$

Using equation (10) together with the expression for $var(\bar{x}_{-i,k})$ derived in Appendix B, we obtain a formula for the magnitude of the bias in a test of random peer assignment without stratification when the true $\beta_1 = 0$:

$$
\begin{aligned}
E(\hat{\beta}_1^{OLS}) &= \frac{cov(\bar{x}_{-i,k}, \epsilon_{ik})}{var(\bar{x}_{-i,k})} \\
&= \frac{\frac{-\sigma_\epsilon^2}{(N-1)}}{\frac{(N-1)(N-K)+(K-1)}{(N-1)^2(K-1)}\sigma_\epsilon^2} \\
&= -\frac{(N-1)(K-1)}{(N-1)(N-K)+(K-1)} \qquad (11)
\end{aligned}
$$

2.2.2 With stratification

We now generalize this finding to allow for the case when the total population Ω is stratified into distinct pools from within which peers are selected. This arises, for instance, when students in a school are first divided into classes, and then assigned a peer group within their class. In such cases, testing for random peer assignment should control for pool fixed effects. This is the case we consider in this sub-section.

Suppose that population Ω is divided into $\frac{N}{L}$ pools of equal size L, indexed by l. We continue to assume that individuals are assigned to peer groups of size K, but now the pool P_i from which i's $(K-1)$ peers are drawn is a subset l of Ω of size $N_P = L$. Testing random peer assignment is achieved by estimating regression (1), which we reproduce here:

$$
x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \epsilon_{ikl} \qquad (12)
$$

Appendix C shows how the pool FE estimate $\hat{\beta}_1^{FE}$ in (12) is biased downwards according to the following expression:

$$
E(\hat{\beta}_1^{FE}) = -\frac{(L-1)(K-1)}{(L-1)(L-K)+(K-1)} \qquad (13)
$$

Equation (13) is similar to equation (11), except that the magnitude of the exclusion bias now depends on L instead of N. In other words, the exclusion bias in a typical test of random peer assignment depends on the size N_P of the pool from which peers are drawn. This result highlights that the exclusion bias is not a small sample property: when the size of each selection pool is fixed, the magnitude of the bias does not decrease as N tends to infinity.

2.2.3 Proposition 1

Proposition 1 summarizes the results of the previous two sections.

1 *Suppose all individuals in a sampled population Ω of N individuals are randomly allocated to groups of K peers from within a cluster $\mathit{\Pi} \subseteq \Omega$ of size N_P, where N_P is the size of the pool from which peers are drawn. Suppose a particular test for random peer assignment is of the following form*

$$\begin{cases} x_{ik} = \beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik} & if\ \Pi = \Omega\ ,\ N_P = N \\ x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \epsilon_{ikl} & if\ \Pi = l \subset \Omega\ ,\ N_P = L \end{cases}$$

where i indexes individuals, k indexes (peer) groups, l indexes selection pools, and δ is a set of relevant pool dummies. Then the OLS estimate of β_1 is downward biased according to the following expression:

$$E(\hat{\beta}_1) = -\frac{(N_P - 1)(K - 1)}{(N_P - 1)(N_P - K) + (K - 1)} \tag{14}$$

where N_P is the size of the pool from which peers are drawn. Specifically, $N_P = N$ if peers are selected from the entire sampled network, or $N_P = L$ if peers are selected at the level of the selection pool l.

Proposition 1 demonstrates that the magnitude of the exclusion bias in the test of random peer assignment depends on two key parameters: the size of the peer group K; and the size N_P of the pool from which each individual's $(K - 1)$ peers are drawn. Specifically we have:

1. $\frac{\triangle |bias|}{\triangle N_P} < 0$: For a given peer group size K, exclusion bias is less severe in datasets with a larger pool N_P of potential peers. This result is consistent with the simulation results reported in Guryan et al. [2009].[3]

[3] Formally, as shown in equation (5), the larger N_P is, the larger is the variance of u. From equation (3) we know that, when σ_u^2 is large, more variation in $\bar{x}_{-i,k}$ is explained by the random component u_i rather than by the mean of the pool of potential peers, $\bar{x}_{-i}$. Intuitively, we recall that the exclusion bias is driven by the fact that one is excluded from the pool from which one's peers are drawn. And because of that, the average of that pool will go up or down depending on whether one is a low or high ability student. As a result, the larger the number of students in the pool, the less sensitive the average of the pool will be to an exclusion of oneself and therefore the less severe the bias will be.

2. $\frac{\triangle|bias|}{\triangle K} > 0$: *Ceteris paribus*, exclusion bias is more severe in datasets with larger peer groups. The magnitude of the bias is not linear in K. If it were, it would indicate that the bias *per additional peer* remains constant. In contrast, Proposition 1 shows that the bias per peer increases with the total number of peers.[4] ,[5]

Exclusion bias is conceptually different from the attenuation bias associated with classical measurement error (CME). First, exclusion bias is not driven by measurement error – i.e., it arises even in the absence of measurement error. Secondly, exclusion bias behaves very differently from CME. Classical measurement bias is multiplicative in β_1. Consequently, its sign and magnitude depends on the true β_1. In particular, CME does not exist if the true $\beta_1 = 0$. In contrast, exclusion bias in the test of random peer assignment is additive instead of multiplicative: it is always negative, and it does not disappear when the true β_1 is zero.

2.3 Simulation results and implications for inference

To illustrate the magnitude of the exclusion bias and its implications for inference in a typical test of random peer assignment, we conduct a set of Monte Carlo simulations. The results presented in Table 1 focus on a setting where peers are randomly assigned and thus the true $\beta_1 = 0$. Simulations vary in pool size L and peer group size K. We assume throughout an integer number L/K of groups within each pool. For each simulation we generate 1000 observations and we randomly assign a normally distributed i.i.d. characteristic $x_i \sim N(1,1)$ to each of them. We then randomly assign the N individuals to $\frac{N}{L}$ pools of L persons each and we introduce a pool specific shock to x_{li} to allow for differences across pools.[6] Finally, within each pool we randomly allocate individuals to peer groups of equal size K.

For each dataset generated in this fashion, we estimate the pool fixed effect regression (1). We repeat this process 100 times for each vector of K and L. Given that our data generation process randomly assigns individuals to peer groups, the true β_1 is equal to

[4]From Proposition 1 we see that the bias per peer is $\frac{N_P}{N_P(N_P+1-K)+(K-1)}$

[5]Formally, as shown above in equation (3), as K becomes larger, the variance σ_u^2 becomes smaller. From equation (4) we know that, as σ_u^2 becomes smaller, less of the variation in $\bar{x}_{-i,k}$ is explained by the random component u_i rather than being governed by the mean of the pool of potential peers $\bar{x}_{-i}$. This explains the finding. Intuitively, the larger the number of peers you draw from the pool, the more similar your peer group will look like the pool itself, and the more closely the average of the peer group will follow the average of the pool from which you draw the peers rather than being governed by some random noise. And therefore, the more severe it will be affected by the exclusion bias effect on the expected value of the pool of potential peers.

[6]Note that the introduction of a pool specific shock is not required for our results to hold.

Table 1: Simulation results - exclusion bias in the test for random peer assignment

		$L = 20$	$L = 50$	$L = 100$
$K = 2$	Predicted E($\hat{\beta}_1$)	-0.06	-0.02	-0.01
	Average $\hat{\beta}_1^s$	-0.05	-0.01	-0.01
	% rejected at 1% level	19%	10%	6%
	% rejected at 5% level	36%	20%	19%
	% rejected at 10% level	44%	26%	27%
$K = 5$	Predicted E($\hat{\beta}_1$)	-0.26	-0.09	-0.04
	Average $\hat{\beta}_1^s$	-0.25	-0.09	-0.05
	% rejected at 1% level	71%	21%	8%
	% rejected at 5% level	85%	37%	26%
	% rejected at 10% level	87%	46%	33%
$K = 10$	Predicted E($\hat{\beta}_1$)	-0.86	-0.22	-0.10
	Average $\hat{\beta}_1^s$	-0.84	-0.23	-0.13
	% rejected at 1% level	95%	35%	16%
	% rejected at 5% level	98%	47%	33%
	% rejected at 10% level	99%	58%	45%

Note: $\beta_1 = 0$; $N = 1000$; Pool fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions.

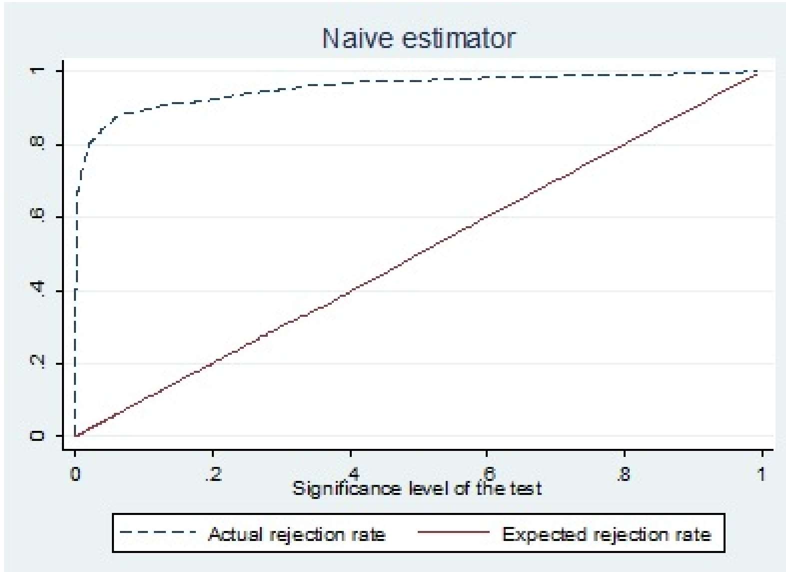
zero. The average of estimated $\widehat{\beta}_1$'s over 100 replications is summarized in Table 1 for different values of K and L. For purpose of comparison, we also report the theoretical $E(\hat{\beta}_1)$ summarized in Proposition 1. These results confirm our theoretical predictions: the exclusion bias is large in magnitude, increases in K, and decreases in L.

Table 1 indicates, for each parameter vector, the proportion of artificially generated sample for which we reject the (true) null hypothesis at the 1%, 5% and 10% significance levels. To illustrate the implication for inference graphically using one particular case as an example ($N = 1000$, $L = 20$ and $K = 5$), we plot in Figure 1 the rate at which naive OLS rejects the null hypothesis at various significance levels and we compare this rate to the rejection rate that we would expect if the test was unbiased (i.e. the 45 degree line). It is clear that ignoring exclusion bias has a dramatic effect on inference: a researcher relying on regression (1) to test random peer assignment erroneously rejects the null in a substantial proportion of cases.

2.4 Inference correction

We now discuss some simple methods that can be used to obtain correct inference when testing random peer assignment, i.e, the null hypothesis of $\beta_1 = 0$. It is important to

Figure 1: Illustration implications for inference - N = 1000, L = 20, K = 5



Note: Pool fixed effects added in all regressions; Simulations under the null ($\beta_1 = 0$) over 100 Monte Carlo repetitions

recognize that none of these methods is able to provide a correct point estimate of β_1. For that, the reader has to wait for Section 3 where we offer a method that yields a consistent estimate of β_1. Other methods are discussed in Section 5.

2.4.1 GKN method

To correct for exclusion bias in a test of random peer assignment, Guryan, Kroft, and Notowidigdo [2009] propose to control for differences in mean characteristic across selection pools. To this effect, they suggest adding to equation (1) the mean characteristic $\bar{x}_{-i,l}$ of individuals other than i in selection pool l. We denote this method by the GKN method. The estimating equation they propose is the following:

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \varphi \bar{x}_{-i,l} + \epsilon_{ikl} \tag{15}$$

where φ is an additional parameter to be estimated.

To see how, under specific conditions, this effectively deals with exclusion bias when the true $\beta_1 = 0$, we substitute equation (15) in for equation (3) and rearrange as follows:

$$\begin{aligned} x_{ikl} &= \beta_0 + \beta_1 \overline{x}_{-i,k,l} + \delta_l + \varphi \bar{x}_{-i,l} + \epsilon_{ikl} \\ &= \beta_0 + \beta_1 (\overline{x}_{-i,l} + u_{ikl}) + \delta_l + \varphi \bar{x}_{-i,l} + \epsilon_{ikl} \\ &= \beta_0 + (\beta_1 + \varphi) \overline{x}_{-i,l} + \beta_1 u_{ikl} + \delta_l + \epsilon_{ikl} \end{aligned}$$

We see that the inclusion of the proxy variable $\overline{x}_{-i,l}$ soaks up the non-random component of $\overline{x}_{-i,k,l}$. As a result, if the true $\beta_1 = 0$ (and therefore in the absence of reflection bias), the coefficient estimate $\hat{\beta}_1$ measures the partial effect of the random component u_{ikl}. Since $E(u_{ikl}\epsilon_{ikl}) = 0$ under the assumption of random peer selection, $E(\hat{\beta}_1) = \beta_1$ and OLS yields consistent estimates of the peer effect β_1.

This method has some limitations, however. First, as already noted by Guryan et al. [2009], parameters β_1 and φ are separately identified only if there is variation in pool size. If every selection pool has the same number of individuals L, then $x_{ikl} = L\,\bar{x}_l - (L-1)\bar{x}_{-i,l}$ and the model is unidentified. Secondly, even when there is some variation in L across pools, this variation may be limited, leading to multicollinearity and quasi-underidentification of β_1 and φ. Thirdly, this method requires precise knowledge of each selection pool. Such knowledge may be not available, especially when peers form arbitrary social networks, a point we revisit later in this paper.

2.4.2 Joint F-test

Wang [2009] has suggested an alternative test of random peer assignment. It involves running an F-test of joint significance of peer group dummies in a model of the form:

$$x_{ikl} = \beta_0 + \beta_1 C_k + \delta_l + \epsilon_{ikl}$$

where C_k is a set of group dummies (excluding a base category). The authors argue that, if individuals are randomly assigned to groups, then all group means should be statistically similar and therefore the coefficients included in vector β_1 should jointly not be significantly different from zero. This method has recently been criticized by Stevenson [2015a] who argues, based on simulation results, that the method fails to reject the null hypothesis if peers are negatively correlated.

2.4.3 Split-sample method

Stevenson [2015b] and Stevenson [2015a] propose a 'split-sample' method which, as the term suggests, involves splitting the original sample to break the mechanical negative correlation introduced by exclusion bias. The approach recognizes the fact that exclusion bias manifests itself if and only if (i) individuals are excluded from their own peer groups *and* (ii) if they are included in the peer groups of other individuals in the sample. If each individual in the study sample only appears on one side of the peer effect estimation equation, then there is no problem.

The split-sample method exploits this feature, as follows:

1. In the first step the researcher randomly selects one observation from each peer group in the original dataset;

2. Next the researcher calculates the average outcome of the peers of those individuals selected in Step 1, excluding the selected individuals themselves;

3. Finally, the researcher regresses the outcomes of the sub-sample of the individuals selected in Step 1 on the average peer group outcomes constructed in Step 2.

Hence, the method effectively creates a dataset - derived from the original dataset of the study - where (i) individuals are excluded from their own peer group but where (ii) they are also excluded from the peer groups of other individuals in the sample. This breaks the source of the exclusion bias.

One obvious downside of this approach is the large loss of efficiency that results from the reduction in sample size. The efficiency of the split sample can be improved by performing multiple iterations but this is cumbersome, especially with large datasets.

2.4.4 Transformed model estimation

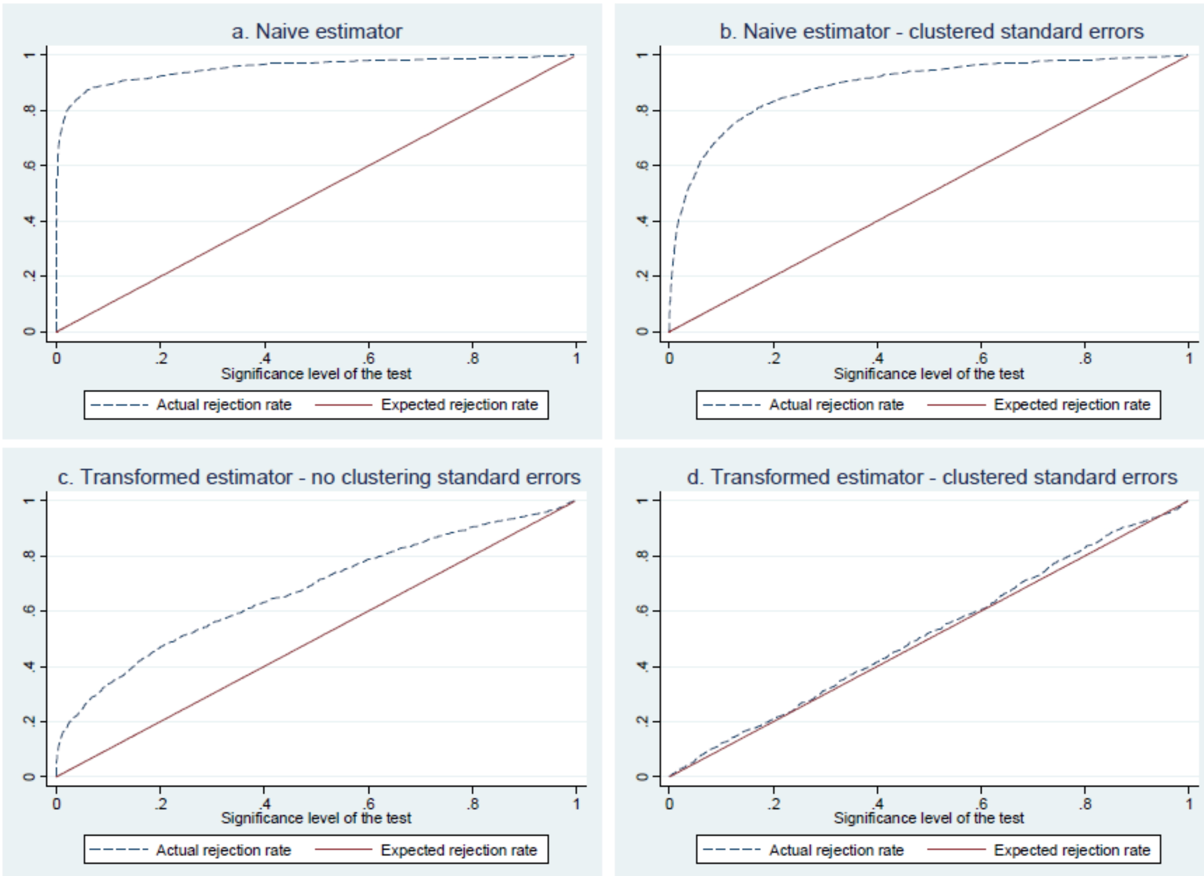
We propose a new simpler approach that draws on Proposition 1. In particular, we use the bias formula 13 to derive a transformation of equation 12 that can easily be shown to yield a consistent point estimate of the true β_1 under the null:

$$\widetilde{x}_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \epsilon_{ikl} \tag{16}$$

where $\widetilde{x}_{ikl} = x_{ikl} - bias * \bar{x}_{-i,k,l}$.

In practice, the transformed model above can be estimated by estimating regression (16) instead of regression (1). Testing the null hypothesis of random peer assignment then proceeds in the usual fashion, using OLS reported standard errors clustered at the pool level. Clustering is important. As illustrated by simulation results presented in Figure 2,

Figure 2: Inference when $N = 1000$, $L = 20$, $K = 5$



Note: Pool fixed effects added in all regressions; Simulations under the null ($\beta_1 = 0$) over 100 Monte Carlo repetitions

transforming the model to correct the point estimate for bias under the null is necessary but not sufficient to obtain unbiased inference. Only when standard errors are clustered at the pool level do we obtain correct inference.

2.4.5 Permutation method

The above method requires to be able to calculate the bias using formula 13. It is not applicable in cases where an algebraic formula cannot be computed – e.g., for network data or partially overlapping groups. In such cases the permutation method, initially suggested by Fisher (1925) and applied to network data by Krackhardt, 1988, can be used to test random peer assignment. The object of this method is to simulate, using the data at hand, the distribution of $\widehat{\beta}_1$ under the null hypothesis – here, random peer assignment.[7]

Table 2: Illustration permutation method

i	k	l	x_{ikl}	$\widetilde{x}_{ikl}$
1	1	1	x_{111}	x_{211}
2	1	1	x_{211}	x_{521}
3	2	1	x_{321}	x_{111}
4	2	1	x_{421}	x_{321}
5	2	1	x_{521}	x_{421}
6	3	2	x_{632}	x_{842}
7	3	2	x_{732}	x_{632}
8	4	2	x_{842}	x_{942}
9	4	2	x_{942}	x_{1052}
10	5	2	x_{1052}	x_{732}

To illustrate how this method works in the case of mutually exclusive groups, imagine the researcher has observational data x_{ikl} partitioned in groups of varying size K_i coming from pools of varying size N_{Pi}. This is illustrated in Table 2 for the simple case of a fixed pool size. The object is to test that individuals are randomly assigned to groups within pools using regression (1). We can simulate the distribution of $\widehat{\beta}_1$ under the null hypothesis of random assignment by reordering individuals within pools and reassigning them into groups.

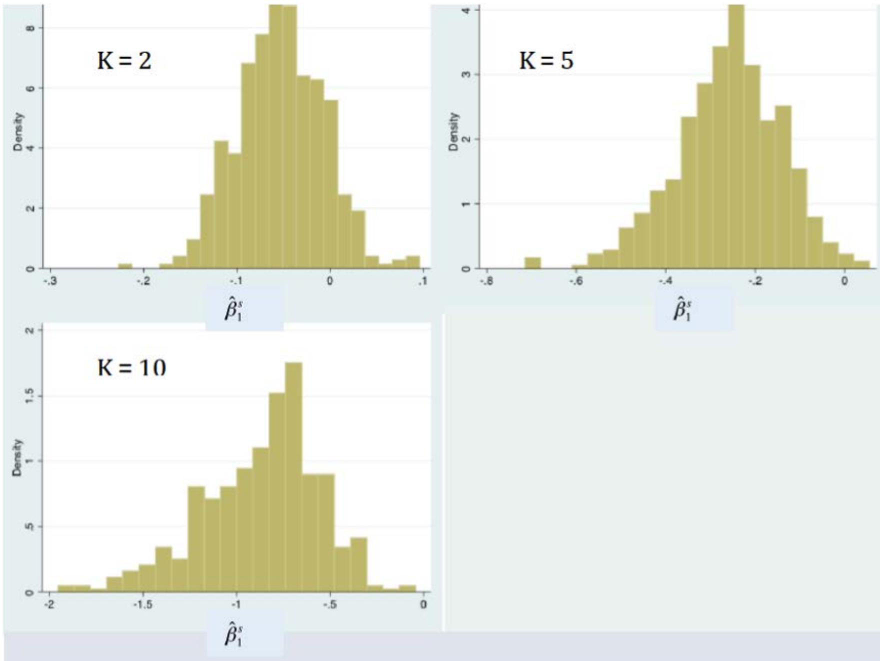
Formally, let us denote the vector of group sizes in pool l as $K^l \equiv [K_1^l, K_2^l, ..., K_m^l]$ where m is the number of groups in pool l. Let observations be sorted by group within each pool, as shown in the x_{ikl} column of Table 2. To mimic random assignment within pools, we create a new variable $\widetilde{x}_{ikl}$ that is obtained by reordering x_{ikl} at random within pools, as shown in column $\widetilde{x}_{ikl}$ of Table 2. We then estimate regression (1) using $\widetilde{x}_{ikl}$ in lieu of x_{ikl} – and $\widetilde{\bar{x}}_{-i,k} = \sum_{j \neq i, j \in k} \widetilde{x}_{j,k,l}$ in lieu of $\bar{x}_{-i,k}$.

By repeating this process many times, we can trace the distribution of $\widehat{\beta}_1$ in the data if the null hypothesis is true. Each repetition yields a separate estimate of $\widehat{\beta}_1^s$. The mean of the distribution of $\widehat{\beta}_1^s$ is an estimate of the bias under the null. More importantly, the empirical distribution of $\widehat{\beta}_1^s$ over the simulated samples can be used to obtain a corrected p-value for the test that $\beta_1 = 0$. This is accomplished in the same way as in other bootstrapping procedures, e.g., by taking the proportion of $\widehat{\beta}_1^s$ that are above the

[7]Simulations can also be used to obtain a close approximation of the distribution of $\widehat{\beta}_1$ (and thus of the bias) under more complicated random assignment processes, e.g., multi-level stratification, and the like.

absolute value of $\widehat{\beta}_1$ or below minus the absolute value of $\widehat{\beta}_1$.[8]

Figure 3: Histogram $\hat{\beta}_1^s$ under null ($N = 1000$, $L = 20$)



Note: Pool fixed effects added in all regressions; Simulations over 100 Monte Carlo repetitions.

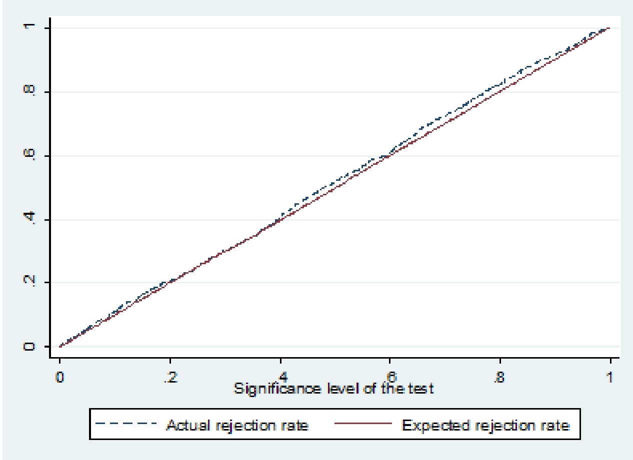
To illustrate the performance of this procedure, we generate three artificial samples of 1000 observations under the null hypothesis of random assignment, one for each three values of $K = \{2, 5, 10\}$. As before, we set the size of each pool $L = 20$ and we posit $\epsilon_{ik} \sim N(1, 1)$. Figure 3 shows the distribution of 100 simulated $\widehat{\beta}_1^s$ under the null hypothesis that $\beta_1 = 0$, for different peer group sizes. The histograms are clearly centered around the naive estimates for β_1 under the null (shown in Table 1), not around the true $\beta_1 = 0$. The permutation method corrects p-values by taking this distributional shift into consideration when calculating the likelihood of observing the naive $\hat{\beta}_1$ under the null rather than relying on OLS-reported standard errors. Figure 4 shows how the permutation method successfully corrects inference for a particular case where $N = 1000$, $L = 20$ and $K = 5$.

2.5 Implication of adding cluster fixed effects

In this section we discuss more in detail the implication of adding cluster fixed effects (FE) for the magnitude of the exclusion bias. Here, clusters refer to any partition of the sample

[8]Note that the simulated distribution of $\widehat{\beta}_1^s$ need not be symmetric.

Figure 4: Inference using permutation method when $N = 1000$, $L = 20$, $K = 5$



Note: Pool fixed effects added in all regressions; Simulations under the null ($\beta_1 = 0$) over 100 Monte Carlo repetitions.

population Ω into mutually exclusive sets. So far we have focused on the case when fixed effects are added at the level of the peer selection pool. In that case, assignment to peer groups happens within each cluster l. Including pool FEs is a reasonable approach when testing random peer assignment, since pool FEs control for any pool-specific characteristic on which randomization is conditioned. We have already discussed how the inclusion of pool FEs affects exclusion bias.

Studies of endogenous peer effects - to which we turn later - often include FEs at levels other than the selection pool. For instance, common shocks can generate a positive correlation in outcomes even in the absence of peer effects. FEs may be included to deal with common shocks introduced after random assignment. Since common shocks need not occur at the level of the selection pool, the estimated model often includes cluster FEs other than for selection pools. For instance, students in a school cohort may be randomly allocated to rooms – in which case the selection pool is the school cohort. But the researcher adds dormitory fixed effects to control for possible shocks common to those in the same dormitory. In other cases, FEs are added at a higher level than the selection pool, or are not included at all, for instance because the selection pool is not clearly defined in the data.

We compare two estimators: $\hat{\beta}_1^{POLS}$ which is obtained by estimating equation (2) using pooled OLS; and $\hat{\beta}_1^{FE}$ which is obtained by estimating equation (12) with cluster fixed effects δ_l added at the level of the cluster $l \subset \Omega$. We consider two possibilities: either (i) peers are selected at the level of the entire population Ω; or (ii) peers are selected within clusters $l \subset \Omega$. For illustration purposes, we focus on the case where all clusters are of the same size L. In Appendix D we derive the following proposition:

2 *When peers are selected among the entire population* Ω*:*

$$E(\hat{\beta}_1^{FE}) = E(\hat{\beta}_1^{POLS})$$

When peer group formation occurs at the cluster level $l \subset \Omega$*:*

$$E(\hat{\beta}_1^{FE}) < E(\hat{\beta}_1^{POLS})$$

The intuition is as follows. The pooled OLS regression estimate is a weighted average of the FE estimate and the between group estimate. The FE pool estimator essentially estimates the extent to which variation in outcomes within a pool is explained by variation in average peer group outcomes within that same pool. Such variation picks up exclusion bias, for the reasons we explained earlier: individuals with a higher-than-average expected outcome are matched with peers that have a lower expected outcome, and vice versa. The between group estimate, on the other hand, is not affected by exclusion bias - it essentially measures the extent to which the variation in a pool average outcome is correlated with the variation in a pool's average peer group outcome. As long as peers are drawn from within the pool, the average outcome of a pool is the same as the average of peer group outcomes within the pool. This yields a strong positive correlation within the two.[9] Given that the pooled OLS estimate is a weighted average of the negatively biased *within* estimate and the positive *between* estimate, the negative exclusion bias is reduced by the positive between-pool correlation.[10]

Table 3 demonstrates the effect of adding cluster FE relative to pooled OLS (POLS) in the case where peers are selected at the level of the cluster l of size $L = 20$. A corrolary of the above proposition is that, keeping the size of each selection pool constant, $\lim_{N\to\infty} E(\hat{\beta}_1^{FE})$ is a negative constant that does not vary with N. The value of this constant is given by formula (14). In contrast, it can be shown that $\lim_{N\to\infty} E(\hat{\beta}_1^{OLS}) = 0$. The simulation results in Table 3 clearly illustrate this corrolary. This implies that

[9]When all peer groups are of equal size, this correlation is perfect.

[10]For the sake of illustration, imagine the following simple example: People are put in pairs from a selection pool of size two. Imagine that within each pool of two the difference in outcomes is always 1, that is, if one person has outcome=100, then the other has outcome=99 or 101. But mean outcomes differ across pools. If we estimate a pooled OLS regression, variation in mean outcome across pools introduces a strong positive correlation. But if we include pool FEs, the correlation becomes -1 since when we demean outcomes, one observation is always -0.5 while the other is +0.5, and hence they are perfectly negatively correlated.

omitting cluster FEs when testing random assignment leads to an asymptotic elimination of the exclusion bias when the null hypothesis of random peer assignment is true.

Table 3: Simulation results - Proposition 2 ($L = 20$)

	N = 500		N = 1000		N = 2000	
	(1)	(2)	(3)	(4)	(5)	(6)
	FE	POLS	FE	POLS	FE	POLS
K = 2	-0.05	-0.00	-0.05	-0.00	-0.06	-0.00
K = 5	-0.28	-0.03	-0.26	-0.02	-0.26	-0.01
K = 10	-0.95	-.07	-0.85	-0.04	-0.86	-0.02

Note: $\beta_1 = 0$; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions

3 Estimating endogenous peer effects

In this section we allow the true β_1 to be different from 0 and we illustrate how exclusion bias and reflection bias interact to jointly affect the estimation of endogenous peer effects. In addition to the inference correction methods discussed in Section 2.4, we also discuss methods that can be used to correct point estimates for both exclusion bias and reflection bias.

To make this illustration as clear as possible, we start by focusing on one pool of potential peers (thus eliminating the need for pool fixed effects in the presentation of the model) and we abstract from exogenous peer effects. We generalize the model below. The linear-in-means peer effects model that we seek to estimate has the following form:

$$y_{ik} = \beta_0 + \beta_1 \bar{y}_{-i,k} + \epsilon_{ik} \tag{17}$$

If the true $\beta_1 = 0$ the results would be exactly the same as those derived in the previous section and summarized by Proposition 1. Here, we focus on the case where the true $\beta_1 \neq 0$.

We begin with a simple example in which group size $K = 2$. For this example, the exact value of the reflection bias and exclusion bias can be derived algebraically if we assume away unobserved common shocks and correlated effects within peer groups. We then generalize the approach to an arbitrary group size and show how non-linear least squares can be used to obtain an estimate of β_1 that is free of both reflection and exclusion bias.

3.1 Simple model with group size $K = 2$

3.1.1 Reflection bias

We start by ignoring exclusion bias so as to derive a precise formula of the reflection bias in our model. This will allow us to distinguish the exclusion bias from the reflection bias later on. We focus on a special case in which there are no correlated effects. Ignoring exclusion bias, this assumption implies i.i.d. errors. We thus have $E[\epsilon_{ik}] = 0$, $E[\epsilon_{ik}^2] = \sigma_\epsilon^2$, $E[\epsilon_{ik}\epsilon_{jm}] = 0$ for all $i \neq j$ and all $k \neq m$, and $E[\epsilon_{ik}\epsilon_{jk}] = 0$ for all k and all $i \neq j$. The $E[\epsilon_{ik}\epsilon_{jm}] = 0$ equality assumes away correlated effects across groups.[11] The $E[\epsilon_{ik}\epsilon_{jk}] = 0$ equality is far from innocuous since it assumes away the presence of what Manski (1993) calls correlated effects, that is, correlated ϵ_{ik} between individuals belonging to the same peer group. With this assumption, any correlation in outcomes between members of the same peer group is interpreted as evidence of endogenous peer effects. This assumption can thus be used for identification purposes, in spite of the well-known existence of a reflection bias.

To show this formally, we consider a system of equations similar to that of Moffit (2001). We ignore control variables, contextual effects and cluster fixed effects, to make the demonstration easier to follow. In Section 5.1 we discuss an extension of this model to include other explanatory variables. For now we have, in each group:

$$
\begin{aligned}
y_1 &= \alpha + \beta y_2 + \epsilon_1 \\
y_2 &= \alpha + \beta y_1 + \epsilon_2
\end{aligned}
$$

where $0 < \beta < 1, E[\epsilon_1] = E[\epsilon_2] = 0$ and $E[\epsilon^2] = \sigma_\epsilon^2$. Solving this system of simultaneous linear equations yields the following reduced forms:

$$
\begin{aligned}
y_1 &= \frac{\alpha(1+\beta)}{1-\beta^2} + \frac{\epsilon_1 + \beta\epsilon_2}{1-\beta^2} \\
y_2 &= \frac{\alpha(1+\beta)}{1-\beta^2} + \frac{\epsilon_2 + \beta\epsilon_1}{1-\beta^2}
\end{aligned}
$$

which shows that y_1 and y_2 are correlated even if ϵ_1 and ϵ_2 are not. None of the ϵ's from other groups enter this pair of equation since we have assumed no spillovers across groups.

[11] This model can be generalized to allow for correlated effects at the cluster level, in which case cluster fixed effects can be added to the model. The same reasoning then applies to the de-meaned model.

We have:

$$E[y_1] = E[y_2] = \frac{\alpha(1+\beta)}{1-\beta^2} \equiv \overline{y}$$

If ϵ_1 and ϵ_2 are independent from each other, $E[\epsilon_1\epsilon_2] = 0$ and we can write:

$$\begin{aligned} E[(y_1-\overline{y})^2] &= E\left[\left(\frac{\epsilon_1+\beta\epsilon_2}{1-\beta^2}\right)^2\right] \\ &= \sigma_\epsilon^2 \frac{1+\beta^2}{(1-\beta^2)^2} \end{aligned}$$

where we have used the fact that $E[\epsilon_1\epsilon_2] = 0$. The latter assumption will be relaxed in the next section, when we introduce exclusion bias. For now, the covariance between y_1 and y_2 is given by:

$$\begin{aligned} E[(y_1-\overline{y})(y_2-\overline{y})] &= E\left[\left(\frac{\epsilon_1+\beta\epsilon_2}{1-\beta^2}\right)\left(\frac{\epsilon_2+\beta\epsilon_1}{1-\beta^2}\right)\right] \\ &= \frac{2\beta\sigma_\epsilon^2}{(1-\beta^2)^2} \end{aligned}$$

where we have again used the assumption that $E[\epsilon_1\epsilon_2] = 0$. The correlation coefficient r between y_1 and y_2 is thus:

$$\begin{aligned} r &= \frac{E[(y_1-\overline{y})(y_2-\overline{y})]}{E[(y_1-\overline{y})^2]} \\ &= \frac{2\beta}{1+\beta^2} \end{aligned} \tag{18}$$

We can now illustrate the magnitude of the reflection bias on its own. Suppose that we regress y_1 on y_2, i.e., we estimate a model of the form:

$$y_1 = a + by_2 + v_1 \tag{19}$$

Since equation 19 is univariate, we have:

$$\widehat{b} = \widehat{r}\frac{\sigma_{y_1}}{\sigma_{y_2}} = \widehat{r} \text{ since } \sigma_{y_1} = \sigma_{y_2}$$

Hence we have:

$$E[\widehat{b}] = \frac{2\beta}{1+\beta^2} \neq \beta \tag{20}$$

This expression gives a closed-form solution for the reflection bias in this simple example where we have ignored exclusion bias. We first note that, based on this formula, $E[\widehat{b}] = 0$ iff $\beta = 0$. This means we can in principe test whether $\beta = 0$ by testing whether $\widehat{b} = 0$ in regression (19).

Formula (20) forms a quadratic equation that can be solved to recover an estimate of β from the naive $\widehat{b}$. We get:[12]

$$\widehat{\beta} = \frac{1 - \sqrt{1 - \widehat{b}^2}}{\widehat{b}}$$

This demonstrates that, in this simple example, identification can be achieved solely from the assumption of independence of ϵ_1 and ϵ_2. In spite of the reflection problem, we have not had to use any instrument.

3.1.2 Exclusion bias

So far we have assumed that ϵ_1 and ϵ_2 are uncorrelated with each other. This is not, however, strictly true because of the presence of exclusion bias. To see why, consider a simple model in which we ex ante assign to each individual i a value y_i from an i.i.d. distribution ϵ_i:

$$y_i = \epsilon_i$$

We then randomly assign individuals to pairs. As was discussed in Section 2, because assignment is done without replacement, someone with a high ϵ_i is, on average, assigned a peer with a lower ϵ_j – and vice versa if i has a low ϵ_i. It follows that errors ϵ_i are negatively correlated within matched pairs. The value of this correlation is given by Proposition 1. Specifically, from Proposition 1 we know that if we regress ϵ_1 on ϵ_2 under the null hypothesis that they have been randomly assigned to groups, we obtain a regression coefficient that is on average:

$$E[\widehat{b}] = -\frac{N_P - 1}{N_P^2 - 3N_P + 3} \equiv \rho \tag{21}$$

As per formula (18) above, this is also the correlation coefficient ρ between ϵ_1 on ϵ_2 that is due to exclusion bias since $\sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sigma_\epsilon$. The sample covariance between ϵ_1 and ϵ_2 is thus:

$$Cov[\epsilon_1, \epsilon_2] = E[\epsilon_1 \epsilon_2] = \rho\sigma_\epsilon^2 < 0$$

We can now calculate the covariance between y_1 and y_2 that results from the combina-

[12]The other root can be ignored because it is always > 1 and peer effects in a linear-in-means model cannot exceed 1.

tion of both the reflection bias and the exclusion bias. We need to recalculate everything above. The expectation of y is unchanged. The variance of y_1 now is:

$$E[(y_1 - \overline{y})^2] = \frac{\sigma_\epsilon^2(1 + \beta^2 + 2\beta\rho)}{(1-\beta^2)^2}$$

The covariance is:

$$E[(y_1 - \overline{y})(y_2 - \overline{y})] = \frac{\sigma_\epsilon^2(2\beta + (1+\beta^2)\rho)}{(1-\beta^2)^2}$$

Equipped with the above results, we can now derive an expression for the combined bias from reflection bias and exclusion bias that would result if we estimate model:

$$y_1 = a + by_2 + v_1 \tag{22}$$

As before, we use the fact that $\widehat{b}_e = \widehat{r}$. Hence we have:

$$E[\widehat{b}] = \frac{2\beta + (1+\beta^2)\rho}{1 + \beta^2 + 2\beta\rho} \neq \frac{2\beta}{1+\beta^2} \neq \beta \tag{23}$$

Table 4: Simulation results - Exclusion bias in estimation of endogenous peer effects

$K = 2$; $N_P = 10$; $N = 1000$					
	(1)	(2)	(3)	(4)	(5)
True β_1	Predicted reflection bias	Prediction exclusion bias	Total predicted bias	Predicted E($\hat{\beta}_1$)	Simulated E($\hat{\beta}_1$)
0.00	0.00	-0.12	-0.12	-0.12	-0.12
0.02	0.02	-0.12	-0.10	-0.08	-0.08
0.04	0.04	-0.12	-0.09	-0.04	-0.04
0.06	0.06	-0.12	-0.064	0.00	0.00
0.08	0.08	-0.12	-0.04	0.04	0.04
0.10	0.10	-0.12	-0.02	0.08	0.08
0.12	0.12	-0.12	-0.00	0.12	0.12
0.14	0.14	-0.12	0.02	0.16	0.16
0.16	0.15	-0.12	0.04	0.20	0.20
0.18	0.17	-0.11	0.06	0.24	0.24
0.20	0.19	-0.11	0.07	0.27	0.28

Note: Pool fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions.

We present in Table 4 simple calculations based on the above formula to illustrate the magnitude of the reflection and exclusion bias for various values of β. We see that, when β is zero or is small, the total predicted bias is dominated by the exclusion bias and is thus negative. As β increases, the reflection bias takes over and leads to coefficient estimates that over-estimate the true β. What is striking is that the combination of reflection bias and exclusion bias produces coefficient estimates that diverge dramatically from the true β, sometimes under-estimating it and sometimes over-estimating it. The direction of the bias nonetheless has a clear pattern that can be summarized as follows:

1. If $\beta = 0$, we get $E[\widehat{b}] = \rho < 0$ which is the size of the exclusion bias. We cannot draw correct inference about $\beta = 0$ by looking directly at $\widehat{b}$. This is because $\widehat{b}$ can be negative even when β is positive.

2. It is possible for $E[\widehat{b}]$ to be negative even though $\beta > 0$. This arises when ρ is large in absolute value, for instance if $N_p = 10$ and $K = 2$ as in Table 4.

3. Since the exclusion bias is always negative, $\widehat{b} > 0$ can only arise if $\beta > 0$. It follows that a positive $\widehat{b}$ is unambiguously indicative of the presence of peer effects.

3.1.3 Correcting point estimates and inference

The inference correction methods we discussed in Section 2.4 allow us to correct p-values but they are not able to yield a consistent point estimate of β if the true peer effect differs from zero. Under the assumption of independent errors, we can recover an estimate of β using formula (23). Taking roots, we obtain a consistent estimate $\widehat{\beta}$ of the true β using the value of ρ from (21) and the coefficient estimate $\widehat{b}$ from regression (22):

$$\widehat{\beta} = \frac{1 - \widehat{b}\rho - \sqrt{1 + \widehat{b}^2\rho^2 - \widehat{b}^2 - \rho^2}}{\widehat{b} - \rho} \tag{24}$$

This formula confirms that the parameter ρ *does* affect parameter estimates and parameter recovery. Ignoring exclusion bias leads to incorrect point estimates and biased inference about endogenous peer effects. While the reflection bias pushes $\widehat{b}$ to exceed β, the exclusion bias pushes in the other direction. As shown in Table 4, the exclusion bias easily dominates for reasonably moderate values of β. It is only for very large values of β that the reflection bias dominates and leads to an over-estimation of β. The rest of the time, regression (22) is biased towards finding no significantly positive peer effects.

While formula (24) can be used to obtain a corrected estimate of the peer effect coefficient β, there remains the important question of inference: how can we test whether $\widehat{\beta}$ is significantly different from 0. In order to obtain correct inference, we need to correct p-values for the standard test of significance that $\beta = 0$. The solutions are essentially the same as those discussed in Section 2.4. We suggest using the permutation method for this purpose to simulate, using the sample data, the distribution of $\widehat{b}$ that would arise under the null. This is achieved by replicating the random assignment of peers in the sample data to form counterfactual pairs. Since these pairs are formed at random, we expect no correlation in the y's other than that due to exclusion bias. The variation across counterfactual samples mimics the variation that would naturally arise in the data, given the random assignment procedure and other features of the sample.

Table 5: Correction exclusion bias in estimation of endogenous peer effects - $K = 2$

$K = 2$; $L = 10$; $N = 1000$				
	(1)	(2)	(3)	(4)
β_1	Simulated E($\hat{b}$)	Simulated p-value	Corrected E($\hat{\beta}$)	Corrected p-value
0.00	-0.12	0.06	0.00	0.47
0.02	-0.08	0.17	0.02	0.38
0.04	-0.04	0.44	0.04	0.21
0.06	0.00	0.36	0.06	0.09
0.08	0.04	0.25	0.08	0.02
0.10	0.08	0.11	0.10	0.00
0.12	0.12	0.04	0.12	0.00
0.14	0.16	0.00	0.14	0.00
0.16	0.20	0.00	0.16	0.00
0.18	0.24	0.00	0.18	0.00
0.20	0.28	0.00	0.20	0.00

Note: Pool fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions; Column (3) reports the corrected E($\hat{\beta}_1$) obtained using equation (24); Column (4) reports the p-value obtained using the permutation method over 500 replications.

To illustrate, we present the results of a Monte Carlo study in Table 5. We created random samples of 1000 observations following the data generating process described above but for different values of β. We then used bootstrapping to obtain correct p-values using 500 replications per regression. Reported p-values are for two-sided tests. We also report the simulated distribution of $E(\widehat{b})$ for different values of β. We set $K = 2, L = 10$ and $N = 1000$. In columns 1 and 2 we report the estimates of $E(\widehat{b})$ and the corresponding p-value as reported by OLS. In column 3 we report the corrected estimate $\widehat{\beta}$ obtained using formula (24). The last column presents the corrected p-values obtained from 500 bootstrapping replication of the null hypothesis of no peer effect. Results confirm that $\widehat{b}$ is dramatically biased, sometimes yielding a significantly negative estimate of β when the true β is close to zero, sometimes yielding an inflated estimate of β when reflection bias dominates. Corrected estimates $\widehat{\beta}$ do not display this pattern, however: they are centered on the true β. We also note that using corrected p-values eliminates the risk of incorrectly concluding that $\beta < 0$. When the true value of β is positive but small, we are unable to reject that $\beta = 0$, an indication that power may not always be sufficient to identify the presence of peer effects. As a whole, however, the method we propose produces a massive improvement in inference in this case.

3.2 General group model

We used the $K = 2$ case to illustrate how reflection and exclusion bias combine to affect coefficient estimates. In this case, we were able to derive a formula to correct the estimate of β. Obtaining a closed-form formula becomes more difficult if not impossible once we generalize to a larger group size K or to groups of varying size. But provided that we are willing to assume i.i.d. errors conditional on selection pool fixed effects, it remains possible to obtain an estimate of the true β and to bootstrap its p-value.

To illustrate, we consider a general structural model of the form:

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i \tag{25}$$

where Y is vector of all Y_i, vector G_i identifies all the peers of individual i, X_i is a vector of individual characteristics that affect Y_i directly, and X is the matrix of all X_i. Parameter γ captures the effect of the characteristics of individual i on Y_i, β captures endogenous peer effects as before, and δ captures so-called exogenous peer effects, that is, the effect of the characteristics of peers that affect i directly without the need to influence the behavior of the peers. Matrix G is the matrix of all G_i vectors. In the linear-in-means model (1), G_i is a vector of 0's and $1/(K-1)$ so that $G_i Y$ is equal to $\bar{y}_{-i,k,l}$. But this can be generalized to other influence models by varying G_i, for instance by letting G be a network adjacency matrix (see below).

Regression model (25) can be written in matrix form as:

$$Y = \beta GY + \gamma X + \delta GX + \epsilon$$

Simple algebra yields the following reduced-form model:

$$Y = (I - \beta G)^{-1}(\gamma X + \delta GX + \epsilon)$$

from which we obtain:

$$\begin{aligned} E[YY'] &= E[(I-\beta G)^{-1}(\gamma X+\delta GX+\epsilon)\ (\gamma X+\delta GX+\epsilon)'(I-\beta G')^{-1}] \\ &= (I-\beta G)^{-1}E[(\gamma X+\delta GX)(\gamma X+\delta GX)'](I-\beta G')^{-1} \\ &\quad +(I-\beta G)^{-1}E[\epsilon\ \epsilon'](I-\beta G')^{-1} \end{aligned} \tag{26}$$

where we have assumed that the G matrix is non-stochastic. The covariance matrix of the X's is identified from the data. If the ϵ's are i.i.d, we have:

$$E[\epsilon\ \epsilon'] = \sigma_\epsilon^2 I$$

as before. With this assumption, expression (26) can be used as starting point for estimation. With exclusion bias,

$$E[\epsilon\ \epsilon'] \neq \sigma_\epsilon^2 I$$

Formula (21) can be used to derive the covariance matrix of the ϵ's. To illustrate, suppose that all observations are arranged so that the observations from the first group come first, then the observations from the second group, etc. In this case $E[\epsilon\ \epsilon']$ is a block-diagonal matrix:

$$E[\epsilon\epsilon'] = \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & B & 0 & 0 \\ 0 & 0 & B & 0 \\ 0 & 0 & 0 & B \end{bmatrix} \tag{27}$$

For $K = 2$, each block B is of the form:

$$B = \begin{bmatrix} E[\epsilon_1^2] & E[\epsilon_1\epsilon_2] \\ E[\epsilon_2\epsilon_1] & E[\epsilon_2^2] \end{bmatrix} \tag{28}$$

We have shown earlier that, for two individuals i and j in the same selection pool of size N_P, we have $E[\epsilon_i\epsilon_j] = \rho\sigma_\epsilon^2$ with $\rho = -\frac{N_P-1}{N_P^2-3N_P+3}$ for $i \neq j$. Hence B can be rewritten as:

$$B = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \equiv \sigma_\epsilon^2 A \tag{29}$$

If $K > 2$, A becomes a $K \times K$ matrix but its form remains the same: 1 on the diagonal and ρ off the diagonal. The value of ρ is given by formula (21). What is important is that ρ is known and does not need to be estimated.

Equation (21), combined with (27) and (29), provides a characterization of the data generating process that can be used to estimate structural parameters β, γ, δ and σ^2. Identification is achieved from the assumption that, conditional on network (cluster) fixed effects, errors are independent across observations from the same peer group – except for exclusion bias. With this assumption, instruments are not required in spite of the presence of reflection bias. Inference can be conducted in the same way as before, that is, by simulating the distribution of estimated coefficients under the null hypothesis of no peer effects.

One approach to estimate (26) is to rely on a method of moments estimation strategy. That is, we choose the parameter β that provides the best fit to the observed data $E[YY']$. This is achieved using a search algorithm. For each guess $\beta^{(n)}$ that the algorithm makes about β, we solve for the corresponding values of γ and δ by calculating $Y - \beta^{(n)}GY$ and regressing it on X and GX to obtain estimates of $\widehat{\gamma}^{(n)}$ and $\widehat{\delta}^{(n)}$. This process also yields an estimate of the variance of errors $\widehat{\sigma}_\epsilon^{2(n)}$. Using $\beta^{(n)}, \widehat{\gamma}^{(n)}, \widehat{\delta}^{(n)}$ and $\widehat{\sigma}_\epsilon^{2(n)}$ we compute the value of each element of the right hand side of equation (26). Subtracting each value from the corresponding $y_i y_j$, taking squares, and summing over all ij pairs yields the value of the 'fit' for guess $\beta^{(n)}$. We then search over possible values of β to achieve the best fit/lowest sum of squared residuals.

Table 6: Correction exclusion bias in estimation of endogenous peer effects - Groups

	K = 2			K = 5		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	0.00	0.10	0.20	0.00	0.10	0.20
$\hat{\beta}_1^{Naive}$ - no corrections	-0.05	0.15	0.34	-0.25	-0.02	0.19
Naive p-value	0.21	0.01	0.00	0.07	0.33	0.10
$\hat{\beta}_1^{Ref}$ - correction for reflection bias only	-0.02	0.07	0.18	-0.11	0.00	0.09
$\hat{\beta}_1^{Corr}$ - correction for reflection bias + exclusion bias	0.00	0.10	0.20	0.01	0.09	0.20
Corrected p-value (permutation method)	0.54	0.02	0.00	0.45	0.15	0.00

Note: $L = 20$; $N = 1000$; Pool fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions. Permutations over 500 replications.

To illustrate the effectiveness of this approach, we estimate model (25) on simulated data, using 100 Monte Carlo replications, the results of which are shown in Table 6. We keep the number of observations in each sample constant at $N = 1000$ but we vary K and β. Cluster fixed effects are included throughout. In the first row we report the uncorrected $\widehat{\beta}^{Naive}$ obtained by regressing Y_i on G_iY and cluster fixed effects. Results confirm that the uncorrected β is biased. As before this bias combines two sources of bias: reflection and exclusion. When β is small, the exclusion bias dominates and the naive β underestimates the true β. The naive β is more likely to overestimate the true β when exclusion bias is small, which occurs when the selection pools are large. In the second row in Table 6, we report the $\widehat{\beta}^{Ref}$ estimate corrected for reflection bias but ignoring the exclusion bias. This is the estimate derived from model (26) with $E[\epsilon\ \epsilon'] = \sigma_\epsilon^2 I$. In all cases, the estimate is closer to the true β, but the failure to eliminate exclusion bias results in an underestimation of the true β on average. The last column reports the average $\widehat{\beta}^{Corr}$ estimate derived from model (26) with $E[\epsilon\ \epsilon']$ given by (29). The $\widehat{\beta}^{Corr}$ is centered around its true value in all cases.

Table 6 also shows the naive p-values reported with $\widehat{\beta}^{Naive}$ as well as the corrected p-values obtained using the permutation method described in Section 2.4.5.

3.3 Network data

Until now we have considered situations in which peers form groups, i.e., such that if i and j are peers and j and k are peers, then i and k are peers as well. Exclusion bias also arises when peers form more general networks, i.e., such that i and k need not be peers. To illustrate this, we go back to the canonical case considered in Section 2, namely, let us assume that individuals in selection pool l are randomly assigned peers within that pool. The only difference with Section 2 is that we no longer restrict attention to peer groups but allow links between peers to take an arbitrary (including directed or undirected) network shape within each pool l. Groups of unequal size are handled in the same manner.

The approach developed to estimate general group models with uncorrelated errors can be applied to network data virtually unchanged. Equation (26) remains the same. Formally let $g_{ijl} = 1$ if i and j in cluster l are peers, and 0 otherwise. We follow convention and set $g_{ii} = 0$ always. The network matrix in cluster l is written $G_l = [g_{ijl}]$ and G is a block diagonal matrix of all G_l matrices. To estimate network models in levels, we use G directly.

If the model we wish to estimate is linear-in-means, let n_{il} denote the number of peers (or degree) or i. The value of n_{il} typically differs across individuals. Let us define vector $\widehat{g}_{il}$ as a vector formed by dividing i's row of G_l by n_{il}, i.e.:

$$\widehat{g}_{il} = [\frac{g_{i1l}}{n_{il}}, ..., \frac{g_{iLl}}{n_{il}}]$$

where, as before, L denotes the size of the selection pool.[13] The average outcome of i's

[13]To illustrate, let $L = 4$ and assume that individual 1 has individuals 2 and 4 as peers. Then

peers can then be written as $\widehat{g}_{il} y_l$ where y_l is the vector of all outcomes in selection pool l. The peer effect model that we aim to estimate is:

$$y_{ikl} = \beta_0 + \beta_1 \widehat{g}_{il} y_l + \delta_l + \epsilon_{ikl} \tag{30}$$

We define $\widehat{G}_l$ as the $L_l \times L_l$ matrix obtained by stacking all $\widehat{g}_{il}$ in pool l. Similarly define $\widehat{G}$ as the block-diagonal matrix of all $\widehat{G}_l$ matrices. The linear-in-means network autoregressive model can thus be written in matrix form as:

$$Y = \beta \widehat{G} Y + \gamma X + \delta \widehat{G} X + \epsilon \tag{31}$$

As in the previous section, equation (21) combined with (27) and (29) can be used to estimate structural parameters β, γ, δ and σ^2. The only difference is that G is now a network matrix rather than a block-diagonal matrix. It is intuitively clear that exclusion bias affects model (30) as well: individual i is excluded from the selection pool of its own peers, and this generates a mechanical negative correlation between i's outcome and that of its peers. To correct for this in a way that automatically accommodates any network matrix, we redefine expression $E[\epsilon\epsilon']$ to be a block-diagonal matrix B_l of size $L_l \times L_l$ where L_l is the number of individuals in selection pool l. Parameter ρ is computed as before. Pre- and post-multiplying matrix $E[\epsilon\ \epsilon']$ by $(I - \beta G)^{-1}$ in expression (26) picks the relevant off-diagonal elements of B to construct the needed correction for exclusion bias. Estimation proceeds using the same iterative algorithm described above.

We illustrate this approach for network data in Table 7. We generate each adjacency matrix G_l as a Poisson random network with linking probability p. In other words, p is the probability that a link exists between any two individuals i and j within the same pool. When $p = 0.1$ and $L = 20$, each individual has two peers on average; when $p = 0.25$ (0.5) each individual has on average 5 (10) peers, respectively. Table 7 provides simulation results and shows how our suggested method of moments correction method is able to correct the estimate of β_1 to be close to the true β_1 even after only 100 simulation repetitions.

The permutation method can similarly be used to correct p-values. To recall, we want to simulate the counterfactual distribution of $\widehat{\beta}_1$ under the null hypothesis of zero peer effects. In contrast with Section 3, peers are no longer selected by randomly partitioning individuals into groups within clusters, but rather by randomly assigning peers within clusters. In order to simulate the distribution of $\widehat{\beta}_1$ in regression model (30) under the

$\widehat{g}_{il} = [0, \frac{1}{2}, 0, \frac{1}{2}]$.

null of no peer effect, we keep the network matrices in each selection pool unchanged. But we change who is linked to whom. To achieve this within a cluster l, we scramble matrix G_l in the following way. Say the original ordering individual indices in cluster l is $\{1, ..., i, ..., j, ..., L\}$. We generate a random reordering (k) of these indices, e.g., $\{j, ..., 1, ..., L, ..., i\}$. We then reorganize the elements of G_l according to this reordering to obtain a counter-factual network matrix $G_l^{(k)}$. To illustrate, imagine that i has been mapped into k and j into m. Then element g_{ijl} of matrix G_l becomes element g_{kml} in matrix $G_l^{(k)}$. We then use this matrix to compute the average peer variable $\widehat{g}_{il}^{(k)} y_l$. This approach is known in the statistical sociology literature as Quadratic Assignment Procedure or QAP (e.g., Krackhardt, 1988). For each reordering (k) we estimate model (30) and obtain a counter-factual estimate $\widehat{\beta}_1^{(k)}$. We then use the distribution of the $\widehat{\beta}_1^{(k)}$'s as approximation of the distribution of $\widehat{\beta}_1$ under the null of zero peer effects. We compare in Table 7 the p-values obtained from the naive model and the permutation approach applied to model 31. The performance of our estimation method in the network case is comparable to what it was in the peer group case.

Table 7: Correction exclusion bias in estimation of endogenous peer effects - Networks

	p = 0.10			p = 0.25		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	0.00	0.10	0.20	0.00	0.10	0.20
$\hat{\beta}_1^{Naive}$ - no corrections	-0.09	0.08	0.25	-0.27	-0.09	0.09
Naive p-value	0.13	0.22	0.001	0.03	0.26	0.29
$\hat{\beta}_1^{Ref}$ - correction for reflection bias only	-0.04	0.02	0.11	-0.09	-0.01	0.02
$\hat{\beta}_1^{Corr}$ - correction for reflection bias + exclusion bias	0.00	0.09	0.19	0.01	0.07	0.19
Corrected p-value (permutation method)	0.51	0.06	0.00	0.48	0.24	0.02

Note: p = probability of link between i and j within a cluster ; $L = 20$; $N = 1000$; Pool fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions. Permutations over 500 replications.

3.4 More complex settings

So far we have assumed that peers are randomly drawn from within a well-specified pool of fixed size N_P. In this section we discuss possible extensions to more complex settings.

1. So far we have assumed that all selection pools are of equal size $N_P = L$. If selection pools vary in size, it can be shown that in expectation the bias is a weighted average

of the biases associated with the different cluster sizes:

$$E(Bias) = \sum_{l=1}^{Z} \frac{L_l}{N} E(Bias|L = L_l) \tag{32}$$

where L_l denotes the size of selection pool l and Z is the total number of selection pools.

Similarly, if peer groups differ in size K_k it can be shown that the exclusion bias is a weighted average of the bias associated with the different K_k. That is:

$$E(Bias) = \sum_{k=1}^{S} \frac{K_k}{L} E(Bias|k = K_k) \tag{33}$$

where K_k denotes the size of the peer group k and S is the total number of peer groups within each selection pool.

2. Peers may not be selected from mutually exclusive selection pools. For instance, students tend to befriend mostly classmates. But they may also have friends in other classrooms. To capture this situation, let us now reserve the word 'cluster' to denote the pool from which most but not all friends are selected – so that selection pool and cluster are no longer synonymous. In this example, membership to a peer group is correlated with membership to a cluster, but some peers are selected from outside the cluster. This means that selection pools are not mutually exclusive; they partially overlap.

   To illustrate, consider a specification in which individual i in cluster l selects a proportion θ of her peers from within cluster l and a proportion $(1 - \theta)$ from outside cluster l. In previous sections we have focused on cases where $E(\bar{y}_{-i,k,l}) = \bar{y}_{-i,l}$. Now $E(\bar{y}_{-i,k,l})$ follows:

$$E(\bar{y}_{-i,k,l}) = \theta\bar{y}_{-i,l} + (1 - \theta)\bar{y}_{-l,\Omega} \tag{34}$$

   where $0 \leq \theta \leq 1$ and $\bar{y}_{-l,\Omega}$ is the average outcome over the entire population Ω excluding cluster l. Although a further treatment of this extension is beyond the scope of this paper, we conjecture that, in a group fixed effects model, the exclusion bias becomes less severe when at least one peer is selected from outside the cluster ($\theta < 1$) compared to a situation where all peers are drawn from within the cluster ($\theta = 1$). This is because a model with cluster fixed effects only considers the variation in outcomes within the cluster. The exclusion bias is driven by the negative correlation between i's outcome and the expected outcome $E(\bar{y}_{-i,k,l})$ of i's peer group. This correlation should fall whenever $\theta < 1$, hence reducing exclusion bias.[14]

In practice, a model of this type can be estimated using the general network approach outlined above, defining the network matrix G over the entire population to capture links across clusters. The rest remains unchanged.

3. A third complication is that, in practice, the pool of potential peers is not always well defined. In some studies on peer effects in educational achievement, survey respondents are asked to identify peers from the entire school roster, in which case $\theta < 1$ (e.g. Halliday and Kwak, 2012; Fletcher, 2012).[15] In other studies, people are restricted to identify peers among a list of students in their classroom, thereby ensuring that $\theta = 1$ (e.g. de Melo, 2014). Fletcher and Ross [2012], in an attempt to control for correlated effects, construct 'clusters of observationally equivalent individuals who face the same friendship opportunity set and make the same type of friendship choices' within the school. To the extent that students select peers from within these clusters, it is the size of these groups that determines the magnitude of the exclusion bias. The boundaries of the pools from which peers are selected are seldom well defined, however. It may be impossible to obtain a closed-form expression for the bias in such cases, but estimation can proceed as in case 2 above.

4. Fourth, even if the pool of potential peers is precisely known, it is often the case that peers cannot be considered as randomly drawn within the selection pools. Consequently, the expected value of the outcome of i's peers may differ from the net-i-pool-average $E(\bar{y}_{-i,k,l})$. In such cases, calculating the size of the exclusion bias would require simulating the peer assignment process that generated the data. If the researcher is willing to posit a data generation process for peer assignment, the exclusion bias can be approximated using the same type of randomization inference

[14]Note that, whereas an increase in peer group size unambiguously increases the magnitude of the exclusion bias as long as all peers are drawn from within a cluster, the bias is insensitive to the number of additional peers drawn from outside the cluster.

[15]If it is further assumed that peers are as likely to be selected from within the classroom as from outside the classroom, we would have $\theta = \frac{L}{N}$

that we described earlier. How difficult this would be in practice depends on the nature of the posited peer assignment process. Estimation can proceed as before. But permutation-based inference must be adjusted to correct for non-random peer assignment. If the researcher is willing to posit a predictive model for the link formation process, this model can in principle be estimated from the data and used to simulate counter-factual peer assignment under the null of no peer effects. Working this case out in detail is beyond the scope of this paper.

4 Exclusion bias in practice

4.1 A basic application

A natural starting point to illustrate the practical relevance of the exclusion bias - and its interaction with the reflection bias - is to revisit the results reported by Sacerdote [2001] on estimated peer effects in academic achievement among Darthmouth College roommates. Sacerdote uses a sample of 1589 graduate students ($N = 1589$) whose college application forms are divided into 42 piles ('blocks') based on their revealed housing preferences (e.g. whether they smoke, listen to music while studying, etc.). Within each block students are randomly allocated to dorms and dorm rooms. Hence, in this example the peer selection pool is the block. The study does not provide a breakdown of pool sizes so for the purpose of illustration we assume that pools are of equal size $L = 38$. The breakdown by room group size is reported as follows: 53% of students are in double occupancy rooms, 44% of students in triples, and the remaining 3% of students are in quad rooms.

This gives us all the parameter values that we need to apply equation (1) in order to estimate the magnitude of the exclusion bias in the standard test of random peer assignment used by Sacerdote [2001]. Using formulas (14) and (33) we obtain the following exclusion bias under the null of zero correlation:

$$Bias = 0.53 * (bias|K = 2) + 0.44 * (bias|K = 3) + 0.03 * (bias|K = 4) = -0.04$$

The bias in this particular case is fairly small as the peer group size (2-4 peers per group) is small relative to the peer selection pool size (38); in many other applications the bias is likely to be larger. It is thus all the more interesting to assess whether a bias of this relatively small magnitude is able to affect inference.

The first row in Table 8 shows the estimation results for the test of random peer assignment originally presented in Sacerdote [2001], for four different baseline characteristics of interest (SATH math score, SAT verbal score, high school academic class index and high school academic index). In the second row we present the coefficient estimates corrected for exclusion bias under the null of zero correlation. All coefficient estimates turn positive and, using the standard errors originally reported by Sacerdote [2001], we now reject the null of zero correlation for one outcome at the 10% significance level. This result needs not survive the use of corrected standard errors that are probably larger. Unfortunately we do not have the information needed to cluster standard errors at the dorm level.

Table 8: Evidence of random assignment of roommates in Darthmouth College [Sacerdote, 2001]

	SATH Math	SAT verbal	High school Academic class index	High school academic index
	(1)	(2)	(3)	(4)
$\hat{\beta}_1^{Naive}$ - Sacerdote [2001]	-0.025	-0.009	0.010	-0.032
	(0.028)	(0.029)	(0.028)	(0.028)
$\hat{\beta}_1^{Corr}$	0.015	0.031	0.05*	0.008
	(0.028)	(0.029)	(0.028)	(0.028)

Note: Standard errors (in parantheses) are not clustered at the level of the dorm; All regressions include 41 dummies representing nonempty blocks based on housing preferences.

Next, in the first row of Table 9 we revisit the endogenous peer effect estimation results that Sacerdote [2001] obtains when running regressions similar to (17) using GPA test score as the outcome of interest[16]. Here, the Table in the original study does report correct standard errors that are clustered at the level of the room. The author of the study notes that the coefficient estimate of 0.07 - which is significant at the 5% level - cannot be interpreted as causal given the presence of reflection bias. As we have seen, this estimate is also affected by exclusion bias.

We do not have the data at hand and therefore cannot use the generalized method of moments strategy for $K > 2$ discussed in Section 3.2. To circumvent this difficulty, we proceed as if there were no students in triple or quad rooms and assume that all students are in double occupancy rooms. This has the mechanical effect of understating the exclusion bias but has the advantage of allowing us to apply formula (21) to correct

[16]Sacerdote also controls for high-school test scores of self and peers in his regression. As discussed in Section 5.1 this inclusion could potentially reduce the magnitude of the exclusion bias. But given the small coefficient estimates of these control variables, we do not expect this to affect the magnitude of the exclusion bias much (as confirmed by simulations - not reported).

the estimate for reflection bias only and formula (24) to yield an estimate that corrects for both exclusion bias and reflection bias. As expected, correcting for reflection bias leads to a reduction in the peer effect estimate, which falls to 0.03 and is no longer statistically significant. When accounting for both exclusion and reflection bias, we obtain a corrected peer effect estimate of 0.05, which is statistically significant at the 10% level – but is very similar in magnitude and significance to that obtained for one of the baseline outcomes in Table 8.

Table 9: Peer effects in academic outcome - [Sacerdote, 2001]

	GPA test score
	(1)
$\hat{\beta}_1^{Naive}$ - Sacerdote (2001)	0.07**(0.029)
$\hat{\beta}_1^{Ref}$ - Correction for reflection bias only	0.03(0.029)
$\hat{\beta}_1^{Corr}$ - Correction for reflection bias + exclusion bias	0.05*(0.029)

Note: Standard errors (in parantheses) are clustered at the level of the student room; All regressions include 41 dummies representing nonempty blocks based on housing preferences; The corrected estimates assume $K = 2$ whereas in the Sacerdote application some students resided in triple or quad rooms (so the corrections are conservative, although we do not expect this to affect the results much).

The lesson from this exercise is that, even in this illustrative example where the peer group size is small relative to the peer selection pool, the exclusion and reflection biases jointly have a non-negligible impact on inference. In applications that rely on larger peer groups or smaller peer selection pools, the negative exclusion bias would affect findings even more substantially.

4.2 Caution against peer group comparisons

We have shown that, for a given pool size, exclusion bias becomes more severe as peer group size increases. It follows that comparisons of estimates between models that vary in peer group size can be misleading.

For example, Halliday and Kwak [2012] use the Add Health dataset to compare peer effect estimates on three outcomes (GPA, smoking and drinking) for different definitions of peer groups used in the education literature. The relevant peer selection pool in their setting is the school, which has on average 246 to 374 sampled students (the available sample size depends on the outcome of interest)[17]. As shown in Table 10, the authors find that estimated peer effects are significantly smaller when school grade cohort is used

[17] There are 145 schools in the Add Health dataset. The size of the peer selection pool is calculated by dividing the relevant sample size by 145.

as peer group (containing on average $K = 144$ students) instead of the small circle of friends reported by students (containing on average $K = 5$ students). As we do not have the Add Health dataset at hand we cannot calculate the total bias on the peer effect estimate caused by a combination of the exclusion bias and the reflection bias. Back-of-the-envelope calculations suggest however that the exclusion bias under the null of zero peer effects ranges from -0.01 to -0.02 for friends and -0.62 to -1.25 for school grade cohort as choice of peer group. Although the total bias is likely to be less negative as a result of the reflection bias, the stark differences in the magnitude of the exclusion bias for school grade cohorts as a peer group definition as opposed to that for small circles of friends (for a given peer selection pool) can very well explain the differences observed between the peer effect estimates.

Table 10: Peer effect estimates for different peer group sizes - Halliday and Kwak [2012]

	GPA		Smoking		Drinking	
	(1)	(2)	(3)	(4)	(5)	(6)
Peer group definition	Friends	School grade cohorts	Friends	School grade cohorts	Friends	School grade cohorts
Estimated peer effect	0.45***	0.38***	0.59***	0.25***	0.25***	0.04
	(38.12)	(8.59)	(36.93)	(4.13)	(16,39)	(0.58)
N	35,649	37,423	53,854	54,371	53,736	54,269
Number of schools	145	145	145	145	145	145
Average N_p	246	258	371	375	371	374
Average K	5	144	5	144	5	144
Average exclusion bias under the null	-0.02	-1.25	-0.01	-0.62	-0.01	-0.62

Note: Adapted from Halliday and Kwak [2012]. Authors use Add-Health data. All regressions include grade and, when appropriate, gender dummies. t-statistics are in brackets. Peer selection pool is the school. Given that there are 145 schools in the dataset, average N_P is calculated by dividing sample size N by 145. All standard errors adjust for clustering on schools. All regressions include controls for health status as well as race dummies and parental education. *** significant at 1% level, ** significant at 5% level, * significant at 10% level.

Our findings also caution against naive comparisons between peer effect models that include cluster fixed effects and models that do not. Indeed, exclusion bias is aggravated by the inclusion of cluster fixed effects whenever group formation is correlated with cluster formation. For example, studies adding classroom effects (e.g. de Melo, 2014), dormitory effects (e.g. Sacerdote, 2001), or school effects (e.g. Fletcher, 2012) are more severely affected by exclusion bias if peers are selected - partially or completely - from within theses clusters. When adding cluster fixed effects, the literature tends to interpret a drop in estimated peer effects as evidence of unobserved covariates at the cluster level. Although such correlates often matter, the results presented in this paper demonstrate that such interpretations may be unwarranted since it is confounded by exclusion bias.

4.3 Peer effects on outcome gains or losses rather than outcome levels

So far we have focused on models that measure the influence of peers on outcome *levels*. Exclusion bias is similarly present in models that measure peer effects on the *time variation* of outcomes, as when the researcher introduces individual fixed effects to control for unobserved effects that may bias the estimation of peer effects.

To illustrate the issue, we abstract from other possible controls and focus on the comparison between two simple models in which the dependent variable is expressed in first difference:

$$\Delta y_{ikt} = \beta_0 + \beta_1 \Delta \bar{y}_{-i,k,t} + \epsilon_{ikt} \tag{35}$$

and

$$\Delta y_{ikt} = \beta_0 + \beta_1 \bar{y}_{-i,k,t-1} + \epsilon_{ikt} \tag{36}$$

Such models are often used in practice. Hanushek et al. [2003], for instance, examines the impact of the lagged math test score of peers on the variation in test score between grades.

Exclusion bias is present in models such as (35) for the same reason that it is present in levels: a student with a lower-than-average gain in test score is on average matched with a student with a higher-than-average gain in test score. This creates a mechanical negative relationship between Δy_{ikt} and $\Delta \bar{y}_{-i,k,t}$.

The sign and magnitude of the exclusion bias in (36) is less clearcut, but is nonetheless a serious consideration. If Δy_{ikt} is uncorrelated with $y_{ik,t-1}$ then (36) should be unaffected by exclusion bias. However, if individuals with a low baseline $y_{ik,t-1}$ experience systematically smaller (larger) than average changes in outcome, then a mechanical negative (positive) relationship arises between Δy_{ikt} and $\bar{y}_{-i,k,t-1}$, yielding a downward (upward) bias in $\widehat{\beta}_1$, respectively. Introducing individual fixed effects thus does not, by itself, eliminate exclusion bias.

5 Methodologies unaffected by exclusion bias

5.1 Studies adding particular control variables

In some circumstances, it is possible to eliminate the exclusion bias using control variables. This is best illustrated with an example, namely, the golf tournament studied by Guryan et al. [2009][18].

At $t+1$ golfers participating to tournament l are assigned to a peer group k with whom they play throughout the tournament. The performance of golfer i in tournament l is written as $y_{ikl,t+1}$. The researcher has information on the performance of each golfer i in past golf tournaments. This information is denoted as y_{iklt}. The researcher wishes to test whether performance in a tournament depends on who golfers are paired with. The researcher's objective is thus to estimate coefficient β_1 in a regression of the form:

$$y_{ikl,t+1} = \beta_0 + \beta_1 \bar{y}_{-i,klt} + \delta_l + \epsilon_{ikl,t+1} \tag{37}$$

A key difference with our earlier models is that here $\bar{y}_{-i,klt}$ is calculated using the *past* performance of peers, before the random assignment of peers. Because random assignment to peer groups is done without replacement, $\bar{y}_{-i,klt}$ is negatively correlated with the past performance y_{iklt} of individual i even though, given the random nature of the assignment process, there cannot be peer effects. Since past performance is correlated with unobserved talent, we expect y_{iklt} to be positively correlated with $y_{ikl,t+1}$. This generates a negative exclusion bias in regression (37): coefficient β_1 is biased downward.

This example nonetheless suggests an immediate and easy solution: to include y_{iklt} as additional regressor, since this automatically eliminates the exclusion bias. The model to estimate is thus:

$$y_{ikl,t+1} = \beta_0 + \beta_1 \bar{y}_{-i,klt} + \beta_2 y_{iklt} + \delta_l + \epsilon_{ikl,t+1}$$

where y_{iklt} serves the role of control variable. This is the approach adopted, for instance, in Munshi [2004].

A similar reasoning applies if the researcher is interested in the influence of the pre-existing characteristics of peers $\bar{x}_{-ikl}$ on i's subsequent outcome $y_{ikl,t+1}$. Here too the pre-existing characteristic of peers is negatively correlated with i's pre-existing characteristic x_{ikl}. Hence if the researcher fails to control for x_{ikl} and x_{ikl} is positively correlated with $y_{ikl,t+1}$, then estimating a model of the form:

$$y_{ikl,t+1} = b_0 + b_1 \bar{x}_{-i,kl} + u_{ikl,t+1}$$

will result in a negative exclusion bias.[19] This bias is easily corrected by including x_{ikl} as

[18]Many random pairing experiments, such as the random assignment of students to rooms or to classes, have a similar structure.

[19]If x_{ikl} is negatively correlated with $y_{ikl,t+1}$ then the exclusion bias is positive, i.e., b_1 is estimated to be less negative than it is.

control, as done for instance in Bayer et al., 2009:

$$y_{ikl,t+1} = b_0 + b_1\bar{x}_{-i,kl} + b_2 x_{ikl} + u_{ikl,t+1}$$

If the researcher does not have data on y_{iklt} or x_{ikl}, respectively, the exclusion bias may potentially be reduced by including individual characteristics of i as control variables to soak up some of the variation in $y_{ikl,t+1}$. How successful this approach will be in practice depends on how strongly individual characteristics predict y_{iklt} or x_{ikl}, as the case may be. Simulations (not reported here) indicate that the reduction in exclusion bias is sizable when control variables collectively predict much of the variation in $y_{ikl,t+1}$ (e.g., a correlation of 0.8). The improvement is negligible, however, when the correlation is small (e.g., 0.2).

5.2 2SLS estimation strategies

The use of instrumental variables can - under certain conditions - eliminate exclusion bias. One case that is particularly relevant in practice is when the researcher uses the peer average of a variable z to intrument peer effects, but also includes z_i in the regression. To illustrate this formally, let us assume that the researcher has a suitable instrument $\bar{z}_{-i,kl}$ for $\bar{y}_{-i,kl}$. For instance, $\bar{z}_{-i,kl}$ may be the peer group average of a characteristic z known not to influence y_{ikl}, e.g., because this characteristic has been assigned experimentally. If $\bar{z}_{-i,kl}$ is informative about $\bar{y}_{-i,kl}$, then z_{ikl} should be informative about y_{ikl} as well. For this reason, z_{ikl} is often included in the estimated regression as well. In this case, the first and second stages of this 2SLS estimation strategy can be written as follows:

$$\begin{aligned}\bar{y}_{-i,kl} &= \pi_0 + \pi_1\bar{z}_{-i,kl} + \pi_2 z_{ikl} + \delta_l + v_{ikl} \\ y_{ikl} &= \beta_0 + \beta_1\hat{\bar{y}}_{-i,kl} + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl}\end{aligned}$$

where $E(z_{ikl}\epsilon_{ikl}) = 0$, $E(\epsilon_{ikl}) = 0$ and $\hat{\bar{y}}_{-i,kl} = \hat{\pi}_0 + \hat{\pi}_1\bar{z}_{-i,kl} + \hat{\pi}_2 z_{ikl} + \hat{\delta}_l$ is the fitted value from the first-stage regression.

Expanding the second-stage 2SLS equation and replacing the fitted values by the above expression, it is straightforward to show that $cov(\hat{\bar{y}}_{-i,kl}, \epsilon_{ikl}|z_{ikl}) = 0$ and therefore that $\hat{\beta}_1^{2SLS}$ does not suffer from exclusion bias. Indeed we have:

$$
\begin{aligned}
y_{ikl} &= \beta_0 + \beta_1 \hat{\hat{y}}_{-i,k,l} + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \\
&= \beta_0 + \beta_1(\hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{-i,kl} + \hat{\pi}_2 z_{ikl} + \hat{\delta}_l) + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \qquad (38)
\end{aligned}
$$

If y_{ikl} and z_{ikl} are correlated (i.e., if $\beta_2 \neq 0$), we expect $\bar{z}_{-i,kl}$ to be mechanically correlated with y_{ikl} because $\bar{z}_{-i,kl} = \frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} z_{js}\right] - z_{ikl}}{L-1} + \tilde{u}_{ikl}$, where $\tilde{u}$ is defined in the same manner for z as u was defined for y in equation (3): $z_{-i,k} = z_{-i} + \widetilde{u}_{ik}$. Since equation (38) controls for z_{ikl} directly, this mechanical relationship is preventing from generating an exclusion bias, allowing $\hat{\beta}_1^{2SLS}$ to be an unbiased estimate of the peer effect.

An important implication of this result is that, in the absence of correlated effects and other sources of endogeneity such as measurement error, 2SLS strategies of the type described here yield IV estimates that tend to be *larger* – i.e., more positive – than OLS peer-effect estimates since they do not contain the exclusion bias. The finding that the downward bias present in OLS can be eliminated by 2SLS provides an alternative explanation for the common but counter-intuitive tendency of peer effects studies to obtain 2SLS estimates that are larger than their OLS counterparts (e.g. Goux and Maurin, 2007; Halliday and Kwak, 2012; De Giorgi et al., 2010; de Melo, 2011; Brown and Laschever, 2012; Helmers and Patnam, 2012; Krishnan and Patnam, 2012; Naguib, 2012; Collin, 2013). So far, this counter-intuitive finding has either been ignored, or has been attributed to classical measurement error or to the local average treatment interpretation of 2SLS. Exclusion bias offers another possible explanation for this often observed pattern.

For 2SLS to effectively eliminate exclusion bias, it is necessary to control for i's own value of the instrument z_{ikl} in (38). Estimation strategies employed in Bramouille et al. [2009] and De Giorgi et al. [2010], for instance, satisfy this criterion. Any instrumentation method that fails to do so suffers from exclusion bias in the same way and for the same reason as OLS.

6 Concluding remarks

The objective of this study was to conduct an in-depth and formal analysis of a listed (Guryan et al., 2009) but so far largely undocumented source of downward estimation bias in standard peer effects models. This negative bias - which we call 'exclusion bias' - exists on top of other, well-known sources of bias such as reflection bias and correlated effects. The paper provides important insights into the cause, consequences and solutions of this bias, which has largely been ignored to date.

We have shown that the bias is driven by the exclusion of individuals from the pool from which their peers are drawn and have demonstrated that this negative bias can seriously affect point estimates and inference in standard tests of random peer assignment and in the estimation of endogenous peer effects. The magnitude of the bias is particularly strong in studies that consider large peer groups relative to the size of the peer selection pool (e.g. number of peers considered in a classroom) and those that include cluster fixed effects whenever peers are selected at the level of a sub-cluster (e.g. classroom). A striking result is that when the true peer effect is small or zero, the negative exclusion bias dominates the positive reflection bias yielding an overall negative bias on the peer effect estimate.

Based on this, we suspect that some peer effect studies have never been published – or even never saw the light of day. In the absence of exclusion bias, researchers normally expect an upward bias in peer effects due to reflection bias. If application of simple OLS to data yields an insignificant or even negative peer effect coefficient, a researcher unaware of exclusion bias is likely to conclude that positive peer effects are absent from their data – and thus that the issue is not worthy of further investigation. We suspect that many researchers to date have abandoned research plans to study peer effects because they did not realize that this small or even negative OLS estimate could have been the result of exclusion bias.

The ideas presented here also offer an alternative to the estimation of peer effects using instrumental variables. Valid 2SLS estimation requires the availability of suitable strong instruments. Moreover, 2SLS is biased in finite samples (Bound et al., 1995). Exogenous sources of variation are particularly hard to find in settings that control for cluster fixed effects. Therefore, many studies rely on OLS with cluster fixed effects to identify peer effects – in spite of the obvious shortcomings, i.e., reflection bias and exclusion bias. We propose an alternative method to estimate peer effects that deals with these shortcomings but does not rely on instrumentation. This method comes at a cost, though: it requires assuming away correlated effects between peers. Whether or not this assumption is warranted depends on the specific context of the study. But even when correlated effects cannot be ruled out on a priori grounds, researchers can nonetheless use our method to derive peer effect estimates that are free of reflection and exclusion bias – as an additional check on their results.

A Relationship between σ_u^2 and σ_ϵ^2 in model without cluster FE

We can re-write equation 3 as follows:

$$u_{ik} = \bar{x}_{-i,k} - \bar{x}_{-i} = \frac{\left(\sum_{j=1}^{K} x_{jk}\right) - x_{ik}}{K-1} - \frac{\left(\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} x_{js}\right) - x_{ik}}{N-1}$$
$$= \frac{(N-K)\left[\left(\sum_{j=1}^{K} x_{jk}\right) - x_{ik}\right]}{(N-1)(K-1)} - \frac{\sum_{s \neq k} \sum_{j=1}^{K} x_{js}}{N-1}$$

Using $var(x_{ik}) = \sigma_\epsilon^2$ and the assumption that x_{ik} is i.i.d. (prior to treatment), we derive:

$$var(u_{ik}) = \sigma_u^2 = \frac{(N-K)^2(K-1)}{(N-1)^2(K-1)^2}\sigma_\epsilon^2 - \frac{(N-K)}{(N-1)^2}\sigma_\epsilon^2 = \frac{(N-K)}{(N-1)(K-1)}\sigma_\epsilon^2 < \epsilon_\epsilon^2 \quad (39)$$

B Deriving an expression for $var(\bar{x}_{-i,k})$ in model without cluster FE

Using the reduced form of $\bar{x}_{-i,k}$ provided in equation 9 , we obtain:

$$var(\bar{x}_{-i,k}) = var(\bar{x}_{-i} + u_{ik}) = var(\bar{x}_{-i}) + 2cov(\bar{x}_{-i}, u_{ik}) + var(x_{ik}) = \frac{(N-1)^2\sigma_u^2 + \sigma_\epsilon^2}{(N-1)^2}$$

Using equation (40) and equation (39) we then get:

$$var(\bar{x}_{-i,k}) = \frac{(N-1)(N-K) + (K-1)}{(N-1)^2(K-1)}$$

C Exclusion bias in cluster sampling

The cluster sampling equivalent of equation 7 is

$$x_{ikl} = \beta_0 + \beta_1 \left(\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - x_{ikl}}{L-1} + u_{ikl} \right) + \epsilon_{ikl} \quad (40)$$

Averaging equation 40 over all L observations in group l , we obtain the cluster average outcome:

$$\bar{x}_l = \beta_0 + \beta_1 \left[\frac{\sum_{i=1}^{L} \left(\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - x_{ikl}}{L-1} \right)}{L} + \bar{u}_l \right] + \bar{\epsilon}_l = \beta_0 + \beta_1 \left[\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - \bar{x}_l}{L-1} + \bar{u}_l \right] + \bar{\epsilon}_l \tag{41}$$

Note that cluster fixed effect equation 1 can be rewritten in terms of deviations of outcomes from their respective cluster averages, as follows

$$x_{ikl} - \bar{x}_l = \beta_1 \left(\bar{x}_{-i,k,l} - \bar{\bar{x}}_{-i,l} \right) + \left(\epsilon_{ikl} - \bar{\epsilon}_l \right)$$

Inserting 40 and 41, we derive the following expression for the cluster fixed effects model where peers are drawn from the cluster l:

$$x_{ikl} - \bar{x}_l = \beta_1 \left[\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - x_{ikl}}{L-1} + u_{ikl} - \left(\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - \bar{x}_l}{L-1} \right) - \bar{u}_l \right] + \epsilon_{ikl} - \bar{\epsilon}_l$$

$$\Leftrightarrow x_{ikl} - \bar{x}_l = \beta_1 \left(\frac{\bar{x}_l - x_{ikl}}{L-1} + u_{ikl} - \bar{u}_l \right) + \epsilon_{ikl} - \bar{\epsilon}_l$$

Denoting $\ddot{z} = x_{ikl} - \bar{x}_l$, for $z = x, u, \epsilon$, we have:

$$\ddot{x} = \beta_1 \left(\frac{-\ddot{x}}{L-1} + \ddot{u} \right) + \ddot{\epsilon} \tag{42}$$

Using the properties of the covariance and variance operators, we obtain the following expression for the cluster fixed effects estimate of β_1 when the true $\beta_1 = 0$:

$$E\left(\hat{\beta}_1^{FE}\right) = \frac{cov\left(\frac{-\ddot{x}}{L-1} + \ddot{u}, \ddot{\epsilon}\right)}{var\left(\frac{-\ddot{x}}{L-1} + \ddot{u}\right)} = \frac{cov\left(\frac{-\ddot{x}}{L-1}, \ddot{\epsilon}\right) + cov\left(\ddot{u}, \ddot{\epsilon}\right)}{var\left(\frac{-\ddot{x}}{L-1}\right) + 2cov\left(\frac{-\ddot{x}}{L-1}, \ddot{u}\right) + var\left(\ddot{u}\right)}$$

In order to expand equation 43, we consider:

$$cov\left(\ddot{u}, \ddot{\epsilon}\right) = E\left(\ddot{u}\ddot{\epsilon}\right) = E\left[\left(u_{ikl} - \bar{u}_l\right)\left(\epsilon_{ikl} - \bar{\epsilon}_l\right)\right] = E\left(u_{ikl}\epsilon_{ikl}\right) - E\left(\bar{u}_l\epsilon_{ikl}\right) + E\left(\bar{u}_l\bar{\epsilon}_l\right) - E\left(u_{ikl}\bar{\epsilon}_l\right) = 0 \tag{43}$$

and

$$var\left(\ddot{u}\right) = var\left(u_{ikl} - \bar{u}_l\right) = var\left(u_{ikl}\right) - 2E\left(u_{ikl}\bar{u}_l\right) + var\left(\bar{u}_l\right)$$

Since u_{ik} are self-constructed deviations of $\bar{x}_{-i,k}$ from the pool average $\bar{x}_{-i}$, by construction $\bar{u}_l = 0$. Therefore:

$$\begin{cases} var(\bar{u}_l) & = 0 \\ E\left(u_{ikl}\bar{u}_l\right) & = 0 \end{cases} \Rightarrow var(\ddot{u}) = var(u_{ikl}) = \sigma_u^2 \tag{44}$$

Furthermore, note that:

$$\begin{cases} E\left(\epsilon_{ikl}\bar{\epsilon}_l\right) & = \frac{E\left(\epsilon_{ikl}^2\right)}{L} = \frac{\sigma_\epsilon^2}{L} \\ var\left(\bar{\epsilon}_l\right) & = var\left(\frac{\sum_{i=1}^L \epsilon_{ikl}}{L}\right) = \frac{\sum_{i=1}^L var(\epsilon_{ikl})}{L^2} = \frac{L\sigma_\epsilon^2}{L^2} = \frac{\sigma_\epsilon^2}{L} \end{cases} \tag{45}$$

$$\Rightarrow var\left(\ddot{\epsilon}\right) = \sigma_\epsilon^2 - 2\frac{\sigma_\epsilon^2}{L} + \frac{\sigma_\epsilon^2}{L} = \frac{(L-1)\sigma_\epsilon^2}{L} \tag{46}$$

Using equation 42 , we derive the reduced form of $\left(-\frac{\ddot{x}}{L-1}\right)$:

$$\begin{aligned} \ddot{x} = \beta_1\left(\frac{-\ddot{x}}{L-1} + \ddot{u}\right) + \ddot{\epsilon} \Leftrightarrow \left[\frac{L-1+\beta_1}{L-1}\right]\ddot{x} = \beta_1\ddot{u} + \ddot{\epsilon} \\ \Leftrightarrow -\frac{\ddot{x}}{L-1} = \frac{-\beta_1\ddot{u}}{L-1+\beta_1} - \frac{\ddot{\epsilon}}{L-1+\beta_1} \end{aligned} \tag{47}$$

Using $E(\ddot{\epsilon}) = E(\epsilon_{ikl} - \ddot{\epsilon}_l) = 0$, equations 43 , 46, 47 and $\beta_1 = 0$ (since we consider the case of random peer assignment), we derive:

$$\begin{aligned} cov\left(\frac{-\ddot{x}}{L-1}, \ddot{\epsilon}\right) &= E\left[\left[\frac{-\ddot{x}}{L-1} - E\left(\frac{-\ddot{x}}{L-1}\right)\right]\ddot{\epsilon}\right] \\ &= E\left[\frac{-\ddot{\epsilon}\ddot{\epsilon}}{L-1}\right] \\ &= \frac{-var(\ddot{\epsilon})}{L-1} \\ &= -\frac{1}{L-1}\frac{(L-1)\sigma_\epsilon^2}{L} \\ &= -\frac{\sigma_\epsilon^2}{L} \end{aligned} \tag{48}$$

Similarly,

$$2cov\left(\frac{-\ddot{x}}{L-1}, \ddot{u}\right) = -2\frac{E(\ddot{u}\ddot{\epsilon})}{L-1} = 0 \tag{49}$$

Again using equation 47 , we have:

$$
\begin{aligned}
var\left(\frac{-\ddot{x}}{L-1}\right) &= var\left(-\frac{\ddot{\epsilon}}{L-1}\right)\\
&= \frac{\frac{(L-1)}{L}\sigma_\epsilon^2}{(L-1)^2}\\
&= \frac{\sigma_\epsilon^2}{L(L-1)}
\end{aligned} \tag{50}
$$

Using equations 43 - 50 we obtain:

$$
cov(\frac{-\ddot{x}}{L-1}+\ddot{u},\ddot{\epsilon}) = (-\frac{\sigma_\epsilon^2}{L}) \tag{51}
$$

and

$$
var(\frac{-\ddot{x}}{L-1}+\ddot{u}) = \frac{\sigma_\epsilon^2}{L(L-1)} + \sigma_u^2 \tag{52}
$$

Similar to the procedure used in Appendix (A) we can derive an expression for σ_u^2 but now for a model that includes cluster fixed effects. We obtain:

$$
\sigma_u^2 = \frac{(L-K)}{(L-1)(K-1)} var(\ddot{\epsilon})
$$

Using equation ((46)) we obtain the following relationship between σ_u^2 and σ_ϵ^2 in a model that includes cluster fixed effects:

$$
\sigma_u^2 = \frac{(N-K)(L-1)}{(N-1)(K-1)L}\sigma_\epsilon^2 \tag{53}
$$

Substituting this into equation (52) we obtain:

$$
var(\frac{-\ddot{x}}{L-1}+\ddot{u}) = \frac{\sigma_\epsilon^2}{L(L-1)} + \frac{(N-K)(L-1)}{(N-1)(K-1)L}\sigma_\epsilon^2 = \frac{(K-1)+(L-K)(L-1)}{L(L-1)(K-1)}\sigma_\epsilon^2
$$

Finally, we can expand equation 43 as follows:

$$
\begin{aligned}
E\left(\hat{\beta}_1^{FE}\right) &= \frac{cov\left(\frac{-\ddot{x}}{L-1}+\ddot{u},\ddot{\epsilon}\right)}{var\left(\frac{-\ddot{x}}{L-1}+\ddot{u}\right)} = \frac{(-\frac{\sigma_\epsilon^2}{L})}{\frac{(K-1)+(L-K)(L-1)}{L(L-1)(K-1)}\sigma_\epsilon^2}\\
\Rightarrow E\left(\hat{\beta}_1^{FE}\right) &= -\frac{(L-1)(K-1)}{(L-1)(L-K)+(K-1)}
\end{aligned} \tag{54}
$$

D Proof proposition 2

When data are clustered at a sub-level smaller than the total sample size N, the pooled OLS estimator $\hat{\beta}_1^{OLS}$ (i.e. OLS in a model omitting cluster fixed effects) is a weighted average of the cluster fixed effects estimator $\hat{\beta}_1^{FE}$ (or *within* estimator) and the *between* estimator $\hat{\beta}_1^{BE}$ (Raudenbush and Bryk [2002]):

$$\hat{\beta}_1^{OLS} = \eta^2 \hat{\beta}_1^{BE} + (1 - \eta^2)\hat{\beta}_1^{FE} \tag{55}$$

where $0 < \eta^2 < 1$ is the ratio of the between sum of squares of the independent variable of interest, $\bar{y}_{-i,k,l}$, to its total sum of squares. The cluster fixed effects estimator was derived in Appendix C where we found that $E\left(\hat{\beta}_1^{FE}\right) < 0$. The between estimator is the OLS estimator from a regression of $\bar{y}_l$ on an intercept and $\bar{\bar{y}}_{-i,l}$, where $\bar{y}_l$ denotes the average outcome y_{ikl} over the individuals in group l and $\bar{\bar{y}}_{-i,l}$ denotes the average peer group outcome of the group:

$$\bar{y}_{-i,l} = \beta_0 + \beta_1 \bar{\bar{y}}_{-i,l} + \bar{\epsilon}_l \tag{56}$$

In this Appendix we will derive this estimator and show that $E\left(\hat{\beta}_1^{BE}\right) \geq 0$. Given that $E\left(\hat{\beta}_1^{FE}\right) < 0$ and $E\left(\hat{\beta}_1^{BE}\right) \geq 0$, equation 55 implies that:

1. When peers are selected from within the cluster $l \subset \Omega$, we expect there to be a positive correlation between the cluster average outcome and the average peer group outcome in the cluster, that is, $\beta_1 > 0$ in 56. In this case, $E\left(\hat{\beta}_1^{BE}\right) > 0$ and the OLS estimate will lie somewhere in between the negative FE estimate and the positive between-group estimate and

$$E(\hat{\beta}_1^{FE}) < E(\hat{\beta}_1^{OLS})$$

2. When peers are selected among the entire population Ω, we do not expect there to be any correlation between the cluster average outcome and the average peer group outcome in the cluster. In this case, $\beta_1 = 0$ in 56 and $E\left(\hat{\beta}_1^{BE}\right) = 0$. This implies:

$$E(\hat{\beta}_1^{FE}) = E(\hat{\beta}_1^{OLS})$$

.

This proves proposition 2.

We will now derive an expression for $E\left(\hat{\beta}_1^{BE}\right)$ and prove that $E\left(\hat{\beta}_1^{BE}\right) > 0$ when peers are selected from within the cluster.

The between-group model equivalent of the reduced form equation given in (8) is:

$$\bar{y}_{-i,l} = \frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} y_{jsl}\right] - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\epsilon}_l}{L - 1 + \beta_1}$$
$$= \frac{L\bar{y}_l - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\epsilon}_l}{L - 1 + \beta_1} \quad (57)$$

where $\bar{y}_{-i,l}$ is the average outcome over the individuals in the group l , excluding individual i, and $\bar{y}_l$, $\bar{u}_l$ and $\bar{\epsilon}_l$ denote the group averages of y , u and ϵ , respectively. Under random peer assignment (i.e. $\beta_1 = 0$), this equation reduces to:

$$\bar{y}_{-i,l} = \frac{L\bar{y}_l}{L - 1} - \frac{\bar{\epsilon}_l}{L - 1} \quad (58)$$

Using (58), we have:

$$cov(\bar{\bar{y}}_{-i,l}, \bar{\epsilon}_l) = cov(\bar{y}_{-i,l} + \bar{u}_l, \bar{\epsilon}_l)$$
$$= cov(\bar{y}_{-i,l}, \bar{\epsilon}_l)$$
$$= L\frac{E(\bar{\epsilon}_l^2)}{L - 1} - \frac{E(\bar{\epsilon}_l^2)}{L - 1}$$
$$= var(\bar{\epsilon}_l)$$
$$= \frac{\sigma_\epsilon^2}{L} \quad (59)$$

and

$$
\begin{aligned}
var\left(\bar{\bar{y}}_{-i,l}\right) &= var\left(\frac{\sum_{i=1}^{L} \bar{y}_{-i,k,l}}{L}\right) \\
&= \frac{1}{L^2} var\left(\sum_{i=1}^{L}\left(\frac{\sum_{j=1}^{L} y_{jl} - y_{il}}{L-1}\right) + \sum_{i=1}^{L} u_{il}\right) \\
&= \frac{1}{L^2} var\left(\frac{L\sum_{i=1}^{L} y_{il}}{L-1} - \frac{\sum_{i=1}^{L} y_{il}}{L-1} + \sum_{i=1}^{L} u_{il}\right) \\
&= \frac{1}{L^2} var\left(\sum_{i=1}^{L} y_{il} + \sum_{i=1}^{L} u_{il}\right) \\
&= \frac{\sigma_\epsilon^2 + \sigma_u^2}{L} \\
&= \frac{\sigma_\epsilon^2 + \frac{L(L-K)}{(L-1)^2(K-1)}\sigma_\epsilon^2}{L} \\
&= \frac{(L-1)^2(K-1) + L(L-K)}{L(L-1)^2(K-1)}\sigma_\epsilon^2 \qquad (60)
\end{aligned}
$$

where in the last step we used the result in (5).
Using equation (59) and (60) we obtain:

$$
\begin{aligned}
E\left(\hat{\beta}_1^{BE}\right) &= \frac{cov\left(\bar{\bar{y}}_{-i,l}, \bar{\epsilon}_l\right)}{var\left(\bar{\bar{y}}_{-i,l}\right)} \\
&= \frac{\frac{\sigma_\epsilon^2}{L}}{\frac{(L-1)^2(K-1)+L(L-K)}{L(L-1)^2(K-1)}\sigma_\epsilon^2} \\
&= \frac{(L-1)^2(K-1)}{(L-1)^2(K-1) + L(L-K)} > 0 \qquad (61)
\end{aligned}
$$

This proves that $E\left(\hat{\beta}_1^{BE}\right) > 0$ when peers are selected from within the cluster.

We can also use 55 to prove that corrolary that comes with Proposition 2, i.e. that $\hat{\beta}_1^{OLS}$ tends to zero for large sample sizes. To proceed, we need expressions for $\hat{\beta}_1^{FE}$, $\hat{\beta}_1^{BE}$ and η^2. The within estimator $\hat{\beta}_1^{FE}$ and the between estimator $\hat{\beta}_1^{BE}$ are presented in (14) and 61, respectively. We will now derive an expression for η^2.

Weight parameter η^2 in equation (62) is the ratio of the between-group sum of squares of the independent variable of interest, $\bar{y}_{-i,k,l}$, to its total sum of squares:

$$\eta^2 = \frac{SS^{BG}_{\bar{y}_{-i,k,l}}}{SS^{Total}_{\bar{y}_{-i,k,l}}} = \frac{SS^{BG}_{\bar{y}_{-i,k,l}}}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}} \tag{62}$$

Specifically, $SS^{BG}_{\bar{y}_{-i,k,l}}$ is the sum of all the squared differences between each of the cluster group means and the overall sample mean, multiplied by the number of observations in the group l. In other words:

$$SS^{BG}_{\bar{y}_{-i,k,l}} = SS^{BE}_{\bar{y}_{-i,k,l}} \times L \tag{63}$$

where $SS^{BE}_{\bar{y}_{-i,k,l}}$ is the sum of squares of $\bar{\bar{y}}_{-i,l}$ in the between estimation regression (46) in Appendix (C). Furthermore, using the definition of the variance operator, we know that:

$$var\left(\bar{\bar{y}}_{-i,l}\right) = \frac{SS^{BE}_{\bar{y}_{-i,k,l}}}{\left(\frac{N}{L} - 1\right)} \Rightarrow SS^{BE}_{\bar{y}_{-i,k,l}} = var\left(\bar{\bar{y}}_{-i,l}\right) \times \left(\frac{N}{L} - 1\right) \tag{64}$$

Using equations (62) - (64), we obtain:

$$SS^{BG}_{\bar{y}_{-i,k,l}} = var\left(\bar{\bar{y}}_{-i,l}\right) \times \left(\frac{N}{L} - 1\right) \times L$$

Substituting in for the expression of $var\left(\bar{\bar{y}}_{-i,l}\right)$ given by equation (60) , we finally have:

$$SS^{BG}_{\bar{y}_{-i,k,l}} = \frac{(L-1)^2(K-1)}{(L-1)^2(K-1) + L(L-K)} \times \left(\frac{N}{L} - 1\right) \times \sigma^2_{\epsilon} \tag{65}$$

Next, $SS^{Within}_{\bar{y}_{-i,k,l}}$ is the sum of the squared differences between each individual's average peer group outcome, $\bar{y}_{-i,k,l}$, and its average for the individual's group $\bar{\bar{y}}_{-i,l}$. Similarly to equation (64), we have:

$$var\left(\bar{y}_{-i,l} - \bar{\bar{y}}_{-i,l}\right) = \frac{SS^{Within}_{\bar{y}_{-i,k,l}}}{(N-1)} \Rightarrow SS^{Within}_{\bar{y}_{-i,k,l}} = var\left(\bar{y}_{-i,l} - \bar{\bar{y}}_{-i,l}\right) \times (N-1)$$

From the above we know that $var\left(\bar{y}_{-i,k,l} - \bar{\bar{y}}_{-i,l}\right) = var\left(\frac{-\ddot{y}}{L-1} + \ddot{u}\right)$. Therefore, we can substitute in for the expression of $var\left(\bar{y}_{-i,k,l} - \bar{\bar{y}}_{-i,l}\right)$ by using equations (52):

We have:

$$SS^{Within}_{\bar{y}_{-i,k,l}} = \frac{L + (L-K)(K-1)}{K(K-1)L} \times \frac{N-1}{L}\sigma^2_{\epsilon} \tag{66}$$

Combining equations (62), (65) and (66) , we obtain:

$$\eta^2 = \frac{SS^{BG}_{\bar{y}_{-i,k,l}}}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}}$$

where

$$\begin{cases} SS^{BG}_{\bar{y}_{-i,k,l}} & = \frac{(L-1)^2(K-1)}{(L-1)^2(K-1)+L(L-K)} \times \left(\frac{N}{L}-1\right) \times \sigma^2_\epsilon \\ SS^{Within}_{\bar{y}_{-i,k,l}} & = \frac{L+(L-K)(K-1)}{K(K-1)L} \times \frac{N-1}{L} \times \sigma^2_\epsilon \end{cases}$$

Finally, denoting as constants $A = \frac{(L-1)^2(K-1)}{(L-1)^2(K-1)+L(L-K)}\sigma^2_\epsilon$ and $B = \frac{L+(L-K)(K-1)}{K(K-1)L}$ and taking probability limits, we obtain the following expression for $plim(\eta^2)$:

$$\begin{aligned} plim\left(\eta^2\right) &= plim\left[\frac{A\left(\frac{N}{L}-1\right)}{A\left(\frac{N}{L}-1\right)+B\left(\frac{N-1}{L}\right)}\right] \\ &= \frac{A}{A+B} \end{aligned} \tag{67}$$

Note that this closed form result only holds in the limit, that is, when sample size N tends to infinity. Using (55), (14), (61) and (67) we now derive the large sample property of pooled OLS when peer group formation occurs at group level l and when the true $\beta = 0$:

$$\begin{aligned} plim\left(\hat{\beta}_1^{OLS}\right) &= plim(\eta^2)plim(\hat{\beta}_1^{BE}) + \left[1 - plim(\eta^2)\right] plim(\hat{\beta}_1^{FE}) \\ &= \left(\frac{A}{A+B}\right)\frac{1}{AL} - \left(1 - \frac{A}{A+B}\right)\frac{1}{BL} \\ &= 0 \end{aligned}$$

This proves the corollary that comes with Proposition 2 in Section 2.5.

Finally, we explain formally why in smaller sample sizes the exclusion bias is more present. Note that:

$$\begin{aligned} E\left(\eta^2\right) &= E\left(\frac{SS^{BG}_{\bar{y}_{-i,k,l}}}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}}\right) \\ &= E\left(SS^{BG}_{\bar{y}_{-i,k,l}}\right) E\left(\frac{1}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}}\right) + cov\left(SS^{BG}_{\bar{y}_{-i,k,l}}, \frac{1}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}}\right) \\ &= \frac{LK-2K+1}{L(L-1)} + cov\left(SS^{BG}_{\bar{y}_{-i,k,l}}, \frac{1}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}}\right) \\ &= plim\left(\eta^2\right) + cov\left(SS^{BG}_{\bar{y}_{-i,k,l}}, \frac{1}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}}\right) \end{aligned}$$

It is clear that $cov\left(SS^{BG}_{\bar{y}_{-i,k,l}}, \frac{1}{SS^{BG}_{\bar{y}_{-i,k,l}} + SS^{Within}_{\bar{y}_{-i,k,l}}}\right) < 0$. Therefore, we obtain:

$$0 < E\left(\eta^2\right) < plim\left(\eta^2\right) < 1$$

This means that, *ceteris paribus*, the smaller the sample size the more weight is given to the cluster FE estimator in the estimation of the pooled OLS estimate (see 55) and therefore the larger the effect of the exclusion bias will be in pooled OLS. Similarly, the larger the sample size, the more weight is given to the between-group estimator in the estimation of the pooled OLS estimate and therefore the smaller the effect of the exclusion bias will be.

References

J.D. Angrist. The perils of peer effects. *Labour Economics*, 30:98–108, 2014.

P. Bayer, R. Hjalmarsson, and D. Pozen. Building criminal capital behind bars: peer effects in juvenile corrections. *The quarterly journal of economics*, 124(1):105–147, 2009.

J. Bound, D. Jaeger, and R. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450, 1995.

Y. Bramouille, H. Djebbari, and B. Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55, 2009.

W. Brock and S. Durlauf. *Handbook of Econometrics*, pages 3299–3380. Amsterdam, 2001.

G. De Giorgi, M. Pellizzari, and S. Redaelli. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2 (2):241–275, 2010.

de Melo. Peer effects identified through social networks. evidence from uruguayan schools. Working Paper, Banco de Mexico, 2014.

E. Duflo and E. Saez. Participation and investment decisions in a retirement plan: the influence of colleagues' choices. *Journal of Public Economics*, 85(1):121–148, 2011.

J. Fletcher. Peer influence on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics*, 25(4): 1265–1286, 2012.

J.M. Fletcher and S.L. Ross. Estimating the effects of friendship networks on health behaviors of adolescents. Working Paper, National Bureau of Economic Research, MA, 2012.

E.L. Glaeser, B.I. Sacerdote, and J.A. Scheinkman. The social multiplier. *Journal of the European Economic Association*, 1(2-3):345–353, 2003. URL `http://post.economics.harvard.edu/hier/2002papers/2002list.html`.

J. Guryan, D. Kroft, and N. J. Notowidigdo. Peer effects in the workplace: evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 44(3):289–302, 2009.

T.J. Halliday and S. Kwak. What is a peer? the role of network definitions in estimation of endogenous peer effects. *Applied Economics*, 44(3):289–301, 2012.

E.A. Hanushek, J.F. Kain, J.M. Markman, and S.G. Rivkin. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18:527–544, 2003.

D. Krackhardt. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10:359–381, 1988.

C. Manski. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60(3):531–542, 1993.

C.F. Manski. Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3):115–136, 2000.

R.A Moffitt. *Social Dynamics*, chapter Policy interventions, low level equilibria, and social interactions., pages 45–82. MIT Press, Cambridge, MA, 2001.

K. Munshi. Social learning in a heterogeneous population: Technology diffusion in the indian green revolution. *Journal of Development Economics*, 73(1):185–215, 2004.

S. W. Raudenbush and A. S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, 2002.

B. Sacerdote. Peer effects with random assignment: results for dartmouth roommates. *Quarterly Journal of Economics*, 116(92):681–704, 2001.

M. Stevenson. Tests of random assignment to peers in the face of mechanical negative correlation: an evaluation of four techniques. Mimeo, 2015a. URL `https://www.law.upenn.edu/cf/faculty/mstevens/workingpapers/Tests%20for%20random%20assignment.pdf`.

M. Stevenson. Breaking bad: Mechanisms of social influence and the path to criminality in juvenile jails. Mimeo, 2015b. URL `AvailableatSSRN:http://ssrn.com/abstract=2627394orhttp://dx.doi.org/10.2139/ssrn.2627394`.

L.C. Wang. Peer effects in the classroom: Evidence from a natural experiment in malaysia. Mimeo, 2009.

D. Zimmerman. Peer effects in academic outcomes: evidence from a natural experiment. *The Review of Economics and Statistics*, 85(1):9–23, 2003.