

Exclusion Bias in the Estimation of Peer Effects

Bet Caeyers*

Marcel Fafchamps[†]

January 29, 2016

Abstract

This paper formalizes a listed [Guryan et al., 2009] but unproven source of estimation bias in social interaction models. This bias is driven by the systematic exclusion of an individual from her peer group in the computation of average peer group outcomes. After deriving an exact formula for the magnitude of the bias in models using non-overlapping peer groups, we discuss its underlying parameters. We demonstrate 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. We discuss the conditions under which the exclusion bias is aggravated when adding cluster fixed effects. Simulation results confirm all the theoretical predictions derived in this paper and illustrate how the bias affects inference and the interpretation of estimation results. To achieve consistent inference, we suggest correcting p-values using permutation methods. We provide a characterization of a generalized data generating process that can be used to consistently estimate all structural parameters in the model, both for models using peer groups and models using social network data. We also show the conditions under which two-stage least squares strategies do not suffer from exclusion bias. This may explain the counter-intuitive observation in the social interaction literature that OLS estimates of endogenous peer effects are often larger than their 2SLS counterparts.

1 Introduction

It is widely assumed that randomized peer assignment ensures that no ex ante systematic relationship exists between individuals and their peers. As a result, any observed correlation between outcomes ex post, conditional on common shocks, is attributed to peer effects (e.g. Sacerdote, 2001). However, as Guryan et al. [2009] have argued intuitively, even with random peer assignment a mechanical negative relationship exists between people’s characteristics and those of their peers. This is because individuals cannot be their own peers, so that the pool of potential peers from which an individual’s peers are drawn systematically excludes the individual herself. Consequently, the expected value of the pool of potential peers is negatively correlated with the characteristics of the individual. Using Monte Carlo simulations, Guryan et al. [2009] show how this correlation yields a downward bias in the ordinary least squares (OLS) estimate in a typical test of random peer assignment. This observation is the starting point of the analysis in this paper, of a bias in the estimation of peer effects that has largely been ignored to date - namely the ‘exclusion bias’.

*bet_c@ifis.org.uk, Institute for Fiscal Studies

[†]fafchamp@stanford.edu Stanford University

We demonstrate that the exclusion bias can seriously affect point estimates and inference when estimating peer effects, yielding a downward bias. This negative bias is on top of other well-known sources of (positive) bias such as the reflection bias and correlated effects (Manski, 1993; Brock and Durlauf, 2001; Moffitt, 2001). We show that the exclusion bias is not negligible, that it does not disappear asymptotically in large samples, that it is stronger when peer groups are large relative to the peer selection pool and when cluster fixed effects are added to the model. We offer a number of simple statistical solutions 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 strategy because 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 other 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 in Guryan et al. [2009] and derive the conditions under which OLS estimates in standard linear-in-means tests of random peer assignment are biased downwards. We derive a simple, exact formula for the magnitude of the exclusion bias, and discuss the underlying parameters that determine it. We note that the downward exclusion bias is conceptually different from other well-known sources of bias, such as attenuation bias associated with measurement error.

Second, we derive for random assignment to groups of size 2 an exact formula for the exclusion bias and reflection bias in standard 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 effects estimates. Using simulations, we generalize these findings to a general model of peer effects for groups of size greater than two.

Third, we demonstrate how exclusion bias is affected by the addition of cluster fixed effects. We show that exclusion bias becomes significantly more severe in models that include fixed effects at the level of the pool from which peers are drawn (e.g., classroom dummies). In this type of models, the bias does not disappear as the sample size tends to infinity.

Fourth, we offer several possible solutions to exclusion bias. The applicability of each solution depends on the type of data available to the researcher. We start by illustrating various limitations of the method suggested by Guryan et al. [2009]. We propose instead to rely on a particular type of bootstrapping alternatively called permutation method or randomization inference. We show that this method provides a way of conducting correct inference in spite of the bias. We also provide a characterization of a generalized data generating process that can be used to consistently estimate all structural parameters in the model for a large variety of network structures (e.g. non-overlapping peer groups or more complex social network data).

Finally, we show the conditions under which two-stage least squares (2SLS) estimation procedures do not suffer from exclusion bias. We note that successful 2SLS procedures require the availability of suitable strong instruments that are of a particular type and are consistent only in large samples (Bound et al., 1995). Our correction methods do not require any instruments and are valid even in small finite samples. Nevertheless, if the dataset available allows for consistent 2SLS estimation, then the latter has the additional advantage of potentially being able to eliminate the exclusion bias. This observation further contributes to the literature by possibly accounting for a counter-intuitive yet common finding in peer effects studies. Many studies on social interactions obtain 2SLS estimates of endogenous peer effects that are significantly larger than OLS estimates. This is counter-intuitive: one would expect OLS estimates to be biased upwards as a result of reflection bias (Manski, 2000), endogenous peer group formation, or un-

observed correlated effects (Brock and Durlauf, 2001; Moffitt, 2001). This paper provides a new explanation for this finding by showing that the negative exclusion bias that is present in OLS estimation can be eliminated when valid 2SLS estimation strategies are used.

The paper is organized as follows. In Section 2 we start off by examining the properties of the exclusion bias in a standard test of random peer assignment into groups. We provide the intuition behind the bias, derive an exact formula, discuss the underlying parameters and suggest various methods that can be used to correct for the bias. We conclude Section 2 by discussing the implications of the addition of cluster fixed effects to the model and show how this aggravates the bias compared to pooled OLS models. Section 3 then moves on to a treatment of the exclusion bias in the estimation of endogenous peer effects. In a simple model with group size equal to 2, we derive exact formulas for the exclusion bias and the reflection bias, which can be used to recover correct peer effect estimates from the naive OLS estimate. We illustrate how the permutation method can be used to correct p-values, to allow for correct inference. Next, we generalize the treatment of the exclusion bias in the estimation of endogenous peer effects, first to peer groups of size greater than 2 and then to the addition of control variables. Section 3 concludes by showing why 2SLS estimation strategies do not suffer from the bias. Section 4 and Section 5 consider the importance of the exclusion bias in more complex peer selection pools and in social network data (as opposed to non-overlapping, mutually exclusive peer groups). Section 6 concludes and discusses the implications of our findings for the literature on peer effects.

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 will refer to as ‘individuals’ – are assigned a number of peers. We focus here on assignment to non-overlapping, mutually exclusive groups because this 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 social networks in general. We revisit this point in Section 5 when we discuss a generalized model.

Here we imagine that a researcher has data on peer assignments and wishes to test whether assignment is random based on an observable variable 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 dorm to which the student is assigned, and k is the room in dorm l to which he or she 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), $\bar{x}_{-i,k,l}$. 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., dorm 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 individuals' characteristics and the average characteristics of 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 in the computation of average peer group outcomes. 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. [2002], Zimmerman [2003] and Duflo and Saez [2011], where students *within* a school are randomly assigned to a dormitory or to a work group.

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 – hence the name ‘exclusion bias’. In the first sub-section below, we consider the case where $P_i = \Omega$, i.e., when peers are drawn from the entire set, with no stratification. In the sub-section that follows, we introduce stratification. In that case, we follow current best practice and add pool fixed effects to regression model (1) and we show how this affects the magnitude of the exclusion bias. To keep the exposition simple, we assume that the size of each pool P_i is the same value N_P for all $i \in \Omega$. Section 4 discusses how the results extend to the more general case where selection pools differ in size.

2.2.1 Without stratification

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

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

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 from the pool of individuals from which peers are drawn. Let \bar{x}_{-i} be the average characteristic of the pool of N individuals. 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} :

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

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

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

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 B this relationship is:

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

which indicates that σ_u^2 is decreasing in K – the size of the peer group – and it is increasing in N_P – the size of the pool from which peers are drawn.

Returning to equation (4) for the specific case where $N_P = N$, we note that \bar{x}_{-i} is nothing but:

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

Equation (4) can thus be rewritten as:

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

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} :

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

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

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

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 :

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

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

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

2.2.2 With stratification

We now generalize this finding to 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 the subset l of Ω of size $N_P = L$. Testing random peer assignment is achieved by estimating regression (1), which we reproduce here:

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

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

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

Equation (13) is similar to equation (11), except that the magnitude of the exclusion bias depends on L instead of N . In other words, the exclusion bias in the 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. Proposition 1 summarizes these results.

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

following form

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

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

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

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 sub-cluster 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. $\frac{\Delta|bias|}{\Delta N_P} < 0$: *Ceteris paribus*, 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]. 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} . This accounts for the result.
2. $\frac{\Delta|bias|}{\Delta K} > 0$: *Ceteris paribus*, exclusion bias is more severe in datasets with larger peer groups. The magnitude of the bias is not linear in K , however. 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.¹ Intuitively, the larger the number of people in i 's peer group, the more closely the average peer group characteristic $\bar{y}_{-i,k}$ follows the average \bar{x}_{-i} of the pool from which peers are drawn. This implies that $\bar{x}_{-i,k}$ is more sensitive to the change in \bar{x}_{-i} that results from the exclusion of x_i . 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.

Proposition 1 also demonstrates that 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; is always negative; and it does not disappear when the true β_1 is zero.

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

2.3 Simulations

To illustrate the magnitude of the exclusion bias in a typical test of random peer assignment, we conduct a set of Monte Carlo simulations. The results presented here focus on the case of clustered randomization. Simulations vary in pool group size L and peer group size K . For each simulation we generate a dataset of 1000 observations and we randomly assign a normally distributed 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.² 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 zero.

The average of estimated $\hat{\beta}_1$'s over 100 replications is summarized in Table 1 for different values of K and L . For comparison purpose, we also report the theoretical $E(\hat{\beta}_1)$ from Proposition 1. These results confirm our theoretical predictions: the exclusion bias is large in magnitude, and is larger when the size K of the peer group K is large and when the size L of the pool is small. Table 1 also indicates, for each simulation, the proportion of repeated simulation rounds (100 in total) for which we reject the hypothesis at the 5% and 10% significance levels. Figure 1 illustrates the implications for inference more graphically for the simulation of $N = 1000$, $L = 20$ and $K = 5$ as an example, by plotting the rates at which we reject the null hypothesis for various significance levels and comparing them to the rejection rate that we would expect if the test was unbiased (i.e. the 45 degree line). We see that ignoring exclusion bias has a dramatic effect on inference: a researcher relying on regression (1) to test random peer assignment draws incorrect inference in a large proportion of cases.

2.4 Correction of the exclusion bias

We now discuss possible ways of obtaining correct inference when testing random peer assignment. Guryan et al. [2009] propose to control for differences in mean characteristic across selection pools by adding to equation (1) the mean characteristic $\bar{x}_{-i,l}$ of individuals other than i in selection pool l . The equation they propose is the following:

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

where φ is an additional parameter to be estimated.

To see how, under specific conditions, this effectively deals with exclusion bias, we substitute equation (15) in for equation (3) and rearrange as follows:

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

We see that the inclusion of $\bar{x}_{-i,l}$ in the regression acts as a proxy and soaks up the non-random component of $\bar{x}_{-i,k,l}$. As a result, 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, $plim(\hat{\beta}_1) = \beta_1$ and OLS yields consistent estimates of the peer effect β_1 .

²Note that the allocation of individuals to different clusters does not have to be random (and usually is not in practice) for our results about the exclusion bias to apply.

This method has some limitations, however. First, as noted by Guryan et al. [2009], parameters β_1 and φ are separately identified only if there is variation in pool sizes. 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 insufficient, resulting in multicollinearity and quasi-underidentification.

It is nonetheless possible to draw unbiased inference about random peer assignment without variation in L . We first note that, in the case of fixed pool size L , Proposition 1 can be used to calculate the average bias and thus to correct the point estimate of $\hat{\beta}_1$:

$$\hat{\beta}_1^{\text{corrected}} = \hat{\beta}_1 + \frac{K(K-1)}{L + (L-K)(K-1)}$$

If pool sizes differ, it is possible to use the generalized version of Proposition 1, which we will discuss later on in Section 4.

To conduct inference, we suggest to simulate the distribution of $\hat{\beta}_1$ under the null hypothesis of random peer assignment from the data at hand, just as we did in the simulations above.³ To illustrate how this method works, imagine the researcher has observational data y_{ikl} partitioned in groups of varying size K_i coming from pools of varying size N_{Pi} . This is illustrated in Table 2 below, 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 $\hat{\beta}_1$ under the null hypothesis as follows. To recall, the null hypothesis of random assignment within pools can be mimicked by reordering individuals within pools and by assigning them at random into groups. Formally, let us denote the vector of group sizes in pool l as $K^l \equiv [K_1^l, K_2^l, \dots, K_m^l]$ 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 \tilde{x}_{ikl} that is obtained by reordering x_{ikl} at random within pools, as shown in column \tilde{x}_{ikl} of Table 2. We then estimate regression (1) using \tilde{x}_{ikl} in lieu of x_{ikl} – and $\tilde{x}_{-i,k} = \sum_{j \neq i, j \in k} \tilde{x}_{j,k,l}$ in lieu of $\bar{x}_{-i,k}$. By repeating this process many times, we trace the distribution of $\hat{\beta}_1$ in the data if the null hypothesis is true. Each repetition yields a separate estimate of $\hat{\beta}_1^s$. The mean of the distribution of $\hat{\beta}_1^s$ is an estimate of the bias under the null. More importantly, the empirical distribution of $\hat{\beta}_1^s$ over the many 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 $\hat{\beta}_1^s$ that are above $\hat{\beta}_1$.

To illustrate the performance of this procedure, we generate artificial samples of 1000 observations under the null hypothesis of random assignment for three values of $K = \{2, 5, 10\}$. The size of each pool $L = 20$ and we posit $\epsilon_{ik} \sim N(1, 1)$. In columns 2 and 3 of Table 3, we present the estimate of $\hat{\beta}_1$ obtained in each of the three artificial samples together with the p -value obtained from applying OLS to equation (1). We note that, in all three cases, using the OLS p -value leads us to reject the null hypothesis of random assignment – strongly so when $K = 5$ or 10. Bootstrapping results are reported in the right panel of Table 2. In column 4 we present the corrected point estimate of $\hat{\beta}_1$. This is obtained by subtracting from the OLS estimate the average of the $\hat{\beta}_1^s$ obtained over 500 replications. Column 5 presents the corrected p -value obtained as the proportion of the 500 $\hat{\beta}_1^s$ estimates that are above $\hat{\beta}_1$.

Figure 2 presents these simulation results in graphical form by showing the distribution of the 500 $\hat{\beta}_1^s$ under the null hypothesis that $\beta_1 = 0$, for different peer group sizes. The histograms

³Simulations can also be used to obtain a close approximation of the distribution of $\hat{\beta}_1$ (and thus of the bias) under more complicated random assignment processes, e.g., multi-level stratification, and the like.

are clearly centered around the 'naive' estimates for β_1 shown in Table 2, not around the true $\beta_1 = 0$. The corrected p -values in Table 2 take this distributional shift into consideration.

2.5 Cluster fixed effects

In this section we discuss more in detail the role of cluster fixed effects (FE) in the magnitude of the exclusion bias. So far we have only considered the case when fixed effects are included at the level of the selection pool (see Section 2). This is a reasonable approach when testing random peer assignment, since pool FEs proxy for any variable on which randomization was conditioned. However, studies of peer effects - which we will turn to next - 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 FEs other than for selection pools. For instance, students in a school cohort may be randomly allocated to rooms (i.e., 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.

In this section we show, for the case when $\beta_1 = 0$ (i.e., no peer effect), that the inclusion of cluster fixed effects can exacerbate the exclusion bias. Similar results apply for the case when $\beta_1 > 0$. We compare two estimators, $\hat{\beta}_1^{OLS}$ which is obtained by estimating equation (2) using OLS; and $\hat{\beta}_1^{FE}$ which is obtained by estimating equation (12) with cluster fixed effects δ_l . We consider two possibilities: either (i) peers are selected at the level of the entire population Ω ; or (ii) peers are selected within clusters indexed by $l \subset \Omega$. For illustration purposes, we focus on the case where all clusters are of the same size L . In Appendix D we demonstrate the following proposition:

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

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

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

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

What the proposition states is that when cluster FEs are added at a level lower than the selection pool, the expected exclusion bias is unchanged. But when cluster FEs are included at the level at which peer selection occurs, the exclusion bias worsens. The intuition behind Proposition 2 is the following. Recall that the exclusion bias arises whenever individual i (i) is excluded from her potential peer group (selection without replacement) but (ii) is a potential peer for other observed individuals j . It is this combination of conditions that creates a negative correlation between x_{ikl} and $\bar{x}_{-i,k,l}$: if selection is done with replacement or if i cannot be a peer of her own peers, no correlation can arise between x_{ikl} and $\bar{x}_{-i,k,l}$ if assignment is random. The bias caused by the introduction of cluster FEs is driven by condition (ii). The logic is as follows. The omission of individual i from her selection pool creates a correlation between x_{ikl} and $\bar{x}_{-i,k,l}$ within each selection pool, but since individual i is not part of other selection pools, x_{ikl} is not correlated with the variation in $\bar{x}_{-i,k,l}$ that is due to differences in pool quality. When cluster FEs are omitted, identification of β_1 uses all the variation in x_{ikl} and $\bar{x}_{-i,k,l}$, including

variation in $\bar{x}_{-i,k,l}$ due to different pool quality. When cluster FEs are included, identification of β_1 uses only the variation in $\bar{x}_{-i,k,l}$ within each cluster. This mechanically exacerbates exclusion bias since identification of β_1 no longer relies on the part of the variation in $\bar{x}_{-i,k,l}$ that is not correlated with x_{ikl} . Table 4 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 corollary of the above proposition is that, keeping the size of each selection pool constant, $\lim_{N \rightarrow \infty} E(\hat{\beta}_1^{FE})$ 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 $\lim_{N \rightarrow \infty} E(\hat{\beta}_1^{OLS}) = 0$. The simulation results in Table 4 clearly illustrate this corollary. This implies omitting cluster FEs when testing random assignment leads to an asymptotic elimination of the exclusion bias. Including FEs for selection pools is therefore not necessarily a good idea, especially in large samples where the difference between $E(\hat{\beta}_1^{OLS})$ and $E(\hat{\beta}_1^{FE})$ is largest.

3 Estimating endogenous peer effects

In this section we illustrate how exclusion bias also affects the estimation of endogenous peer effects. To make this illustration as clear as possible, we ignore the possibility of multiple selection clusters (thus eliminating the need for cluster fixed effects) and we abstract from exogenous peer effects. We generalize the model further below. The linear-in-means peer effects model we seek to estimate has the following form:

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

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 will assume that 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 other correlated effects. We then generalize the approach for an arbitrary group size and show how non-linear least square estimation can be used to obtain an estimate of β_1 that is free of both reflection and exclusion bias. While this estimator is useful for illustration purposes, its practical usefulness is limited by the need to assume the absence of correlated effects, an assumption that may not be warranted in practice. We then discuss how other approaches can be used to address the reflection bias while also eliminating the exclusion bias.

3.1 Simple model with group size $K = 2$

3.1.1 Reflection bias

We start by ignoring exclusion bias 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. 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 driven by 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 3.3 we discuss an extension of this model to include other explanatory variables. 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. We have:

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

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

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

where we have used $E[\epsilon_1\epsilon_2] = 0$ to go from line 2 to line 3. The latter assumption will be relaxed in the next section, when we introduce exclusion bias to the equation. The covariance between y_1 and y_2 is given by:

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

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:

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

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:

$$y_1 = a + by_2 + v_1 \quad (16)$$

What is the relationship between the regression coefficient b and the structural parameter β In a univariate regression, there exists a close relationship between coefficient b and the correlation r between the dependent and independent variable:

$$\begin{aligned}
\hat{b} &= \hat{r} \frac{\sigma_{y_1}}{\sigma_{y_2}} \\
&= \hat{r} \text{ since } \sigma_{y_1} = \sigma_{y_2}
\end{aligned} \quad (17)$$

Hence we have:

$$E[\hat{b}] = \frac{2\beta}{1+\beta^2} \neq \beta \quad (18)$$

This expression gives a closed-form solution for the reflection bias in this simple example where we have ignored exclusion bias. Note that, based on this formula, $E[\hat{b}] = 0$ iff $\beta = 0$. This means we can in principle test whether $\beta = 0$ by testing whether $\hat{b} = 0$ in regression (16).

Formula (18) can also be used to recover a correct estimate of β from the naive \hat{b} . We have:

$$\begin{aligned}
\hat{b}\beta^2 - 2\beta + \hat{b} &= 0 \\
\hat{\beta} &= \frac{+2 \pm \sqrt{4 - 4\hat{b}^2}}{2\hat{b}} \\
&= \frac{1 - \sqrt{1 - \hat{b}^2}}{\hat{b}} \text{ because the other root is } > 1
\end{aligned} \quad (19)$$

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 elaborately 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:

$$E[\widehat{b}] = -\frac{1}{N-1} \equiv \rho \quad (20)$$

As per formula (17) 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 combination 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:

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

The covariance is:

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

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:

$$y_1 = a + by_2 + v_1 \quad (21)$$

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

$$\begin{aligned} E[\widehat{b}_e] &= \frac{\frac{\sigma_\epsilon^2(2\beta + (1 + \beta^2)\rho)}{(1-\beta^2)^2}}{\frac{\sigma_\epsilon^2(1 + \beta^2 + 2\beta\rho)}{(1-\beta^2)^2}} \\ &= \frac{2\beta + (1 + \beta^2)\rho}{1 + \beta^2 + 2\beta\rho} \neq \frac{2\beta}{1 + \beta^2} \neq \beta \end{aligned} \quad (22)$$

We present in Table 5 simple calculations based on the above formulas 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, however, the reflection bias takes over and leads to coefficient estimates that over-estimate the true β_1 . 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 $E[\widehat{b}_e] = \rho < 0$ which is the size of the exclusion bias. We cannot draw correct inference about $\beta = 0$ by looking directly at \widehat{b}_e . This is because \widehat{b}_e can be negative even when β is positive.
2. It is possible for $E[\widehat{b}_e]$ 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. However, since the exclusion bias is always negative, $\widehat{b}_e > 0$ can only arise if $\beta > 0$.

3.1.3 Inference

Under the assumption of independent errors, we can recover an estimate of β using the same approach as in (19). We can turn formula (22) into a quadratic expression in β . Taking roots, we obtain an estimate $\widehat{\beta}_e$ of the true β using the value of the exclusion bias ρ from (20) and the coefficient estimate \widehat{b}_e from regression (21):

$$\widehat{\beta}_e = \frac{1 - \widehat{b}_e \rho - \sqrt{1 + \widehat{b}_e^2 \rho^2 - \widehat{b}_e^2 - \rho^2}}{\widehat{b}_e - \rho} \quad (23)$$

This formula confirms that the exclusion bias 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}_e to exceeds 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 (21) is biased towards finding no significantly positive peer effects.⁴

While formula (23) 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 $\widehat{\beta}_e$ is significantly different from 0. In order to do correct inference, we need to correct p -values for the standard test of significance that $\beta = 0$. The solution is essentially the same as that used earlier. Under the null hypothesis that $\beta = 0$, there is no correlation between paired y 's, so coefficient estimates of β_e should be centered around $\rho < 0$. The question is how much sample variation around ρ is 'too much'. The idea is to simulate, using the sample data, the distribution of $\widehat{\beta}_e$ that would arise under the null. This is achieved by replicating the random assignment of peers followed 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.

To illustrate, we present the results of a Monte Carlo study in Table 6. 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 \widehat{b}_e for different values of β . We set $K = 2, L = 10$ and $N = 1000$. In columns 2 and 3 we report the estimates of \widehat{b}_e and the corresponding p -value as reported by OLS. In column 4 we report the corrected estimate $\widehat{\beta}_e$ obtained using formula (23). 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}_e is dramatically biased, sometimes yielding a significantly negative

⁴If the true value of $\beta < 0$, we cannot sign the direction of the bias.

estimate of β when the true β is close to zero, sometimes yielding an inflated estimate of β when reflection bias dominates. Corrected estimates $\widehat{\beta}_e$ do not display this pattern, however: they are largely 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 with uncorrelated errors

We used the $K = 2$ case to illustrate how reflection and exclusion bias combine to affect coefficient estimates. For the simple $K = 2$ 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 larger values of group size K . But provided that we are still willing to assume i.i.d. errors, it is still 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:

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i \quad (24)$$

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.

Regression model (24) can be written in matrix form:

$$Y = \beta G Y + \gamma X + \delta G X + \epsilon$$

Simple algebra yields:

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

from which we obtain:

$$\begin{aligned} E[YY'] &= E[(I - \beta G)^{-1}(\gamma X + \delta G X + \epsilon) (\gamma X + \delta G X + \epsilon)'(I - \beta G')^{-1}] \\ &= (I - \beta G)^{-1}E[(\gamma X + \delta G X)(\gamma X + \delta G X)'](I - \beta G')^{-1} \\ &\quad + (I - \beta G)^{-1}E[\epsilon \epsilon'](I - \beta G')^{-1} \end{aligned} \quad (25)$$

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 (25) can be used as starting point for estimation. With exclusion bias, however,

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

Fortunately, formula (20) can be used to derive the covariance matrix of the u 's. In this case $E[\epsilon \epsilon']$ is a block-diagonal matrix of the form:

$$E[\epsilon \epsilon'] = \begin{bmatrix} E[\epsilon_1^2] & E[\epsilon_1 \epsilon_2] & \dots \\ E[\epsilon_2 \epsilon_1] & E[\epsilon_2^2] & \dots \\ \dots & \dots & \dots \end{bmatrix} \quad (26)$$

We have shown that, for two individuals i and j in the same selection pool of size L , we have $E[\epsilon_i \epsilon_j] = \rho = -\frac{1}{L-1}$. Hence $E[\epsilon \epsilon']$ can be rewritten more simply as:

$$E[\epsilon \epsilon'] = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho & \dots \\ \rho & 1 & \dots \\ \dots & \dots & \dots \end{bmatrix} \equiv \sigma_\epsilon^2 A \quad (27)$$

where the value of ρ is given by formula (20). What is important is that ρ is known, it does not need to be estimated. Pre- and post-multiplying by $(I - \beta G)^{-1}$ in expression (25) only picks within-group off-diagonal elements.

Equation (20) combined with (26) and (27) provides a characterization of the data generating process that can be used to estimate structural parameters β, γ, δ and σ^2 . As in the simple example presented earlier, identification is achieved from the assumption that errors are independent across observations – 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 (25) is to rely on maximum likelihood. The form of the resulting likelihood function is similar to that used in spatial econometrics (e.g. Drukker et al., 2013), the only difference being exclusion bias which we correct for. Another, simpler approach is to rely on non-linear least squares. In this case, we choose the parameters β that provides the best fit to the observed data $E[YY']$. This is achieved using a search algorithm. For each guess $\beta^{(n)}$ the algorithm makes regarding β , we can easily solve for the corresponding values of γ and δ simply by calculating $Y - \beta^{(n)}GY$ and regressing it on X and GX to obtain estimates of γ and δ . This process also yields a corresponding estimate of the variance of errors σ_ϵ^2 .

To illustrate the effectiveness of this approach, we estimate model (24) on simulated data, using 100 Monte Carlo replications. We keep the number of observations in each sample constant – $N = 100$ – but we vary the value of K and the value of β . To better illustrate the role of exclusion bias, we set the number of groups within a cluster to a small value (i.e., 2). Cluster fixed effects are included throughout. In the first row we report the uncorrected $\hat{\beta}^{OLS}$ obtained by regressing Y_i on $G_i Y$ and cluster fixed effects. Results confirm that the uncorrected β is biased. This bias results from the combination of two sources of bias: reflection bias, and exclusion bias. When β is small, the exclusion bias tends to dominate our model and the naive β underestimate the true β . The naive β is more likely to overestimate the true β when exclusion bias is small, which occurs when ρ is small (e.g., when each selection pool is large). In Table 7, the exclusion bias dominates everywhere except in the last column. In the second row, we report the $\hat{\beta}^{RC}$ estimate corrected for reflection bias but ignoring the exclusion bias. This is the estimate derived from model (25) 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 $\hat{\beta}^{EC}$ estimate derived from model (25) with $E[\epsilon \epsilon']$ given by (27). Even with our small N , $\hat{\beta}^{EC}$ is centered around its true value in all cases.

We also note that, when $\beta = 0$, $\hat{\beta}^{OLS}$ and $\hat{\beta}^{RC}$ often yield incorrect inference – e.g., concluding that $\beta < 0$ when it actually is 0, or concluding that $\beta < 0$ when the true $\beta > 0$.

3.3 Control variable approach

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

At $t + 1$ golfers participating to tournament l are assigned to a peer group k with which they play throughout the tournament. The performance of golfer i in tournament l is written $y_{ikl,t+1}$. 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 wishes to estimate coefficient β_1 in a regression of the form:

$$y_{ikl,t+1} = \beta_0 + \beta_1 \bar{y}_{-i,klt} + \delta_l + \epsilon_{ikl,t+1} \quad (28)$$

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, y_{iklt} is positively correlated with $y_{ikl,t+1}$. This generates a negative exclusion bias in regression (28): 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.

If the researcher does not have panel data, the exclusion bias may nonetheless be reduced by using individual characteristics of i as control variables to soak up some of the variation in y_{ikl} . How successful this approach may be in practice is unclear, although simulations (not reported) indicate that the reduction in the magnitude of the exclusion bias in models can be sizable when adding control variables that have a high correlation with y_{ikl} (e.g., 0.8). The improvement is negligible, however, when the correlation is small (e.g., 0.2).

3.4 Instrumental variable approach

So far we have assumed away the possibility of so-called ‘correlated effects’, that is, that errors ϵ_i ’s are correlated ex ante, that is, before introducing reflection and exclusion bias. Correlated effects can arise for a variety of reasons, such as unobserved characteristics and shocks that are common to individuals in the same peer groups. These correlated effects generate a correlation in the outcomes y_i ’s in the same peer group even in the absence of peer effects. If we erroneously assume away the possibility of correlated effects, we will attribute to peer effects what is in fact due to unobserved correlated effects. For this reason, researchers often rely on instrumental variables when estimating peer effect models – which is the method recommended by Manski (1993).

We show here that the use of instrumental variables can - under certain conditions - de facto eliminate the exclusion bias. In particular, exclusion bias disappears if one controls for i 's own value of the instrument whilst using the average peers' value as an instrument for $\bar{y}_{-i,kl}$.

To illustrate this formally, we 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 and is therefore often included in the estimated regression.

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 see that $cov(\hat{\bar{y}}_{-i,kl}, \epsilon_{ikl} | z_{ikl}) = 0$ and therefore $\hat{\beta}_1^{2SLS}$ does not suffer from exclusion bias. We have:

$$\begin{aligned}y_{ikl} &= \beta_0 + \beta_1 \hat{\bar{y}}_{-i,kl} + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \\ &= \beta_0 + \beta_1 (\hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{-i,kl} + \hat{\pi}_2 z_{ikl} + \hat{\delta}_l) + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl}\end{aligned}\tag{29}$$

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 $\bar{z}_{-i,kl} = \frac{\sum_{s=1}^N \sum_{j=1}^K z_{js}}{L-1} - z_{ikl} + \tilde{u}_{ikl}$, where \tilde{u} is defined in the same manner for z as u was defined for y in equation (3): $z_{-i,k} = z_{-i} + \tilde{u}_{ik}$. However, equation (29) controls for z_{ikl} which prevents this mechanical relationship 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; 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 been attributed to classical measurement error, or to the local average treatment interpretation of 2SLS.

Note, however, that for a 2SLS strategy to effectively eliminate the exclusion bias, it is important that one controls for i 's own value of the instrument, i.e. z_{ikl} , in (29). Estimation strategies employed in Bramouille et al. [2009] and De Giorgi et al. [2010], for instance, successfully satisfy these criteria. Any IV method that fails to do so will suffer from exclusion bias in the same way and for the same reasons as OLS does.

Moreover, valid 2SLS estimation strategies require the availability of suitable strong instruments and are 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 use OLS and rely on cluster fixed effects to identify peer effects. In order to address exclusion bias, such settings could make use of the correction methods suggested in this paper, which do not require any instruments and which are valid even in small finite samples.

4 More complex peer selection pools

In practice, determining the magnitude of the exclusion bias is not always as straightforward as in the models considered so far. Until now, 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 selection pools.

1. In the cluster sampling models considered 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 the exclusion bias in a pool fixed effects model is given by the following expression when $\beta_1 = 0$:

$$E(\hat{\beta}_1^{FE}) = \frac{1}{\left[\sum_{l=1}^{\frac{N}{L}} \frac{L_l + (L_l - K)(K-1)}{K(K-1)} \right]} \quad (30)$$

where L_l denotes the size of selection pool l . Intuitively, expression (30) results from assigning different weights to sampled observations, where weights are given by the share $\frac{L_l}{N}$ of i 's selection pool l in the total sample.

2. Peers may not be selected from mutually exclusive selection pools. For instance, students tend 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 friends are selected – so that selection pool and cluster are no longer synonymous. In the example above, 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 his peers from within cluster l and a proportion $(1 - \theta)$ from the population but outside cluster l . In the previous sections we considered cases where $E(\bar{y}_{-i,k,l}) = \bar{y}_{-i,l}$. Now $E(\bar{y}_{-i,k,l})$ follows:

$$E(\bar{y}_{-i,k,l}) = \theta \bar{y}_{-i,l} + (1 - \theta) \bar{y}_{-l,\Omega} \quad (31)$$

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 falls 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 ($\theta = 1$). The intuition is as follows. 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.⁵

⁵Note 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.

3. A third complication is that, in practice, the pool of potential peers is not always well defined, especially in non-randomized studies. 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).⁶ 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, 2011). 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.
4. Fourth, even if the pool of potential peers is precisely known, in non-randomized studies it is seldom the case that peers can 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 that we described earlier. How difficult this would be in practice depends on the nature of the posited peer assignment process.

5 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 do not form groups but more general networks, i.e., such that i and k need not be peers. To illustrate this, let us go back to the canonical case considered in Section 2, namely, let us assume that individuals in cluster l are randomly assigned peers within that cluster. The only difference with Section 2 is that we no longer restrict attention to peer groups but allow the network of links between peers to take an arbitrary shape within each cluster l .

5.1 Estimation

The approach developed to estimate general group models with uncorrelated errors can be further generalized to groups of unequal size and to arbitrary network data. Equation (25) remains the same. The only thing that changes is the shape of the adjacency matrix G . This is true even though, for arbitrary network data, it is in general impossible to obtain a closed-form expression for the exclusion bias when estimating regression (24). Pre- and post-multiplying matrix $E[\epsilon \epsilon']$ by $(I - \beta G)^{-1}$ in expression (25) picks the relevant off-diagonal elements to construct the needed correction for exclusion bias.

We illustrate this approach in Table... [work in progress]

⁶If it is further assumed that peers are as likely to be selected from within the classroom as from outside the classroom, we would have $\theta = \frac{l}{N}$

5.2 Inference

Formally let $g_{ijl} = 1$ if i and j in cluster l are peers and 0 otherwise.⁷ The network matrix in cluster l is written $G_l = [g_{ijl}]$. We follow convention and set $g_{ii} = 0$ always. We need a way of expressing the average outcome of i 's peers. To this effect, let n_{il} denote the number of peers (or degree) of i . The value of n_{il} may differ across individuals. Let us define vector \widehat{g}_{il} as a vector formed by dividing i 's row in G_l by n_{il} , i.e.:

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

where, as before, L denotes the size of the cluster. For instance, let $L = 4$ and assume that individual 1 has individuals 2 and 4 as peers. The $\widehat{g}_{1l} = [0, \frac{1}{2}, 0, \frac{1}{2}]$. With this notation, the average outcome of the peers of i can be written as $\widehat{g}_{il}y_l$ where y_l is the vector of all outcomes in cluster l . With this notation, the peer effect model that we aim to estimate becomes:

$$y_{ikl} = \beta_0 + \beta_1 \widehat{g}_{il}y_l + \delta_l + \epsilon_{ikl} \quad (32)$$

It is intuitively clear that exclusion bias affects model (32) 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. Although it may not be possible to derive a closed-form solution for this correlation, permutation methods can be used to simulate it when the null hypothesis of $\beta_1 = 0$ is satisfied.

To recall, we want to simulate the counterfactual distribution of $\widehat{\beta}_1$ under the null hypothesis of zero peer effects. The difference with Section 3 is that peers are not selected by randomly partitioning individuals into groups within clusters, but rather by randomly assigning peers within clusters. For simulations to yield a $\widehat{\beta}_1$ distribution that can serve as counterfactual for regression model of (32), the network matrices in each cluster must be the same. In other words, the random assignment of peers within clusters must preserve the network matrix structure. 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 $\{1, \dots, i, \dots, j, \dots, L\}$. We generate a random reordering (k) of these indices, e.g. $\{j, \dots, 1, \dots, L, \dots, i\}$. 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 (32) 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 illustrate the outcome of this simulation approach in Table... [work in progress]

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 unproven and unstudied 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

⁷Extension to a directed network is left as an exercise for the reader.

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 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.

The finding that the exclusion bias becomes more severe as the number of peers in the reference group grows implies that comparisons of estimates between models that vary in peer group size can be misleading. For instance, Glaeser, Sacerdote, and Scheinkman [2002] compare estimates of peer effects on test scores of students in Dartmouth College for three different units of peer group aggregation: The dormitory room (containing on average 2.3 students), the dormitory floor (containing on average eight students) and the dormitory (containing on average 27 students). According to this study, the estimated peer effect decreases with the level of aggregation. Although this finding may reflect diminishing peer effects as peers become more distant from each other, it may also - at least partially - be driven by exclusion bias. Similarly, Halliday and Kwak [2012], whose specific aim is to compare peer effects estimates for different definitions of peer groups in the education literature, find that the estimated peer effect is significantly smaller when school grade cohorts are used instead of smaller circles of friends. Note that this result also applies to any heterogeneity analysis of peer effects across cohorts that differ in peer group size. For instance, if sociology students tend to have more friends than economics students, then estimated peer effects that are smaller for sociology students may be confounded by exclusion bias.

Furthermore, these results caution against naive comparisons between peer effects models that include cluster fixed effects and models that do not control for cluster fixed effects. As shown above, the exclusion bias is aggravated in the latter type of models, whenever peer group formation is correlated with cluster formation. For example, studies adding classroom effects (e.g. de Melo, 2011), dormitory effects (e.g. Sacerdote, 2001) or school effects (e.g. Fletcher, 2012) will be more severely affected by the exclusion bias if peers are selected - partially or completely - from within the respective clusters. The literature tends to interpret a drop in the estimate of the endogenous peer effects in models adding cluster fixed effects as evidence of unobserved covariates at the cluster level. Although such correlates will often matter, the results in this paper suggest that such interpretations of the results may be confounded by the presence of the exclusion bias.

The finding that the downward exclusion bias, present in OLS estimation, can be eliminated by certain 2SLS estimation strategies is equally important. It provides an alternative explanation for the common but counter-intuitive tendency of peer effects studies to obtain 2SLS estimates of endogenous peer effects that are larger than their OLS counterparts (e.g. Goux and Maurin, 2007; De Giorgi et al., 2010; de Melo, 2011; Helmers and Patnam, 2014).

We discussed various methods that can be used to draw unbiased inference whenever 2SLS strategies are not applicable. Unlike the method suggested by Guryan et al. [2009], the methods described in this paper can be used in a wide range of applications, allowing for very general network structures (e.g. fixed pool sizes, non-overlapping peer groups, network data, etc.). For simple peer structures such as peer groups of size 2 we showed how one can use the exact

formulas derived in this paper to correct point estimates and permutation methods to correct p-values. For more complex peer structures (e.g. peer groups of size greater than 2 or network data) and assuming that error terms are independent across observations, we characterized the data generation process that can be used to consistently estimate structural parameters, using maximum likelihood estimation or non-linear least squares.

Finally, note that not all peer effects studies are affected by the exclusion bias. The exclusion bias arises whenever (i) an individual i is excluded from her potential peer group, (ii) whilst i is a potential peer of other observed individuals in the sample, and (iii) one considers a peer effects model that regresses individual i 's characteristic on an average characteristic of i 's peer group without controlling for i 's own characteristic. Examples of studies that do not suffer from exclusion bias include studies that exclude i 's own outcome when calculating the average peer group characteristic (e.g. Conley and Udry, 2010), studies that use lagged outcome or pre-determined outcomes of peers rather than contemporaneous outcomes whilst controlling for lagged outcome or pre-determined outcome of individual i herself (e.g. Munshi, 2004; Bayer et al., 2009) or studies that are based on a randomized controlled trial design to estimate peer effects (e.g. Duflo and Saez, 2003).

If present and not properly addressed, however, exclusion bias may significantly affect the results of social interaction studies. Its consequences may be especially problematic for empirical analyses that consider cluster fixed effects, for which the magnitude of the bias will not decrease as the sample size tends to infinity.

Tables and Figures

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

		L = 20	L = 50	L = 100
K = 2	Predicted $E(\hat{\beta}_1)$	-0.053	-0.020	-0.010
	Average $\hat{\beta}_1^s$	-0.054	-0.021	-0.006
	% rejected at 1% sign level	21%	8%	4%
	% rejected at 5% sign level	44%	20%	15%
	% rejected at 10% sign level	53%	30%	21%
K = 5	Predicted $E(\hat{\beta}_1)$	-0.25	-0.087	-0.042
	Average $\hat{\beta}_1^s$	-0.270	-0.083	-0.038
	% rejected at 1% sign level	77%	14%	8%
	% rejected at 5% sign level	92%	28%	18%
	% rejected at 10% sign level	95%	41%	28%
K = 10	Predicted $E(\hat{\beta}_1)$	-0.818	-0.220	-0.099
	Average $\hat{\beta}_1^s$	-0.864	-0.238	-0.109
	% rejected at 1% sign level	98%	42%	14%
	% rejected at 5% sign level	98%	59%	21%
	% rejected at 10% sign level	99%	64%	30%

[†]Note: $\hat{\beta}_0 = 0$; Cluster fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions

Figure 1: Expected versus actual rejection rate of null hypothesis - Example $N = 1000$; $L = 20$; $K = 5$

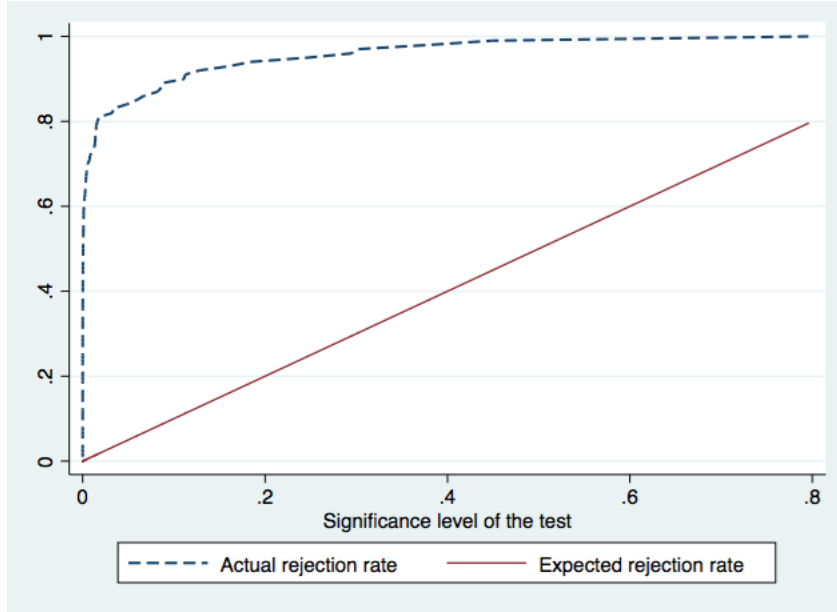


Table 2: Illustration permutation method

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

Table 3: Simulation results - permutation method test of random peer assignment

$L = 20 ; N = 1000 ; \beta_1 = 0$				
Group size	“Naive” $\hat{\beta}_1$	“Naive” p-value	Corrected $\hat{\beta}_1$	Corrected p-value
$K = 2$	-0.059	0.071	0.001	0.428
$K = 5$	-0.310	0.000	-0.047	0.314
$K = 10$	-0.731	0.000	0.129	0.386

Note: $\beta_0 = 0$; Cluster fixed effects added in all regressions; Permutations over 500 replications.

Figure 2: Histogram $\hat{\beta}_1^s$ under null

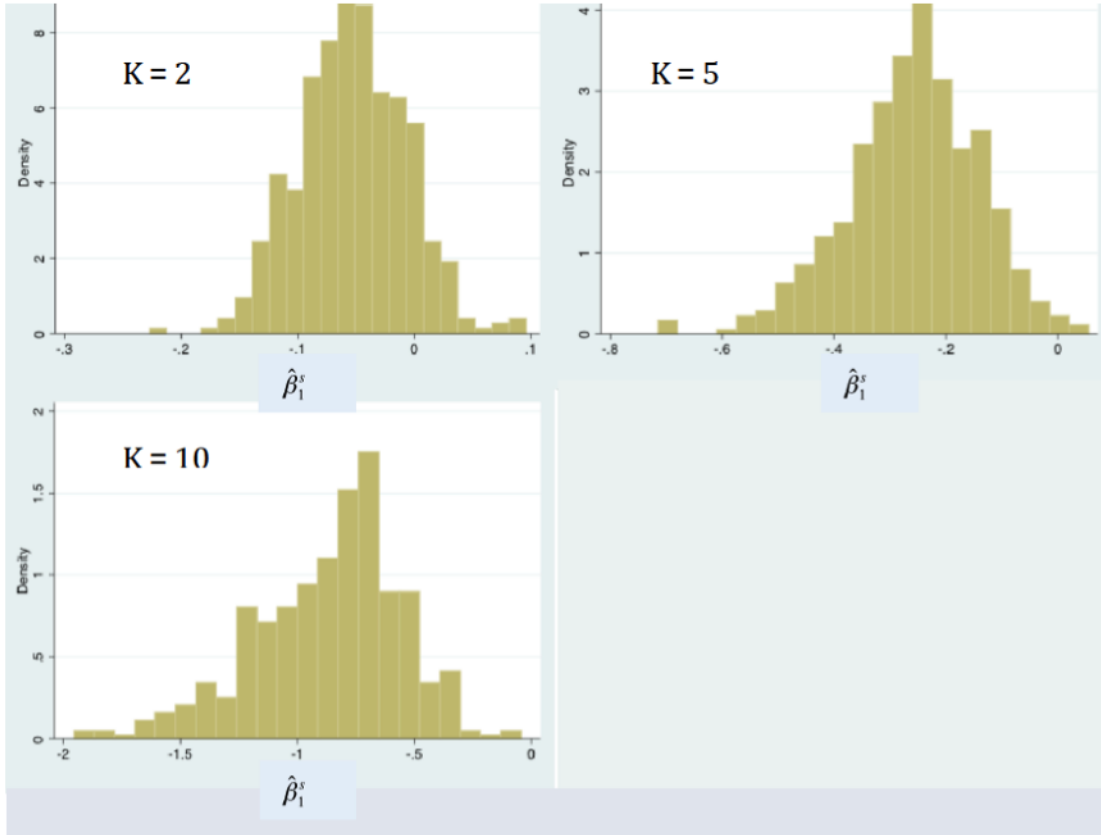


Table 4: Simulation results - Proposition 2 ($L = 20$)

	N = 500		N = 1000		N = 2000		N = 4000	
	FE	POLS	FE	POLS	FE	POLS	FE	POLS
K = 2	-0.053	-0.003	-0.055	-0.003	-0.057	-0.005	-0.052	0.000
K = 5	-0.279	-0.028	-0.250	-0.002	-0.264	-0.013	-0.249	0.000
K = 10	-0.972	-0.078	-0.864	-0.021	-0.832	-0.015	-0.818	-0.001

Table 5: Simulation results - Exclusion bias in the estimation of endogenous peer effects

$K = 2; L = 10; N = 1000$					
True β_1	Predicted reflection bias	Prediction exclusion bias	Total predicted bias	Predicted $\hat{\beta}_1$	Average simulated $\hat{\beta}_1$
0.00	0.000	-0.123	-0.123	-0.123	-0.117
0.02	0.020	-0.124	-0.123	-0.084	-0.077
0.04	0.040	-0.124	-0.123	-0.044	-0.038
0.06	0.060	-0.123	-0.064	-0.004	0.002
0.08	0.079	-0.123	-0.044	0.036	0.042
0.10	0.098	-0.121	-0.023	0.077	0.082
0.12	0.117	-0.120	-0.003	0.117	0.122
0.14	0.135	-0.118	0.017	0.157	0.162
0.16	0.152	-0.116	0.036	0.196	0.201
0.18	0.169	-0.113	0.056	0.236	0.240
0.20	0.185	-0.110	0.074	0.274	0.279

Note: $\beta_0 = 0$; Cluster fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions.

Table 6: Correction exclusion bias in the estimation of endogenous peer effects - $K = 2$

$K = 2; L = 10; N = 1000$				
β_1	“Naive” $\hat{\beta}_1$	“Naive” p-value	Corrected $\hat{\beta}_1$	Corrected p-value
0.00	-0.115	0.0150	0.005	0.472
0.02	-0.075	0.112	0.025	0.290
0.04	-0.035	0.459	0.045	0.116
0.06	0.005	0.913	0.064	0.048
0.08	0.045	0.337	0.084	0.012
0.10	0.086	0.070	0.104	0.004
0.12	0.126	0.008	0.124	0.000
0.14	0.165	0.000	0.144	0.000
0.16	0.205	0.000	0.164	0.000
0.18	0.244	0.000	0.184	0.000
0.20	0.283	0.000	0.204	0.000

Note: $\beta_0 = 0$; Cluster fixed effects added in all regressions; Permutations over 500 replications.

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

	$K = 2; L = 4; N = 100$			$K = 5; L = 10; N = 100$		
True β_1	0.0	0.1	0.2	0.0	0.1	0.2
$\hat{\beta}_1^{OLS}$ - no corrections	-0.33	-0.16	0.04	-0.21	0.07	0.23
$\hat{\beta}_1^{RC}$ - correction for reflection bias only	-0.17	-0.08	0.02	-0.09	0.04	0.17
$\hat{\beta}_1^{EC}$ - correction for reflection bias + exclusion bias	-0.02	0.09	0.19	-0.02	0.09	0.19
ρ	-0.25	-0.25	-0.25	-0.1	-0.1	-0.1

Note: $\beta_0 = 0$; Cluster fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions.

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

When peers are grouped, $var(x_{-i,k})$ is a weighted sum of the within-peer group variance of $x_{-i,k}$, i.e. σ_W^2 and the between-peer group variance of $x_{-i,k}$, i.e. σ_B^2 . Specifically, we have:

$$Var(\bar{x}_{-i,k}) = \frac{1}{N-1} \left[\sigma_W^2 (K-1) \frac{N}{K} + K \left(\frac{N}{K} - 1 \right) \sigma_B^2 \right] \quad (33)$$

The within-peer group variance of $\bar{x}_{-i,k}$ can be derived by use of the reduced form of $\bar{x}_{-i,k}$ for the case when $\beta_1 = 0$, as follows:

$$\begin{aligned} \bar{x}_{-i,k} &= \frac{(\sum_{j=1}^K x_{jk}) - x_{ik}}{K-1} \\ &= \frac{(\sum_{j=1}^K x_{jk}) - (\beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik})}{K-1} \\ \Rightarrow \bar{x}_{-i,k} &= \frac{(\sum_{j=1}^K x_{jk}) - \beta_0}{K-1} - \frac{\epsilon_{ik}}{K-1} \\ \Rightarrow \sigma_W^2 &= Var\left(\frac{\epsilon_{ik}}{K-1}\right) = \frac{\sigma_\epsilon^2}{(K-1)^2} \end{aligned} \quad (34)$$

The between-peer group variance of $\bar{x}_{-i,k}$ is the variance of the peer group mean of $\bar{x}_{-i,k}$:

$$\begin{aligned} \sigma_B^2 &= Var\left(\frac{\sum_{j=1}^K \bar{x}_{-j,k}}{K}\right) \\ &= Var\left(\frac{\sum_{i=1}^K x_{ik}}{K}\right) \end{aligned} \quad (35)$$

$$= \frac{\sigma_\epsilon^2}{K} \quad (36)$$

Using (33), (34) and (35) we obtain an expression for the total $var(\bar{x}_{-i,k})$:

$$Var(\bar{x}_{-i,k}) = \frac{N-K+1}{(N-1)(K-1)} \sigma_\epsilon^2 \quad (37)$$

B Relationship between σ_u^2 and σ_ϵ^2

We can re-write equation 3 as follows:

$$u_{ik} = \bar{y}_{-i,k} - \bar{y}_{-i}$$

Using $var(x_{ik}) = \sigma_\epsilon^2$ and equation (9) and the assumption that x_{ik} is i.i.d. (prior to treatment), we derive

$$\begin{aligned}
\text{var}(u_{ik}) &= \sigma_u^2 = \text{Var}(\bar{x}_{-i,k}) + \text{var}(\bar{x}_{-i}) - 2\text{cov}(\bar{x}_{-i,k}, \bar{x}_{-i}) \\
&= \text{Var}(\bar{x}_{-i,k}) + \text{var}(\bar{x}_{-i}) - 2\text{var}(\bar{x}_{-i}) \\
&= \text{Var}(\bar{x}_{-i,k}) - \text{var}(\bar{x}_{-i}) \\
&= \frac{N - K + 1}{(N_P - 1)(K - 1)} \sigma_\epsilon^2 - \frac{\sigma_\epsilon^2}{(N_P - 1)^2} \\
&= \frac{(N_P)(N_P + 1 - K)}{(N_P - 1)^2(K - 1)} \sigma_\epsilon^2
\end{aligned}$$

C Exclusion bias in cluster sampling

The cluster sampling equivalent of equation 7 is

$$x_{ikl} = \beta_0 + \beta_1 \left(\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^K x_{jsl} \right] - x_{ikl}}{L - 1} + u_{ikl} \right) + \epsilon_{ikl} \quad (38)$$

Averaging equation 38 over all L observations in group l , we obtain the cluster average outcome:

$$\bar{x}_l = \beta_0 + \beta_1 \left[\frac{\sum_{i=1}^L \left(\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^K x_{jsl} \right] - x_{ikl}}{L - 1} \right)}{L} + \bar{u}_l \right] + \bar{\epsilon}_l = \beta_0 + \beta_1 \left[\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^K x_{jsl} \right] - \bar{x}_l}{L - 1} + \bar{u}_l \right] + \bar{\epsilon}_l \quad (39)$$

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

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

Inserting 38 and 39, we derive the following expression for the cluster fixed effects model where peers are drawn from the cluster l :

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

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

$$\ddot{x} = \beta_1 \left(\frac{-\ddot{x}}{L - 1} + \ddot{u} \right) + \ddot{\epsilon} \quad (40)$$

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 $\beta_1 = 0$:

$$\begin{aligned} E\left(\hat{\beta}_1^{FE}\right) &= \frac{cov\left(\frac{-\ddot{x}}{L-1} + \ddot{u}, \ddot{\epsilon}\right)}{var\left(\frac{-\ddot{x}}{L-1} + \ddot{u}\right)} \\ &= \frac{cov\left(\frac{-\ddot{x}}{L-1}, \ddot{\epsilon}\right) + cov(\ddot{u}, \ddot{\epsilon})}{var\left(\frac{-\ddot{x}}{L-1}\right) + 2cov\left(\frac{-\ddot{x}}{L-1}, \ddot{u}\right) + var(\ddot{u})} \end{aligned} \quad (41)$$

In order to expand equation 41, we consider:

$$cov(\ddot{u}, \ddot{\epsilon}) = E(\ddot{u}\ddot{\epsilon}) = E[(u_{ikl} - \bar{u}_l)(\epsilon_{ikl} - \bar{\epsilon}_l)] = E(u_{ikl}\epsilon_{ikl}) - E(\bar{u}_l\epsilon_{ikl}) + E(\bar{u}_l\bar{\epsilon}_l) - E(u_{ikl}\bar{\epsilon}_l) = 0 \quad (42)$$

and

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

Furthermore, since $\bar{u}_l = \frac{\sum_{i=1}^L u_{ikl}}{L}$ and u is assumed independent across individuals (under the Null hypothesis that $\beta_1 = 0$):

$$\begin{cases} E(u_{ikl}\bar{u}_l) &= \frac{E(u_{ikl}^2)}{L} = \frac{\sigma_u^2}{L} \\ var(\bar{u}_l) &= var\left(\frac{\sum_{i=1}^L u_{ikl}}{L}\right) = \frac{\sum_{i=1}^L var(u_{ikl})}{L^2} = \frac{L\sigma_u^2}{L^2} = \frac{\sigma_u^2}{L} \Rightarrow var(\ddot{u}) = \sigma_u^2 - 2\frac{\sigma_u^2}{L} + \frac{\sigma_u^2}{L} = \frac{(L-1)\sigma_u^2}{L} \end{cases} \quad (43)$$

Similarly:

$$var(\ddot{\epsilon}) = \frac{(L-1)\sigma_\epsilon^2}{L} \quad (44)$$

Using equation 40, we derive the reduced form of $\left(-\frac{\ddot{x}}{L-1}\right)$:

$$\begin{aligned} \ddot{x} &= \beta_1 \left(\frac{-\ddot{x}}{L-1} + \ddot{u}\right) + \ddot{\epsilon} \Leftrightarrow \left[\frac{L-1+\beta_1}{L-1}\right] \ddot{y} = \beta_1 \ddot{u} + \ddot{\epsilon} \\ \Leftrightarrow -\frac{\ddot{x}}{L-1} &= \frac{-\beta_1 \ddot{u}}{L-1+\beta_1} - \frac{\ddot{\epsilon}}{L-1+\beta_1} \end{aligned} \quad (45)$$

Using $E(\ddot{\epsilon}) = E(\epsilon_{ikl} - \bar{\epsilon}_l) = 0$, equations 42, 44, 45 and $\beta_1 = 0$ (since we consider the case of random peer assignment), we derive:

$$\begin{aligned}
cov\left(\frac{-\ddot{x}}{L-1}, \ddot{\epsilon}\right) &= E\left[\left[\frac{-\ddot{x}}{L-1} - E\left(\frac{-\ddot{x}}{L-1}\right)\right] \ddot{\epsilon}\right] \\
&= E\left[\frac{-\ddot{\epsilon}\ddot{\epsilon}}{L-1}\right] \\
&= \frac{-var(\ddot{\epsilon})}{L-1} \\
&= -\frac{1}{L-1} \frac{(L-1)\sigma_\epsilon^2}{L} \\
&= -\frac{\sigma_\epsilon^2}{L}
\end{aligned} \tag{46}$$

Similarly,

$$2cov\left(\frac{-\ddot{x}}{L-1}, \ddot{u}\right) = -2\frac{E(\ddot{u}\ddot{\epsilon})}{L-1} = 0 \tag{47}$$

Again using equation 45 , we have:

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

Using equations 42 - 48 and 5 , we obtain:

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

and

$$var\left(\frac{-\ddot{x}}{L-1} + \ddot{u}\right) = \frac{(L-1)^2\sigma_u^2 + \sigma_\epsilon^2}{(L-1)L} = \frac{L + (L-K)(K-1)}{K(K-1)L} \sigma_\epsilon^2 \tag{50}$$

Hence, we can expand equation 41 as follows:

$$\begin{aligned}
E\left(\hat{\beta}_1^{FE}\right) &= \frac{cov\left(\frac{-\ddot{x}}{L-1} + \ddot{u}, \ddot{\epsilon}\right)}{var\left(\frac{-\ddot{x}}{L-1} + \ddot{u}\right)} = \frac{\left(-\frac{\sigma_\epsilon^2}{L}\right)}{\frac{L+(L-K)(K-1)}{K(K-1)L} \sigma_\epsilon^2} \\
\Rightarrow E\left(\hat{\beta}_1^{FE}\right) &= -\frac{K(K-1)}{L + (L-K)(K-1)}
\end{aligned} \tag{51}$$

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]):

$$\hat{\beta}_1^{OLS} = \eta^2 \hat{\beta}_1^{BE} + (1 - \eta^2) \hat{\beta}_1^{FE} \quad (52)$$

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 $E(\hat{\beta}_1^{FE}) < 0$. 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:

$$\bar{y}_{-i,l} = \beta_0 + \beta_1 \bar{\bar{y}}_{-i,l} + \bar{\epsilon}_l \quad (53)$$

In this Appendix we will derive this estimator and show that $E(\hat{\beta}_1^{BE}) \geq 0$. Given that $E(\hat{\beta}_1^{FE}) < 0$ and $E(\hat{\beta}_1^{BE}) \geq 0$, equation 52 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 53. In this case, $E(\hat{\beta}_1^{BE}) > 0$ 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 53 and $E(\hat{\beta}_1^{BE}) = 0$. This implies:

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

This proves proposition 2.

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

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

$$\begin{aligned} \bar{y}_{-i,l} &= \frac{\left[\sum_{s=1}^L \sum_{j=1}^K y_{jsl} \right] - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\epsilon}_l}{L - 1 + \beta_1} \\ &= \frac{L \bar{y}_l - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\epsilon}_l}{L - 1 + \beta_1} \end{aligned} \quad (54)$$

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:

$$\bar{y}_{-i,l} = \frac{L \bar{y}_l}{L - 1} - \frac{\bar{\epsilon}_l}{L - 1} \quad (55)$$

Using (55), we have:

$$\begin{aligned}
cov(\bar{y}_{-i,l}, \bar{\epsilon}_l) &= cov(\bar{y}_{-i,l} + \bar{u}_l, \bar{\epsilon}_l) \\
&= cov(\bar{y}_{-i,l}, \bar{\epsilon}_l) \\
&= L \frac{E(\bar{\epsilon}_l^2)}{L-1} - \frac{E(\bar{\epsilon}_l^2)}{L-1} \\
&= var(\bar{\epsilon}_l) \\
&= \frac{\sigma_\epsilon^2}{L}
\end{aligned} \tag{56}$$

and

$$\begin{aligned}
var(\bar{y}_{-i,l}) &= var\left(\frac{\sum_{i=1}^L \bar{y}_{-i,k,l}}{L}\right) \\
&= \frac{1}{L^2} var\left(\sum_{i=1}^L \left(\frac{\sum_{j=1}^L y_{jl} - y_{il}}{L-1}\right) + \sum_{i=1}^L u_{il}\right) \\
&= \frac{1}{L^2} var\left(\frac{L \sum_{i=1}^L y_{il}}{L-1} - \frac{\sum_{i=1}^L y_{il}}{L-1} + \sum_{i=1}^L u_{il}\right) \\
&= \frac{1}{L^2} var\left(\sum_{i=1}^L y_{il} + \sum_{i=1}^L u_{il}\right) \\
&= \frac{\sigma_\epsilon^2 + \sigma_u^2}{L} \\
&= \frac{\sigma_\epsilon^2 + \frac{L(L-K)}{(L-1)^2(K-1)}\sigma_\epsilon^2}{L} \\
&= \frac{(L-1)^2(K-1) + L(L-K)}{L(L-1)^2(K-1)}\sigma_\epsilon^2
\end{aligned} \tag{57}$$

where in the last step we used the result in (5).

Using equation (56) and (57) we obtain:

$$\begin{aligned}
E(\hat{\beta}_1^{BE}) &= \frac{cov(\bar{y}_{-i,l}, \bar{\epsilon}_l)}{var(\bar{y}_{-i,l})} \\
&= \frac{\frac{\sigma_\epsilon^2}{L}}{\frac{(L-1)^2(K-1) + L(L-K)}{L(L-1)^2(K-1)}\sigma_\epsilon^2} \\
&= \frac{(L-1)^2(K-1)}{(L-1)^2(K-1) + L(L-K)} > 0
\end{aligned} \tag{58}$$

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

We can also use 52 to prove that corollary 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 58, respectively. We will now derive an expression for η^2 .

Weight parameter η^2 in equation (59) 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:

$$\eta^2 = \frac{SS_{\bar{y}_{-i,k,l}}^{BG}}{SS_{\bar{y}_{-i,k,l}}^{Total}} = \frac{SS_{\bar{y}_{-i,k,l}}^{BG}}{SS_{\bar{y}_{-i,k,l}}^{BG} + SS_{\bar{y}_{-i,k,l}}^{Within}} \quad (59)$$

Specifically, $SS_{\bar{y}_{-i,k,l}}^{BG}$ 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:

$$SS_{\bar{y}_{-i,k,l}}^{BG} = SS_{\bar{y}_{-i,k,l}}^{BE} \times L \quad (60)$$

where $SS_{\bar{y}_{-i,k,l}}^{BE}$ is the sum of squares of $\bar{y}_{-i,l}$ in the between estimation regression (44) in Appendix (C). Furthermore, using the definition of the variance operator, we know that:

$$var(\bar{y}_{-i,l}) = \frac{SS_{\bar{y}_{-i,k,l}}^{BE}}{\left(\frac{N}{L} - 1\right)} \Rightarrow SS_{\bar{y}_{-i,k,l}}^{BE} = var(\bar{y}_{-i,l}) \times \left(\frac{N}{L} - 1\right) \quad (61)$$

Using equations (59) - (61), we obtain:

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

Substituting in for the expression of $var(\bar{y}_{-i,l})$ given by equation (57), we finally have:

$$SS_{\bar{y}_{-i,k,l}}^{BG} = \frac{(L-1)^2(K-1)}{(L-1)^2(K-1) + L(L-K)} \times \left(\frac{N}{L} - 1\right) \times \sigma_\epsilon^2 \quad (62)$$

Next, $SS_{\bar{y}_{-i,k,l}}^{Within}$ 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 $\bar{y}_{-i,l}$. Similarly to equation (61), we have:

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

From the above we know that $var(\bar{y}_{-i,k,l} - \bar{y}_{-i,l}) = var\left(\frac{-\bar{y}}{L-1} + \bar{u}\right)$. Therefore, we can substitute in for the expression of $var(\bar{y}_{-i,k,l} - \bar{y}_{-i,l})$ by using equations (50):

We have:

$$SS_{\bar{y}_{-i,k,l}}^{Within} = \frac{L + (L-K)(K-1)}{K(K-1)L} \times \frac{N-1}{L} \sigma_\epsilon^2 \quad (63)$$

Combining equations (59), (62) and (63), we obtain:

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

where

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

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

$$\begin{aligned} plim(\eta^2) &= plim \left[\frac{A \left(\frac{N}{L} - 1\right)}{A \left(\frac{N}{L} - 1\right) + B \left(\frac{N-1}{L}\right)} \right] \\ &= \frac{A}{A+B} \end{aligned} \quad (64)$$

Note that this closed form result only holds in the limit, that is, when sample size N tends to infinity. Using (52), (14), (58) and (64) we now derive the large sample property of pooled OLS when peer group formation occurs at group level l and when the true $\beta = 0$:

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

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:

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

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

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

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 52) 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

- 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), 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.
- T. Conley and C. Udry. Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1), 2010.
- 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., 2011. URL <http://www.econpol.unisi.it/dipartimento/it/node/1606>.
- M. Drukker, I. R. Prucha, and R. Raciborski. A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. *The Stata Journal*, 13(2):287–301, 2013.
- E. Duflo and E. Saez. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics*, 118(3):815–842, 2003.
- 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., 2012.
- E.L. Glaeser, B.I. Sacerdote, and J.A. Scheinkman. The social multiplier., 2002. URL <http://post.economics.harvard.edu/hier/2002papers/2002list.html>.
- D. Goux and E. Maurin. Close neighbours matter: neighbourhood effects on early performance at school. *The Economic Journal*, 117(523):1193–1215, 2007.
- 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.

- C. Helmers and M Patnam. Does the rotten child spoil his companion? spatial peer effects among children in rural india. *Quantitative Economics*, 5(1):67–121, 2014.
- 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 Per*, 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(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.
- D. Zimmerman. Peer effects in academic outcomes: evidence from a natural experiment. *The Review of Economics and Statistics*, 85(1):9–23, 2003.