NBER WORKING PAPER SERIES

EXCLUSION BIAS IN THE ESTIMATION OF PEER EFFECTS

Bet Caeyers
Marcel Fafchamps

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

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2016, Revised February 2020

We benefitted from useful comments from three anonymous referees, Matt Jackson, Aureo de Paula, Arun Chandrasekhar, Ben Golub, Pau Milan, Steve Durlauf, Emily Breza, Markus Mobius, Giacomo De Giorgi, Matt Elliott, Xu Tan, Adam Szeidl, Michael Gechter, Bernard Fortin and Elena Manresa as well as from participants to seminars at Gothenburg University, Norwegian School of Economics, Institute for Fiscal Studies, LICOS, UC Davis and the International Food Policy Research Institute, participants to the November 2016 NEUDC conference, the March 2016 RES conference, the May 2016 Stanford conference on Social Networks, the March 2016 CSAE conference on African development, the April 2016 Laval University conference on Social Networks, and the 2019 Barcelona GSE Summer forum, as well as students at Stanford University, the 2016 PODER Summer School, and the 2016 Nova University Summer School. Finally, we are grateful to the ESRC Centre for Microeconomic Analysis of public policy (CPP) at the Institute for Fiscal Studies for funding parts of this research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Exclusion Bias in the Estimation of Peer Effects
Bet Caeyers and Marcel Fafchamps
NBER Working Paper No. 22565
August 2016, Revised February 2020
JEL No. C31

ABSTRACT

We examine a largely unexplored source of downward bias in peer effect estimation, namely, exclusion bias. We derive formulas for the magnitude of the bias in tests of random peer assignment, and for the combined reflection and exclusion bias in peer effect estimation. We show how to consistently test random peer assignment and how to estimate and conduct consistent inference on peer effects without instruments. The method corrects for the presence of reflection and exclusion bias but imposes restrictions on correlated effects. It allows the joint estimation of endogenous and exogenous peer effects in situations where instruments are not available and cannot be constructed from the network matrix. We estimate endogenous and exogenous peer effects in two datasets where instrumental approaches fail because peer assignment is to mutually exclusive groups of identical size. We find significant evidence of positive peer effects in one, negative peer effects in the other. In both cases, ignoring exclusion bias would have led to incorrect inference. We also demonstrate how the same approach applies to autoregressive models.

Bet Caeyers
FAIR/Norwegian School of Economics (NHH)
Bet.Caeyers@nhh.no

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

1 Introduction

The estimation of peer effects is fraught with difficulties that, since Manski (1993), have customarily been divided into reflection bias, selection bias, and unobserved correlated effects (see also Brock and Durlauf 2001, Moffitt 2001). Reflection bias refers to the fact that, if i influences j, then j also influences i, creating a multiplier effect. The standard solution to reflection bias is to instrument. However, because reflection bias is just a multiplier effect, inference about the *presence* of peer effects is widely believed to be unaffected by reflection bias – only the magnitude of the coefficient is. We demonstrate that a largely unknown source of bias invalidates this reasoning.

The estimation of peer effects is also affected the presence of unobserved correlated effects within the pool from which peers are selected. This has led researchers to include fixed effects to absorb them. We show that including selection pool fixed effects is the main contributor to the bias we study here. The bias only disappears when each selection pool gets large enough. The source of bias is similar to what arises with autoregressive models in short panels: introducing fixed effects generates a bias that only disappears when T, the number of periods, gets large enough (Nickell 1981). In time series, this problem has been successfully addressed using lagged values as instruments (e.g., Arellano and Bond 1991, Arellano and Bover 1995, Blundell and Bond 1998). Such instruments are not available in peer effect models because of reflection.

Selection bias arises when peers share unobserved characteristics or proclivities that affect the outcome variables of interest. Efforts to eliminate selection bias have focused on random peer assignments, using either natural (e.g., Sacerdote 2001) or controlled experiments (e.g., Carrell et al. 2013, Fafchamps and Quinn 2017, Cai and Szeidl 2016). Random assignment is widely believed to eliminate the correlation of residuals due to peer selection. We show this to be untrue: random assignment produces a negative correlation between peer outcomes in standard tests of random assignment with selection pool fixed effects.

This paper centers on a recently discovered source of bias in the estimation of peer effects. This bias, which we call exclusion bias, was first mentioned by Guryan et al. (2009). It arises from the fact that the assignment of peers is done without replacement: i cannot be his own peer. When including selection pool fixed effects, the exclusion of i from the pool of i's peers creates a small sample negative relationship between i's characteristics and that of his peers: if i is above average,

the average of those remaining in the pool is lower than i; conversely, if i is below average, the average of those remaining in the pool is higher than i. Hence i's characteristics are negatively correlated with the expected value of the remaining peers in the pool. This is true irrespective of whether peers self-select each other or peers are randomly assigned.

Through Monte Carlo simulations and the use of basic algebra, Guryan et al. (2009) and Angrist (2014) show how this correlation produces a negative bias in ordinary least squares (OLS) estimates of peer effects. The purpose of this paper is to move beyond these basic observations and offer insights into the properties, causes, and consequences of exclusion bias, all of which have largely been ignored to date. Although exclusion bias is also present when peers are self-selected, here we focus our analysis on randomly assigned peers.

In the first part of the paper we focus on the presence of exclusion bias in standard tests of random peer assignment. We quantify the magnitude of the bias, derive its properties and show how tests of random peer assignment can be corrected for exclusion bias in various settings (including unequal group sizes and/or unequal pool sizes). Exclusion bias is also present in regressions without fixed effects when sample size is small. Next we demonstrate how exclusion bias combines with reflection bias to distort coefficient estimates, and we offer a general solution to the estimation of peer effects that corrects for both sources of bias – provided that some restrictions are put on correlated effects. We combine this new estimation method with a simple approach to conduct hypothesis testing and obtain correct inference. We use simulations and two empirical applications using data from published papers (Guryan et al 2009 and Fafchamps and Mo 2018) to demonstrate the importance of our findings in practice.

This paper contributes to the literature in a number of ways. First, we derive an exact formula for the magnitude of the exclusion bias in standard linear-in-means tests of random peer assignment. Unlike top-level expressions of the bias provided for instance in Angrist (2014), all formulas presented in this paper are expressed as functions of the core parameters driving the bias: the size of the peer group and the size of the pool from which peers are selected. Second we derive, for groups of size two, exact formulas for the exclusion and reflection bias in standard linear-in-means model with or without pool fixed effects. While reflection bias tends to inflate peer effect estimates, exclusion bias always operates in the negative direction. We identify conditions under which the exclusion bias dominates the reflection bias, changing the sign of OLS estimates. Using simulations,

we generalize these findings to peer groups of size greater than two. Third, we show that exclusion bias is more severe in models that include cluster fixed effects at levels other than the selection pool, whenever peer group membership is correlated with cluster fixed effects. In these models exclusion bias does not disappear as the sample size tends to infinity.

Fourth, we offer several solutions for obtaining estimates of peer effects that correct for exclusion bias. The applicability of each solution depends on the objective of the researcher and the type of data available. To correct inference, we propose to rely on one of two methods: if appropriate, use our formula to correct point estimates and use standard errors clustered by peer selection pool to draw inference; or, if this is inappropriate, rely on randomization inference based on the permutation method of Krackhardt (1988).

Fifth, we show how to obtain consistent point estimates of endogenous and exogenous peer effects in linear-in-means models with a large variety of network structures – including non-overlapping peer groups, partly overlapping peer groups, and arbitrary network data. This solution is general and simple to apply and as such dominates alternative approaches proposed by Guryan et al. (2009) and in unpublished work by Wang (2009) and Stevenson (2015a, 2015b). It does not require instruments but relies on the assumption of zero correlated shocks between peers.[1] Whether or not this assumption is appropriate depends on the context. When it is satisfied (for instance, when common shocks can credibly be absorbed by control variables and/or by the addition of cluster fixed effects), our approach allows the consistent estimation of peer effects in situations that are not amenable to the instrumental variable approaches proposed by Bramoulle et al. (2009) and De Giorgi et al. (2010) and extended by Lee (2007). In particular, it can estimate peer effects in non-overlapping groups of similar size, something that the above-mentioned approaches cannot do. This includes: the assignment of students to dorms (e.g., Sacerdote 2001, Carrell et al. 2013) or work groups (e.g., Carrell et al. 2016); the assignment of workers to teams (e.g., Bandiera et al. 2009); and the assignment of entrepreneurs to social groups (e.g., Fafchamps and Quinn 2017, Cai and Szeidl 2016).

Since the assumption of zero correlated effects does not suit all settings, we identify conditions under which two-stage least squares (2SLS) does not suffer from exclusion bias. When these con-

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

ditions are satisfied, they can account for a counter-intuitive yet common finding. Many studies of social interactions obtain positive 2SLS estimates of endogenous peer effects that are significantly larger than OLS estimates. This is counter-intuitive: positive OLS estimates ought to be biased upwards due to reflection bias[2] (e.g., Manski 1993, Brock and Durlauf 2001, Moffitt 2001). This paper provides a new explanation for this pattern: the negative exclusion bias that affects OLS disappears when a valid 2SLS estimation strategy is used.

2 Testing random peer assignment

The nature of exclusion bias is best illustrated in a situation where peers are assigned at random and there are no peer effects. In this case we would normally expect peer characteristics to be uncorrelated across peers. Yet, because of exclusion bias, they are. To illustrate this, we begin by demonstrating how exclusion bias affects tests of random peer assignment.

2.1 Intuition

We are interested in the properties of a data generating process in which individuals are assigned a number of peers selected from a finite selection pool of size L. In this section we focus on non-overlapping, mutually exclusive peer groups because it is a commonly observed form of random peer assignment (e.g., assignment to a room, a class, a team). In Section 3.3 we demonstrate how the analysis extends to other networks. We assume throughout that individuals from a pool only have peers from that pool, without overlap across pools. Thus if each peer group has size K and the number of groups in a pool is 3, then the pool size $L = 3K$. Similarly, if each pool has size L and the number of pools is N, then the total sample size is $N \times L$.

Suppose a researcher has data on peer assignment and wishes to test whether assignment is random using an observable pre-treatment characteristic x_{ikl} of individual i in peer group k from pool l. Random peer assignment is typically verified by testing whether $\beta_1 = 0$ in a linear-in-means model of the form:

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

where $\bar{x}_{-ikl}$ denotes the average of i's peers in group k (excluding i herself) and where selection

[2] In addition to other sources of bias such as positive assorting in peer selection and unobserved correlated effects.

pool dummies δ_l control for randomization strata fixed effects. Model (1) is typically estimated using ordinary least squares (OLS).

One example is the study of Dartmouth college freshmen by Sacerdote (2001), who exploits the random allocation of students to roommates to study peer effects. In that study, i denotes an individual student, l is the pool to which the student is assigned based on her stated housing preferences ('block'), and k is the shared room within pool l to which the student is randomly assigned. Sacerdote tests random peer assignment by regressing freshman i's pre-treatment test score (e.g., SAT math score) on the average pre-treatment test score of his/her roommates and a set of block dummies.

2.2 Formula

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 argued by Guryan et al. (2009), this is incorrect: in small samples or when using pool fixed effects, a mechanical *negative* relationship exists between i's characteristics and those of i's peers prior to treatment. Given that individuals cannot be their own peers, they are excluded from the pool from which their peers are drawn.

We now provide a formula for the exclusion bias that affects $\widehat{\beta}_1$ in regression (1). The proof is provided in Appendix A. We start by considering the case when N pools of L individuals are each randomly partitioned into non-overlapping groups of K peers – for instance, when students in a school cohort l are randomly assigned to a dormitory or work group k (e.g., Sacerdote 2001, Glaeser et al. 2003, Zimmerman 2003, and Duflo and Saez 2011). Below we extend our main result to the more general case when selection pools and peer groups differ in size. If $N = 1$, pool dummies δ_l drop out of regression (1). We have the following Proposition:

Proposition 1: *Estimates of* β_1 *in model* (1) *satisfy the following properties:*

$$Part\,1:\ plim_{N\to\infty}[\hat{\beta}_1] = -\frac{(L-1)(K-1)}{(L-K)L+(K-1)} < 0\ \ for\ L, K\ fixed \tag{2}$$

$$Part\,2:\ plim_{L\to\infty}[\hat{\beta}_1] = 0\ \ for\ N = 1\,and\,K\ fixed \tag{3}$$

$$Part\,3:\ E\left[\hat{\beta}_1|N\right] < plim_{N\to\infty}\left[\hat{\beta}_1\right] \leq 0\ for\ L, K\ fixed \tag{4}$$

Proof: see Appendix A.1 and Appendix A.2.

Equation (2) in Proposition 1 provides a formula for the magnitude of the exclusion bias in tests of random peer assignment in the most common case when peers are drawn from separate selection pools and $L < N$. It demonstrates that, for a sufficiently large number of pools of fixed size L, the magnitude of the exclusion bias depends on only two key parameters: the size of peer groups K; and the size L of the pools from which peers are drawn. More specifically we have:

1. $\frac{\triangle|plim_{N\to\infty}[\hat{\beta}_1]|}{\triangle L} < 0$: For a given peer group size K, the asymptotic exclusion bias falls as L increases. This is similar to what happens in autoregressive models with panel fixed effects, where the bias falls as T, the number of periods, increases.[3]

2. $\frac{\triangle|plim_{N\to\infty}[\hat{\beta}_1]|}{\triangle K} > 0$: For a given pool size L, the asymptotic exclusion bias is more severe with large peer groups or, equivalently, with a smaller number of groups in each pool.[4]

Equation (3) extends formula (2) to the special case when all peers come from the same selection pool and this peer selection pool equals the sample population. In this case, the exclusion bias disappears asymptotically as L grows. A more detailed discussion is presented in Appendix A.2. This property too is reminiscent of what happens in autoregressive regressions with fixed effects, where the bias disappears as T increases.

Equations (2) and (3) only apply in the limit, that is, when sample size tends to infinity. Can we say something about exclusion bias in small samples? The last part of the Proposition, equation (4), provides an additional result, obtained using Taylor approximations and Monte Carlo simulations. It shows that, for a given pool size L and a given number of pools N, the expectation of the exclusion bias is more negative than its asymptotic value. In the next section, we illustrate this with a simulation analysis. We also confirm that the expected bias converges to its asymptotic

[3]Proof: Since $\frac{(L-1)(K-1)}{(L-K)L+(K-1)} = \frac{(K-1)}{\frac{L-K}{L-1}L+(K-1)}$, the derivative only depends on how the first term in the denominator varies with L: if it increases with L, the absolute value of the bias falls. It is easy to see that $\frac{L-K}{L-1}$ increases with L since $L > K$ by construction. Hence the result. QED.

[4]Proof: Since $\frac{(L-1)(K-1)}{(L-K)L+(K-1)} = \frac{L-1}{\frac{L-K}{K-1}L+1}$, the derivative only depends on how the first term in the denominator varies with K: if it falls with K, the absolute value of the bias increases. We have $\frac{\partial\frac{L-K}{K-1}}{\partial K} = -\frac{L-1}{(K-1)^2} < 0$ since both L and K are larger than 1 by construction. Hence the result. QED.

value (2) as the sample size grows larger, keeping the sizes of selection pools L and peer groups K constant. A similar result applies to the situation where $N = 1$, in which case $E\left[\hat{\beta}_1|L\right] < plim_{L\to\infty}\left[\hat{\beta}_1\right] \leq 0 \; for\, N, K \; fixed.$ When the number of peer groups is small, the magnitude of the exclusion bias can be large even though L is large, something we illustrate in the next section as well.

So far we have assumed that all selection pools are of equal size L and groups of equal size K. If selection pools vary in size, it can be shown (and confirmed through simulations) that the $plim_{N\to\infty}$ of the exclusion bias for a given group size K is a weighted average of the biases associated with the different pool sizes:

$$plim_{N\to\infty}\left[\hat{\beta}_1^{FE}\right] = \frac{\sum_{i=1}^{N} L_i \, plim_{N\to\infty}\left[\hat{\beta}_1^{FE}|L_i\right]}{\sum_{i=1}^{N} L_i} \tag{5}$$

where L_i denotes the size of selection pool l and, as before, N is the total number of selection pools. Similarly, given a pool size L, if peer groups differ in size K_k it can be shown that exclusion bias is a weighted average of the bias associated with the different K_k. That is:

$$plim_{N\to\infty}\left[\hat{\beta}_1^{FE}\right] = \frac{\sum_{i=1}^{S} K_k \, plim_{N\to\infty}\left[\hat{\beta}_1^{FE}|K_k\right]}{\sum_{i=1}^{S} K_k} \tag{6}$$

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

Proposition 1 demonstrates that exclusion bias is conceptually different from the attenuation bias associated with classical measurement error (CME). First, it is not driven by measurement error – i.e., it arises even in the absence of measurement error. Secondly, it behaves 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 bias $\widehat{\beta}_1$ if the true $\beta_1 = 0$. In contrast, exclusion bias is additive, always negative, and does not disappear when the true β_1 is zero. This makes exclusion bias a more serious adversary in terms of inference: by ignoring exclusion bias, a researcher can wrongly conclude to the presence of (negative) assorting when $\beta_1 = 0$; or conclude that peer assignment is random when in fact the true $\beta_1 > 0$ and there is positive assorting.

2.3 Simulation results

Results from a Monte Carlo simulation are presented in Table 1 to illustrate the magnitude of the exclusion bias in random peer assignment. Simulations vary in pool size L and peer group size K while keeping an integer number of groups L/K. For each simulation we generate a random sample of $N \times L = 1000$ observations. Each observation is assigned one realization of a standard normally distributed i.i.d. characteristic $x_i \sim N(0,1)$. The $N \times L$ observations are then randomly assigned to pools of L individuals each, and subsequently randomly assigned to a group of size K within each pool. A pool-specific shock is added to simulate differences across pools δ_l.

We repeat this process 1000 times for a particular vector $\{K, L\}$ and for each generated sample we estimate regression (1) and collect the estimated $\widehat{\beta}_1$. The average $\widehat{\beta}_1$ for each vector $\{K, L\}$ is summarized in Table 1. For comparison purposes, we also report the predicted $plim_{N\to\infty}[\hat{\beta}_1]$ derived in Proposition 1.1. Results verify Proposition 1.1: the average bias over 1000 replications is reasonably close to its predicted asymptotic value; it increases in K; and decreases in L. Simulated values differ slightly from $plim_{N\to\infty}[\hat{\beta}_1]$ because of the sample size $N \times L = 1000$ is not quite large enough to converge to their asymptotic value for particular values of L and K. In Table 2 below we illustrate this point more clearly.

Table 1 also shows the proportion of artificially generated samples for which we falsely reject the null hypothesis that $\beta_1 = 0$ at the 1%, 5% and 10% significance levels. Results indicate that random assignment is rejected in a surprisingly large fraction of simulations, especially when K is large relative to L. To illustrate this graphically for one particular example ($L = 20$ and $K = 5$), we plot in Figure 1 the rate at which OLS rejects the null hypothesis that $\beta_1 = 0$. If the test is unbiased, the rejection rate should lie along the 45 degree line. This is clearly not what we observe: the rejection rate is well above the 45 degree line, confirming that testing whether $\beta_1 = 0$ in regression (1) over-rejects the null of random assignment in a substantial proportion of cases. In other words, that test is strongly biased, and the magnitude of the bias in large samples is well predicted by formula (2).

Finally, simulation results presented in Table 2 show for a given pool size $L = 50$ and separately for $K = 5$ and $K = 10$, what happens to the exclusion bias when N increases. The results confirm that the bias is larger in small samples and that it converges to the value predicted by (2) as N

increases (predicted values for $K = 5$ and $K = 10$ are shown in, respectively, the middle and bottom panel of column 2 in Table 1).

2.4 Inference

Guryan et al (2009), Wang (2009) and Stevenson (2015a, 2015b) have already proposed alternative methods to test the null hypothesis of random peer assignment. We discuss these methods and their limitations in some detail in Appendix B. In particular, the method proposed by Guryan et al (2009), henceforth GKN (described in Appendix B.1), uses the average of the selection pool as control variable to eliminate exclusion bias. While the method is simple to implement, it only identifies the parameter of interest β_1 if there is sample variation in the size of peer selection pools; if every selection pool has the same number of individuals (which is common in practice), the model is strictly unidentified. By extension, limited variation in pool size leads to quasi-underidentification. Secondly, the GKN method requires precise knowledge of each peer selection pool. Such knowledge is not always available, as for instance arises when peers form arbitrary social networks. We offer instead two simple, and more generally applicable ways of testing random peer assignment.

The idea behind the first method is to circumvent GKN's identification problem by netting out the asymptotic exclusion bias using the results from Proposition 1, rather than adding one additional parameter to estimate. Specifically, we use formula (2) – or its extension to cases of varying group and pool sizes that is provided above – to transform the dependent variable in model (1) so as to obtain a consistent point estimate of the true β_1 under the null. To this effect, we apply OLS to estimate:

$$\widetilde{x}_{ikl} = \beta_0 + \beta_1 \bar{x}_{-ikl} + \delta_l + \epsilon_{ikl} \tag{7}$$

where $\widetilde{x}_{ikl} \equiv x_{ikl} - plim_{N\to\infty}[\hat{\beta}_1]\bar{x}_{-ikl}$ with $plim_{N\to\infty}[\hat{\beta}_1]$ given by formula (2).[5] Random peer assignment is verified by testing whether $\widehat{\beta}_1 = 0$ in model (7) using OLS standard errors clustered at the pool level. As illustrated by simulation results presented in Figure 2, only when standard errors are clustered by selection pool does the method yield correct inference. We should point out that regression model (7) does not yield a consistent estimate of β_1 when the true $\beta_1 \neq 0$ – more

[5] Under the null of $\beta_1 = 0$, this transformed model is obtained as follows: $x_{ikl} = \beta_0 + \left(\beta_1 + plim_{N\to\infty}[\hat{\beta}_1]\right)\bar{x}_{-ikl} + \delta_l + \epsilon_{ikl} \Leftrightarrow x_{ikl} - plim_{N\to\infty}[\hat{\beta}_1]\bar{x}_{ikl} = \beta_0 + \beta_1\bar{x}_{-ikl} + \delta_l + \epsilon_{ikl}$. It immediately follows that $plim_{N\to\infty}[\tilde{\beta}_1] = \beta_1$ where $\tilde{\beta}_1$ denotes the estimate obtained from estimating (7).

about this in Section 3.[6]

The above method only works when formula (2) can be calculated. It does not apply to partially overlapping groups, or arbitrary network data. In such cases randomization inference can be used instead (e.g., Fisher 1925). The idea is to simulate, using the data at hand, the distribution of $\widehat{\beta}_1$ under the null hypothesis of random peer assignment.[7] The application of this idea to networks goes back to Krackhardt (1988). It is more general and simpler to use than the method proposed by Athey et al. (2015), which re-randomizes treatments across peers.

We illustrate how the method works with mutually exclusive groups of different sizes. Imagine that the researcher has observational data x_{ikl} partitioned in groups of varying size K_i within pools of size L_i. The first four columns of Table 3 gives an example of such data structure. We wish to test random assignment within pools using regression (1). We start by estimating the model on the data to obtain the OLS estimate of $\widehat{\beta}_1$. We wish to know how likely it is to obtain value $\widehat{\beta}_1$ under the null of random assignment within pools. To this effect, we simulate the distribution of $\widehat{\beta}_1$ under the null. This is accomplished by keeping individuals within their selection pool but reassigning them to counterfactual groups. This is illustrated in column 5 of Table 3. For each reassignment we estimate regression (1) and obtain a counterfactual realization of $\widehat{\beta}_1^s$ for simulation sample s under the null. By repeating this process a large enough number of times, we obtain an approximation of the distribution of $\widehat{\beta}_1$ under the null. The mean of the distribution of $\widehat{\beta}_1^s$ is the average bias under the null. We then compare our $\widehat{\beta}_1$ estimate to the distribution of $\widehat{\beta}_1^s$. To obtain the p-value of the test of random peer assignment, we proceed in the same way as in other bootstrapping procedures, e.g., by taking the proportion of $\widehat{\beta}_1^s$ that are either above the absolute value of $\widehat{\beta}_1$ or below minus the absolute value of $\widehat{\beta}_1$.[8]

To visualize the performance of this procedure, we generate artificial samples of 1000 observations for three values of $K = \{2, 5, 10\}$. As before, we set the size of each pool $L = 20$ and we posit $\epsilon_{ik} \sim N(0, 1)$. Figure 3 shows the distribution of 1000 simulated $\widehat{\beta}_1^s$ under the null of random peer

[6]It also does not work when other regressors are included. If the researcher wishes to add regressors w_{ikl} when testing randomized assignment (e.g., to control for stratification variables), it is then necessary to first partial out w_{ikl} from x_{ikl} and $\bar{x}_{-ikl}$. Practically, this means doing the following: (1) regress x_{ikl} on w_{ikl} and keep the residuals, which we denote as $\hat{u}_{ikt}$; (2) regress $\bar{x}_{-ikl}$ on $\overline{w}_{-ikl}$ (the leave-out mean of w_{ikl} for peers) and keep the residuals, which we denote as $\hat{v}_{-ikl}$; and (3) construct $\tilde{\hat{u}}_{ikt} \equiv \hat{u}_{ikt} - \rho\hat{v}_{-ikl}$; and (4) regress $\tilde{\hat{u}}_{ikt}$ on $\hat{v}_{-ikl}$.

[7]Permutation methods can also approximate the distribution of $\widehat{\beta}_1$ under more complicated random assignment processes, such as multi-level stratification.

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

assignment for $K = \{2, 5, 10\}$. The histograms are centered on the *plim* of $\hat{\beta}_1$ under the null that were shown in Table 1, not around $\beta_1 = 0$. The permutation method corrects p-values by taking this distributional shift into consideration when calculating the probability of observing $\widehat{\beta}_1$ under the null. Figure 4 illustrates, for one particular example (i.e., $N = 1000$, $L = 20$ and $K = 5$), that the permutation method yields correct inference.

2.5 Cluster fixed effects

So far we have focused on the case when FEs are added at the level of the peer selection pool, which is appropriate for typical tests of random peer assignment. Before moving to the estimation of endogenous peer effects in Section 3, we discuss the implication of not including such fixed effects (FE), or of including cluster FEs at a more aggregated level than the selection pool. Researchers typically add cluster fixed effects to a regression to soak up unobserved common shocks correlated with regressors of interest. As a result, estimates obtained with fixed effects are regarded as more conservative/reliable than those without. In this section we show that this reasoning does not apply in the presence of exclusion bias: when testing for random peer assignment, adding cluster fixed effects increases the magnitude of the bias.[9]

Some tests of random assignment do not include any cluster fixed effect – e.g., because the researcher does not observe the level at which peer selection occurs and cannot control for it. Other studies include FEs at levels other than selection pools in an effort to control for unobservables. Of particular concern is the presence of common shocks that generate a positive correlation between peers even in the presence of random assignment. Since these shocks need not occur at the level of the selection pool, it is not uncommon for researchers to estimate models that include FEs other than for selection pools.

To demonstrate what cluster fixed effects do to exclusion bias, we start by comparing two estimators of β_1 in model (1): $\hat{\beta}_1^{POLS}$ obtained using pooled OLS without any δ_l fixed effects; and $\hat{\beta}_1^{FE}$ obtained by estimating equation (1) with pool fixed effects δ_l. Table 4 simulates what happens for different values of N, L and K. In each simulation, peers are randomly assigned within selection pools of size $L = 50$ and $\delta_l = 0$ for all l. The true value of β_1 in the data generation process (DGP) is thus 0. The first and the last column of each panel report the simulated averages of $\hat{\beta}_1^{FE}$ and

[9] This finding is again reminiscent of what happens when adding fixed effects to an autoregressive model.

$\hat{\beta}_1^{POLS}$. We see that adding pool FEs induces a dramatic increase in exclusion bias. For instance, with $N = 500$ and $K = 10$, the average $\widehat{\beta}_1^{POLS} = -0.04$ while the average $\hat{\beta}_1^{FE} = -0.25$. We also include a third column in which model (1) is estimated with cluster FEs δ_c at a level of aggregation larger than the selection pool, i.e., with $C = 4L$. The resulting estimator is denoted $\widehat{\beta}_1^{CL}$. We see that on average $\widehat{\beta}_1^{CL}$ suffers less exclusion bias than $\hat{\beta}_1^{FE}$ but more than $\hat{\beta}_1^{POLS}$.

The simulation results presented in Table 4 also illustrate what happens to exclusion bias as we increase sample size. As noted before, exclusion bias tends to be largest in small samples. When the sample size N increases, $\hat{\beta}_1^{FE}$ converges to the value given in formula (2) in Proposition 1, $\hat{\beta}_1^{POLS}$ converges to zero and $\widehat{\beta}_1^{CL}$ converges to a value in between the two. These patterns are not an artifact of a particular choice of parameter vector. They are general results, something we prove in Appendix A.4 and summarize in the following proposition:

Proposition 2 *Consider a population of N individuals indexed by i, divided into selection pools l of size L, and randomly partitioned into groups of size K within their pool. Each individual is assigned an outcome x_{ikl} potentially subject to a pool fixed effect but devoid of group fixed effect. Let Ω denote a finite, non-overlapping partition of the selection pools such that each cluster in Ω contains at least one pool, no two clusters contain the same pool, and the set Ω contains at least two clusters. Let $\delta_c = 1$ if an observation belongs to cluster c in Ω, and zero otherwise. Define three estimators $\widehat{\beta}_1^{POLS}$, $\widehat{\beta}_1^{CL}$, and $\widehat{\beta}_1^{FE}$ obtained from the following three OLS regressions, respectively:*

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-ikl} + \epsilon_{ikl} \tag{8}$$

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-ikl} + \delta_c + \epsilon_{ikl} \tag{9}$$

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-ikl} + \delta_l + \epsilon_{ikl} \tag{10}$$

Then:

$$plim[\hat{\beta}_1^{FE}] < plim[\hat{\beta}_1^{CL}] < plim[\hat{\beta}_1^{POLS}]$$

where $plim[\hat{\beta}_1^{FE}]$ is given by formula (2) in Proposition 1. Furthermore $plim[\hat{\beta}_1^{POLS}] = 0$ if $\delta_l = \delta_c = 0$ for all l, c.

Proof: see Appendix A.4.

The intuition behind the proof is as follows. The pooled OLS estimator is a weighted average

of the FE (within) estimator and the between estimator. The pool-FE estimator captures the extent to which variation in individual outcomes within a pool is explained by variation in peer outcomes within that pool. This correlation is affected by exclusion bias, which is always negative. The between group estimator, in contrast, measures the correlation between average individual outcomes and average peer outcomes across pools. This correlation is unaffected by exclusion bias and is naturally positive, even when $\delta_l = \delta_c = 0$ for all l, c.[10] Since $\hat{\beta}_1^{POLS}$ combines the negative exclusion bias contained in $\hat{\beta}_1^{FE}$ with the positive correlation across pools of the between estimator, $\hat{\beta}_1^{POLS}$ is less negative than $\hat{\beta}_1^{FE}$. The same reasoning applies to $\hat{\beta}_1^{CL}$ except that, by combining multiple selection pools within a cluster, $\hat{\beta}_1^{CL}$ captures part of the positive correlation across pools that is inherent to the between estimator. This explains the result. It also implies that, when $\delta_l = \delta_c = 0$ for all l, c, omitting cluster and pool FEs when testing random assignment leads to an asymptotic elimination of the exclusion bias under the null.

3 Estimating endogenous peer effects

In this section we allow the true β_1 to differ from 0 and we demonstrate how exclusion bias and reflection bias interact to jointly affect the estimation of endogenous peer effects. We also present an estimation method that corrects point estimates for both exclusion bias and reflection bias and does not require the presence of instruments. It does, however, rule out group correlated effects that are not absorbed by the inclusion of selection pool fixed effects.

Methods exist that, under certain conditions, allow peer effects to be estimated even in the presence of correlated effects – e.g., common shocks. One method applies to models of the form $y_i = \beta_0 + \beta_1 g_i y + \gamma x_i + \epsilon_i$ where g_i is the row of the network adjacency matrix identifying the neighbors of i and x_i is a vector of variables affecting only y_i. Correlated effects are allowed, e.g., $Cov(u_i, u_j) = \omega$ if i, j are neighbors and 0 otherwise. In this model, the peer effect coefficient β_1 can be estimated by using $g_i x$ as instrument for $g_i y$. This approach has long been criticized for requiring that y_i not be directly affected by $g_i x$, i.e., it rules out exogenous peer effects in x.

Bramoulle et al. (2009) estimate a more general model of the form $y_i = \beta_0 + \beta_1 g_i y + \gamma x_i + \theta g_i x + \epsilon_i$ by using the x values of the neighbors of the neighbors of i as instrument for $g_i y$.[11] While this

[10] When all peer groups are of equal size K, the average individual outcome in each pool is also the average of peer outcomes in that pool, and the correlation equals 1.

[11] The method can be extended to include further network lags as instruments. See for instance, Lee et al. (2012),

approach allows the joint estimation of endogenous and exogenous peer effects $g_i y$ and $g_i x$, it does not help in situations where individuals are partitioned into mutually exclusive groups. In this case the network matrix is block-diagonal, there are no neighbors of neighbors, and thus no neighbors of neighbors. Identification of β_1 through this method also requires that the set of neighbors of neighbors be significantly smaller than the rest of the selection pool. In small enough selection pools (e.g., a classroom), this is often not the case – in which case identification is, de facto, infeasible.

We know of only one paper that offers a method for estimating peer effects in mutually exclusive groups with correlated within-group effects, namely, Lee (2007). The approach suggested by the author is to rely on variation in the size of groups to distinguish peer effects (which operate like multipliers and thus are stronger in large groups) from correlated effects (which are constant with group size). Successful identification requires having sufficient variation in the size of peer groups.

The method we propose here allows peer groups to be of equal size – but rules out correlated effects within peer groups. It does, however, allow correlation within selection pools. While it applies to any network structure, its practical usefulness is highest when the above cited methods fail, e.g., for non-overlapping peer groups of equal (or similar) size, or when Bramoulle et al. (2009) instruments in principle exist but are too weak to achieve identification. Whether or not it is reasonable to assume away within-group correlated effects depends on the context. In the empirical section of this paper we offer two illustrations in which the assumption is defensible given the controlled nature of the experimental or quasi-experimental environment.

Formally, we consider a data generating process similar to that of Moffit (2001). To make the demonstration easier to follow we start by ignoring control variables and contextual effects. In Section 4.1 we extend the model to include regressors other than endogenous peer effects. We begin with a simple example in which group size $K = 2$. For this example, the exact value of the reflection and exclusion biases can be derived algebraically if we assume away unobserved correlated effects within peer groups. We then generalize the approach to any group size and we show how non-linear least squares/GMM can be used to obtain an estimate of β_1 that is free of both reflection and exclusion bias.

Appendix 5 for examples.

3.1 Model with group size $K = 2$

We start with a model where exclusion bias is absent so as to derive a precise formula of the reflection bias in our model. This formula allows us to conceptually distinguish the reflection bias from exclusion bias later on. For simplicity, we assume homoskedastic i.i.d. errors. We have $E[\epsilon_{ikl}] = 0$, $E[\epsilon_{ikl}^2] = \sigma_\epsilon^2$, $E[\epsilon_{ikl}\epsilon_{jml}] = 0$ for all $i \neq j$ and all $k \neq m$, and $E[\epsilon_{ikl}\epsilon_{jkl}] = 0$ for all k and all $i \neq j$. The $E[\epsilon_{ikl}\epsilon_{jkl}] = 0$ equality is far from innocuous since it assumes away the presence of what Manski (1993) calls correlated effects, that is, correlated ϵ_{ikl} between individuals belonging to the same peer group.[12] With this assumption, correlation in outcomes between members of the same peer group constitutes evidence of endogenous peer effects. The can be used for identification purposes as follows.

Following Moffit (2001), the estimating equations for any two individuals 1 and 2 in the same group can be written as:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 y_2 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 y_1 + \epsilon_2 \end{aligned}$$

where $0 < \beta_1 < 1, E[\epsilon_1] = E[\epsilon_2] = 0$ and $E[\epsilon^2] = \sigma_\epsilon^2$. We estimate:

$$y_1 = a + by_2 + v_1 \tag{11}$$

by OLS. Note that selection pool fixed effects are omitted. This means that exclusion bias disappears as sample size increases. Using part 2 of Proposition 1, we can show that the magnitude of the reflection bias is given by the following proposition:

Proposition 3: *[Proof in Appendix A.5]: If $E[\epsilon_1\epsilon_2] = 0$ (i.e., there are no correlated effects), the bias in model (11) is given by:*

$$plim_{N\to\infty}[\widehat{b}^{OLS}] = \frac{2\beta_1}{1+\beta_1^2} \tag{12}$$

An immediate corollary is that $plim_{N\to\infty}[\widehat{b}^{OLS}] = 0$ iff $\beta_1 = 0$, implying that the existence of

[12] As we show later, the model can easily accommodate FEs to capture correlated effects at the level of a cluster or selection pool.

peer effects can be investigated by testing whether $b = 0$. Moreover, formula (12) can be solved to recover an estimate of β_1 from the naive $\widehat{b}$, yielding:[13]

$$\widehat{\beta}_1 = \frac{1 - \sqrt{1 - \widehat{b}^2}}{\widehat{b}} \tag{13}$$

This demonstrates that identification can be achieved solely from the assumption of independence of ϵ_1 and ϵ_2, without the need for instrument.

As shown in Part 1 of Proposition 1, exclusion bias arises when selection pool fixed effects are added to model *(11)* and the size L of each selection pool is fixed. The estimated model is now $y_1 = a + by_2 + \delta_l + v_1$, which we can rewrite in deviation from the pool mean to eliminate the fixed effect δ_l:

$$\ddot{y}_1 = a + b\ddot{y}_2 + \ddot{\epsilon}_1 \tag{14}$$

where the notation $\ddot{z}_{ikl} \equiv z - \bar{z}_l$ where $\bar{z}_l$ is the sample mean of z in pool l. As demonstrated in part 1 of Proposition 1, for any i.i.d. variable z, there exists a sample correlation between any demeaned z_{ikl} and the demeaned average of a set of peers $\overline{z}_{ikl}$ in the same pool. This correlation is given by formula (2). This formula, with $K = 2$ also applies to demeaned errors $\ddot{\epsilon}_{ikl}$ and $\ddot{\epsilon}_{jkl}$:

$$\rho \equiv plim_{N\to\infty} SampleCorr(\ddot{\epsilon}_{ikl}\ddot{\epsilon}_{jkl}) = -\frac{L-1}{(L-2)L+1} = -\frac{1}{L-1} \tag{15}$$

Integrating this result into the algebra leading to Proposition 3, we obtain a formula for the size of the combined reflection and exclusion bias as follows:

Proposition 4: *[Proof in Appendix A.6] The bias in model (14) is given by:*

$$plim_{N\to\infty}[\widehat{b}^{FE}] = \frac{2\beta_1 + (1+\beta_1^2)\rho}{1 + \beta_1^2 + 2\beta_1\rho} \tag{16}$$

where $\rho = -\frac{1}{L-1}$.

[13]The other root can be ignored because it is always > 1 and peer effects in a linear-in-means model cannot exceed 1. Furthermore, in the simple model presented here, the maximum value that $\hat{b}$ can take is 1, which arises when y_1 and y_2 are perfectly positively correlated. Similarly, the smallest value it can take is -1, which arises if they are perfectly negatively correlated. It is thus impossible for the absolute value of $\hat{b}$ to exceed 1, which guarantees the generality of the formula.

As for Proposition 3, we can take roots of formula (16) to obtain a consistent estimate $\widehat{\beta_1}$ as: [14]

$$\widehat{\beta}_1^{FE} = \frac{1 - \widehat{b}\rho - \sqrt{1 + \widehat{b}^2\rho^2 - \widehat{b}^2 - \rho^2}}{\widehat{b} - \rho} \tag{17}$$

We present in Table 5 calculations based on formula (17) and simulation of $\widehat{b}^{FE}$ over 100 replications to illustrate the magnitude of the reflection and exclusion bias for various values of β_1 and for $N = 500$, $L = 20$ and $K = 2$.[15] We see that, when β_1 is zero or small, the total predicted bias is dominated by the exclusion bias and is thus negative. As β_1 increases, 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_1, 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 $\beta_1 = 0$, then $plim_{N\to\infty}[\widehat{b}^{FE}] = \rho < 0$ which is the size of the exclusion bias.

2. It is possible for $plim_{N\to\infty}[\widehat{b}^{FE}]$ to be negative even though $\beta_1 > 0$. This arises when ρ is large in absolute value, for instance if $L = 20$ and $K = 2$ as in Table 5.

3. Since the exclusion bias is always negative, $\widehat{b}^{FE} > 0$ can only arise if $\beta_1 > 0$. It follows that, in this model, a positive $\widehat{b}^{FE}$ unambiguously indicates the presence of peer effects.

While formula (17) can be used to obtain a corrected estimate of the peer effect coefficient β_1, there remains the important question of inference: how can we test whether $\widehat{\beta}_1^{FE}$ is significantly different from 0. In order to obtain correct inference, we need to correct p-values for the standard test of significance that $\beta_1 = 0$. The solution is to use one of the methods discussed in Section 2.4 since the null hypothesis is the same.

To illustrate, we present the results of a Monte Carlo study in Table 6. We create random samples of $N \times L$ =10,000 observations following the data generating process described above but

[14] There are two roots, but one of them is larger than one and can thus be ignored as a realistic value for β_1. Indeed, in a linear-in-means such as the one here, $\beta_1 > 1$ implies an explosive solution for the y_1 and y_2 system of equation, i.e., $y_1 = \infty = y_2$ – or possibly a corner solution (not modeled here). As long as the researcher observes interior values of y, we can ignore the $\beta_1 > 1$ root as plausible value.

[15] We use a large sample size of $N \times L = 10,000$ to show convergence of the simulation results to the predicted values. Given that each replication takes a long time for such a large sample, we restrict the number of replications to 100 in this exercise, which is sufficient to illustrate this point for samples of size $N \times L = 10,000$.

for different values of β_1. We then use randomization inference to obtain correct p-values using 500 permutation replications for each regression. We set $K = 2, L = 20$ and $N = 500$. In columns 2 and 3 we report the mean simulated $\widehat{b}^{FE}$ and the corresponding p-value as reported by OLS. Reported p-values are for two-sided tests. In column 4 we report the corrected estimate $\widehat{\beta}_1^{FE}$ obtained using formula (17). The last column presents the corrected p-values obtained from 500 bootstrapping replication of the null hypothesis of no peer effect. Results confirm that $\widehat{b}^{FE}$ is dramatically biased, sometimes yielding a significantly negative estimate when the true β_1 is close to zero, sometimes yielding an inflated estimate when reflection bias dominates. Corrected estimates $\widehat{\beta}_1^{FE}$ do not display this pattern: they are centered on the true β_1. We also note that using corrected p-values eliminates the risk of incorrectly concluding that $\beta_1 < 0$. When the true value of β_1 is positive but small, we are unable to reject that $\beta_1 = 0$, an indication that power is not always sufficient to identify the presence of peer effects. As a whole, however, the method produces a massive improvement in inference.

3.2 General group model

In the case where $K = 2$ we were able to derive an algebraic formula to correct the estimate of β_1. Obtaining a closed-form formula becomes difficult if not impossible when we generalize to a larger group size K or to groups of varying size. But provided that we are willing to assume i.i.d. errors conditional on selection pool fixed effects, it remains possible to obtain an estimate of the true β_1 and to bootstrap its p-value.

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

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i \tag{18}$$

where Y is the 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_iY is equal to $\bar{y}_{-ikl}$. But this can be generalized to other influence models by varying G_i, for instance by letting G be a network adjacency matrix (see below).

Regression model (18) can be written in matrix form as:

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

Simple algebra yields the following reduced-form model:

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

from which we obtain:

$$\begin{aligned} E[YY'] &= E[(I-\beta G)^{-1}(\gamma X + \delta GX + \epsilon)\ (\gamma X + \delta GX + \epsilon)'(I - \beta G')^{-1}] \\ &= (I-\beta G)^{-1}E[(\gamma X + \delta GX)(\gamma X + \delta GX)'](I-\beta G')^{-1} \\ &\quad +(I-\beta G)^{-1}E[\epsilon\ \epsilon'](I-\beta G')^{-1} \end{aligned} \tag{19}$$

where we have assumed that the G matrix is non-stochastic. As in Liu (2017), 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 (19) can be used as starting point for estimation.

With exclusion bias – e.g., if all variables in the above model are expressed in deviation from their pool mean – $E[\ddot{\epsilon}\ \ddot{\epsilon}'] \neq \sigma_\epsilon^2 I$. Formula (15) can then be used to construct the asymptotic covariance matrix of the $\ddot{\epsilon}$'s.

To illustrate, suppose that all observations are arranged so that the observations from the first pool come first, then the observations from the second pool, etc. In this case $E[\ddot{\epsilon}\ \ddot{\epsilon}']$ is a block-diagonal matrix:

$$E[\ddot{\epsilon}\ \ddot{\epsilon}'] = \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & B & 0 & 0 \\ 0 & 0 & B & 0 \\ 0 & 0 & 0 & B \end{bmatrix} \tag{20}$$

where each block B is an $L \times L$ matrix of the form:

$$B = \begin{bmatrix} E[\ddot{\epsilon}_1^2] & E[\ddot{\epsilon}_1 \ddot{\epsilon}_2] & ... \\ E[\ddot{\epsilon}_2 \ddot{\epsilon}_1] & E[\ddot{\epsilon}_2^2] & ... \\ ... & ... & ... \end{bmatrix} \tag{21}$$

From equation (15), we know that, for two individuals i and j in the same selection pool of size L, we have $E[\ddot{\epsilon}_i \ddot{\epsilon}_j] = \rho \sigma_\epsilon^2$ with $\rho = -\frac{1}{L-1}$ for $i \neq j$. Hence B can be rewritten as:

$$B = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho & ... \\ \rho & 1 & ... \\ ... & ... & ... \end{bmatrix} \equiv \sigma_\epsilon^2 A \tag{22}$$

What is important is that the asymptotic value of ρ is known and need not be estimated.

Equation (19), combined with (20) and (22), provides a characterization of the data generating process that can be used to estimate structural parameters β, γ, δ and σ^2. Identification is achieved from the assumption that, conditional on pool fixed effects, errors ϵ_{ikl} are independent across observations from the same peer group. With this assumption, instruments are not required in spite of the presence of reflection bias.[16] 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 (19) is to rely on the method of moments to choose the parameter β that provides the best fit to the observed data $E[YY']$. This is achieved using a search algorithm. For each guess $\beta^{(n)}$ that the algorithm makes about β, we solve for the corresponding values of γ and δ by calculating $Y - \beta^{(n)} GY$ and regressing it on X and GX to obtain estimates of $\widehat{\gamma}^{(n)}$ and $\widehat{\delta}^{(n)}$. This process also yields an estimate of the variance of errors $\widehat{\sigma}_\epsilon^{2(n)}$. Using $\beta^{(n)}, \widehat{\gamma}^{(n)}, \widehat{\delta}^{(n)}$ and $\widehat{\sigma}_\epsilon^{2(n)}$ we compute the value of each element of the right hand side of equation (19). Subtracting each value from the corresponding $y_i y_j$, taking squares, and summing over all ij pairs yields the value of the 'fit' for guess $\beta^{(n)}$. We then search over possible values of β to achieve the best fit/lowest

[16] The observation that the cross product of the errors can be used to estimate a network or spatial autocorrelation parameter has been made before, e.g., in Kelijian and Prucha (1999) for spatial models. But, to our knowledge, it has not been proposed as a way of overcoming the issue of reflection bias, as done here.

sum of squared residuals in equation (19).

To illustrate the effectiveness of this approach, we estimate model (18) on simulated data. The average results from 1000 Monte Carlo replications are shown in Table 7. We keep the number of observations in each sample constant at $N = 1000$ but we vary K and β_1 across simulation exercises. Pool fixed effects are included throughout. In the first two rows we report the uncorrected $\widehat{\beta_1}^{OLS}$ and its p-value obtained by regressing Y_i on G_iY and pool dummies. Results confirm that the uncorrected $\widehat{\beta_1}^{OLS}$ and the inference based on it is biased. As before this bias comes from two sources: reflection and exclusion bias. When β_1 is small, the exclusion bias dominates and the naive $\widehat{\beta_1}^{OLS}$ underestimates the true β_1. On average, $\widehat{\beta_1}^{OLS}$ is more likely to overestimate the true β_1 when exclusion bias is small, which occurs when L is large. The third row shows the proportion of times the simulated naive p-value is smaller or equal to 0.05. In the third row of column 1 and column 4 (where $\beta_1 = 0$), this statistic essentially gives us the likelihood of making a type II error, that is, the probability of rejecting the null hypothesis when it is in fact true. If the estimator is unbiased then we would expect this statistic to be close to 5%. In the third row of columns 2-3 and columns 5-6 (where $\beta_1 > 0$), this statistic is indicative of the statistical power of the test, that is, the probability of rejecting the null hypothesis when it is not true. If the estimator is unbiased then we would expect this statistic to be close to 100%. Combined these result show that the probability of making a type II error and the statistical power of the test are very high for the naive estimator, particularly for $K = 5$. In the fourth row in Table 7, we report the average of $\widehat{\beta_1}^{Ref}$ estimates corrected for reflection bias but ignoring the exclusion bias. This is estimated using model (19) with $E[\epsilon\ \epsilon'] = \sigma_\epsilon^2 I$. In all cases, the average estimate is closer to the true β_1, but the failure to eliminate exclusion bias results in an underestimation of the true β_1 on average. The fifth row reports the average $\widehat{\beta_1}^{Corr}$ derived from model (19) with $E[\ddot{\epsilon}\ \ddot{\epsilon}']$ given by (20). The $\widehat{\beta_1}^{Corr}$ is centered around its true value in all cases. The sixth row in Table 7 shows the corrected p-values obtained using the permutation method described earlier. We see that the method yields unbiased inference. Moreover, the last row in the first column shows that the permutation-based inference method has relatively high statistical power and a low probability of rejecting the null hypothesis when it is in fact true.

In the form presented here, the method does not accommodate heteroskedastic ϵ errors. Borrowing from Liu (2017), it may nonetheless be possible to generalize the approach to heteroskedastic

errors by relying on a root estimator instead. This would require considering a moment condition of the form $E[YGY']$ and using the consistent root of this equation to estimate peer effects when instruments are not available. The advantage of using this approach is that $E[\epsilon G\epsilon'] = 0$ even if the errors are heteroskedastic (continuing to rule out correlated effects within peers). Thus the estimator is heteroskedasticity robust. Using this approach while correcting for exclusion bias is left for future research. Thanks to an anonymous referee for making this suggestion.

3.3 Network data

Until now we have considered situations in which peers form mutually exclusive groups, i.e., such that if i and j are peers and j and k are peers, then i and k are peers as well. Exclusion bias also arises when peers form more general networks, i.e., such that i and k need not be peers. To illustrate this, let us consider the canonical case examined in Section 3.2 and assume that individuals in selection pool l are randomly assigned peers within that pool. The only difference with Section 3.2 is that we no longer restrict attention to mutually exclusive peer groups but allow links between peers to take an arbitrary (including directed or undirected) network shape within each pool l. Partially overlapping groups and mutually exclusive groups of unequal size can be handled in the same manner.

The approach developed to estimate general group models with uncorrelated errors can be applied to network data virtually unchanged. Equation (19) remains the same. Formally let $g_{ijl} = 1$ if i and j in cluster l are peers, and 0 otherwise. We follow convention and set $g_{ii} = 0$ always. The network matrix in cluster l is written $G_l = [g_{ijl}]$ and G is a block diagonal matrix of all G_l matrices.

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

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

where, as before, L denotes the size of the selection pool.[17] The average outcome of i's peers can

[17] To illustrate, let $L = 4$ and assume that individual 1 has individuals 2 and 4 as peers. Then $\widehat{g}_{il} = [0, \frac{1}{2}, 0, \frac{1}{2}]$.

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

$$y_{ikl} = \beta_0 + \beta_1 \widehat{g}_{il} y_l + \delta_l + \epsilon_{ikl} \tag{23}$$

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

$$Y = \beta \widehat{G} Y + \gamma X + \delta \widehat{G} X + \epsilon \tag{24}$$

As in the previous section, equation (19) combined with (15), (20) and (22) can be used to estimate structural parameters β, γ, δ and σ^2. The only difference is that G is now a network matrix rather than a block-diagonal matrix. It is intuitively clear that exclusion bias affects model (23) as well: individual i is still excluded from the selection pool of its own peers, and this continues to generate a mechanical negative correlation between i's outcome and that of its peers. The same asymptotic formula is used to substitute for parameter ρ as before. Pre- and post-multiplying matrix $E[\ddot{\epsilon}\ \ddot{\epsilon}']$ by $(I - \beta G)^{-1}$ in expression (19) picks the relevant off-diagonal elements of B to construct the needed correction for exclusion bias. Estimation proceeds using the same iterative algorithm as described above.

We illustrate this approach for network data in Table 8. We generate each adjacency matrix G_l as a Poisson random network with linking probability p. In other words, p is the probability that a link exists between any two individuals i and j within the same pool. When $p = 0.1$ and $L = 20$, each individual has two peers on average; when $p = 0.25$ (0.5) each individual has on average 5 (10) peers, respectively. Table 8 provides simulation results and shows how our suggested method of moments correction method is able to correct the estimate of β_1 to be close to the true β_1.

The permutation method can be adapted to correct p-values for this case as well. To recall, we want to simulate the counterfactual distribution of $\widehat{\beta}_1$ under the null hypothesis of zero peer effects. In contrast with Section 3, peers are no longer selected by randomly partitioning individuals into groups within pools, but rather by randomly assigning peers within pools. In practice, we keep the

network matrices in each selection pool unchanged but we change who is linked to whom. This approach is known in the statistical sociology literature as Quadratic Assignment Procedure or QAP (e.g., Krackhardt 1988).

To implement this approach within pool l, we scramble matrix G_l in the following way. Say the original ordering individual indices in l is $\{1, ..., i, ..., j, ..., L\}$. We generate a random reordering (k) of these indices, e.g., $\{j, ..., 1, ..., L, ..., i\}$. We then reorganize the elements of G_l according to this reordering to obtain a counter-factual network matrix $G_l^{(k)}$. To illustrate, imagine that i has been mapped into k and j into m. Then element g_{ijl} of matrix G_l becomes element g_{kml} in matrix $G_l^{(k)}$. We then use this matrix to compute the average peer variable $\widehat{g}_{il}^{(k)} y_l$. For each reordering (k) we estimate model (23) and obtain a counter-factual estimate $\widehat{\beta}_1^{(k)}$ corresponding to the null hypothesis of zero peer effects. 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.

In Table 8 we compare the p-values obtained from the naive model and the permutation approach applied to model (23). We find that the performance of the estimation method in the network case is comparable to what it was in the peer group case.

4 Avoiding exclusion bias

4.1 Exogenous peer effects

When estimating exogenous peer effects, it is possible to eliminate the exclusion bias by using control variables. A good example is 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 whom they play throughout the tournament. The performance of golfer i in tournament l is written as $y_{ikl,t+1}$. The researcher has information on the performance of each golfer i in past golf tournaments. This information is denoted as y_{iklt}. The researcher wishes to test whether the performance of golfer i in tournament l depends on the past performance of the golfers i is paired with. The researcher's

objective is thus to estimate coefficient β_1 in a regression of the form:

$$y_{ikl,t+1} = \beta_0 + \beta_1 \bar{y}_{-iklt} + \delta_l + \epsilon_{ikl,t+1} \tag{25}$$

where $\bar{y}_{-iklt}$ denotes the average past performance of i's assigned peers. A key difference with the models discussed earlier is that here $\bar{y}_{-iklt}$ is calculated using the *past* performance of peers in other tournaments, before being assigned to be i's peers. Because of exclusion bias, $\bar{y}_{-iklt}$ is mechanically negatively correlated with y_{iklt} due to the presence of pool fixed effects. Since i's past performance is correlated with i's unobserved talent, we expect y_{iklt} to be positively correlated with $y_{ikl,t+1}$. This generates a negative correlation between $\bar{y}_{-iklt}$ and the omitted variable y_{iklt} which is part of the error term. The result is a negative bias for β_1 in regression (25).

The example suggests an immediate solution: include y_{iklt} as additional regressor to eliminate the exclusion bias:

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

where y_{iklt} serves as control variable. This is the approach adopted, for instance, in Munshi (2004).

A similar reasoning applies if the researcher wishes to test the influence of the pre-existing characteristics of peers $\bar{x}_{-ikl}$ on i's subsequent outcome $y_{ikl,t+1}$ and includes pool fixed effects.[18] Here too the pre-existing characteristics of peers are negatively correlated with i's pre-existing characteristic x_{ikl}. Hence if the researcher fails to control for x_{ikl} and x_{ikl} is positively correlated with $y_{ikl,t+1}$, then estimating a model of the form:

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

will result in a negative exclusion bias.[19] This bias can be corrected by including x_{ikl} as control, as done for instance in Bayer et al. (2009):

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

[18] As discussed in Proposition 1 Part 3, even if the researcher does not include pool fixed effects, there is still an exclusion bias if the pool size L is small enough.

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

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

4.2 Endogenous peer effects

When estimating endogenous peer effects, the use of instrumental variables can – under certain conditions – eliminate exclusion bias. One case that is particularly relevant in practice is when the researcher uses the peer average of a variable z to instrument peer effects, but also includes z_i in the regression. To illustrate this formally, let us assume that the researcher has a suitable instrument $\bar{z}_{-ikl}$ for $\bar{y}_{-ikl}$. For instance, $\bar{z}_{-ikl}$ 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 $\bar{z}_{-ikl}$ is informative about $\bar{y}_{-ikl}$, then z_{ikl} should be informative about y_{ikl} as well. For this reason, z_{ikl} is often included in the estimated regression as well. In this case, the first and second stages of this 2SLS estimation strategy can be written as follows:

$$\bar{y}_{-ikl} = \pi_0 + \pi_1 \bar{z}_{-ikl} + \pi_2 z_{ikl} + \delta_l + v_{ikl} \tag{26}$$

$$y_{ikl} = \beta_0 + \beta_1 \hat{\bar{y}}_{-ikl} + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \tag{27}$$

where $E(z_{ikl}\epsilon_{ikl}) = 0$, $E(\epsilon_{ikl}) = 0$ and $\hat{\bar{y}}_{-ikl} = \hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{-ikl} + \hat{\pi}_2 z_{ikl} + \hat{\delta}_l$ is the fitted value from the first-stage regression.[20]

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

$$\begin{aligned} y_{ikl} &= \beta_0 + \beta_1 \hat{\bar{y}}_{-ikl} + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \\ &= \beta_0 + \beta_1(\hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{-ikl} + \hat{\pi}_2 z_{ikl} + \hat{\delta}_l) + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \end{aligned} \tag{28}$$

If y_{ikl} and z_{ikl} are correlated (i.e., if $\beta_2 \neq 0$), we expect $\bar{z}_{-ikl}$ to be mechanically correlated with y_{ikl} because $\bar{z}_{-ikl} = \frac{\left[\sum_{s=1}^{N}\sum_{j=1}^{K} z_{js}\right] - z_{ikl}}{L-1} + \tilde{u}_{ikl}$, where $\tilde{u}_{ikl} \equiv \bar{z}_{-ikl} - \bar{z}_{-il}$. Since equation (28) controls for z_{ikl} directly, this mechanical relationship is prevented from generating an exclusion bias.

Since such 2SLS strategies eliminate the negative exclusion bias, they yield peer effect estimates that are *larger* – i.e., more positive – than OLS estimates. This counter-intuitive finding is often attributed to classical measurement error or some other cause (e.g., Goux and Maurin 2007, Halliday and Kwak 2012, De Giorgi et al. 2010, de Melo 2014, Brown and Laschever 2012, Helmers and Patnam 2012, Krishnan and Patnam 2012, Naguib 2012). Exclusion bias offers an alternative, mechanical explanation.

The above examples serve to illustrate that for 2SLS to effectively eliminate exclusion bias, it is necessary to control for i's own value of the instrument z_{ikl} in equation (26). This condition is satisfied, for instance, by the estimation strategies employed by Bramoulle et al. (2009), Di Giorgi et al. (2010) or Lee (2007). Any instrumentation method that fails to do so suffers from exclusion bias in the same way and for the same reason as OLS.

5 Empirical applications

We now illustrate how our method can be used to estimate endogenous peer effects in situations where the IV methods suggested by Bramoulle et al. (2009), De Giorgi et al. (2010) and Lee . (2007) all fail. To this effect we revisit two data sets in which subjects within a selection pool are randomly assigned to groups of fixed size. The first dataset comes from golf tournaments in which participants are randomly assigned to groups of three players within their qualification category. It is the same dataset as that used by Guryan et al. (2009).[21] The second dataset comes from Fafchamps and Mo (2018) and includes Chinese primary school students randomly paired within their classroom for a computer-assisted course lasting the entire academic year.[22] In both cases, we limit our data to groups of the same size – three in the golfer data and two in the student data.[23] This is done to demonstrate that our method provides identification even with groups of fixed size.

[21] Verification of the data reveals that some players (25% of all observations) had been assigned to groups of size larger than $K = 3$. We drop these observations from the sample.

[22] Using results from an earlier version of our current paper, Fafchamps and Mo (2018) report p-values corrected for exclusion bias in their test of random peer assignment using our method described in Section 2.4. In their main peer effect estimations, they abstract from reflection bias and exclusion bias by focusing on the estimation of contextual peer effects only. In this section we use the methods described in the current paper to yield estimates of endogenous peer effects that correct for both reflection bias and exclusion bias.

[23] In the golfer data, players assigned to groups of two players account for 25% of observations – some tournaments are played in pairs instead of triads.

In both cases we estimate a model of the following form:

$$y_{ikl,t+1} = \beta_0 + \beta_1 \bar{y}_{-ikl,t+1} + \beta_2 y_{ikl,t} + \beta_3 \bar{y}_{-ikl,t} + \delta_l + \epsilon_{ikl,t+1} \tag{29}$$

where, as before, $y_{ikl,t+1}$ denotes an outcome of interest for individual i in group k from selection pool l at time $t+1$ and $\bar{y}_{-ikl,t+1}$ is the average value of $y_{kl,t+1}$ for the peers of i in group k from selection pool l. Coefficient β_1 is the endogenous peer effect. Regressors y_{iklt} and $\bar{y}_{-iklt}$ measure the past performance of i and of his/her peers. Coefficient β_3 estimates what is commonly referred to as an exogenous peer effect (or contextual peer effect). We include pool fixed effects δ_l and assume that the residuals $\epsilon_{ikl,t+1}$ are not correlated within groups. The suitability of this assumption depends on the context but, given the inclusion of pool fixed effects, it is unproblematic in the two datasets we have selected. This assumption means that any correlation in outcomes within group must come either from endogenous or exogenous peer effects. As noted earlier, regressor y_{iklt} needs to be included to avoid exclusion bias in β_3.

Model (29) is estimated by expressing it in the same form as in equation (19) and using the GMM-based iterative algorithm developed in Section 3.2. To recall, this algorithm iterates on β_1 guesses to find the best fit to the data. For each guess about $\beta_1^{(n)}$, we estimate a regression of the form:

$$\widetilde{y}_{ikl,t+1} = \beta_0 + \beta_2 y_{iklt} + \beta_3 \bar{y}_{-iklt} + \delta_l + \epsilon_{ikl,t+1} \tag{30}$$

where $\widetilde{y}_{ikl,t+1} \equiv y_{ikl,t+1} - \beta_1^{(n)} \bar{y}_{-ikl,t+1}$. This regression is then demeaned and combined with equation (20) to compute the right-hand side of equation (19) that corresponds to that particular guess $\beta_1^{(n)}$. The algorithm then seeks the value of $\beta_1^{(n)}$ that minimizes the distance between the constructed matrix and the data matrix $E[YY']$. The p-value for $\widehat{\beta}_1$ is obtained using randomizing inference as in Section 2.4 – that is, by constructing artifactual samples in which groups are formed at random within selection pools and simulating the distribution of $\widehat{\beta}_1$ under the null hypothesis of no endogenous or exogenous peer effects. Estimates for β_2 and β_3 are those given by model (30) at the optimal value of $\widehat{\beta}_1$; their standard errors are clustered by selection pool.

Results for golfers using data from Guryan et al (2009) are presented in the first column of Table 9. To keep the estimation as transparent as possible, we restrict our attention to the first

round of each tournament and we drop observations from the second round that could potentially provide an additional source of identification. We also drop observations involving golfers assigned to a group outside their selection pool (6% of observations) since they do not fit our postulated data generation process. To demonstrate the efficiency of the approach, we also reduce the number of observations by focusing on a random sub-set of 100 out of 302 selection pools. This makes the sample size comparable to the student data and speeds up simulation-based inference later. We also focus on groups of size 3 ($K = 3$) which consists of 95% of all observations. This leaves a sample of 2,517 observations from 100 pools of roughly 25 golfers each, organized in groups of three.

The estimates for regression model (29) are presented in the first panel of Table 9. In this empirical application, the outcome of interest is the golfer's score. Unsurprisingly, the golfer's past tournament performance is a strong predictor of current performance: $\widehat{\beta}_2$ is large and significant. Regarding peer effects, we find a positive endogenous peer effect $\hat{\beta}_1^{Corr}$ significant at the 1% level. The magnitude of the coefficient is large: i's performance increases by 5.8% of the average performance of the two golfers in i's group, conditioning for their average past performance. Given the multiplier effect induced by reflection, the *total* impact on performance is even larger. This suggests that emulation between players helps performance in golf tournaments: when one player in a group plays above its own average, the other players in that group also tend to play better than normal. The opposite holds as well: when a golfer in a group plays worse than normal, this has a negative ripple effect on the other golfers in that group. Importantly, this effect is not a result of matching: the exogenous peer effect coefficient $\widehat{\beta}_3$ is not significant and, if anything, it is negative. The results presented here therefore suggest that emulation comes from play during the tournament, not from who golfers are grouped with.

To illustrate how these results compare with alternative estimation strategies, we report in the second panel of Table 9 the point estimate $\hat{\beta}_1^{OLS}$ obtained by OLS and the point estimate $\hat{\beta}_1^{Ref}$ obtained from the GMM estimator to correct for reflection, but ignoring exclusion bias. In practice, $\hat{\beta}_1^{Ref}$ is estimated by erroneously setting $\rho = 0$ in matrix (22). As predicted in Section 3, the OLS point estimate 'shrinks' when correcting for reflection bias: $\widehat{\beta}_1$ drops from 0.022 to 0.010. We also note that the naive p-value of $\hat{\beta}_1^{OLS}$ wrongly concludes that there are no endogenous peer-effects. However, if we use randomization inference to obtain a consistent p-value for $\hat{\beta}_1^{OLS}$ we get $p = 0.014$, indicating the presence of endogenous peer effects. The reason for this is illustrated in

Figure 5 where we plot the simulated distribution of $\hat{\beta}_1^{OLS}$ under the null of no endogenous peer effects: this distribution is centered well below 0, compared to the simulated distribution of $\hat{\beta}_1^{Corr}$ under the null, which is centered on $\beta_1 = 0$. This is yet another illustration of the fact that, as shown in Section 2.4, it is possible to test for the presence of endogenous peer effects by applying randomization inference directly on OLS estimates.[24] We also see that the simulated estimator $\hat{\beta}_1^{Corr}$ has a smaller variance than $\hat{\beta}_1^{OLS}$ under the null. This is because each $\hat{\beta}_1^{OLS}$ estimate is magnified by reflection and thus varies more across samples. A similar reasoning applies to $\hat{\beta}_1^{Ref}$, shown in the third panel of Table 9 together with its p-value obtained by randomization inference. Here too randomization inference yields a p-value that indicates the presence of endogenous peer effects, in spite of the fact that $\hat{\beta}_1^{Ref}$ is quite small in magnitude. The explanation is illustrated in Figure 5: under the null of $\beta_1 = 0$, the simulated distribution of $\hat{\beta}_1^{Ref}$ is tighter than the distribution of $\hat{\beta}_1^{OLS}$ since reflection bias has been eliminated, but it is shifted to the left of 0 due to exclusion bias.

The second column of Table 9 presents similar estimates for the student data of Fafchamps and Mo (2018). Here too we find that, as expected, the past math score is a strong and significant predictor of the future score: $\widehat{\beta}_2$ is large and significant and, amusingly, of same magnitude as in the golfer data. The fact that $\widehat{\beta}_2$ is well below one indicates strong reversion to the mean among our primary school student population. Results for $\hat{\beta}_1^{Corr}$ are quite different to those we obtained for golfers, however: the point estimate is negative and significant, indicating that endogenous peer effects are negative – suggesting for instance congestion effects in computer usage. We also find some evidence of positive exogenous peer effects: a pupil assigned to share a computer with a stronger math student tends to learn slightly more from computer-assisted learning. The latter result is reminiscent of what Fafchamps and Mo (2018) conclude in their own analysis, but the negative endogenous peer effect is a new result.

Estimates $\hat{\beta}_1^{OLS}$ and $\hat{\beta}_1^{Ref}$ are reported in the bottom half of Table 9. We find that $\hat{\beta}_1^{OLS}$ is negative and unrealistically large in magnitude. The OLS p-value suggests the presence of negative endogenous peer effects. The $\hat{\beta}_1^{Ref}$ shrinks towards zero, due to correction for reflection bias. Applying randomization inference to both estimators yields significant p-values, again confirming that inference about the presence of endogenous peer effects can be undertaken without estimating

[24] This is basically what the quadratic assignment procedure (QAP) of Krackhardt (1999) does.

$\hat{\beta}_1^{Corr}$directly.

Given that $K = 2$ in the student data, we can use formulas (13)-(17) to obtain exact predictions about the plim of $\hat{\beta}_1^{OLS}$, $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{Corr}$ under the null. These predictions are shown in Table 10 and compared to the means of the simulated distributions of $\hat{\beta}_1^{OLS}$, $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{Corr}$ shown in Figure 6. As predicted by Proposition 1, $\hat{\beta}_1^{OLS}$ is centered around -0.059 instead of being centered around the true $\beta_1 = 0$. Under the null, formula (13) predicts $\hat{\beta}_1^{Ref}$ to be centered around -0.029, which is close to the average of -0.026 yielded by the simulations shown in Figure 6. Similarly, by applying formulas (15) and (17), we expect $\hat{\beta}_1^{Corr}$ to be centered on zero. The simulation average of $\hat{\beta}_1^{Corr}$ is 0.001. Finally, formulas (13)-(17) predict an exact linear relationship between $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{OLS}$, and between $\hat{\beta}_1^{Corr}$ and $\hat{\beta}_1^{OLS}$. Given this relationship, Columns 1 and 3 in Table 11 show the predicted constant and coefficient estimate of a regression of simulated $\hat{\beta}_1^{Ref}$ on simulated $\hat{\beta}_1^{OLS}$, and of simulated $\hat{\beta}_1^{Corr}$ on simulated $\hat{\beta}_1^{OLS}$, respectively. Column 2 and column 4 in turn show the actual estimation results. Notwithstanding small differences due to Monte Carlo approximation error, the predicted values are strikingly similar to the actual simulation results.

6 Application to time series autoregressive models

The methodological approach proposed here can in principle be applied to autoregressive models other than those operating on network or group data. We illustrate this with a time series autoregressive model with fixed effects of the form:

$$x_{it} = \beta_1 x_{it-1} + \delta_i + \epsilon_{it} \tag{31}$$

where T is small and N is large. Here T serves the same role as L in peer effect models: it is the size of the pool from which peers (here, the $t-1$ neighbor of t) are drawn. Such models are known to suffer from bias (Nickell 1981) and various instrumentation strategies have been proposed to estimate them (e.g., Arellano and Bond 1991, Arellano and Bover 1995, Blundell and Bond 1998).

Using an approach similar to Proposition 1, the asymptotic bias in β_1 under the null can easily be derived as:

Proposition 5: *When the true $\beta_1 = 0$, estimates of β_1 in model (31) satisfy:*

$$plim_{N\to\infty}(\hat{\beta}_1) = -\frac{1}{T-1} = \rho \tag{32}$$

See Appendix A.7 for a proof. Interestingly, the limit given by formula (32) is the same as that given by Proposition 1 Part 1 for $K = 2$ and it is equal to the value of ρ in equation (15). Formula (32) shows how large the Nickell bias is at the null: for $T = 3$, the shortest panel for which instruments exist, the *plim* of $\hat{\beta}_1$ under the null of $\beta_1 = 0$ is -0.5; for $T = 10$, the asymptotic bias under the null is still -0.111.[25]

The good news is that the different approaches proposed here also work for model (31). For instance, if the researcher is solely interested in testing whether $\beta_1 = 0$, this is easily achieved by creating a variable $\tilde{x}_{it} \equiv x_{it} - \rho x_{it-1}$ and regressing it on x_{it-1}, as indicated in equation (7). The GMM estimation model (19) can similarly be used by setting network matrix G to have 1's immediately to the left of the diagonal, and 0's everywhere else, so as to pick the lagged value of the dependent variable in lieu of the 'average of peers'. Everything we said about inference applies as well. While this approach allows the estimation of β_1 in model (31) without recourse to instruments, it does impose the fairly strict requirement that errors ϵ_{it} be i.i.d. within each pool, which precludes autocorrelated errors.

7 Concluding remarks

This paper has examined an under-studied source of downward bias in the estimation of peer effects. This bias exists on top of other, well-known problems such as reflection bias and correlated effects, and it arises even if peers are randomly assigned. We provided a comprehensive treatment of its causes and consequences and offered ways to correct it in the estimation of endogenous peer effects.

We first have shown that, with selection pool fixed effects, a negative correlation in peer outcomes mechanically arises because individuals cannot be their own peers: i.e., they are excluded from the pool from which their peers are drawn – hence its moniker 'exclusion bias'. We have demonstrated that the exclusion bias can seriously affect point estimates and inference in standard tests of random peer assignment and in the estimation of endogenous peer effects. The magnitude of the bias is most prevalent in studies that include pool fixed effects as well in studies with large

[25] See Nickel (1981) and Arellano (2003) for simulations of the bias when $\beta_1 \neq 0$.

peer groups relative to the size of the peer selection pool.

In contrast to exclusion bias, the widely-publicized reflection bias is little more than a multiplier effect. It follows that, if exclusion bias did not exist and we are willing to assume zero correlated effects within groups, inference about the *presence* of endogenous effects can be conducted using OLS: the reflection bias simply magnifies OLS estimates of endogenous peer effects. In this paper, however, we have shown that when the true peer effect is small and pool fixed effects are included, the negative exclusion bias dominates the reflection bias, yielding an overall negative bias in OLS estimates of peer effects. Hence if OLS yields an insignificant or even negative estimate of endogenous peer effects, a researcher unaware of exclusion bias will conclude that (positive) peer effects are absent and the issue is not worthy of further investigation. Because of this, we suspect that many peer effect studies have never seen the light of day – creating a so-called 'file drawer problem'.

To demonstrate that exclusion bias is not simply an irrelevant statistical oddity, we provide two empirical examples based on published papers in which lack of awareness about exclusion bias leads to incorrect inference. In the first example, we show that the OLS estimate of peer effects is close to zero and not statistically significant, which would normally be interpreted as prima facie evidence against peer effects. When we correct for exclusion bias, however, we find significant evidence of endogenous peer effects. In a second example, the OLS estimate of endogenous peer effects is negative and large in magnitude. Correcting for exclusion bias reduces the size of the estimate but confirms the presence of negative peer effects. These two examples illustrate the policy relevance of the method: finding no peer effects where these are actually present could have serious implications for policy makers.

We have also presented an alternative to the estimation of peer effects using instrumental variables. Methods that rely on network structure to identify suitable instruments (e.g., Bramoulle et al. 2009, Di Giorgi et al. 2010, and Lee 2007) are unsuitable for mutually exclusive peer groups. Even when they are applicable, they can yield weak instruments, especially when pool fixed effects are included. Because suitable instruments are hard to find, many studies rely on OLS with pool fixed effects to test for peer effects. As just noted, this approach often yields misleading inference due to the presence of exclusion bias. We offer an alternative estimation method that deals with these shortcomings but does not rely on instrumentation. Although the method allows the inclusion of selection pool fixed effects, it assumes away correlated effects within peer groups. Whether or

not this assumption is reasonable depends on the specific context of the study. But even when correlated effects cannot be ruled out on a priori grounds, researchers can still use the method as a robustness check free of reflection and exclusion bias. More importantly, the method offers a way of estimating endogenous peer effects when peer groups are mutually exclusive and have equal size, in which case the instrumentation methods of Bramoulle et al. (2009), Di Giorgi et al. (2010), and Lee (2007) all fail. There is an abundance of peer effect studies that have this data structure – most notably the assignment of students to rooms, dorms, and study groups. Controlled experiments on peer effects also often have a fixed-size, non-overlapping peer group structure. In all these cases, our method is capable of offering a viable alternative for the estimation of endogenous peer effects.

References

Angrist, J.D. (2014). "The Perils of Peer Effects", *Labour Economics*, 30: 98-108

Arellano, M., and S. Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". Review of Economic Studies 58: 277-297.

Arellano, M, and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-component models", Journal of Econometrics, 68(1): 29-51.

Arellano, M. (2003). *Panel Data Econometrics*, Oxford University Press, Oxford, 2003.

Athey, S., D. Eckles and G. W. Imbens (2015). "Exact P-values for Network Interference", Stanford University. Mimeo.

Bandiera, O., I. Barankay and I. Rasul (2009). "Social Connections and Incentives in the Workplace: Evidence from Personnel Data", *Econometrica*, 77(4): 1047-94.

Bayer, P., R. Hjalmarsson and D. Pozen (2009). "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections", *Quarterly Journal of Economics*, 124(1): 105-47.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models", Journal of Econometrics, 87: 115-43.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). "Identification of Peer Effects through Social Networks", *Journal of Econometrics*, 150(1): 41-55.

Brock, W. and S. Durlauf (2001). "Interactions Based Models", in *Handbook of Econometrics*, 5: 3299-380, North Holland, Amsterdam.

Brown, K. M. and R. Laschever (2012). "When They're Sixty-Four: Peer Effects and the

Timing of Retirement", *American Economic Journal: Applied Economics*, 4(3): 90-115.

Cai, J. and A. Szeidl (2016). "Interfirm Relationships and Business Performance", NBER Working Paper No. 22951

Carrell, S., B. Sacerdote and J. West (2013). "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*. 81(3): 855-82.

Carrell, S., M. Hoekstra and J. West (2016). "The Impact of College Diversity on Behavior Toward Minorities," NBER Working Paper 20940.

Chandrasekhar, A. and R. Lewis (2016). "Econometrics of Sampled Networks", Stanford University Working Paper.

De Giorgi, G. , M. Pellizzari and S. Redaelli (2010). "Identification of Social Interactions through Partially Overlapping Peer Groups", *American Economic Journal: Applied Economics*, 2(2): 241-75.

de Melo, J. (2014). "Peer Effects Identified through Social Networks. Evidence from Uruguayan Schools", Banco de México, Working Paper No. 2014-05.

Duflo, E. and E. Saez (2011). "Participation and Investment Decisions in a Retirement Plan: The Influence of Colleagues' Choices", *Journal of Public Economics*, 85(1): 121-48.

Elandt-Johnson, C.E. and N. L. Johnson (1980). *Survival Models and Data Analysis*, John Wiley & Sons NY, p. 69.

Fafchamps, M and D. Mo (2018). "Peer effects in computer assisted learning: evidence from a randomized experiment", *Experimental Economics*, 21(2): 355-382.

Fafchamps, M and S. Quinn (2017). "Networks and Manufacturing Firms in Africa: Results from a Randomized Field Experiment", *World Bank Economic Review*, Forthcoming.

Fisher, R.A. (1925). "Theory of Statistical Estimation", *Proceedings of the Cambridge Philosophical Society*, 22: 700-25.

Glaeser, E. L., B. I. Sacerdote, and J. A. Scheinkman (2003). "The Social Multiplier", *Journal of the European Economic Association*, 1(2-3): 345-53.

Goux, D. and E. Maurin (2007). "Close Neighbors Matter: Neighborhood Effects on Early Performance at School," *Economic Journal*, 117(523): 1193-215.

Guryan, J. , D. Kroft, and N. J. Notowidigdo (2009). "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments", *American Economic Journal: Applied*

Economics, 44(3): 289-302.

Halliday, T. J. and S. Kwak (2012). "What Is a Peer? The Role of Network Definitions in Estimation of Endogenous Peer Effects", *Applied Economics*, 44(3): 289-301.

Helmers, C. and M. Patnam (2011). "The Formation and Evolution of Childhood Skill Acquisition: Evidence from India," *Journal of Development Economics*, 95(2): 252-66.

Kelejian, H, H. and I. R. Prucha (1999). "A generalized moments estimator for the autoregressive parameter in a spatial model", *International Economic Review*, 40: 509-533.

Krackhardt, D. (1988). "Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data", *Social Networks*, 10: 359-81.

Krishnan, P. and M. Patnam (2012). "Neighbors and Extension Agents in Ethiopia: Who Matters More for Technology Diffusion?", Department of Economics, University of Cambridge. Mimeo.

Lee, L. F. (2007). "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects", *Journal of Econometrics*, 140(2): 333–74.

Liu, X., E. Patacchini, Y. Zenou and L. F. Lee (2012). "Criminal Networks: Who Is the Key Player?", Nota di Lavoro, Fondazione Eni Enrico Mattei, 39.2012.

Liu, X. (2017). "Identification of Peer Effects via a Root Estimator", *Economic Letters*, 156: 168-71.

Manski, C. (1993). "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60(3): 531-42.

Moffitt, R. A. (2001). "Policy Interventions, Low Level Equilibria, and Social Interactions", *Social Dynamics*, 45-82, MIT Press, Cambridge, MA.

Munshi, K. (2004). "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution", *Journal of Development Economics*, 73(1): 185-215.

Naguib, K. (2012). "The Effects of Social Interactions on Female Genital Mutilation: Evidence from Egypt", Department of Economics, Boston University. Mimeo.

Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49: 1417-26.

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

Sacerdote, B. (2001). "Peer Effects with Random Assignment: Results for Dartmouth Room-

mates", *Quarterly Journal of Economics*, 116(92): 681-704.

Stevenson, M. (2015a). "Tests of Random Assignment to Peers in the Face of Mechanical Negative Correlation: An Evaluation of Four Techniques", University of Pennsylvania, Mimeo,

Stevenson, M. (2015b). "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails", University of Pennsylvania, Mimeo.

Stuart, A. and Ord, K. (1998). *Kendall's Advanced Theory of Statistics*, Arnold, London, 1998, 6th Edition, Volume 1, p. 351.

Wang, L.C. (2009). "Peer Effects in the Classroom: Evidence from a Natural Experiment in Malaysia", Department of Economics, UC San Diego, Mimeo.

Zimmerman, D. (2003). "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", *Review of Economics and Statistics*, 85(1): 9-23.

Appendix

A Proofs of propositions

The notation is as follows. In a sampled population Ω, each individual $i \in \Omega$ is randomly assigned to a group of K_i people. Let $\Pi_i \subseteq \Omega$ be the pool of people from which i's $(K_i - 1)$ peers are drawn at random. When the pool Π_i is the entire sample, $\Pi_i = \Omega$. The pool Π_i can also be a subset of the sample of size L_i, with $\Pi_i \subset \Omega$ and $L_i < N$. Section A.1 deals with cases with multiple peer selection pools, i.e., $\Pi_i \subset \Omega$ (Part 1 of Proposition 1). Section A.2 deals with $\Pi_i = \Omega$ (Part 2 of Proposition 1). Section A.3 discusses the magnitude of the exclusion bias in small samples (Part 3 of Proposition 1). These first three sections focus on cases with a constant pool size L and peer group size K. Sections A.4, A.5, A.6, and A.7 prove Propositions 2, 3, 4 and 5, respectively.

A.1 Proposition 1 part 1: Multiple peer selection pools of fixed size L and peer groups of fixed size K

Let the sampled population Ω be partitioned into N distinct pools of size L. Individuals in each pool are partitioned into mutually exclusive groups of size K – which implies that L is an integer multiple of K. Each individual is assigned a realization of a random variable x with the following data generating process:

$$x_{ikl} = \delta_l + \epsilon_{ikl} \tag{33}$$

where x_{ikl} is the value of x for individual i in group k of pool l, δ_l is a pool fixed effect, and ϵ_{ikl} is an i.i.d. random variable with mean 0 and variance σ^2_ϵ.

To test random peer assignment on these data, the researcher estimates regression (1), reproduced here:

$$x_{ikl} = \beta_1 \bar{x}_{-ikl} + \delta_l + \epsilon_{ikl} \tag{34}$$

where $\bar{x}_{-ikl}$ is the sample mean of x_{ikl} for individuals other than i who are in the same group k as

i, i.e.:

$$\bar{x}_{-ikl} = \frac{\left[\sum_{j=1}^{K} x_{jkl}\right] - x_{ikl}}{K-1}$$

Regression (34) can be expressed in deviation from the pool mean so as to eliminate the pool fixed effect δ_l:

$$x_{ikl} - \bar{x}_l = \beta_1(\bar{x}_{-ikl} - \bar{x}_l) + (\epsilon_{ikl} - \bar{\epsilon}_l) \tag{35}$$

where $\bar{x}_l$ is the pool sample mean of x_{ikl}, $\bar{\epsilon}_l$ is the pool sample mean of ϵ_{ikl}, and we have used the fact that the pool sample mean of $\bar{x}_{-ikl}$ is $\bar{x}_l$.

We note that, by construction, $\bar{x}_l \equiv \delta_l + \bar{\epsilon}_l$. It follows that the demeaned regressor $\bar{x}_{-ikl} - \bar{x}_l$ is mechanically correlated with the demeaned error term $\epsilon_{ikl} - \bar{\epsilon}_l$, resulting in a bias in the estimation of β_1 using equation (35). This problem has long been noted in the estimation of autoregressive models with fixed effects and need not be further discussed here. In that literature, the proposed solution has been to first-difference regression (34) and instrument x_{ikl} with lagged values. This approach does not apply here since peer effects are reflexive.

In the rest of this Section, we derive a formula for the asymptotic bias of β_1 for our specific case of a constant pool and group size. This bias is present even when the true $\beta_1 = 0$, leading to incorrect inference when using model (35) to test random peer assignment. We start by defining $u_{ikl} \equiv \bar{x}_{-ikl} - \bar{x}_{-il}$ where $\bar{x}_{-il}$ is the sample mean of x_{ikl} for individuals other than i who are in the same pool l as i, i.e.:

$$\bar{x}_{-il} \equiv \frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - x_{ikl}}{L-1} \tag{36}$$

With this new notation, $\bar{x}_{-ikl} = \bar{x}_{-il} + u_{ikl}$ and equation ((35)) can be rewritten as:

$$x_{ikl} - \bar{x}_l = \beta_1 \left[\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - x_{ikl}}{L-1} + u_{ikl} - \left(\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - \bar{x}_l}{L-1} \right) - \bar{u}_l \right] + \epsilon_{ikl} - \bar{\epsilon}_l \tag{37}$$

where $\bar{u}_l$ is the pool sample mean of u_{ikl} and is identically 0 by construction. The above equation thus simplifies to:

$$x_{ikl} - \bar{x}_l = \beta_1 \left(\frac{\bar{x}_l - x_{ikl}}{L-1} + u_{ikl} - \bar{u}_l \right) + \epsilon_{ikl} - \bar{\epsilon}_l \tag{38}$$

If we define the notation $\ddot{z} \equiv z - \bar{z}_l$, for $z = x, \epsilon, u$, we can further simplify equation (35) as:

$$\ddot{x} = \beta_1 \left(\frac{-\ddot{x}}{L-1} + \ddot{u} \right) + \ddot{\epsilon} \tag{39}$$

from which it is immediately apparent that the regressor used to identify β_1 is mechanically correlated with the error term since it contains the dependent variable itself.

Next we apply the standard formula for calculating the *plim* of the OLS estimator for β_1, which takes the following form :

$$plim_{N\to\infty} \left(\hat{\beta}_1^{FE} \right) = \beta_1 + \frac{cov \left(\frac{-\ddot{x}}{L-1} + \ddot{u}, \ddot{\epsilon} \right)}{var \left(\frac{-\ddot{x}}{L-1} + \ddot{u} \right)} \tag{40}$$

where $\hat{\beta}_1^{FE}$ stands for the fixed effect estimator obtained using regression (39). Since $\beta_1 = 0$ by construction, we can write:

$$plim_{N\to\infty} \left(\hat{\beta}_1^{FE} \right) = \frac{cov \left(\frac{-\ddot{x}}{L-1}, \ddot{\epsilon} \right) + cov \left(\ddot{u}, \ddot{\epsilon} \right)}{var \left(\frac{-\ddot{x}}{L-1} \right) + 2cov \left(\frac{-\ddot{x}}{L-1}, \ddot{u} \right) + var \left(\ddot{u} \right)} \tag{41}$$

With some algebra, equation (41) will now enable us to calculate the asymptotic value of the bias in $\hat{\beta}_1^{FE}$. We start by noting that, since $\bar{u}_l \equiv 0$ by construction, we have:

$$\begin{aligned} cov \left(\ddot{u}, \ddot{\epsilon} \right) &= E \left(\ddot{u}\ddot{\epsilon} \right) = E \left[\left(u_{ikl} - \overline{u}_l \right) \left(\epsilon_{ikl} - \bar{\epsilon}_l \right) \right] \\ &= E \left(u_{ikl} \epsilon_{ikl} \right) - E \left(u_{ikl} \bar{\epsilon}_l \right) = 0 \end{aligned} \tag{42}$$

by definition of the average. Similarly we can write:

$$var \left(\ddot{u} \right) = var \left(u_{ikl} - \bar{u}_l \right) = \sigma_u^2 \tag{43}$$

To tackle the three remaining terms in equation (41), we start by transforming equation (39) to obtain an expression for $-\frac{\ddot{x}}{L-1}$. By simple manipulation of equation (39), we obtain:

$$\left[\frac{L - 1 + \beta_1}{L - 1} \right] \ddot{x} = \beta_1 \ddot{u} + \ddot{\epsilon}$$

which leads to:

$$-\frac{\ddot{x}}{L-1} = \frac{-\beta_1 \ddot{u}}{L-1+\beta_1} - \frac{\ddot{\epsilon}}{L-1+\beta_1} \tag{44}$$

Next we note that:

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

from which we obtain

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

Using the facts that $E(\ddot{\epsilon}) = E(\epsilon_{ikl} - \ddot{\epsilon}_l) = 0$ and that $\beta_1 = 0$ by assumption, and combining these with equations (42), ((46), and (44), we obtain:

$$\begin{aligned} cov\left(\frac{-\ddot{x}}{L-1}, \ddot{\epsilon}\right) &= E\left[\left[\frac{-\ddot{x}}{L-1} - E\left(\frac{-\ddot{x}}{L-1}\right)\right]\ddot{\epsilon}\right] \\ &= E\left[\frac{-\ddot{\epsilon}\ddot{\epsilon}}{L-1}\right] \\ &= \frac{-var(\ddot{\epsilon})}{L-1} = -\frac{\sigma_\epsilon^2}{L} \end{aligned} \tag{47}$$

This gives the value of the first term in the numerator of equation (41).

Next, we use equation (42) and (44) to get the value of the middle term in the denominator of (41):

$$2cov\left(\frac{-\ddot{x}}{L-1}, \ddot{u}\right) = -2\frac{E(\ddot{u}\ddot{\epsilon})}{L-1} = 0 \tag{48}$$

For the first term in the denominator of (41), we again use equation (44) to get:

$$\begin{aligned} var\left(\frac{-\ddot{x}}{L-1}\right) &= var\left(-\frac{\ddot{\epsilon}}{L-1}\right) \\ &= \frac{\sigma_\epsilon^2}{L(L-1)} \end{aligned} \tag{49}$$

Summarizing these different results, we can write the numerator and denominator of (40) as

follows:

$$cov(\frac{-\ddot{x}}{L-1}+\ddot{u},\ddot{\epsilon}) = -\frac{\sigma_\epsilon^2}{L} \quad (50)$$

$$var(\frac{-\ddot{x}}{L-1}+\ddot{u}) = \frac{\sigma_\epsilon^2}{L(L-1)}+\sigma_u^2 \quad (51)$$

We now need an expression for σ_u^2. Recall that $u_{ikl} \equiv \bar{x}_{-ikl} - \bar{x}_{-il}$. Therefore:

$$\sigma_u^2 = Var(u) = Var\left[\bar{x}_{-ikl} - \bar{x}_{-il}\right] = Var\left[\frac{\left[\sum_{j=1}^K x_{jkl}\right] - x_{ikl}}{K-1} - \frac{\left[\sum_{s=1}^{\frac{L}{K}}\sum_{j=1}^K x_{jsl}\right] - x_{ikl}}{L-1}\right]$$

$$= Var\left[\frac{(L-1)\left[\left(\sum_{j=1}^K x_{jkl}\right) - x_{ikl}\right]}{(L-1)(K-1)} - \frac{(K-1)\left[\left(\sum_{j=1}^K x_{jkl}\right) - x_{ik}\right]}{(L-1)(K-1)} - \frac{\sum_{s\neq k}^{\frac{L}{K}}\sum_{j=1}^K x_{jsl}}{L-1}\right]$$

$$= Var\left[\frac{(L-K)\left[\left(\sum_{j=1}^K x_{jkl}\right) - x_{ikl}\right]}{(L-1)(K-1)} - \frac{\sum_{s\neq k}^{\frac{L}{K}}\sum_{j=1}^K x_{jsl}}{L-1}\right]$$

Using $var(x_{ikl}) = \sigma_\epsilon^2$ and the assumption that x_{ikl} is i.i.d., we obtain the following relationship between σ_u^2 and σ_ϵ^2:

$$\sigma_u^2 = \frac{(L-K)^2(K-1)}{(L-1)^2(K-1)^2}\sigma_\epsilon^2 + \frac{(L-K)}{(L-1)^2}\sigma_\epsilon^2 = \frac{(L-K)}{(L-1)(K-1)}\sigma_\epsilon^2 < \epsilon_\epsilon^2 \quad (52)$$

Substituting this into equation (51) the denominator of (40) can be written:

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

$$= \frac{(K-1)+(L-K)L}{L(L-1)(K-1)}\sigma_\epsilon^2$$

Combining these results we get:

$$plim_{N\to\infty}\left(\hat{\beta}_1^{FE}\right) = \frac{(-\frac{\sigma_\epsilon^2}{L})}{\frac{(K-1)+(L-K)L}{L(L-1)(K-1)}\sigma_\epsilon^2}$$

$$= -\frac{(L-1)(K-1)}{(K-1)+(L-K)L} \quad (53)$$

which is obviously negative. This proves the first part of Proposition 1.

A.2 Proposition 1 part 2: one single peer selection pool $\Pi_i = \Omega$ and $L = N$

We now turn to the second part of Proposition 1 when peers are randomized at the level of the sampled population Ω and there is a single peer selection pool $\Pi_i = \Omega$ and $L = N$. In this case, the estimated regression does not include pool fixed effects δ_l.

The first part of Proposition 1 (summarized by formula (2) and derived in Section A.1) states that the magnitude of the exclusion bias depends on the size of the peer selection pool L: for a given peer group size K, a larger pool size is associated with a smaller exclusion bias. From the same formula (2) it immediately follows that as L converges to infinity, the exclusion bias converges to zero. Formally, if $\Pi_i = \Omega$, then

$$plim_{L\to\infty}\left(\hat{\beta}_1^{OLS}\right) = 0 \tag{54}$$

However, in samples that are small relative to the peer group size K, the magnitude of the exclusion bias can be large, even when there is only one peer selection pool $\Pi_i = \Omega$.

A.3 Proposition 1 Part 3: Small sample exclusion bias

Formula (53) only holds in the limit, that is, for large sample sizes N. The computation of $E(\hat{\beta}_1^{FE})$ that applies in small sample sizes is not as straightforward, because $E\left[\frac{samplecov\left(\frac{-\ddot{x}}{L-1}+\ddot{u},\ddot{\epsilon}\right)}{samplevar\left(\frac{-\ddot{x}}{L-1}+\ddot{u}\right)}\right] \neq \frac{E\left[samplecov\left(\frac{-\ddot{x}}{L-1}+\ddot{u},\ddot{\epsilon}\right)\right]}{E\left[samplevar\left(\frac{-\ddot{x}}{L-1}+\ddot{u}\right)\right]}$. We can however use a Taylor expansion to sign the bias.

Stuard and Ord (1998) and Elandt-Johnson and Johnson (1980) have shown that for two random variables R and S, where S either has no mass at 0 (discrete) or has support $[0,\infty)$, a Taylor expansion approximation for $E[A/B]$ is as follows:

$$E\left(\frac{R}{S}\right) \simeq \frac{\mu_R}{\mu_S} - \frac{Cov(R,S)}{\mu_S^2} + \frac{Var(S)\mu_R}{\mu_S^3}$$

In our application $R = SampleCov\left(\frac{-\ddot{x}}{L-1}+\ddot{u},\ddot{\epsilon}\right)$, $S = SampleVar\left(\frac{-\ddot{x}}{L-1}+\ddot{u}\right)$, μ_R is the mean of R and μ_S is the mean of S. The first term, $\frac{\mu_R}{\mu_S}$, is expression (53). We know from equation (50) and equation (51) that $\mu_R < 0$ and $\mu_S > 0$. While an expression for $Cov(R,S)$ is harder to derive,

simulation results indicate that $Cov(R, S) < 0$. Given that $Var(S) > 0$, it follows that:

$$E\left[\hat{\beta}_1^{FE}|L\right] < plim_{N\to\infty}\left[\hat{\beta}_1^{FE}\right] \tag{55}$$

a finding that is also confirmed through numerous simulations. Hence, we see that for a given size of the selection pool L and a given size of the peer group K, the negative exclusion bias shrinks from below towards its *plim* as sample size $N \times L$ increases.

A.4 Proof Proposition 2

We assume N pools of fixed size L each partitioned into peer groups of size K. Let, as before, $\hat{\beta}_1^{FE}$ denote the pool fixed effect estimator and let and $\hat{\beta}_1^{OLS}$ denote the pooled OLS estimator without pool fixed effects. We want to show that:

$$plim_{N\to\infty}(\hat{\beta}_1^{FE}) < plim_{N\to\infty}(\hat{\beta}_1^{OLS})$$

As is well known, the pooled OLS estimator $\hat{\beta}_1^{OLS}$ is a weighted average of the within estimator $\hat{\beta}_1^{FE}$ and the between estimator $\hat{\beta}_1^{BE}$:

$$\hat{\beta}_1^{OLS} = \eta^2\hat{\beta}_1^{BE} + (1-\eta^2)\hat{\beta}_1^{FE} \tag{56}$$

where $0 < \eta^2 < 1$ is the ratio of the between sum of squares of $\bar{y}_{-ikl}$ to its total sum of squares (e.g., Raudenbush and Bryk 2002). From Proposition 1, we know that $plim_{N\to\infty}\left(\hat{\beta}_1^{FE}\right) < 0$. Thus if we can prove that $plim_{N\to\infty}\left(\hat{\beta}_1^{BE}\right) \geq 0$, we will have proven that $plim_{N\to\infty}(\hat{\beta}_1^{FE}) < plim_{N\to\infty}(\hat{\beta}_1^{OLS})$.

The between estimator is the OLS estimator from a regression of $\bar{x}_l$ on an intercept and $\bar{\bar{x}}_{-il}$, where $\bar{x}_l$ denotes the average outcome x_{ikl} over the individuals in pool l and $\bar{\bar{x}}_{-il}$ denotes the average peer group outcome of the pool:

$$\bar{x}_l = \beta_0 + \beta_1\bar{\bar{x}}_{-il} + \bar{\epsilon}_l \tag{57}$$

where

$$\bar{\bar{x}}_{-il} = \bar{x}_{-il} + \bar{u}_l \tag{58}$$

and $\bar{x}_{-il}$ is the average outcome over the individuals in the pool l, excluding individual i and $\bar{u}_l$ denotes the pool average of u.

The between-group model reduced form equation is:

$$\begin{aligned}\bar{x}_{-il} &= \frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} x_{jsl}\right] - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\epsilon}_l}{L - 1 + \beta_1} \\ &= \frac{L\bar{x}_l - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\epsilon}_l}{L - 1 + \beta_1}\end{aligned} \tag{59}$$

where $\bar{x}_l$, $\bar{u}_l$ and $\bar{\epsilon}_l$ denote the pool averages of x , u and ϵ , respectively. Under random peer assignment (i.e. $\beta_1 = 0$), this equation reduces to:

$$\bar{x}_{-il} = \frac{L\bar{x}_l - \beta_0}{L - 1} - \frac{\bar{\epsilon}_l}{L - 1} \tag{60}$$

Using (58) and (60), we have:

$$\begin{aligned}cov(\bar{\bar{x}}_{-il}, \bar{\epsilon}_l) &= cov(\bar{x}_{-il} + \bar{u}_l, \bar{\epsilon}_l) \\ &= cov(\bar{x}_{-il}, \bar{\epsilon}_l) \\ &= L\frac{E(\bar{\epsilon}_l^2)}{L - 1} - \frac{E(\bar{\epsilon}_l^2)}{L - 1} \\ &= var(\bar{\epsilon}_l) = \frac{\sigma_\epsilon^2}{L}\end{aligned} \tag{61}$$

and

$$
\begin{aligned}
var\left(\bar{\bar{x}}_{-il}\right) &= var\left(\frac{\sum_{i=1}^{L}\bar{x}_{-ikl}}{L}\right) \\
&= \frac{1}{L^2}var\left(\sum_{i=1}^{L}\left(\frac{\sum_{j=1}^{L}x_{jl}-x_{il}}{L-1}\right)+\sum_{i=1}^{L}u_{il}\right) \\
&= \frac{1}{L^2}var\left(\sum_{i=1}^{L}x_{il}+\sum_{i=1}^{L}u_{il}\right) \\
&= \frac{\sigma_\epsilon^2+\sigma_u^2}{L} \\
&= \frac{(L-1)(K-1)+(L-K)}{L(L-1)(K-1)}\sigma_\epsilon^2
\end{aligned} \tag{62}
$$

where in the last step we used the result in (52). Using equation (61) and (62) we then obtain:

$$
plim_{N\to\infty}\left(\hat{\beta}_1^{BE}\right) = \frac{cov\left(\bar{\bar{x}}_{-il},\bar{\epsilon}_l\right)}{var\left(\bar{\bar{x}}_{-il}\right)} \tag{63}
$$

$$
\begin{aligned}
&= \frac{\frac{\sigma_\epsilon^2}{L}}{\frac{(L-1)(K-1)+(L-K)}{L(L-1)(K-1)}\sigma_\epsilon^2} \\
&= \frac{(L-1)(K-1)}{(L-1)(K-1)+(L-K)} > 0
\end{aligned} \tag{64}
$$

This proves that $plim_{N\to\infty}\left(\hat{\beta}_1^{BE}\right) > 0$.

We can also use (56) to prove the corollary 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 $\hat{\beta}_1^{BE}$ were presented in (2) and (64), respectively. We now derive an expression for η^2.

Weight parameter η^2 in equation (65) is the ratio of the between-group sum of squares of $\bar{x}_{-ikl}$ relative to its total sum of squares:

$$
\eta^2 = \frac{SS^{BG}_{\bar{x}_{-ikl}}}{SS^{Total}_{\bar{\bar{x}}_{-ikl}}} = \frac{SS^{BG}_{\bar{x}_{-ikl}}}{SS^{BG}_{\bar{x}_{-ikl}}+SS^{Within}_{\bar{x}_{-ikl}}} \tag{65}
$$

Specifically, $SS^{BG}_{\bar{y}_{-i,k,l}}$ is the sum of all squared differences between cluster group means and the

overall sample mean, multiplied by the number of observations in the pool l. In other words:

$$SS^{BG}_{\bar{\bar{x}}_{-ikl}} = SS^{BE}_{\bar{\bar{x}}_{-ikl}} \times L \tag{66}$$

where $SS^{BE}_{\bar{\bar{x}}_{-ikl}}$ is the sum of squares of $\bar{\bar{x}}_{-il}$ in the between regression (46). Furthermore, using the definition of the variance, we know that:

$$var\left(\bar{\bar{x}}_{-il}\right) = \frac{SS^{BE}_{\bar{\bar{x}}_{-ikl}}}{\left(\frac{N}{L}-1\right)} \Rightarrow SS^{BE}_{\bar{\bar{x}}_{-ikl}} = var\left(\bar{\bar{x}}_{-il}\right) \times \left(\frac{N}{L}-1\right) \tag{67}$$

By combining equations (65) - (67) we obtain:

$$SS^{BG}_{\bar{\bar{x}}_{-ikl}} = var\left(\bar{\bar{x}}_{-il}\right) \times \left(\frac{N}{L}-1\right) \times L$$

Substituting in for $var\left(\bar{\bar{x}}_{-il}\right)$ given by equation (62), we have:

$$SS^{BG}_{\bar{\bar{x}}_{-ikl}} = \frac{(L-1)(K-1)+(L-K)}{(L-1)(K-1)}\sigma^2_\epsilon \times \left(\frac{N}{L}-1\right) \times \sigma^2_\epsilon \tag{68}$$

Next, $SS^{Within}_{\bar{\bar{x}}_{-ikl}}$ is the sum of the squared differences between each individual's average peer group outcome, $\bar{x}_{-ikl}$, and its average for the individual's group $\bar{\bar{x}}_{-il}$. Similarly to equation (67), we have:

$$var\left(\bar{x}_{-il} - \bar{\bar{x}}_{-il}\right) = \frac{SS^{Within}_{\bar{\bar{x}}_{-ikl}}}{(N-1)} \Rightarrow SS^{Within}_{\bar{\bar{x}}_{-ikl}} = var\left(\bar{x}_{-il} - \bar{\bar{x}}_{-il}\right) \times (N-1)$$

From the above we know that $var\left(\bar{x}_{-ikl} - \bar{\bar{x}}_{-il}\right) = var\left(\frac{-\ddot{x}}{L-1} + \ddot{u}\right)$. Therefore, we can substitute in for the expression of $var\left(\bar{x}_{-ikl} - \bar{\bar{x}}_{-il}\right)$ by using equations (51). We have:

$$SS^{Within}_{\bar{\bar{y}}_{-ikl}} = \frac{L+(L-K)(K-1)}{K(K-1)L} \times \frac{N-1}{L}\sigma^2_\epsilon \tag{69}$$

Combining equations (65), (68) and (69) , we obtain:

$$\eta^2 = \frac{SS^{BG}_{\bar{\bar{x}}_{-ikl}}}{SS^{BG}_{\bar{\bar{x}}_{-ikl}} + SS^{Within}_{\bar{\bar{x}}_{-ikl}}}$$

where

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

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

$$\begin{aligned} plim_{N\to\infty}(\eta^2) = plim_{N\to\infty}\left[\frac{A\left(\frac{N}{L}-1\right)}{A\left(\frac{N}{L}-1\right)+B\left(\frac{N-1}{L}\right)}\right] \\ = \frac{A}{A+B} \end{aligned} \tag{70}$$

This closed form result only holds when sample size N tends to infinity. Using (56), (2), (64) and (70) we now derive the large sample property of pooled OLS when peers are selected at the pool level l and when the true $\beta = 0$:

$$\begin{aligned} plim_{N\to\infty}\left(\hat{\beta}_1^{OLS}\right) &= plim(\eta^2)plim(\hat{\beta}_1^{BE}) + \left[1 - plim(\eta^2)\right] plim(\hat{\beta}_1^{FE}) \\ &= \left(\frac{A}{A+B}\right)\frac{1}{AL} - \left(1-\frac{A}{A+B}\right)\frac{1}{BL} = 0 \end{aligned} \tag{71}$$

This proves the corollary.

Finally, we illustrate formally why the exclusion bias is more present in smaller samples. We first note that:

$$\begin{aligned} E\left(\eta^2\right) &= E\left(\frac{SS^{BG}_{\bar{x}_{-ikl}}}{SS^{BG}_{\bar{x}_{-ikl}} + SS^{Within}_{\bar{x}_{-ikl}}}\right) \\ &= E\left(SS^{BG}_{\bar{x}_{-ikl}}\right) E\left(\frac{1}{SS^{BG}_{\bar{x}_{-ikl}} + SS^{Within}_{\bar{x}_{-ikl}}}\right) + Cov\left(SS^{BG}_{\bar{x}_{-ikl}}, \frac{1}{SS^{BG}_{\bar{x}_{-ikl}} + SS^{Within}_{\bar{x}_{-ikl}}}\right) \\ &= \frac{LK-2K+1}{L(L-1)} + Cov\left(SS^{BG}_{\bar{x}_{-ikl}}, \frac{1}{SS^{BG}_{\bar{x}_{-ikl}} + SS^{Within}_{\bar{x}_{-ikl}}}\right) \\ &= plim_{N\to\infty}\left(\eta^2\right) + Cov\left(SS^{BG}_{\bar{x}_{-ikl}}, \frac{1}{SS^{BG}_{\bar{x}_{-ikl}} + SS^{Within}_{\bar{x}_{-ikl}}}\right) \end{aligned}$$

It is clear that $Cov\left(SS^{BG}_{\bar{x}_{-ikl}}, \frac{1}{SS^{BG}_{\bar{x}_{-ikl}} + SS^{Within}_{\bar{x}_{-ikl}}}\right) < 0$. Therefore, we obtain:

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

Hence, *ceteris paribus*, as N gets smaller, more weight is given to the pool FE estimator in the estimation of pooled OLS (see (56)). This in turn magnifies exclusion bias in the pooled OLS. In contrast, as sample size increases, more weight is given to the between estimator in the pooled OLS, which reduces exclusion bias.

A similar logic applies to $\hat{\beta}_1^{CL}$ which is obtained by adding fixed effects at a cluster level that combines multiple peer selection pools. The only difference is that, by combining multiple selection pools within a cluster, $\hat{\beta}_1^{CL}$ captures part of the positive correlation across pools that is inherent to the between estimator. This yields the final result that

$$plim[\hat{\beta}_1^{FE}] < plim[\hat{\beta}_1^{CL}] < plim[\hat{\beta}_1^{POLS}]$$

A.5 Proof of Proposition 3

To recall, 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 – this is the reflection bias. None of the ϵ's from other groups enter this pair of equations since we have assumed no spillovers across groups. We have $E[y_1] = E[y_2] = \frac{\alpha(1+\beta)}{1-\beta^2} \equiv \overline{y}$. If ϵ_1 and ϵ_2 are independent from each other,

$E[\epsilon_1\epsilon_2] = 0$ and we can write:

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

The covariance between y_1 and y_2 is given by:

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

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

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

We estimate a model of the form:

$$y_1 = a + by_2 + v_1$$

Since equation (11) is univariate, we have $\widehat{b} = \widehat{r}\frac{\sigma_{y_1}}{\sigma_{y_2}} = \widehat{r}$ since $\sigma_{y_1} = \sigma_{y_2}$. Hence it follows that:

$$plim_{N\to\infty}[\widehat{b}] = \frac{2\beta}{1+\beta^2} \neq \beta$$

A.6 Proof of Proposition 4

We have shown in the text that, starting from Proposition 1 with $K = 2$, if we regress $\ddot{\epsilon}_{ikl}$ on $\ddot{\epsilon}_{ikl}$, the regression coefficient converges to:

$$\rho \equiv plim_{N\to\infty} SampleCorr(\ddot{\epsilon}_{ikl}\ddot{\epsilon}_{jkl}) = -\frac{1}{L-1} \tag{72}$$

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. The variance and covariance of y are now:

$$\begin{aligned} plim_{N\to\infty}[(\ddot{y}_1 - \overline{\ddot{y}})^2] &= \frac{\sigma_\epsilon^2(1+\beta^2+2\beta\rho)}{(1-\beta^2)^2} \\ plim_{N\to\infty}[(\ddot{y}_1 - \overline{\ddot{y}})(\ddot{y}_2 - \overline{\ddot{y}})] &= \frac{\sigma_\epsilon^2(2\beta+(1+\beta^2)\rho)}{(1-\beta^2)^2} \end{aligned}$$

Equipped with the above results, we can now derive an expression for the combined reflection and exclusion bias in model (11). As before, we use the fact that $\widehat{b}^{FE} = \frac{SampleCov[(\ddot{y}_1 - \bar{\ddot{y}})(\ddot{y}_2 - \bar{\ddot{y}})]}{SampleVar[(\ddot{y}_1 - \bar{\ddot{y}})^2]}$. Simple algebra along the same lines as Proposition 3 yields:

$$plim_{N\to\infty}[\widehat{b}^{FE}] = \frac{2\beta + (1+\beta^2)\rho}{1 + \beta^2 + 2\beta\rho} \tag{73}$$

A.7 Proof of Proposition 5

Let the sampled population Ω be partitioned into N distinct pools of size T. Observations in each pool refer to a given individual i and are ordered chronologically by $t = \{1, ...T\}$. Each individual observation is assigned a realization of a random variable x with the following data generating process:

$$x_{it} = \delta_i + \epsilon_{it} \tag{74}$$

where x_{it} is the value of x for individual i at time t, δ_i is an individual fixed effect, and ϵ_{it} is an i.i.d. random variable with mean 0 and variance σ^2_ϵ. Note that here the individual index i corresponds to the pool index l in the network data. Under the null, the variance of x_{it} is the same as the variance of ϵ_{it} and the two variables are perfectly correlated.

To test whether variable x_{it} is autoregressive, the researcher estimates the following regression:

$$x_{it} = \beta_1 x_{it-1} + \delta_i + \epsilon_{it} \tag{75}$$

where x_{it-1} is the lagged value of x_{it}. Note that the above regression is estimated using observations $t = \{2, ...T\}$ on variable x_{it} while observations $t = \{1, ..., T-1\}$ of x_{it} are used for regressor. Regression ((75)) can be expressed in deviation from the individual mean so as to eliminate the individual fixed effect δ_l:

$$x_{it} - \bar{x}_i = \beta_1(x_{it-1} - \bar{x}'_i) + (\epsilon_{it} - \bar{\epsilon}_i) \tag{76}$$

where $\bar{x}_i$ is the pool sample mean of x_{it}, $\bar{x}'_i$ is the pool sample mean of x_{it-1}, and $\bar{\epsilon}_l$ is the pool sample mean of ϵ_{it}. Specifically we have:

$$\bar{x}_i = \frac{1}{T-1}\sum_{t=2}^{T} x_{it}$$

$$\bar{x}_i' = \frac{1}{T-1}\sum_{t=1}^{T-1} x_{it}$$

$$\bar{\epsilon}_i = \frac{1}{T-1}\sum_{t=2}^{T} \epsilon_{it}$$

When T is large, $\bar{x}_i \simeq \bar{x}_i'$ but when T is small the difference matters. We can rewrite the demeaned model more concisely as:

$$\ddot{x}_{it} = \beta_1 \ddot{x}_{it}' + \ddot{\epsilon}_{it} \tag{77}$$

The $plim_{N\to\infty}(\hat{\beta}_1^{FE})$ is thus:

$$plim_{N\to\infty}\left(\hat{\beta}_1^{FE}\right) = \beta_1 + \frac{cov\left(\ddot{x}_{it}', \ddot{\epsilon}_{it}\right)}{var\left(\ddot{x}_{it}'\right)} \tag{78}$$

We now derive an expression for $cov\left(\ddot{x}', \ddot{\epsilon}\right)$; it is not equal to 0, implying a systematic bias in $\hat{\beta}_1^{FE}$. The basic reason is that observations for $\ddot{x}', \ddot{\epsilon}$ overlap except for observation 1, which only appears in $\ddot{x}'$, and observation T, which only appears in $\ddot{\epsilon}$. To simplify the algebra, we use equation 75 to replace x with ϵ throughout. We have:

$$\bar{x}_i = \delta_i + \frac{1}{T-1}\sum_{t=2}^{T} \epsilon_{it}$$

$$\bar{x}_i' = \delta_i + \frac{1}{T-1}\sum_{t=1}^{T-1} \epsilon_{it}$$

$$\bar{\epsilon}_i = \frac{1}{T-1}\sum_{t=2}^{T} \epsilon_{it}$$

$$\bar{\epsilon}_i' = \frac{1}{T-1}\sum_{t=1}^{T-1} \epsilon_{it}$$

$$\ddot{x}_{it}' = \epsilon_{it-1} - \frac{1}{T-1}\sum_{t=1}^{T-1} \epsilon_{it}$$

$$\ddot{\epsilon}_{it} = \epsilon_{it} - \frac{1}{T-1}\sum_{t=2}^{T} \epsilon_{it}$$

By construction we have that $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}^2) = \sigma_e^2$, and, by independence of the errors,

$E(\epsilon_{it}\epsilon_{is}) = 0$ for all $s \neq t$. By extension, $E(\ddot{\epsilon}_{it}) = 0$ and $E(\ddot{x}'_{it}) = 0$ as well. We also note that the variance of a sample means $\bar{\epsilon}_i$ and $\bar{\epsilon}'_i$ is simply $\frac{\sigma_e^2}{T-1}$. Hence we have:

$$cov\left(\ddot{x}'_{it}, \ddot{\epsilon}_{it}\right) = E(\ddot{x}'_{it}\ddot{\epsilon}_{it}) = E(\epsilon_{it-1} - \frac{1}{T-1}\sum_{t=1}^{T-1}\epsilon_{it})(\epsilon_{it} - \frac{1}{T-1}\sum_{t=2}^{T}\epsilon_{it})$$

$$= E(\epsilon_{it-1}\epsilon_{it} - \frac{\epsilon_{it-1}}{T-1}\sum_{t=2}^{T}\epsilon_{it} - \frac{\epsilon_{it}}{T-1}\sum_{t=1}^{T-1}\epsilon_{it} + \frac{1}{(T-1)^2}(\sum_{t=1}^{T-1}\epsilon_{it})(\sum_{t=2}^{T}\epsilon_{it}))$$

$$= -\frac{2(T-2)\sigma_e^2}{(T-1)^2} + \frac{T-2}{(T-1)^2}\sigma_e^2 = -\frac{T-2}{(T-1)^2}\sigma_e^2$$

The first term on the second line drops out because errors are iid across observations by assumption. Regarding the second term, for observation 2 the cross-term $E(\frac{\epsilon_{it-1}}{T-1}\sum_{t=2}^{T}\epsilon_{it}) = 0$ since ϵ_{i1} does not appear in $\sum_{t=2}^{T}\epsilon_{it}$. Similarly for observation T in the cross-term $E(\frac{\epsilon_{it}}{T-1}\sum_{t=1}^{T-1}\epsilon_{it}) = 0$. Hence, over $T-1$ observations, these cross-terms are equal to $\frac{\sigma_e^2}{T-1}$ only $T-2$ times. Hence, in expectations, each cross-term is equal to $\frac{\sigma_e^2}{T-1}$ only $\frac{T-2}{T-1}$ of the time.

Turning to the denominator, we have:

$$var\left(\ddot{x}'_{it}\right) = E(\epsilon_{it-1} - \frac{1}{T-1}\sum_{s=1}^{T-1}\epsilon_{is})(\epsilon_{it-1} - \frac{1}{T-1}\sum_{s=1}^{T-1}\epsilon_{is})$$

$$= E(\epsilon_{it-1}^2 - 2\frac{\epsilon_{it-1}^2}{T-1} + \frac{1}{(T-1)^2}(\sum_{s=1}^{T-1}\epsilon_{is}^2))$$

$$= \frac{T-2}{T-1}\sigma_e^2$$

It follows that:

$$plim\left(\hat{\beta}_1^{FE}\right) = -\frac{1}{T-1}$$

B Comparison to other tests of random peer assignment

Various methods have recently been proposed to test random peer assignment. We discuss them briefly in turn.

B.1 GKN method

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

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-ikl} + \delta_l + \varphi \bar{x}_{-i,l} + \epsilon_{ikl} \tag{79}$$

where φ is an additional parameter to be estimated.

To see how, under specific conditions, this effectively deals with exclusion bias when the true $\beta_1 = 0$, we substitute equation (79) in for equation (**??**) and rearrange as follows:

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

The inclusion of the proxy variable $\overline{x}_{-il}$ soaks up the non-random component of $\overline{x}_{-ikl}$. As a result, if $\beta_1 = 0$, the coefficient estimate $\hat{\beta}_1$ measures the partial effect of the random component u_{ikl}. Since $E(u_{ikl}\epsilon_{ikl}) = 0$ under the assumption of random peer selection, $E(\hat{\beta}_1) = \beta_1$ and OLS yields a consistent estimate of the peer effect β_1.

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

B.2 Joint F-test

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

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

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

B.3 Split-sample method

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

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

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

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

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

Hence, the method effectively creates a dataset – derived from the original data – where (i) individuals are excluded from their own peer group but where (ii) they are also excluded from the peer

groups of other individuals in the sample. This eliminates the source of the exclusion bias. One obvious downside of this approach is the large loss of efficiency that results from the reduction in sample size. The efficiency of the approach can in principle be improved by performing multiple iterations, but this is cumbersome, especially with large datasets.

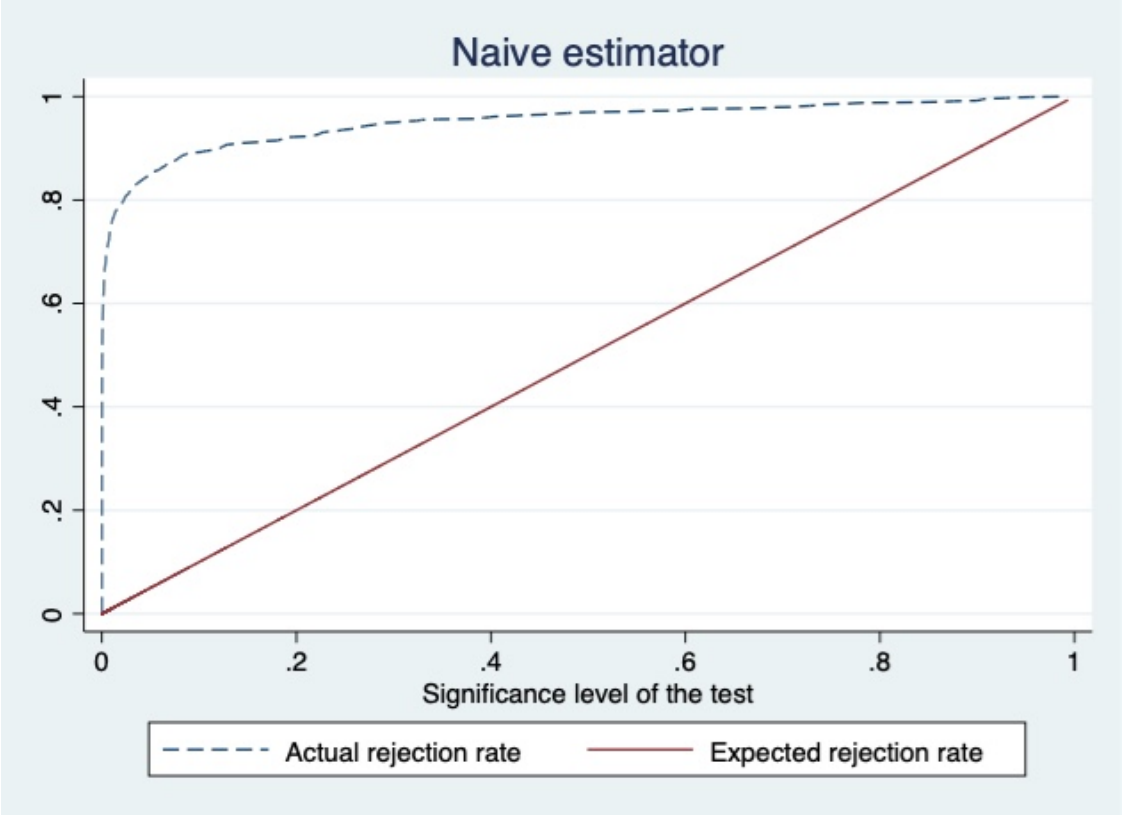
TABLES AND FIGURES

Table 1: Simulated exclusion bias with random peer assignment

		$L = 20$	$L = 50$	$L = 100$
		(1)	(2)	(3)
$K = 2$	Predicted plim$[\hat{\beta}_1]$	-0.05	-0.02	-0.01
	Average $\hat{\beta}_1^s$	-0.05	-0.02	-0.01
	% of $\hat{\beta}_1^s = 0$ rejected at 1% level	26%	10%	8%
	% of $\hat{\beta}_1^s = 0$ rejected at 5% level	43%	21%	18%
	% of $\hat{\beta}_1^s = 0$ rejected at 10% level	52%	29%	24%
$K = 5$	Predicted plim$[\hat{\beta}_1]$	-0.25	-0.09	-0.04
	Average $\hat{\beta}_1^s$	-0.26	-0.10	-0.04
	% of $\hat{\beta}_1^s = 0$ rejected at 1% level	75%	22%	9%
	% of $\hat{\beta}_1^s = 0$ rejected at 5% level	85%	38%	21%
	% of $\hat{\beta}_1^s = 0$ rejected at 10% level	89%	48%	31%
$K = 10$	Predicted plim$[\hat{\beta}_1]$	-0.82	-0.22	-0.10
	Average $\hat{\beta}_1^s$	-0.86	-0.25	-0.11
	% of $\hat{\beta}_1^s = 0$ rejected at 1% level	97%	42%	17%
	% of $\hat{\beta}_1^s = 0$ rejected at 5% level	99%	58%	27%
	% of $\hat{\beta}_1^s = 0$ rejected at 10% level	100%	65%	36%

Notes: The Table reports simulation results from 1000 Monte Carlo replications for different values of K and L. Each simulation includes $N \times L = 1000$ observations generated with a true $\beta_1 = 0$. In each simulated sample s, coefficient $\hat{\beta}_1^s$ is estimated using fixed effects at the level of the selection pool. The predicted $plim_{N\to\infty}[\hat{\beta}_1]$ is obtained using Proposition 1. The average $\hat{\beta}_1^s$ is the average of $\hat{\beta}_1^s$ estimates over all replications. The percentage of rejections is the proportion of replications for which a standard t-test rejects the null that $\beta_1 = 0$ for different critical levels of the test.

Figure 1: Performance of the standard t-test under the null of random assignment



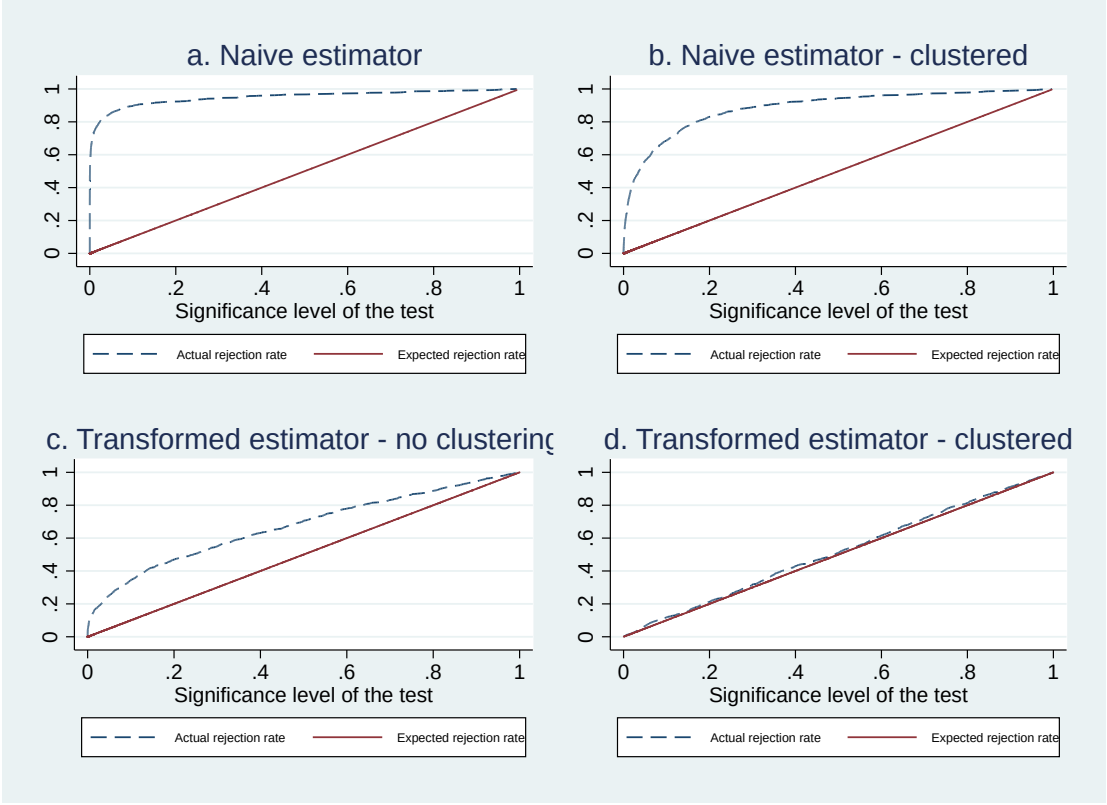
Notes: The Figure shows the simulated performance of a standard t-test to evaluate whether $\beta_1 = 0$ under the null hypothesis of random assignment that it is true. The expected rejection rate is a 45 degree line. The actual performance of the test under the null is simulated using 1000 Monte Carlo replications with N=50, L=20 and K=5. Pool fixed effects are included in each replication. An actual rejection rate above the 45 degree line indicates over-rejection: the probability of rejecting the null of random assignment is larger than the critical value of the test.

Table 2: Simulated exclusion bias with random peer assignment: Different sample sizes N

	$N = 2, L = 50$	$N = 4, L = 50$	$N = 10, L = 50$	$N = 20, L = 50$	$N = 40, L = 50$	$N = 80, L = 50$	$N = 120, L = 50$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
K = 5	-0.14	-0.12	-0.10	-0.09	-0.09	-0.09	-0.09
K = 10	-0.46	-0.33	-0.25	-0.24	-0.23	-0.22	-0.22

Notes: The Table reports simulation results from 1000 Monte Carlo replications for different values of K and N. Each simulation considers pool size $L = 50$, with N pools and considers observations generated with a true $\beta_1 = 0$. In each simulated sample s, coefficient $\hat{\beta}_1^s$ is estimated using fixed effects at the level of the selection pool.

Figure 2: Performance of the corrected model with different standard error estimators

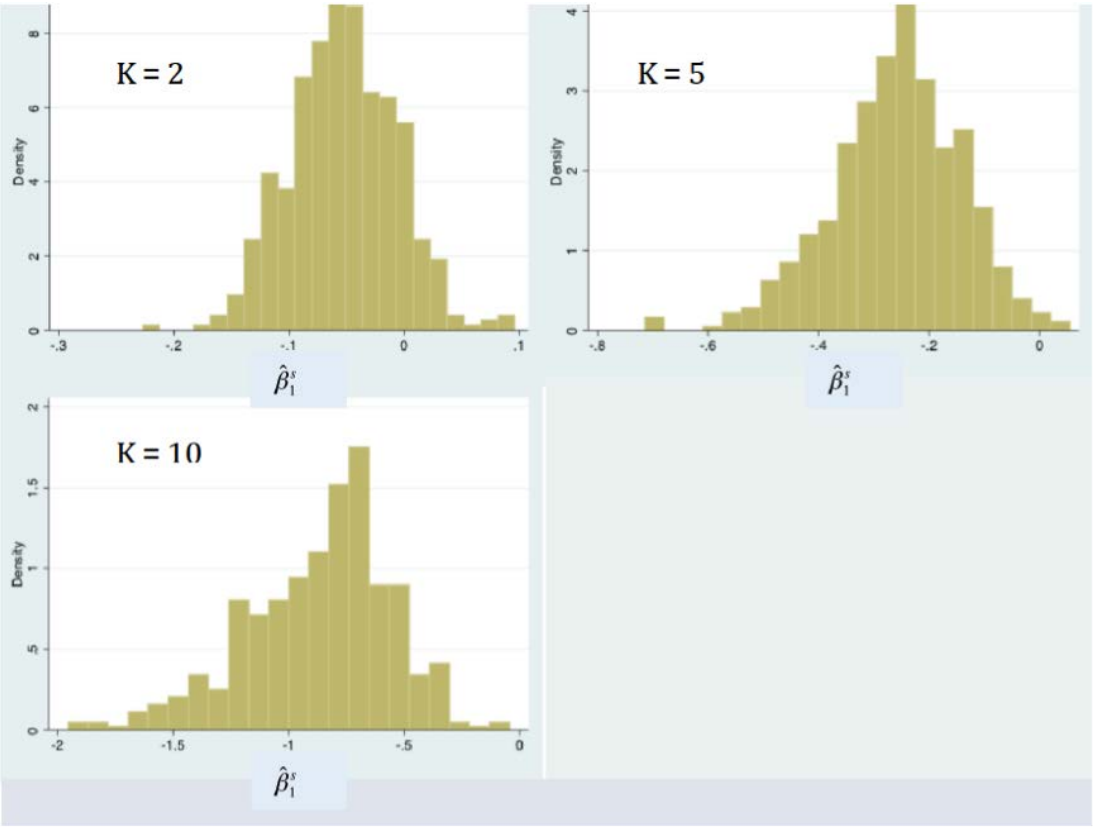


Notes: Figure shows for different estimators the simulated performance of a standard t-test to evaluate whether $\beta_1 = 0$ under the null hypothesis of random assignment that it is true. The upper two panels show this for the 'naive' model (1) for different standard error estimators: One without clustering at the selection pool level (left) and one with standard errors clustered at the selection pool level (right). Using model (3) with a corrected dependent variable, the bottom two panels show the results without (left) and with (right) clustering of standard errors at the selection pool level. The expected rejection rate is a 45 degree line. The actual performance of the test under the null is simulated using 1000 Monte Carlo replications with N=50, L=20 and K=5. Pool fixed effects are included in each replication. An actual rejection rate above the 45 degree line indicates over-rejection: the probability of rejecting the null of random assignment is larger than the critical value of the test.

Table 3: An illustration of the permutation method

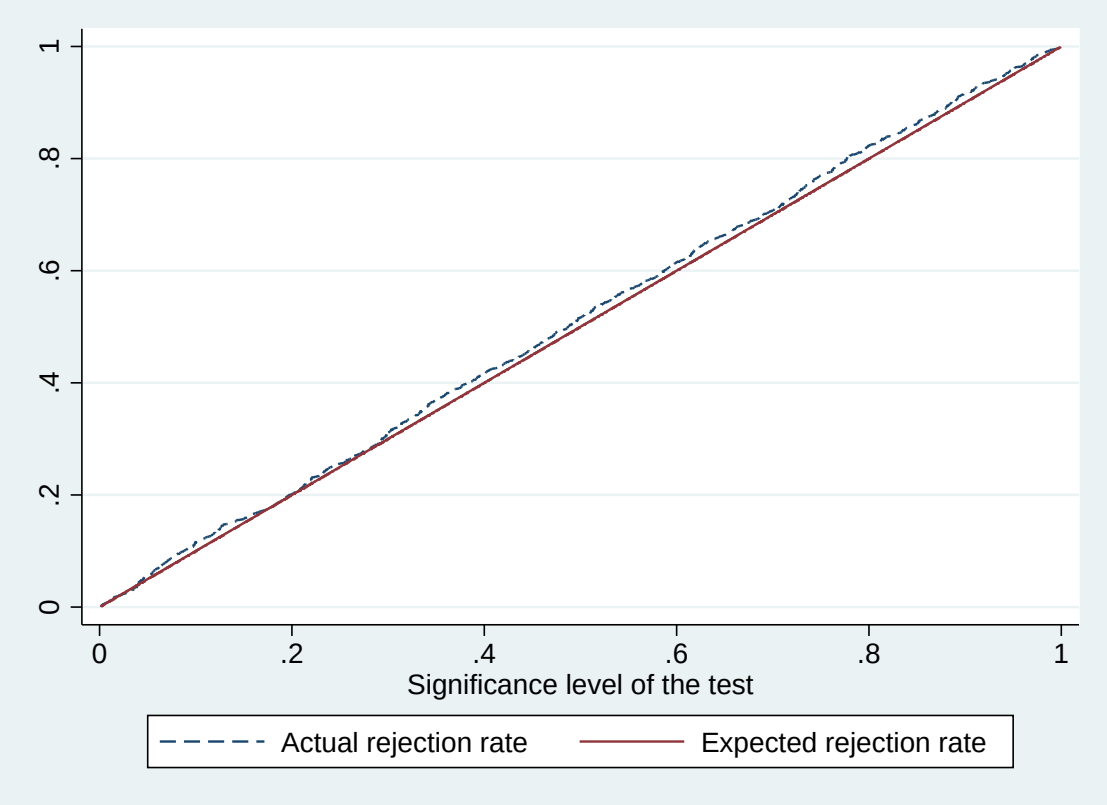
i	k	l	x_{ikl}	$\tilde{x}_{ikl}$
1	1	1	x_{111}	x_{211}
2	1	1	x_{211}	x_{521}
3	2	1	x_{321}	x_{111}
4	2	1	x_{421}	x_{321}
5	2	1	x_{521}	x_{421}
6	3	2	x_{632}	x_{842}
7	3	2	x_{732}	x_{632}
8	4	2	x_{842}	x_{942}
9	4	2	x_{942}	x_{1052}
10	5	2	x_{1052}	x_{732}

Figure 3: Histogram of $\hat{\beta}_1^s$ under the null



Notes: This Figure shows the distribution of simulated $\hat{\beta}_1^s$ using 1000 Monte Carlo replications with random assignment for different group sizes K. We set N=50 and L=20. Each histogram presents the frequency distribution of $\hat{\beta}_1^s$ under the null. Pool fixed effects are included in all regressions.

Figure 4: Performance of the permutation test under the null



Notes: The Figure shows the simulated performance of a permutation test to evaluate whether $\beta_1 = 0$ under the null hypothesis of random assignment. The expected rejection rate is a 45 degree line. The actual performance of the test under the null is simulated using 1000 Monte Carlo replications with N=50, L=20 and K=5. Pool fixed effects are included in each replication. An actual rejection rate above the 45 degree line indicates over-rejection: the probability of rejecting the null of random assignment is larger than the critical value of the test.

Table 4: Comparing Pool FE, Cluster FE and Pooled OLS under the null

	$N = 10, C = 200, L = 50$			$N = 20, C = 200, L = 50$			$N = 40, C = 200, C = 50$			$N = 80, C = 200, L = 50$			$N = 120, C = 200, L = 50$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	$\hat{\beta}_1^{FE}$	$\hat{\beta}_1^{CL}$	$\hat{\beta}_1^{POLS}$	$\hat{\beta}_1^{FE}$	$\hat{\beta}_1^{CL}$	$\hat{\beta}_1^{POLS}$	$\hat{\beta}_1^{FE}$	$\hat{\beta}_1^{CL}$	$\hat{\beta}_1^{POLS}$	$\hat{\beta}_1^{FE}$	$\hat{\beta}_1^{CL}$	$\hat{\beta}_1^{POLS}$	$\hat{\beta}_1^{FE}$	$\hat{\beta}_1^{CL}$	$\hat{\beta}_1^{POLS}$
K = 5	-0.10	-0.03	-0.02	-0.09	-0.03	-0.01	-0.09	-0.02	-0.01	-0.09	-0.02	0.00	-0.09	-0.02	0.00
K = 10	-0.25	-.008	-0.04	-0.24	-0.06	-0.02	-0.23	-0.05	-0.01	-0.22	-0.05	0.00	-0.22	-0.05	0.00

Notes: Each cell of this Table gives, for different sample sizes and different values of K, the mean value of the estimated β_1 under the null, for $L = 50$ and $C = 200$ – which implies that each cluster contains four selection pools. Each average is simulated using 1000 Monte Carlo replications with $\beta_1 = 0$ and with no correlated effects at the cluster level. For each replication $\hat{\beta}_1^{FE}$ is estimated using OLS with pool fixed effects, $\hat{\beta}_1^{CL}$ is estimated using OLS with cluster fixed effects, with a cluster covering four pools, while $\hat{\beta}_1^{POLS}$ is estimated using OLS without any fixed effects.

Table 5: Bias in the estimation of endogenous peer effects

$K = 2$; $L = 20$; $N = 500$

(1)	(2)	(3)
True β_1	Predicted plim($\hat{b}$)	Mean simulated $\hat{b}$
0.00	-0.06	-0.06
0.01	-0.04	-0.04
0.02	-0.02	-0.02
0.03	0.01	0.01
0.04	0.03	0.03
0.05	0.05	0.05
0.06	0.07	0.07
0.07	0.09	0.09
0.08	0.11	0.11
0.09	0.12	0.12
0.10	0.14	0.14

Notes: Each row of the Table corresponds to a different Monte Carlo simulation. The first column gives the value of β_1 used to generate each simulated sample. The second column gives the predicted plim($\hat{b}$) from formula (12) in the text. The third column reports the average value of the estimated $\hat{b}$ over 100 Monte Carlo replications with N=500, L=20 and K=2. Pool fixed effects are included in all regressions.

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

$K = 2$; $L = 20$; $N = 500$				
(1)	(2)	(3)	(4)	(5)
β_1	Mean simulated $\hat{b}$	Simulated p-value	Corrected $\hat{b}$	Corrected p-value
0.00	-0.06	0.005	0.00	0.474
0.01	-0.04	0.044	0.01	0.294
0.02	-0.02	0.320	0.02	0.060
0.03	0.01	0.340	0.03	0.007
0.04	0.03	0.110	0.04	0.000
0.05	0.05	0.016	0.05	0.000
0.06	0.07	0.000	0.06	0.000
0.07	0.09	0.000	0.07	0.000
0.08	0.11	0.000	0.08	0.000
0.09	0.12	0.000	0.09	0.000
0.10	0.14	0.000	0.10	0.000

Notes: Each row of the Table corresponds to a different Monte Carlo simulation over 100 Monte Carlo replications with N=500, L=20 and K=2. The first column gives the value of β_1 used to generate each simulated sample. Columns 2 and 3 report, respectively, the estimates for $plim(\hat{b})$ and the corresponding p-value as reported by OLS. The fourth column reports the corrected estimate $\hat{\beta_1}$ obtained using formula (15). The last column presents the corrected p-values obtained from 500 bootstrapping replication of the null hypothesis of no peer effect. Pool fixed effects are included in all regressions.

Table 7: Correction bias in the estimation of endogenous peer effects - Groups

	$K = 2$			$K = 5$		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$
$\hat{\beta_1}^{OLS}$ - no corrections	-0.05	0.15	0.34	-0.27	-0.04	0.18
Mean of p-value of $\hat{\beta_1}^{OLS}$	0.22	0.01	0.00	0.04	0.35	0.08
Proportion of naive p-value ≤ 0.05	39.8%	96.4%	100.0%	86.7%	22.5%	75.7%
$\hat{\beta_1}^{Ref}$ - corrected for reflection bias only	-0.02	0.07	0.16	-0.11	-0.01	0.09
$\hat{\beta_1}^{Corr}$ - corrected for reflection bias + exclusion bias	0.00	0.09	0.19	-0.01	0.09	0.18
Mean of p-value of $\hat{\beta_1}^{Corr}$ (using permutation method)	0.50	0.00	0.00	0.50	0.15	0.00
Proportion of p-value ≤ 0.05	4.0%	99.2%	100.0%	5.8%	49.9%	98.6%

Notes: Each column corresponds to a different Monte Carlo simulation over 1000 replications. We keep the number of observations in each sample and selection pool constant at N=1000 and L=20, but we vary β_1 and group size K. Cluster fixed effects are included throughout. Row 1 and row 2 report, respectively, the uncorrected $\hat{\beta_1}^{OLS}$ and its p-value obtained by regressing Y_i on G_iY and pool fixed effects. The third row reports the proportion of times the simulated naive p-value is smaller or equal to 0.05. For column 1 and column 4 this statistic essentially tells us what is the likelihood to make a type II error, that is, rejecting the null hypothesis when it is in fact true. For columns 2-3 and columns 5-6 this statistic essentially gives us the statistical power of the test. The fourth row presents the average of $\hat{\beta_1}^{Ref}$ estimates corrected for reflection bias but ignoring exclusion bias. This is estimated using model (15) with $E[\epsilon\epsilon'] = \sigma_\epsilon^2 I$. The fifth row reports the average $\hat{\beta_1}^{Corr}$ derived from model (15) with $E[\epsilon\epsilon']$ given by (16). The last two rows show the corrected p-value obtained using the permutation method and a statistic related to the power of the permutation inference method (similarly computed as in the third row).

Table 8: Correction bias in the estimation of endogenous peer effects - Networks

	p = 0.10			p = 0.25		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$
$\hat{\beta}_1^{OLS}$ - no corrections	-0.09	0.08	0.25	-0.26	-0.09	0.10
Mean of p-value of $\hat{\beta}_1^{OLS}$	0.18	0.18	0.00	0.03	0.32	0.26
Proportion of naive p-value ≤ 0.05	51.1%	41.7%	99.9%	88.8%	27.9%	36.3%
$\hat{\beta}_1^{Ref}$ - correction for reflection bias only	-0.05	0.04	0.12	-0.10	-0.03	0.03
$\hat{\beta}_1^{Corr}$ - correction for reflection bias + exclusion bias	0.00	0.10	0.19	0.00	0.09	0.19
Mean of p-value of $\hat{\beta}_1^{Corr}$ (using permutation method)	0.51	0.04	0.00	0.50	0.18	0.01
Proportion of p-value ≤ 0.05	6.3%	88.9%	100.0%	4.9%	47.3%	96.5%

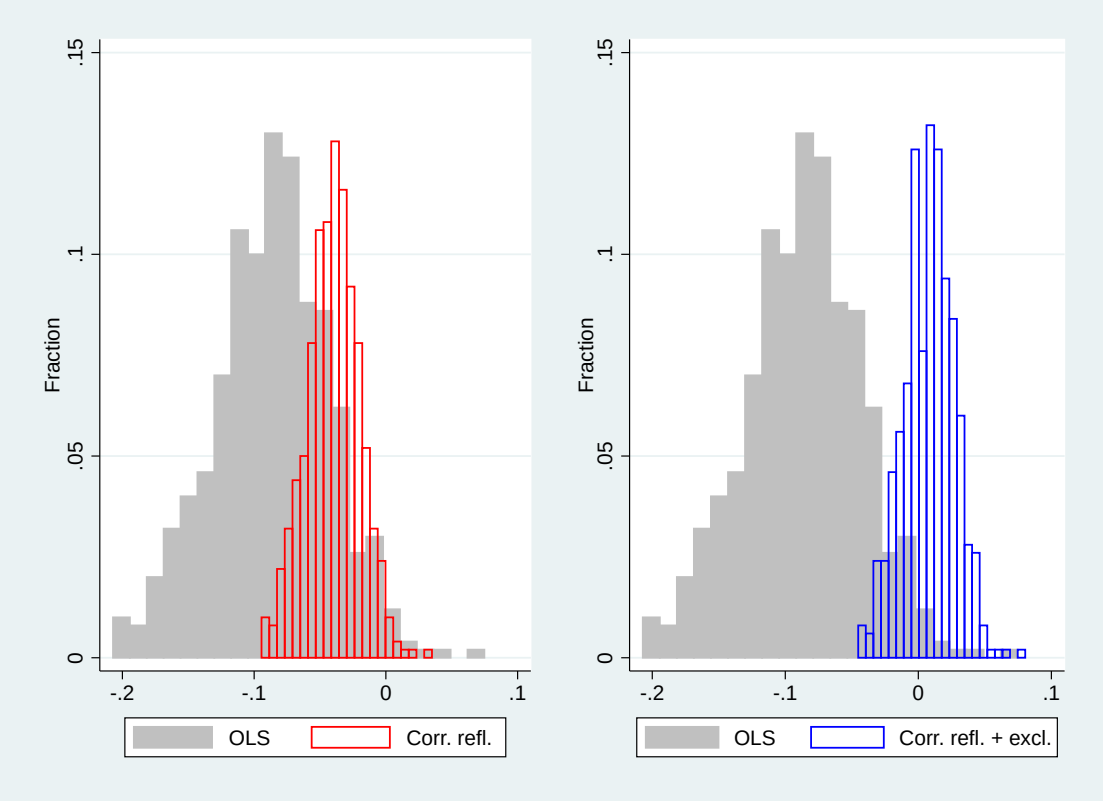
Notes: Each column corresponds to a different Monte Carlo simulation over 1000 replications. We keep the number of observations in each sample and selection pool constant at N=50 and L=20, but we vary β_1 and the linking probability p. Cluster fixed effects are included throughout. Row 1 and row 2 report, respectively, the uncorrected $\hat{\beta_1}^{OLS}$ and its p-value obtained by regressing Y_i on G_iY and pool fixed effects. The third row reports the proportion of times the simulated naive p-value is smaller or equal to 0.05. For column 1 and column 4 this statistic essentially tells us what is the likelihood to make a type II error, that is, rejecting the null hypothesis when it is in fact true. For columns 2-3 and columns 5-6 this statistic essentially gives us the statistical power of the test. The fourth row presents the average of $\hat{\beta_1}^{Ref}$ estimates corrected for reflection bias but ignoring exclusion bias. This is estimated using model (15) with $E[\epsilon\epsilon'] = \sigma_\epsilon^2 I$. The fifth row reports the average $\hat{\beta_1}^{Corr}$ derived from model (15) with $E[\epsilon\epsilon']$ given by (16). The last two rows show the corrected p-value obtained using the permutation method and a statistic related to the power of the permutation inference method (similarly computed as in the third row).

Table 9: Empirical applications

	Golfer data (1)		Student data (2)	
Final estimates:				
Endogenous peer effect (β_1^{Corr})	0.058	***	-0.023	**
p-value obtained through randomization inference	0.008		0.02	
Lagged own effect (β_2)	0.481	***	0.477	***
Standard error	0.079		0.020	
Exogenous peer effect (β_3)	-0.077		0.027	*
Standard error	0.130		0.015	
OLS estimates:				
Endogenous peer effect (β_1^{OLS})	0.022		-0.113	**
Naïve p-value	0.439		0.000	
p-value obtained through randomization inference	0.014		0.022	
GMM estimates correcting for reflection bias only:				
Endogenous peer effect (β_1^{Ref})	0.010		-0.051	**
p-value obtained through randomization inference	0.008		0.020	
Number of observations:	2517		2960	
Number of selection pools	100		155	
Group size	3		2	

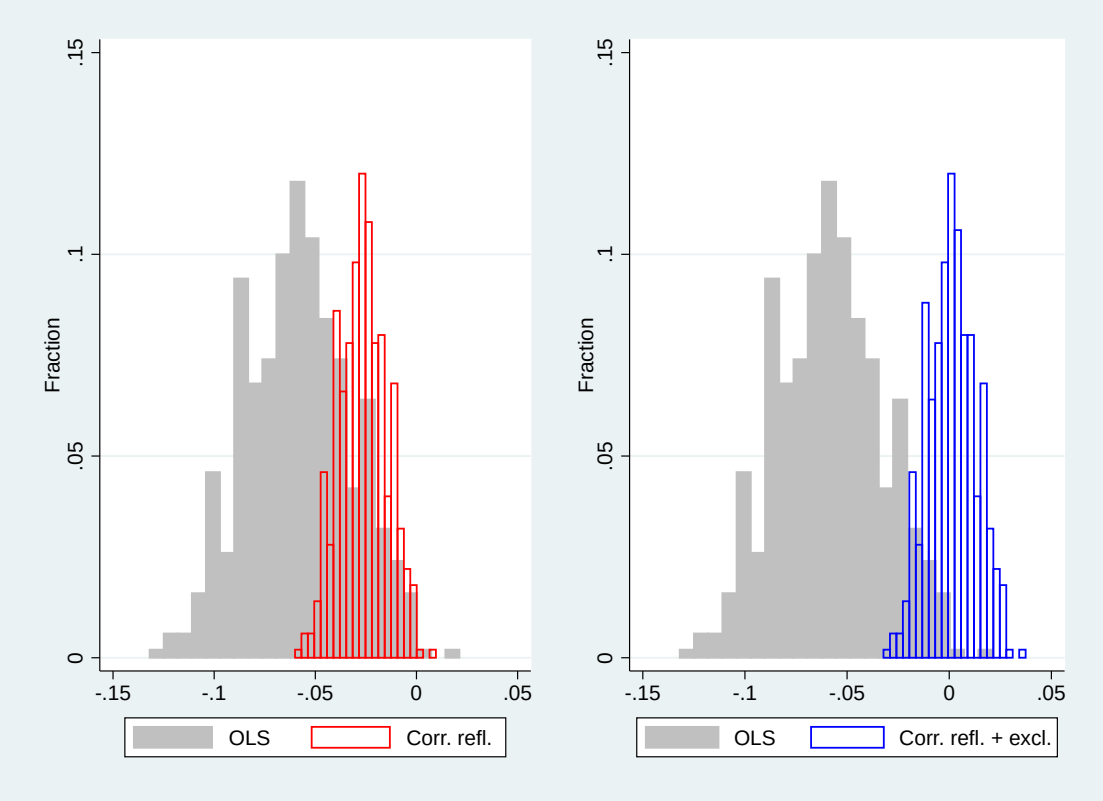
Notes: The golfer data are from Guryan et al (2009) and the student data are from Fafchamps and Mo (2018). For demonstration purpose, we restrict the golfer sample to the first tournament round, and to a random sub-set of N=100 out of 302 pools, making the overall sample size more comparable to the student application. We also focus on groups of size 3 ($K = 3$) which consist of 75% of all observations. We drop some observations which in the original dataset had erroneously been assigned to one or more players from a different pool than the one assigned to them (6% of all observations). The variable of interest in the golfer application is golf player's score and the lagged variable is a measure of ability of skill for every player (which is constructed based on lagged test scores). For the student application, we drop a few observations for which we observed inconsistencies in the indication of peers within a pair (16%). The variable of interest is the math score of the students. All regressions include pool fixed effects. In the golfer application the pool is the qualification category to which each player is assigned within each tournament. In the student application the pool is the classroom. Corrected P-values for each estimate are obtained using the permutation method, using 500 iterations.

Figure 5: Simulated $\hat{\beta}_1^{OLS}$, $\hat{\beta}_1^{Ref}$, and $\hat{\beta}_1^{Corr}$ under $H_0 : \hat{\beta}_1 = 0$ - Golfer data



Notes: These Figures plot for the Guryan et al (2009) application the simulated distribution of the naive $\hat{\beta}_1^{OLS}$ under the null of no endogenous peer effects (obtained after 500 repetitions of randomly reshuffling observations to different peers through Monte Carlo simulations) and compares this distribution (i) in the left panel to the distribution of simulated $\hat{\beta}_1^{Ref}$, i.e. the coefficient estimate which corrects for reflection bias but not for exclusion bias, and (ii) in the right panel to the distribution of simulated $\hat{\beta}_1^{Corr}$, i.e. the coefficient estimate correcting for both reflection and exclusion bias.

Figure 6: Simulated $\hat{\beta}_1^{OLS}$, $\hat{\beta}_1^{Ref}$, and $\hat{\beta}_1^{Corr}$ under $H_0 : \hat{\beta}_1 = 0$ - Student data



Notes: These Figures plot for the Fafchamps and Mo (2018) application the simulated distribution of the naive $\hat{\beta}_1^{OLS}$ under the null of no endogenous peer effects (obtained after 500 repetitions of randomly reshuffling observations to different peers through Monte Carlo simulations) and compares this distribution (i) in the left panel to the distribution of simulated $\hat{\beta}_1^{Ref}$, i.e. the coefficient estimate which corrects for reflection bias but not for exclusion bias, and (ii) in the right panel to the distribution of simulated $\hat{\beta}_1^{Corr}$, i.e. the coefficient estimate correcting for both reflection and exclusion bias.

Table 10: Mean $\hat{\beta}_1^{OLS}$, $\hat{\beta}_1^{Ref}$, and $\hat{\beta}_1^{Corr}$ under $H_0 : \hat{\beta}_1 = 0$ - Student data

	$\hat{\beta}_1^{OLS}$		$\hat{\beta}_1^{Ref}$		$\hat{\beta}_1^{Corr}$	
	Prediction	Simulation	Prediction	Simulation	Prediction	Simulation
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	-0.059	-0.058	-0.029	-0.026	0.000	0.001

Notes: This Table compares for the Fafchamps and Mo (2018) application (where $K = 2$) the mean of the simulated $\hat{\beta}_1^{OLS}$, $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{Corr}$ (shown in Figure 6), to the exact predictions made by formulas (11)-(13) about the plim of $\hat{\beta}_1^{OLS}$, $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{Corr}$ under the null of no endogenous peer effects.

Table 11: Relationship between the different estimators under $H_0 : \hat{\beta}_1 = 0$ - Student data

	$\hat{\beta}_1^{Ref}$		$\hat{\beta}_1^{Corr}$	
	Prediction	Simulation	Prediction	Simulation
	(1)	(2)	(3)	(4)
Constant	0.000	0.000	0.029	0.028
$\hat{\beta}_1^{OLS}$	0.500	0.453	0.500	0.454
N	1000	1000	1000	1000

Notes: This Table shows for the Fafchamps and Mo (2018) application (where $K = 2$) results of regressions of simulated $\hat{\beta}_1^{Ref}$ on simulated $\hat{\beta}_1^{OLS}$ and of simulated $\hat{\beta}_1^{Corr}$ on simulated $\hat{\beta}_1^{OLS}$ and compares these results to the predicted relationships between these different estimators, obtained using equations (11)-(13).