Exclusion Bias and the Estimation of Peer Effects*

Bet Caeyers† Marcel Fafchamps‡

April 2023

Abstract

We estimate peer effects in two datasets with non-overlapping peer groups: golfers who play tournaments randomized in groups of three; and students who are randomly paired for in-class computer-assisted learning. In such data, existing instrumental variable methods to address bias in peer effect estimation do not apply. Alternative estimation methods exist that do not require instruments, but they fail to correct for one understudied but important source of bias which we call 'exclusion bias'. We provide formulas for the magnitude of this bias when fixed effects are included at the level of selection pools. We then derive a consistent estimator that corrects for this bias and propose a simple method for testing the presence of endogenous peer effects. Using this novel method, we find positive peer effects in the first case – consistent with emulation between golfers during the tournament – and negative peer effects in the other – consistent with congestion or wasteful competition for the computer between students. These results differ markedly from existing methods in terms of magnitude, significance, and inference.

Keywords: Exclusion bias, Peer effects, Reflection bias, Random peer assignment, Social interactions, Autoregressive Models

*We benefitted from comments from anonymous referees, Ingmar Prucha, 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 audiences at Gothenburg University, FAIR/NHH, IFS, LICOS, UC Davis, Peking University, IFPRI, NEUDC conference, RES conference, Stanford conference on Social Networks, CSAE conference on African development, Laval University conference on Social Networks, Barcelona GSE Summer forum, PODER Summer School, and Nova University Summer School. We thank the Stanford Rural Education Research Program (REAP) for sharing the student data with us. Finally, we are grateful to the ESRC Centre for Microeconomic Analysis of public policy (CPP) at the IFS for funding parts of this research. A proof of Proposition 1 first appeared in Bet Caeyers' (2013) PhD thesis. The original version of this paper came out in August 2016 as NBER Discussion Paper 22565. In February 2020, the original was replaced on the NBER website with a new version containing the empirical application and most of the other improvements included here.

†bet.caeyers@cmi.no, FAIR (Chr. Michelsen Institute, FAIR/Norwegian School of Economics)

‡fafchamp@stanford.edu Stanford University, NBER and IZA

1 Introduction

Since Manski's (1993) seminal article, the estimation of endogenous peer effects has raised considerable interest among economists. Social contact may spur emulation, cooperation, and social learning between peers. But it can also lead to congestion, competition, or conflict. Promising environments in which to identify such peer effects are situations in which agents are randomly assigned to a group within which they have to complete a task, either individually or as a team. Random assignment takes care of possible self-selection on common ability and interest; and the partition of individuals into mutually exclusive groups reduces contamination across groups and should therefore improve causal identification. Examples of such studies include: the assignment of students to classes (e.g., Carrell et al. 2019), dorm rooms (e.g., Sacerdote 2001, Carrell et al. 2013, Corno et al. 2022), and study groups (e.g., Carrell et al. 2019; Fafchamps and Mo 2018); the assignment of worker to teams (e.g., Bandiera et al 2009); the assignment of athletes to groups in a tournament (e.g., Guryan et al. 2009); and the assignment of entrepreneurs to social groups (e.g., Fafchamps and Quinn 2018, Cai and Szeidl 2018). In all these examples, individuals are assigned to a group from within distinct selection pools – e.g., a cohort, school, or classroom. Since selection pools typically differ in important ways, any analysis of endogenous peer effects must include selection pool fixed effects.

In this paper, we use a novel method to revisit the estimation of endogenous peer effects in two separate such applications. In the first application, we re-examine golfers randomly divided in groups of three for (the first part of) a tournament (e.g., Guryan et al. 2009). Our empirical relationship of interest is whether a player plays *better* or *worse* than predicted by their past performance when they are matched with players who play better than predicted by their past performance - and vice versa. Because tournaments differ in the pool of participating players, it is essential that we control for the average quality of players in a tournament, which we do by including tournament fixed effects. In the second application, we re-examine students randomly divided in pairs to work on computer assisted learning in their classroom (e.g., Fafchamps and Mo 2018). Our empirical relationship of interest is whether a student performs better at the exam than predicted by their previous score when their assigned peer also performed better than predicted by their own previous score. We include classroom fixed effects to account for variation in average student ability across classes. In both cases, we assume the absence of correlated effects within groups other than those occurring at the level of the selection pool – i.e., the tournament in the case of golfers, and the classroom in the case of students. We believe this assumption is reasonable given the controlled nature of each setting: in the tournament, all golfers play the same course in the same conditions; in the classroom, all students are taught in the same way by the same teachers.

In both cases, a positive relationship would imply mutually reinforcing peer effects, indicating strategic complements – which could be caused, for instance, by emulation or mutual assistance between peers. In contrast, a negative relationship would imply strategic substitutes – for instance driven by congestion, noxious competition, or emotional drain. In our empirical analysis, we find positive endogenous peer effects among golfers, consistent with an emulation interpretation, and we find negative endogenous peer effects among students, consistent with a congestion or competition interpretation. Our findings differ markedly from results obtained using existing endogenous peer effect estimation methods, in terms of magnitude, significance, and inference (e.g., Guryan et al. 2009 and Fafchamps and Mo 2018).

The main innovation of our paper is the novel methodology that we develop to obtain these estimates. The literature has essentially developed two main approaches for estimating endogenous peer effects. Both formalize peer effects as taking place on a network where individuals are nodes and peer effects are links. Non-overlapping peer groups are represented as a block diagonal adjacency matrix.

The first of the two existing estimation methods relies on instrumental variables to address endogeneity. One popular approach uses as instruments the exogenous characteristics of the peers of my peers (who are not my peers). These characteristics satisfy the exclusion restriction since they directly affect the behavior of my peers but not mine (e.g., Bramoullé et al. 2009, De Giorgi et al. 2010). This intuition can be generalized to more distant peers (e.g., Kelejian and Prucha 1998, 1999; Lee et al. 2021). While all these approaches allow the estimation of endogenous and exogenous peer effects in general network data, they do not help in situations where individuals are partitioned into mutually exclusive groups since my peers do not have peers that are not also my peers. This means that these methods produce no instruments. Rose (2017) also applies the idea of using neighbors of neighbors, but achieves identification not by using instruments but by using fluctuations in the variances and covariances of the dependent variable in the social network. This method also fails in non-overlapping group settings. Another instrumental approach, proposed by Lee (2007), relies on variation in group size for identification. The logic behind this approach is that larger groups generate larger multiplier effects. Graham (2008) also uses variation in peer group size for identification, but relies on variances and covariances instead. In both cases, successful identification requires having sufficient variation in the size of peer groups (e.g., Davezies et al. 2009). This does not apply to either of our two applications, for which group size is constant.

The second approach, discussed by Anselin (1988) and turned into a Stata command by Drukker et al. (2013), draws from spatial econometrics (see also Anselin and Bera 1998). The method uses a maximum likelihood approach to estimate endogenous peer effects directly from the network covariance matrix between observations of the dependent variable, without the need for instruments.[1] In this more structural method, identification of endogenous peer effects is achieved from correlation patterns between peers. Endogenous peer effects due to reflection (i.e., peers influence each other) create a network covariance matrix that decays, in a mathematically precise way, as the network distance between two individuals increases. In contrast, correlated shocks among peers manifest themselves as correlation patterns between directly linked individuals, without spilling over through the network. For groups, these take the form of group-level random effects. While this method is applicable to many social networks (and to geographical data), its application to non-overlapping peer groups requires assuming away correlated effects. This is because, in this case, endogenous peer effects are not identified separately from correlated effects: peer effects do not spread beyond the group and, consequently, they are indistinguishable from group-level random/correlated effects.[2]

We propose a novel approach to estimate endogenous peer effects that applies more generally, including in settings with fixed group sizes. This approach shares a key similarity with the spatial ML estimator (e.g., Anselin 1988): it relies on the covariance matrix between observations to

[1] The Stata command developed by Drukker et al. (2013) also implements the GMM instrumental variable approach of Kelejian and Prucha (1998). Like the IV estimators of Bramoullé et al. (2009), De Giorgi et al. (2010), and Lee et al. (2021), this estimator cannot deal with non-overlapping groups.

[2] The reader familiar with time series analysis will recognize the analogy between endogenous peer effects, which resemble AR1 autocorrelation in the sense that they create a covariogram that decays with distance; and group random effects, which resemble a moving average process, in the sense that they create a covariogram that is positive for linked nodes, and zero otherwise. With non-overlapping groups, there is no AR1 decay beyond the group and thus both AR1 and moving average DGP produce the same covariogram, ruling out separate identification through the network covariance of the observations. See Online Appendix B.5 for more details.

identify peer effects. But it differs in one essential new feature: we correct for exclusion bias, an important source of bias that, to date, has only received limited attention. Exclusion bias affects all estimators relying on the covariance matrix for identification when they include pool fixed effects. But it is most problematic in situations with non-overlapping peer groups of fixed size, for which alternative IV estimators are not available.

We start by showing that exclusion bias arises from the fact that the assignment of peers is done without replacement: i cannot be his own peer. When fixed effects are included at the level of selection pools, the exclusion of i from the pool of i's peers creates a small sample negative relationship between i's observation 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. After netting out the pool average via fixed effects, this implies that i's observation is negatively correlated with the sample average of the remaining peers in the pool.[3] Guryan et al. (2009) were the first to introduce the notion of exclusion bias when testing for random peer assignment. They however ignore the presence of exclusion bias in the estimation of endogenous peer effects.

We illustrate the magnitude of this bias in our two datasets, first by implementing a test of random assignment of individuals into groups. We find that, had we ignored exclusion bias, we would have incorrectly concluded that, in both datasets, peers were negatively assorted. Next, we derive a formula for the asymptotic bias itself, which shows that exclusion bias disappears when the size of each selection pool tends to infinity.[4] 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. We then use this formula to construct a consistent estimator of endogenous peer effects. This estimator resembles the spatial maximum likelihood estimator of Anselin (1988), except that it corrects for exclusion bias. One limitation of this estimator is that identification requires assuming away correlated effects when peer groups are non-overlapping – a limitation that

[3] See Online Appendix A for a formal derivation of this statement.

[4] The source of bias is similar to what arises with autoregressive models in short panels: in such models, 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 reverse feedback.

also affects the spatial ML estimator in this case. Our new estimator is then used to estimate endogenous peer effects in the golfer and student data. We find that, had we failed to correct for exclusion bias, we would have erroneously concluded that peer effects were smaller in both cases. In the golfer data, positive peer effects would have been rejected, and in the student data, negative peer effects would have been overestimated. This means that ignoring exclusion bias presents a real risk of drawing incorrect inference.

At the end of the paper, we briefly discuss how a correction for exclusion bias could be added to the implementation of the ML estimator by Drukker et al. (2013). This correction would also make it possible to obtain consistent estimates of peer effects in data with overlapping peer groups and more general network or spatial configurations, while at the same time allowing for correlated effects across peers.

The paper is organized as follows. In Section 2 we briefly introduce the two datasets and the testing strategy used in our analysis. We then conduct a test of random assignment to peer groups in Section 3. In Section 4 we derive an asymptotic formula for exclusion bias and use it to construct a consistent estimator of endogenous peer effects in Section 5. The performance of our estimator is illustrated using simulations. This estimator is applied to our two datasets in Section 6 to obtain consistent estimates of peer effects in non-overlapping peer groups. Section 7 concludes. In Online Appendix A we discuss in more detail the source of the exclusion bias, how the reflection bias and exclusion bias combine, and the intuition behind the methodology that we propose in the paper to simultaneously address both biases. Online Appendix B presents extensions of the method.ology as well as a discussion of ways to avoid exclusion bias in IV regressions. All proofs are gathered into Online Appendix C.

2 Data and testing strategy

When estimating peer effects, there are two types of causal identification that researchers may have in mind. They may want to estimate the causal effect that an exogenous rise in the behavior of one agent may have on the peers. This is the domain of diffusion experiments in which the researcher 'treats' a random or carefully selected set of nodes and observes how the treatment affects peers (e.g., Banerjee et al. 2013). Alternatively, the researcher may want to study the nature of peer interactions itself, for instance to determine whether a particular activity is characterized by strategic complements or substitutes: e.g., does a sporting event create emulation among groups of players; or does a particular environment generate wasteful competition or congestion between peers. Our empirical applications belong to the second type of study: we are not seeking to identify the 'causal' effect that treating one agent would have on its peers. Rather we want to determine the nature of strategic interactions between agents by observing whether their performance in a task is more correlated within than across groups.

To this effect, we investigate two existing experiments in which subjects within a selection pool are randomly assigned to groups of fixed size and asked to perform a task in a tightly controlled environment. The first dataset comes from golf tournaments in which participants are randomly assigned to groups of players within their qualification category. Here we are using data made publicly available by Guryan et al. (2009). 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. In both cases, we limit our data to groups of the same size – three in the golfer data and two in the student data. This is done to ensure entire comparability across groups, but it also serves to demonstrate that our method works with groups of fixed size for which instrumental variables do not exist. In the Guryan et al. (2009) dataset, we 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 speed up an execution time that increases exponentially with the size of matrix G later in our analysis, we reduce the number of observations by focusing on a random sub-set of 100 out of 302 selection pools. This leaves a sample of 2,517 observations from 100 pools of 25 golfers each, on average, organized in groups of three.

In both cases we want to estimate a standard linear-in-means model of peer effects (Manski, 1993):

$$y_{ikl,t+1} = \beta_1 \bar{y}_{-ikl,t+1} + \beta_2 y_{iklt} + \beta_3 \bar{y}_{-iklt} + \delta_l + \epsilon_{ikl,t+1} \tag{2.1}$$

where $y_{ikl,t+1}$ 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. The intercept is subsumed into the pool fixed effects δ_l. Coefficient β_1 is the endogenous peer effect that measures the nature and extent of strategic interactions: if effort and performance are strategic complements, we expect $\beta_1 > 0$; if they are strategic substitutes, $\beta_1 < 1$. Regressors y_{iklt} and $\bar{y}_{-iklt}$ measure the past performance of i and of his/her peers. Since past performance of individual i is bound to affect their performance in our data, we expect $\beta_2 > 0$. Coefficient β_3 estimates what is commonly referred to as an exogenous peer effect (or contextual peer effect): $\beta_3 > 0$ means that i's performance is higher when matched with peers who have performed well in the past, and lower if they have performed poorly; in contrast, $\beta_3 < 0$ means that i's performance suffers when matched with peers who have done well in the past.

We include selection pool fixed effects δ_l to capture possible correlation in residuals within *selection pools*, as is likely. But we assume that, conditional on δ_l, the residuals $\epsilon_{ikl,t+1}$ are not correlated within *groups*. The suitability of this assumption depends on the context. Given the inclusion of pool fixed effects and the controlled nature of both study environments, it is a reasonable assumption in the two datasets we have selected. It implies that correlation in performance between individuals in the same group must come either from endogenous or exogenous peer effects.

3 Testing for random peer assignment

When estimating model (2.1), peer self-selection is a major threat to identification because, if individuals were left to their own device, they would sort differently depending on the nature of the strategic interactions (e.g., Legros and Newman 2007). Hence, if assignment into peer groups was not random, we may falsely ascribe a correlation in performance to strategic effects when they are in fact due to positive or negative assorting. In the two empirical settings that we include in our analysis, individuals were supposed to be assigned to peer groups in a random fashion. We need to verify that this was indeed the case.

Since Sacerdote (2001), in applied economics random peer assignment is typically verified by testing whether $\alpha_1 = 0$ in a linear-in-means model of the following form:

$$y_{iklt} = \alpha_1 \bar{y}_{-iklt} + \delta_l + \epsilon_{iklt} \tag{3.1}$$

where y_{iklt} denotes the past performance of individual i in group k from pool l. The intercept is subsumed in the selection pool dummies. Regressor $\bar{y}_{-iklt}$ denotes the average of i's peers in group

k (excluding i herself). Selection pool dummies δ_l are included to control for randomization strata fixed effects. – e.g., in the golfer data, the quality of contestants varies across tournaments; and in the student data, the school performance of students varies across classes and schools. In our data, each peer group has size $K = 3$ in the golfer data and $K = 2$ in the student data. If the number of groups in a selection pool is N_p, then the pool size $L = N_p \times K$. If the total number of pools in the sample is N, then the total sample size $S = N \times L$.

Model (3.1) is typically estimated using ordinary least squares (OLS). Researchers proceed as if random assignment of peers implies that the OLS estimate of the coefficient α_1 in regression (3.1) should be 0. As shown through simulations 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. This can be shown easily if we rewrite model (3.1) in deviation from the pool mean to eliminate the fixed effects δ_l:

$$\ddot{y}_{ikl} = \alpha_1 \ddot{\overline{y}}_{-ikl} + \ddot{\epsilon}_{ikl}$$

where $\ddot{y}_{ikl} \equiv y_{ikl} - \overline{y}_l$ where $\overline{y}_l$ is the sample mean of variable y_{ikl} in selection pool l . Variables $\ddot{\overline{y}}_{-ikl}$ and $\ddot{\epsilon}_{ikl}$ are similarly defined. The time subscript t has been omitted to improve clarity. Now let $\overline{y}_{-il} \equiv \sum_{j \neq i, j \in l} y_{jl}$ denote the leave-out mean of y, that is, the sample mean of the y observations that belong to the same selection pool as i, but does not include the i observation. Let similarly denote the deviation of this variable from its pool mean as $\ddot{\overline{y}}_{-il}$. It follows immediately that the correlation between $\ddot{y}_{ikl}$ and $\ddot{\overline{y}}_{-il}$ is -1: if $\ddot{y}_{ikl}$ deviate from the pool sample mean by a value d, then $\ddot{\overline{y}}_{-il}$ must mechanically deviate from the pool mean by an equivalent but opposite amount.[5] Since the peers assigned to i in group k are selected randomly from the observations that form $\ddot{\overline{y}}_{-il}$, the mean $\ddot{\overline{y}}_{-ikl}$ will, on average, be negatively correlated with y_{ikl}. We refer to this phenomenon as 'exclusion bias' since it mechanically arises from the fact that i is excluded from being her own peer.

Guryan et al. (2009), Wang (2009) and Stevenson (2015, 2017) have proposed methods to test the null hypothesis of random peer assignment while correcting for exclusion bias. The method proposed by Guryan et al. (2009), henceforth GKN, uses the average of the selection pool as control

[5] Online Appendix A provides some simple examples.

variable to eliminate exclusion bias. While the method is simple to implement, it only identifies the parameter of interest α_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 unidentified. By extension, limited variation in pool size results in weak identification.

Wang (2009) suggests an alternative approach that involves running an F-test of joint significance of peer group dummies in a model of the form:

$$y_{ikl} = \alpha_1 C_k + \delta_l + \epsilon_{ikl} \tag{3.2}$$

where C_k is a set of group dummies. The author argues that, if individuals are randomly assigned to groups, all group means should be statistically similar and the coefficients included in vector α_1 should jointly not be significantly different from zero. This method has been criticized by Stevenson (2015) who shows that, based on simulation results, the method fails to reject the null hypothesis if peers are negatively correlated.

Stevenson (2015, 2017) 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, there is no problem. The split-sample method exploits this feature, as follows. The researcher first randomly selects one observation from each peer group in the original dataset. The researcher then calculates the average outcome of the peers of those individuals selected in Step 1, excluding the selected individuals themselves. 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. The method effectively creates a new dataset – derived from the original data – where individuals are excluded from their own peer group but also 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.

In contrast, randomization inference through the permutation method offers a simple and generally applicable way of testing random peer assignment that overcomes the limitations of these other methods (e.g., Fisher 1925, Guryan et al. 2009). The idea is to simulate, using the data at hand, the distribution of $\widehat{\alpha}_1$ under the null hypothesis of random peer assignment.[6] 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. (2018), which re-randomizes treatments across peers.

Before applying the method to our data, we illustrate how it would work when the researcher has observational data y_{iklt} partitioned in groups of varying size K_i within pools of size L_i. The first four columns of Table 1 give an example of such data structure. We test random assignment within pools using regression (3.1) and applying Krackardt's (1988) permutation method to generate each synthetic sample.[7] To visualize the performance of the proposed testing method, we generate artificial samples of 1000 observations for three values of $K = \{2, 5, 10\}$. We set the size of each pool to $L = 20$ and we posit $\epsilon_{iklt} \sim N(0, 1)$. Figure 1 shows the distribution of 1000 simulated $\widehat{\alpha}_1^s$ under the null of random peer assignment for $K = \{2, 5, 10\}$. The striking finding is that the histograms are not centered around $\alpha_1 = 0$. They are all shifted to the left due to exclusion bias. The permutation method corrects p-values by taking this distributional shift into consideration when calculating the probability of observing $\widehat{\alpha}_1$ under the null. Figure 2 illustrates, for one particular example (i.e., $S = 1000$, $N = 50, L = 20$ and $K = 5$), that the permutation method yields correct inference.

Having validated the approach, we apply it to test for random peer assignment in our two datasets. Results are shown in Table 2 . We see that OLS point estimates for $\widehat{\alpha}_1$ are well below 0 in both cases, and that OLS p-values reject random peer assignment in both cases. We then use randomization inference to correct for this and find that random assignment of peers is not rejected

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

[7] We start by estimating the model on the data to obtain the OLS estimate $\widehat{\alpha}_1$. We wish to know how likely it is to obtain value $\widehat{\alpha}_1$under the null of random assignment within pools. To this effect, we simulate the distribution of $\widehat{\alpha}_1$ 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 1. For each reassignment we estimate regression (3.1) and obtain a counterfactual realization of $\widehat{\alpha}_1^s$ 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 $\widehat{\alpha}_1$ under the null. The mean of the distribution of $\widehat{\alpha}_1^s$ is the average bias under the null. We then compare our $\widehat{\alpha}_1$ estimate to the distribution of $\widehat{\alpha}_1^s$. 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 $\widehat{\alpha}_1^s$ that are either above the absolute value of $\widehat{\alpha}_1$ or below minus the absolute value of $\widehat{\alpha}_1$.

in either of the two datasets. This is the first evidence we provide in this paper that neglecting exclusion bias can lead to incorrect inference. For this reason, we devote the next section to a thorough investigation of the source of this bias.

4 Exclusion bias

In this section, we derive a formula for exclusion bias in tests of random peer assignment and demonstrate its validity using simulations. This analysis is an essential stepping stone towards the development of our peer effect estimator in Section 5.

4.1 Formula

We now provide a formula for the exclusion bias that affects $\widehat{\alpha}_1$ in regression (3.1). The formula is applicable to cases assuming homoskedastic error terms and cases where we observe all individuals in each of the non-overlapping pools of potential peers. The proof is provided in Appendix C. 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 (3.1). We obtain the following Proposition:

Proposition 1: *Let the errors in model* (3.1) *be i.i.d. with variance* σ_ϵ^2*, let peers be assigned randomly (*$\beta_1 = 0$*). Then the estimate of* α_1 *obtained by estimating model* (3.1) *with pool fixed effects satisfies the following properties:*

$$plim_{N\to\infty}[\hat{\alpha}_1^{FE}] = -\frac{(L-1)(K-1)}{(L-K)L+(K-1)} < 0 \;\; for\; L, K\, fixed \tag{4.1}$$

$$plim_{L\to\infty}[\hat{\alpha}_1] = 0 \;\; for\; N = 1\, and\, K\, fixed \tag{4.2}$$

$$E\left[\hat{\alpha}_1^{FE}|N\right] < plim_{N\to\infty}\left[\hat{\alpha}_1^{FE}\right] \leq 0\; for\, L, K\, fixed \tag{4.3}$$

Proof: see Appendix C.1, Appendix C.2 and Appendix C.3

Equation (4.1) 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. 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. $\frac{\triangle|plim_{N\to\infty}[\hat{\alpha}_1^{FE}]|}{\triangle L} < 0$: For a given peer group size K, the asymptotic exclusion bias falls as pool size L increases.[8] This property is similar to what happens in autoregressive models with panel fixed effects, where the OLS-FE bias falls as T, the number of periods, increases.

2. $\frac{\triangle|plim_{N\to\infty}[\hat{\alpha}_1^{FE}]|}{\triangle K} > 0$: 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.[9]

Equation (4.2) extends formula (4.1) 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 C.2.

Equations (4.1) and (4.2) 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 (4.3), provides an additional result, obtained using Taylor approximations and Monte Carlo simulations in Appendix C.3. 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 value (4.1) 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 $E[\hat{\alpha}_1|L] < plim_{L\to\infty}[\hat{\alpha}_1] \leq 0 \; for\, N, K\, fixed.$ When the number of peer groups is small, the magnitude of the exclusion bias can be large even with a large L, something we also illustrate in the next section.

[8] 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.

[9] 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.

So far we have assumed that all selection pools are of equal size L and that groups are of equal size K. Proposition 2 generalizes formula (4.1) for any arbitrary combination of group and pool sizes.

Proposition 2: *Let K_k denote the size of group k and let L_k be the size of its pool. Let the errors in model* (3.1) *be i.i.d. with variance σ_ϵ^2, let peers be assigned randomly ($\alpha_1 = 0$). Then the plim of $\hat{\alpha}_1$ in model* (3.1) *with pool fixed effects is given by the following formula:*

$$plim_{N\to\infty}[\hat{\alpha}_1^{FE}] = \sum_k \frac{K_k}{M}\frac{s_{z_k}^2}{s_z^2} plim_{N\to\infty}[\hat{\alpha}_{1k}] \text{ where} \tag{4.4}$$

$$plim_{N\to\infty}[\hat{\alpha}_{1k}] = -\frac{(L_k-1)(K_k-1)}{(L_k-K_k)L_k+(K_k-1)}$$

$$s_{z_k}^2 = \frac{(K_k-1)+(L_k-K_k)L_k}{L_k(L_k-1)(K_k-1)} \ \textit{and}\ s_z^2 = \sum_k \frac{K_k}{M} s_{z_k}^2$$

and $M \equiv \sum_k K_k$ is the total number of observations in the estimation sample.

Proof: see Appendix C.4.

Proposition 2 shows that $plim_{N\to\infty}[\hat{\alpha}_1^{FE}]$ is nothing but a weighted sum of *plim*'s from formula (4.1) with weights derived from a simple covariance decomposition.[10] It should be noted that the results in Propositions 1 and 2 hold for a model having the form of equation (3.1), that is, is the natural form for a test of random assignment. They do not apply if the regression includes additional regressors. If the researcher wishes to add regressors w_{ikl} when testing for randomized assignment, it is necessary to first partial out w_{ikl} from y_{ikl} and $\bar{y}_{-ikl}$.[11] Propositions 1 and 2 and the other results from this section apply to these partialled-out regressions.

[10] In the case where groups in a pool l are all the same size K_l but group size K_l and pool size L_l vary across pools, the formula simplifies to:

$$plim_{N\to\infty}[\hat{\alpha}_1^{FE}] = -\frac{\sum_l \frac{M_l}{L_l}}{\sum_l \frac{M_l}{L_l}\frac{(K_l-1)+(L_l-K_l)L_l}{(L_l-1)(K_l-1)}}$$

where M_l is the size of the sample of all pools of size l. In contrast, if pool size is constant but group size K_k varies, the formula simplifies to:

$$plim_{N\to\infty}[\hat{\alpha}_1^{FE}] = -\frac{1}{\sum_k \frac{K_k}{L}\frac{(K_k-1)+(L-K_k)L}{(L-1)(K_k-1)}}$$

[11] Practically, this means doing the following: (1) take out the pool fixed effect by expressing all variables in deviation from their selection pool mean – e.g., $\ddot{y}_{ikl} \equiv y_{ikl} - \frac{1}{L_k}\sum_{ik\in l} y_{ikl}$; (2) regress the demeaned $\ddot{y}_{ikl}$ on $\ddot{w}_{ikl}$ and keep the residuals, which we denote as $\hat{u}_{ikt}$; (3) regress $\ddot{\bar{y}}_{-ikl}$ on $\ddot{\bar{w}}_{-ikl}$ (the de-meaned leave-out mean of w_{ikl} for peers) and keep the residuals, which we denote as $\hat{v}_{-ikl}$; and (4) construct $\ddot{\hat{u}}_{ikt} \equiv \hat{u}_{ikt} - \rho\hat{v}_{-ikl}$; and (5) regress $\ddot{\hat{u}}_{ikt}$ on $\hat{v}_{-ikl}$. This is the partialled-out regression.

In Appendix B.1 we illustrate how formulas (4.1) and (4.4) can be used in order to transform model (3.1) to obtain a consistent estimate of α_1. This solution offers an alternative approach to correcting inference in standard tests of random peer assignment, instead of using randomization inference described in Section 3.

4.2 Simulation results

We start by noting that Proposition 1 correctly predicts the magnitude of exclusion bias found in Section 3: Figure 1 is centered on the *plim* of $\hat{\alpha}_1$ under the null that is given by (4.1) in Proposition 1. We now present simulation evidence to demonstrate that this is a general property of the formula.

Results from a Monte Carlo simulation are presented in Table 3. Simulations vary 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 $y_i \sim N(0,1)$. 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 (3.1) and collect the estimated $\widehat{\alpha}_1$. The average $\widehat{\alpha}_1$ for each vector $\{K, L\}$ is summarized in Table 3. For comparison purposes, we also report the predicted $plim_{N\to\infty}[\hat{\alpha}_1]$ 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. Table 3 also shows the proportion of artificially generated samples for which we falsely reject the null hypothesis that $\alpha_1 = 0$ at the 1%, 5% and 10% significance levels. Random assignment is falsely 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 3 the rate at which OLS rejects the null hypothesis that $\alpha_1 = 0$. 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 $\alpha_1 = 0$ in OLS regression (3.1) over-rejects the null of random assignment in a substantial proportion of cases. To summarize, the test is strongly biased and the magnitude of the bias in large samples is well predicted by formula

(4.1).

Simulation results presented in Table 4 show for a given pool size $L = 50$ and separately for $K = 5$ and $K = 10$, what happens to the exclusion bias when N, the number of selection pools, increases. The results confirm that the bias is larger in small samples and that it converges to the value predicted by (4.1) as N increases (predicted values for $K = 5$ and $K = 10$ are shown in the middle and bottom panel of column 2 in Table 3, respectively).

5 Endogenous peer effect estimator

Equipped with a better understanding of exclusion bias, we are now in a position to derive a consistent estimator of endogenous peer effects in regression (2.1). In Appendix A we discuss in more detail the source of the exclusion bias, how the reflection bias and exclusion bias combine and the intuition behind the methodology that we propose in the paper to simultaneously address both biases. As explained in the introduction, our estimation approach cannot rely on instrumental variables, since they are not available in non-overlapping peer groups. We rely instead on the structure of the covariance matrix between observations, like the ML estimator of Anselin (1988).

Formally, we consider a data generating process similar to that of Moffit (2001). In Appendix A we use a simple $K = 2$ setting to provide an illustration for which the exact value of the reflection and exclusion biases can be derived algebraically. In Section 5.1 we generalize the approach to any group size and we show how a simple sequential algorithm can be used to obtain an estimate of β_1 in regression (2.1) that is free of both reflection and exclusion bias. Throughout the formal presentation we assume homoskedasticity of the errors. We do, however, conduct inference by re-randomizing peer assignment within pools, which de facto corrects for the clustering of standard errors within pools – and thus also for heteroskedasticity.

5.1 Deriving the estimator

To simplify the algebra, we start by rewriting model (2.1) in a more general, matrix-oriented form:

$$y_{il} = \beta_1 G_{il} Y_l + \beta_2 x_{il} + \beta_3 G_{il} X_l + \delta_l + \epsilon_{il} \tag{5.1}$$

where: Y_l is the vector of all y_{jl} in pool l; vector G_{il} picks all the peers of individual i in pool l and assigns them a weight $1/(K-1)$ to construct the peer group's mean;[12] x_{il} is an individual characteristic that affect y_{il} directly – past performance y_{iklt} in our case;[13] X_l is the vector of all X_{jl} in pool l; and δ_l is a selection-pool fixed effect. Parameter β_1 captures endogenous peer effects; parameter β_2 captures the effect of the characteristics of individual i on y_{il}; and β_3 captures so-called exogenous peer effects, that is, characteristics of peers that affect i directly without the need to influence peers' behavior.

Combining all selection pools, regression model (5.1) can be rewritten in matrix form as:

$$Y = \beta_1 GY + \beta_2 X + \beta_3 GX + \triangle + \epsilon$$

where: Y and X are vectors containing all observations on y_{il} and x_{il}, respectively; and G is a weighting matrix that picks relevant peers and averages them. The model can be further simplified by expressing all variables in deviation from their l pool mean to eliminate the pool fixed effect δ_l. We obtain the following expression:

$$\ddot{Y} = \beta_1 G\ddot{Y} + \beta_2 \ddot{X} + \beta_3 G\ddot{X} + \ddot{\epsilon}$$

where, as before, the notation $\ddot{Z} \equiv Z - \overline{Z}_l$ where $\overline{Z}_l$ stacks the element-by-element sample mean of vector Z_{il} in selection pool l. Simple algebra yields the following reduced-form model:

$$\ddot{Y} = (I - \beta_1 G)^{-1}(\beta_2 \ddot{X} + \beta_3 G\ddot{X} + \ddot{\epsilon})$$

from which we obtain an identifying relationship similar to that used by Rose (2017, equation 3.4):

$$\begin{aligned} E[\ddot{Y}\ddot{Y}'] &= E[(I - \beta_1 G)^{-1}(\beta_2 \ddot{X} + \beta_3 G\ddot{X} + \ddot{\epsilon})\ (\beta_3 \ddot{X} + \beta_3 G\ddot{X} + \ddot{\epsilon})'(I - \beta_1 G')^{-1}] \\ &= (I - \beta_1 G)^{-1} E[(\beta_2 \ddot{X} + \beta_3 G\ddot{X})(\beta_2 \ddot{X} + \beta_3 G\ddot{X})'](I - \beta_1 G')^{-1} \\ &\quad + (I - \beta_1 G)^{-1} E[\ddot{\epsilon}\ \ddot{\epsilon}'](I - \beta_1 G')^{-1} \qquad (5.2) \end{aligned}$$

[12] This can be generalized to linear-in-level peer effect models by letting each G_{il} be a vector from a network adjacency matrix.

[13] This can easily be generalized to allow multiple exogenous characteristics.

where we have assumed that the G matrix is non-stochastic. As in Liu (2017), the covariance matrix of the $\ddot{X}$'s is identified from the data. If the $\ddot{\epsilon}$'s are i.i.d, we have $E[\ddot{\epsilon}\ \ddot{\epsilon}'] = \sigma_\epsilon^2 I$. In this case expression (5.2) can be used as starting point for estimation as suggested, for instance, by Anselin (1988) for spatial models and by Rose (2017) for peer effect models. This is also the formula behind the ML estimator implemented by Drukker et al. (2013).

Since there is exclusion bias, however, $E[\ddot{\epsilon}\ \ddot{\epsilon}'] \neq \sigma_\epsilon^2 I$. This means that all estimators that ignore this fact are inconsistent when they include pool fixed effects. Formula (4.1) can nonetheless be used to construct the asymptotic covariance matrix of the $\ddot{\epsilon}$'s. To demonstrate, let us arrange all observations 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:

$$E[\ddot{\epsilon}\ \ddot{\epsilon}'] = \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & B & 0 & 0 \\ 0 & 0 & B & 0 \\ 0 & 0 & 0 & B \end{bmatrix} \tag{5.3}$$

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

$$B = \begin{bmatrix} E[\ddot{\epsilon}_1^2] & E[\ddot{\epsilon}_1\ddot{\epsilon}_2] & ... \\ E[\ddot{\epsilon}_2\ddot{\epsilon}_1] & E[\ddot{\epsilon}_2^2] & ... \\ ... & ... & ... \end{bmatrix} \tag{5.4}$$

From equation (A.5) in Appendix A, we know that, for any two individuals i and j in the same selection pool of size L, we have $plim[\ddot{\epsilon}_i\ddot{\epsilon}_j] = \rho\sigma_\epsilon^2$ with $\rho = -\frac{1}{L-1}$ for $i \neq j$. We use this fact to replace, in the estimation, each block matrix B by its probability limit:

$$plimB = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho & ... \\ \rho & 1 & ... \\ ... & ... & ... \end{bmatrix} \equiv \sigma_\epsilon^2 A \tag{5.5}$$

where the asymptotic value of ρ is *known* and need not be estimated.

Equation (5.2), combined with (5.3) and (5.5), provides a characterization of the data generating process that can be used to estimate structural parameters $\beta_{1,}\beta_2, \beta_3$ and σ^2. Our approach to estimation is to rely on the method of moments (MM) to choose the parameter β_1 that provides the best fit to the observed data $E[\ddot{Y}\ddot{Y}']$. The resulting estimator inherits the consistency properties of method of moments estimators. Estimation is achieved using a search algorithm. For each guess $\beta_1^{(n)}$ that the algorithm makes about β_1, we solve for the corresponding values of β_2 and β_3 by calculating $\ddot{Y} - \beta_1^{(n)} G\ddot{Y}$ and regressing it on $\ddot{X}$ and $G\ddot{X}$ to obtain estimates of $\widehat{\beta_2}^{(n)}$ and $\widehat{\beta_3}^{(n)}$. In our data, this is achieved by estimating a regression of the form:

$$\widetilde{y}_{ikl,t+1} = \beta_0 + \beta_2 y_{iklt} + \beta_3 \bar{y}_{-iklt} + \delta_l + \epsilon_{ikl,t+1} \tag{5.6}$$

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 (5.3) to compute the right-hand side of equation (5.2) that corresponds to that particular guess $\beta_1^{(n)}$. This process also yields an estimate of the variance of errors $\widehat{\sigma}_\epsilon^{2(n)}$. Using $\beta_1^{(n)}, \widehat{\beta_2}^{(n)}, \widehat{\beta_3}^{(n)}$ and $\widehat{\sigma}_\epsilon^{2(n)}$ we compute the value of each element of the right hand side of equation (5.2). 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 $\beta_1^{(n)}$. The algorithm then seeks the value of $\beta_1^{(n)}$ that minimizes the distance between this constructed matrix and the data matrix $E[\ddot{Y}\ddot{Y}']$, that is, the lowest sum of squared residuals in equation (5.2). This algorithm is intuitive and reasonably fast.[14]

Inference is conducted using the permutation method described in Section 3 to generate the distribution of the estimates under the null. The p-value for $\widehat{\beta_1}$ is obtained by constructing artifactual samples in which groups are formed at random within selection pools and by simulating the distribution of $\widehat{\beta_1}$ under the null hypothesis of no endogenous peer effects. Estimates for β_2 and β_3 are those given by model (5.6) at the optimal value of $\widehat{\beta_1}$; their standard errors are clustered by selection pool.

5.2 Demonstrating the performance of the estimator

To illustrate the effectiveness of this approach, we estimate model (5.1) on simulated data using this algorithm. The average results from 1000 Monte Carlo replications are shown in Table 5. We

[14] It also resembles, in spirit, the concentrated likelihood function method that Drukker et al. (2013) use for the ML estimator – except that we rely on least squares instead of a likelihood function as objective function. This approach, combined with the fact that we rely on randomization inference (see below), obviates the need for making assumptions about the functional form of the distribution of disturbances.

keep the number of pools and pool size in each sample constant at $N = 50$ and $L = 20$ but we vary K and β_1 across simulation exercises. Pool fixed effects are included throughout.

In Panel A, we report the simulated expected value of the naive $\widehat{\beta_1}^{FE}$ and its p-value obtained by regressing Y_i on G_iY and pool dummies. Results confirm that the naive $\widehat{\beta_1}^{FE}$ and the inference based on it are biased. This bias comes from two sources: reflection and exclusion bias. When β_1 is small, the exclusion bias dominates and the naive $\widehat{\beta_1}^{FE}$ underestimates the true β_1. On average, $\widehat{\beta_1}^{FE}$ is more likely to overestimate the true β_1 when exclusion bias is small, which occurs when L is large. The third row of Panel A shows the proportion of times the simulated naive p-value is smaller or equal to 0.05. In columns 1 and 4 (when $\beta_1 = 0$), this statistic gives the likelihood of making a type II error, that is, the probability of rejecting the null hypothesis when it is true. If the estimator is unbiased then we would expect this statistic to be close to 5%. The actual figures are much higher: around 18.4% when $K = 2$ and 56.9% when $K = 5$, confirming that inference based on the naive model is seriously biased. This will come as a surprise to those who only consider reflection bias: indeed this bias operates as a multiplier – i.e., it only multiplies the true value of β_1 – and, consequently, it does not bias estimates when $\beta_1 = 0$. Based on this belief, researchers would have a high probability of erroneously concluding that negative peer effects are present in columns 1 and 4 when in fact they are absent.

In columns 2-3 and 5-6 of Table 5 (where $\beta_1 > 0$), we see that, when $K = 2$, the null hypothesis of $\beta_1 = 0$ is rejected in 87.8% to 100% of the cases – which is good – but the point estimates are massively over-estimated. When $K = 5$, we see instead that when $\beta_1 = 0.10$, the researcher would, 6.2% of the time, reject the null in favor of $\beta_1 < 0$, i.e., would infer the wrong sign. When $\beta_1 = 0.20$, the naive estimator happens to do better, but its power remains well below 100%. These results further confirm that exclusion bias would often lead a researcher to misinterpret results if only considering the inflation bias caused by reflection.

In Panel B we report the $\hat{\beta_1}^{ML}$ estimates obtained by using the ML estimator obtained using the spreg Stata command (see Drukker et al. 2013). This estimator corrects for reflection bias – but ignores exclusion bias. As anticipated, this estimator 'shrinks' the estimates of β_1 to correct for a multiplier effect – but it does not correct its sign. As a result, it still over-rejects the null when it is true, (columns 1 and 4) and it occasionally rejects the null with the wrong sign (column 5).

In Panel C we report estimates obtained using our algorithm described in the previous section. We first report the MM estimator $\widehat{\beta_1}^{Ref}$ obtained from our algorithm but erroneously assuming that $E[\ddot{\epsilon}\,\ddot{\epsilon}'] = \sigma_\epsilon^2 I$. The $\widehat{\beta_1}^{Ref}$ point estimates are very similar to those obtained with the ML estimator in Panel B. Next we present results from our preferred estimator, namely, the average $\widehat{\beta_1}^{Corr}$ derived from model (5.2) with $E[\ddot{\epsilon}\,\ddot{\epsilon}']$ given by (5.3). The $\widehat{\beta_1}^{Corr}$ is centered around its true value in all cases, albeit with a small downward bias. The next line shows the corrected p-values obtained using the permutation method described earlier. We see that the method yields unbiased inference when $\beta_1 = 0$: p-values are centered on 0.50; and the probility of falsely rejecting the null is around 5%. We also note in columns 2-3 and 5-6 that the estimator has high statistical power when $\beta_1 \neq 0$, the only exception being column 5.

So far we have shown that our MM estimator is consistent and that applying permutation inference to it yields consistent inference. This leaves open the question of whether the estimator is efficient. To address this question, we present in Figure 4 simulations illustrating the power of our MM estimator. The simulated model is of the form $y_{il} = \beta_1 G_{il} Y_l + \delta_l + \epsilon_{il}$ where δ_l is the pool fixed effect. There are no other regressors. We choose a moderate total sample size of 1000 observations, divided into 50 pools of 20 observations each. Within each pool there are three groups of size 2, three groups of size 3, and one group of size 5. We generate 100 samples of 1000 observations for values of β_1 ranging from -0.3 to 0.3 in increments of 0.1, and we estimate $\hat{\beta}_1^{Corr}$ for each of them. We then use permutation inference and compare each estimate to the distribution of $\hat{\beta}_1^{Corr}$ under the null of $\beta_1 = 0$.[15] This yields a p-value for each $\hat{\beta}_1^{Corr}$ estimate. Finally we calculate, for each value of the true β_1 separately, the proportion of the corresponding 100 simulated $\hat{\beta}_1^{Corr}$ estimates that tests different from 0 with a p-value of 5% or better. This provides an approximation of the power of rejecting $\beta_1 = 0$ when the true $\beta_1 \neq 0$. As benchmark, we present the power of an OLS univariate regression with the same sample size and the same standard deviation of the dependent variable and the regressor.[16] The simulations show that $\hat{\beta}_1^{Corr}$ achieves high power at values of β_1 larger than 0.1 in absolute value. They also show that the estimator out-performs the power

[15] This is achieved using 100 permutated samples to generate an approximation of the distribution of $\hat{\beta}_1^{Corr}$ under the null of $\beta_1 = 0$. To save computation time, the same permutated sample is used across the simulations.

[16] Power for the OLS univariate regressions are calculated using a Stata command of the form "`power oneslope 0 -0.3, n(1000) sdy(1.063) sdx(0.793) alpha(0.05)`". To ensure comparability with our MM estimator, the values of `sdy` and `sdx` are set equal to the average of the standard deviations of the pool de-meaned values of the dependent variable and the peer effect variable in the corresponding MM simulations. Because of the magnification effect induced by reflection, these standard deviations are larger for values of β_1 further away from 0.

of a univariate OLS with an equivalent sample size and variance of errors. This provides ample reassurance that our proposed MM estimator works well under our maintained assumption of no within-group correlated effect.

By performing permutation inference within selection pools, our inference method yields robust inference equivalent to a wild bootstrap (e.g., Cameron et al. 2008). But, the method does not accommodate heteroskedastic ϵ errors in the *estimation* of $\hat{\beta}_1^{Corr}$ itself. 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 $E[\ddot{Y}G\ddot{Y}']$ 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 $E[\ddot{\epsilon}G\ddot{\epsilon}'] = 0$ 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.

6 Main empirical results

Having constructed a consistent estimator suitable for our data and having demonstrated that it produces consistent point estimates and inference, we now apply it to model (2.1), which we reproduce here for convenience:

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

Results for golfers using data from Guryan et al. (2009) are presented in Table 6. In this empirical application, the outcome of interest is the golfer's score in the tournament and the coefficient of most interest is the endogenous peer effect in golfer score. 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 provide an additional source of identification – but are potentially affected by what happened in the first round. The first column of the Table presents the naive OLS-FE estimate $\hat{\beta}_1^{FE}$. It suggests the presence of small positive but non-significant peer effects. As expected, based on our simulation results, we observe a noticeable shrinkage (i.e., halving) of the estimated coefficient when we correct for reflection bias with the ML estimator

$\hat{\beta}_1^{ML}$ implemented by Drukker et al. (2013) (column 2)[17]. Inference remains the same, though: no evidence of peer effects. This means that someone expecting reflection bias to magnify the peer effect coefficient would conclude to the absence of peer effects in these data. This is indeed the conclusion reached by Guryan et al. (2009), who do not correct for exclusion bias when estimating peer effects.

Our results presented in column 3 of Table 6 demonstrate, however, that this conclusion is wrong. In column 3, we report the MM estimate $\hat{\beta}_1^{Corr}$ that corrects for both reflection and exclusion bias. We now find a positive endogenous peer effect $\hat{\beta}_1^{Corr}$ that is significant at the 1% level. The magnitude of the coefficient is large: i's performance increases by 5.5% of the average performance of the two golfers in i's group, conditioning on i's own past performance and that of the other two players in i's group. 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 better than predicted by his/her own past performance, the other players in that group also tend to play better than their own past performance predicts. 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. Unsurprisingly, the golfer's past tournament performance is a strong predictor of current performance: $\widehat{\beta}_2$ is large and significant. More importantly, we also see that $\widehat{\beta}_3$ is not significant and, if anything, is negative. This means that being matched with a better or worse set of peers does not affect players' performance in the tournament. Emulation comes purely from play during the tournament, not from who golfers are grouped with.

The right panel of Figure 5 provides a visual illustration of how the distributions of $\hat{\beta}_1^{Corr}$ and $\hat{\beta}_1^{FE}$ compare to each other under the null of $\beta_1 = 0$, i.e., no endogenous peer effects. We see that the distribution of $\hat{\beta}_1^{FE}$ is centered well below 0 while the simulated distribution of $\hat{\beta}_1^{Corr}$ is correctly centered on $\beta_1 = 0$. The Figure also illustrates how, as discussed in Section 3, it is possible to test for the presence of endogenous peer effects by applying randomization inference directly to the OLS-FE estimate. The Figure indeed shows that the randomized-inference p-value of $\hat{\beta}_1^{FE}$ is around 0.008 and thus statistically significant. This approach, however, does not yield

[17] As illustrated in Table 5, our $\widehat{\beta_1}^{Ref}$ MM estimator only correcting for reflection bias but not exclusion bias performs very similarly to the ML estimator $\hat{\beta}_1^{ML}$ implemented by Drukker et al. (2013) and is therefore not presented in Table 6 and Table 7.

a point estimate for β_1.[18] We also see that the simulated estimator $\hat{\beta}_1^{Corr}$ has a smaller variance than $\hat{\beta}_1^{FE}$ under the null. This is because each $\hat{\beta}_1^{FE}$ estimate is magnified by reflection and thus varies more across samples. In the left panel of Figure 5, we perform the same comparison under the null, but this time between $\hat{\beta}_1^{FE}$ and the MM estimator $\hat{\beta}_1^{Ref}$ that accounts for reflection bias but ignores exclusion bias. Under the null of $\beta_1 = 0$, the simulated distribution of $\hat{\beta}_1^{Ref}$ is tighter than the distribution of $\hat{\beta}_1^{FE}$ since reflection bias has been eliminated. But it is no longer centered on 0 due to exclusion bias. This provides further confirmation that non-instrumental methods that correct for reflection bias without also correcting for exclusion bias (e.g., Graham 2009, Rose 2017) lead to incorrect point estimates when they include pool fixed effects.

Table 7 presents similar estimates for the student data of Fafchamps and Mo (2018). Results for $\hat{\beta}_1^{Corr}$ are different to those we obtained for golfers: in this dataset, the point estimate is negative and significant, indicating that endogenous peer effects are negative – suggesting for instance congestion effects in computer usage. A similar conclusion would have been reached by using the naive $\hat{\beta}_1^{FE}$ or the reflection-corrected $\hat{\beta}_1^{ML}$ – a significantly negative peer effect – but the magnitude of this effect would have been greatly overestimated: by more than twice with the reflection-corrected ML estimator, and by nearly five times with the FE estimator. There exist situations, however, in which $\hat{\beta}_1^{FE}$ is negative solely due to exclusion bias. In those cases, a researcher unaware of exclusion bias would be led to the wrong inference, i.e., concluding that there are negative peer effects when there are none. Figure 6 is similar to Figure 5: it shows that correcting for reflection bias alone yields biased estimates under the null, while our estimator $\hat{\beta}_1^{Corr}$ has a distribution correctly centered on the null.

Turning to the other coefficients, we again find that, as expected, the past math score of the student is a strong and significant predictor of their 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. 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

[18] Since $\hat{\beta}_1^{FE}$ itself is negative, a hurried researcher unaware of exclusion bias may erroneously conclude that peer effects are significantly negative, which is of course not the case.

endogenous peer effect is a new result.

7 Concluding remarks

We have estimated endogenous and exogenous peer effects in two datasets with non-overlapping peer groups: golfers in tournaments; and students in computer-assisted learning. In such data, existing instrumental variable methods do not apply. Alternative estimation methods exist that do not require instruments, but they fail to correct for one understudied but important source of bias which we call 'exclusion bias'. We derive a consistent estimator that corrects for this bias. Using this novel method, we find positive peer effects in the first case – consistent with emulation between golfers during the tournament – and negative peer effects in the other – consistent with congestion or wasteful competition for the computer between students. These results differ markedly from existing methods in terms of magnitude, significance, and inference.

We also make a methodological contribution. We first show that, with selection pool fixed effects, a negative correlation in peer outcomes mechanically arises because individuals cannot be their own peers. This 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 with large 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'.

The estimation method presented here is 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 in network data 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 the *presence* of peer effects. As just noted, this approach often yields misleading inference due to the presence of exclusion bias. Like Graham (2008) and Rose (2017), we offer an alternative estimation method that deals with these shortcomings but does not rely on instrumentation – except that, unlike these authors, our estimator is consistent because it corrects for exclusion bias. The method allows the inclusion of selection pool fixed effects, but 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.

The correction we propose for the particular case of non-overlapping peer groups can be applied more generally to ML analysis of network data of any kind, as we demonstrate in Appendix B.4. Exclusion bias is present in such data as well and, depending on the context, can be a source of severe bias there as well. We suspect, for instance, that the ML estimator included in the Stata spreg command implemented by Drukker et al. (2013) can easily be amended to incorporate a correction for exclusion bias. This is left for future work.

References

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

Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*, Springer Studies in Operational Regional Science (Volume 4)

Anselin, L. and Bera, A.K. (1998). “Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics”, in *Handbook of Applied Economic Statistics*, A. ullah and D.E.A. Giles (eds), Marcel Dekker, NY, pp. 237-289.

Athey, S., D. Eckles and G. W. Imbens (2018). "Exact P-values for Network Interference", *Journal of the American Statistical Association,* 113(521):230-40

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

Banerjee, Abhijit, Arun Chandrasekhar, Esther Duflo, and Matthew O. Jackson (2013). “The diffusion of microfinance”, *Science*, 341(6144): 1236498 doi:10.1126/science.1236498

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

Cai, J. and A. Szeidl (2018). "Interfirm Relationships and Business Performance", *Quarterly Journal of Economics*, 133(3):1229-82. August

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90(3):414-27.

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 (2019). "The Impact of College Diversity on Behavior Toward Minorities," *American Economic Journal: Economic Policy*, 11(4): 159-82. November

Corno, Lucia, Eliana La Ferrara, and Justine Burns (2022). “Interaction, Stereotypes, and Performance: Evidence from South Africa”, American Economic Review,, 112(12): 3848-75, December

Davezies, Laurent, Xavier D’Haultfoeuille, and Denis Fougere (2009). “Identification of peer effects using group size variation”, *Econometrics Journal*, 12: 397–413.

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.

Drukker, David M., Ingmar R. Prucha, and Rafal Raciborski (2013). “Maximum likelihood and generalized spatial two-state least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances”, *The Stata Journal,* 13(2): 221-41

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.

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 (2018). "Networks and Manufacturing Firms in Africa: Results from a Randomized Field Experiment", *World Bank Economic Review*, 32(3):656-75

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.

Graham, Bryan S. (2008). “Identifying Social Interactions Through Conditional Variance Restrictions”, *Econometrica*, 76(3): 643-60

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.

Kelejian, H. H. and I. R. Prucha (1998): “A generalized spatial twostage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances,” *The Journal of Real Estate Finance and Economics,* 17,:99–121.

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.

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.

Legros, Patrick and Andrew F. Newman (2007). “Beauty Is a Beast, Frog Is a Prince: Assortative Matching with Nontransferabilities”, *Econometrica,* 75(4): 1073-102

Lee, Lung-Fei, Xiaodong Liu, Eleonora Patacchini, and Yves Zenou (2021). “Who is the Key

Player? A Network Analysis of Juvenile Delinquency”, *Journal of Business and Economic Statistics*, 39(3): 849-57

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.

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.

Rose, Christiern D. (2017). “Identification of peer effects through social networks using variance restrictions”, *Econometrics Journal*, 20(3): S47-S60.

Sacerdote, B. (2001). "Peer Effects with Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*, 116(92): 681-704.

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

Stevenson, M. (2017). "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails", *Review of Economics and Statistics*, 99(5): 824-838.

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.

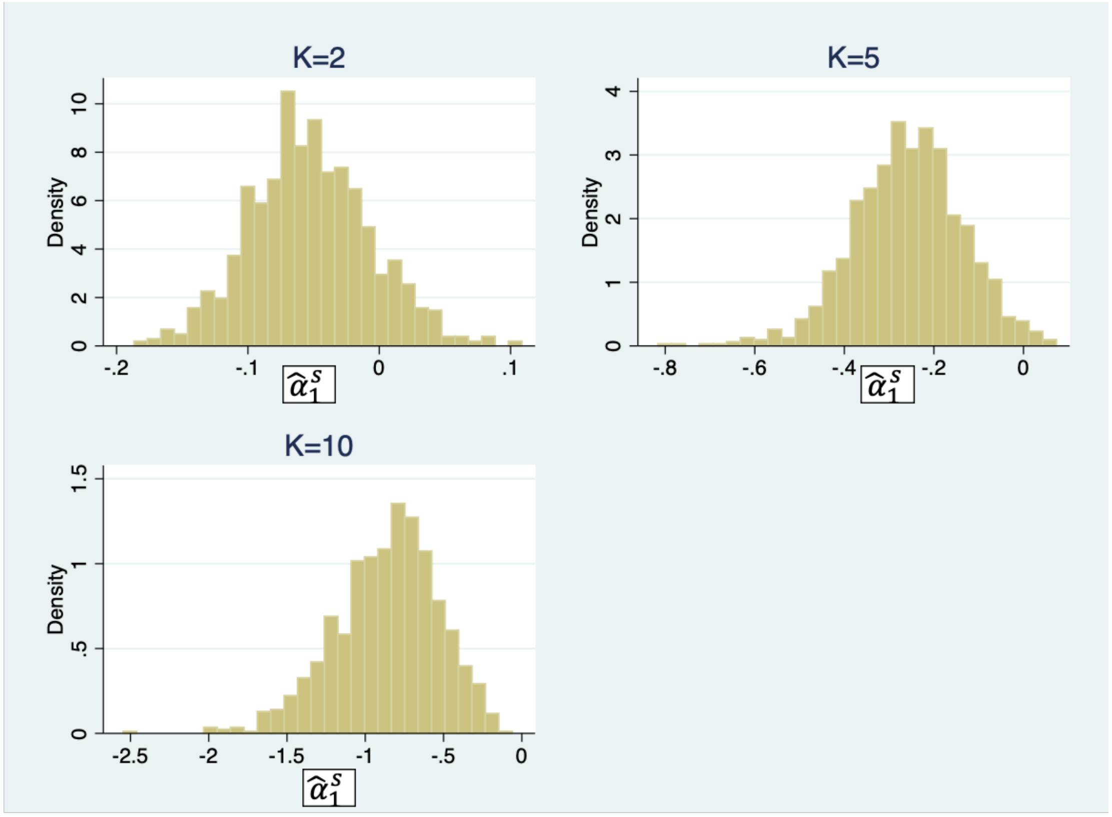
Zimmerman, D. (2003). "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", *The Review of Economics and Statistics*, 85(1): 9-23.

TABLES AND FIGURES

Table 1: An illustration of the permutation method

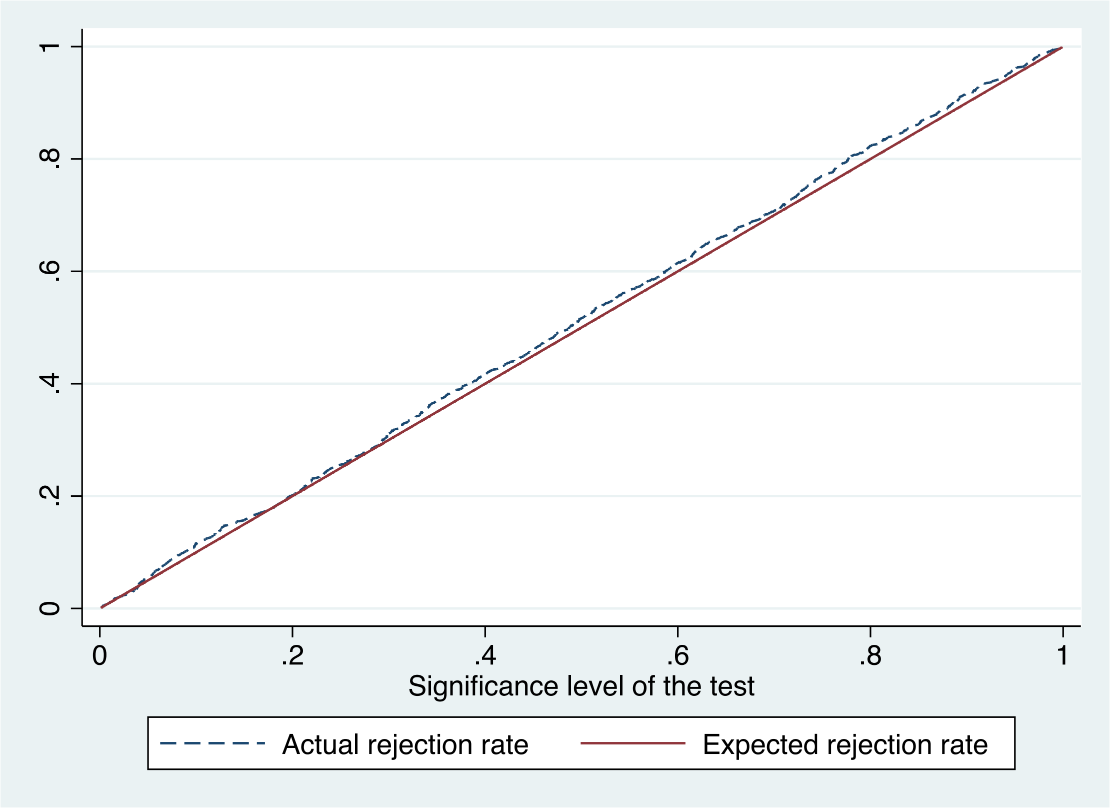
i	k	l	y_{iklt}	$\tilde{y}_{iklt}$
(1)	(2)	(3)	(4)	(5)
1	1	1	y_{111}	y_{211}
2	1	1	y_{211}	y_{521}
3	2	1	y_{321}	y_{111}
4	2	1	y_{421}	y_{321}
5	2	1	y_{521}	y_{421}
6	3	2	y_{632}	y_{842}
7	3	2	y_{732}	y_{632}
8	4	2	y_{842}	y_{942}
9	4	2	y_{942}	y_{1052}
10	5	2	y_{1052}	y_{732}

Figure 1: Histogram of $\hat{\alpha}_1^s$ under the null



Notes: This Figure shows the distribution of simulated $\hat{\alpha}_1^s$ 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 $\hat{\alpha}_1^s$ under the null. Pool fixed effects are included in all regressions.

Figure 2: Performance of the permutation test under the null



Notes: The Figure shows the simulated performance of a permutation test to evaluate whether $\alpha_1 = 0$ 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 2: Empirical applications: Testing for random peer assignment

	Golfer data	Student data
	(1)	(2)
Pool Fixed Effects OLS estimate $\hat{\alpha}_1^{FE}$	-0.126	-0.043
OLS p-value	0.000	0.023
Corrected p-value obtained using permutation method (Krackardt, 1988)	0.540	0.546
Number of observations	2517	2960
Number of selection pools	100	155
Group size	3	2

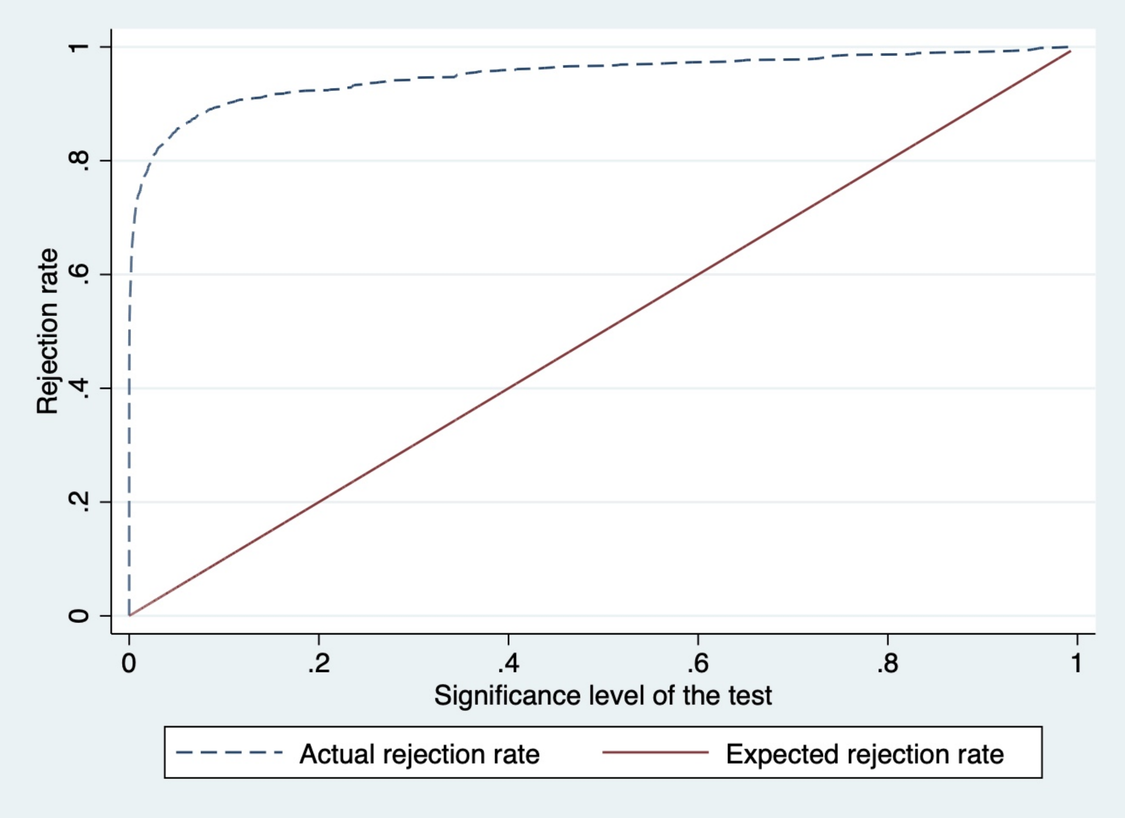
Notes: The golfer data are from Guryan et al. (2009) and the student data are from Fafchamps and Mo (2018). For demonstration purpose, we restrict the golfer sample to the first tournament round, and to a random sub-set of N=100 out of 302 pools, making the overall sample size more comparable to the 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 measure of ability of skill (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 lagged 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. The corrected p-value for the MM estimate is obtained using the permutation method, using 500 iterations.

Table 3: Simulated exclusion bias with random peer assignment

		$L = 20$	$L = 50$	$L = 100$
		(1)	(2)	(3)
$K = 2$	Predicted plim$[\hat{\alpha}_1]$	-0.05	-0.02	-0.01
	Average $\hat{\alpha}_1$	-0.05	-0.02	-0.01
	% of $\hat{\alpha}_1 = 0$ rejected at 1% level	26%	10%	8%
	% of $\hat{\alpha}_1 = 0$ rejected at 5% level	43%	21%	18%
	% of $\hat{\alpha}_1 = 0$ rejected at 10% level	52%	29%	24%
$K = 5$	Predicted plim$[\hat{\alpha}_1]$	-0.25	-0.09	-0.04
	Average $\hat{\alpha}_1$	-0.26	-0.10	-0.04
	% of $\hat{\alpha}_1 = 0$ rejected at 1% level	75%	22%	9%
	% of $\hat{\alpha}_1 = 0$ rejected at 5% level	85%	38%	21%
	% of $\hat{\alpha}_1 = 0$ rejected at 10% level	89%	48%	31%
$K = 10$	Predicted plim$[\hat{\alpha}_1]$	-0.82	-0.22	-0.10
	Average $\hat{\alpha}_1$	-0.86	-0.25	-0.11
	% of $\hat{\alpha}_1 = 0$ rejected at 1% level	97%	42%	17%
	% of $\hat{\alpha}_1 = 0$ rejected at 5% level	99%	58%	27%
	% of $\hat{\alpha}_1 = 0$ 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 $\alpha_1 = 0$. In each simulated sample s, coefficient $\hat{\alpha}_1$ is estimated using fixed effects at the level of the selection pool. The predicted $plim_{N\to\infty}[\hat{\alpha}_1]$ is obtained using Proposition 1. The average $\hat{\alpha}_1$ is the average of $\hat{\alpha}_1$ estimates over all replications. The percentage of rejections is the proportion of replications for which a standard t-test rejects the null that $\alpha_1 = 0$ for different critical levels of the test.

Figure 3: Performance of the OLS estimator under the null



Notes: Figure shows for the OLS estimator 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 4: Simulated exclusion bias with random peer assignment: Different sample sizes

	$N = 2, L = 50$	$N = 4, L = 50$	$N = 10, L = 50$	$N = 20, L = 50$	$N = 40, L = 50$	$N = 80, L = 50$	$N = 120, L = 50$
	(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

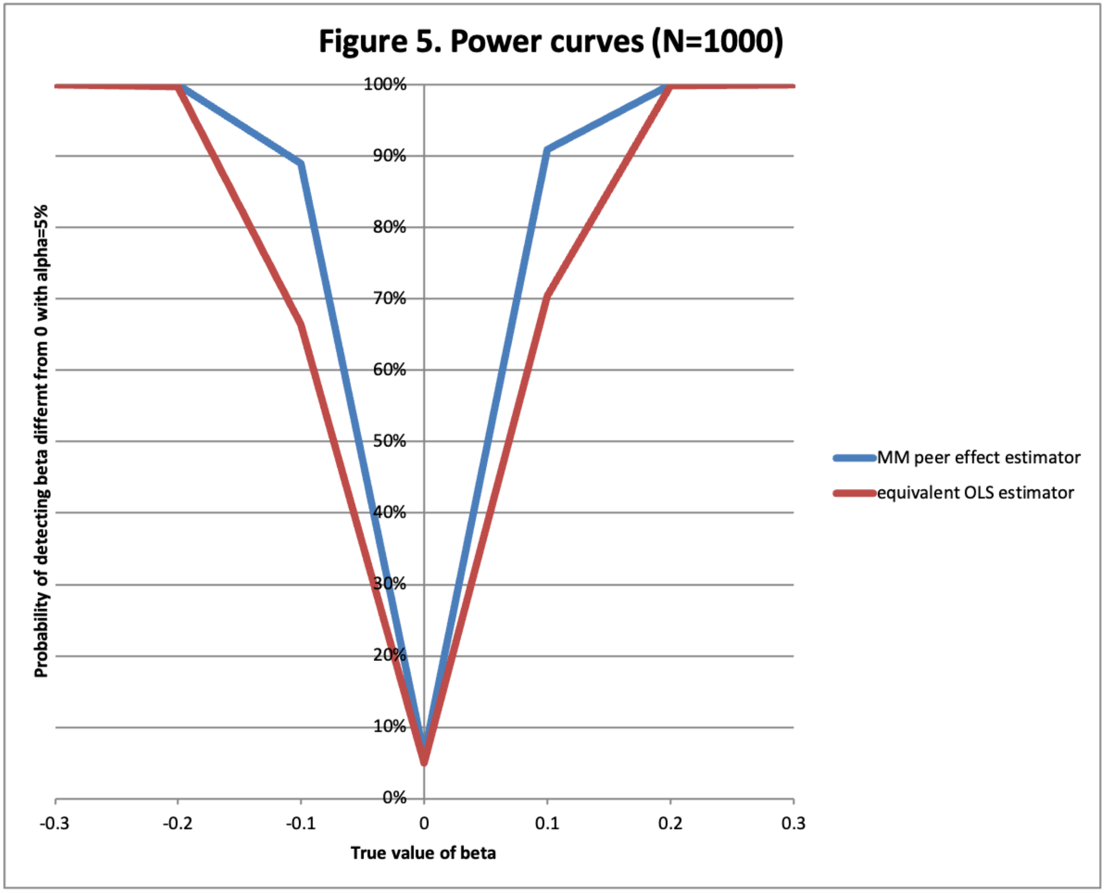
Notes: The Table reports simulation results from 1000 Monte Carlo replications for different values of K and N. Each simulation considers pool size $L = 50$, with Npools and considers observations generated with a true $\alpha_1 = 0$. In each simulated sample s, coefficient $\hat{\alpha}_1^s$ is estimated using fixed effects at the level of the selection pool.

Table 5: Correction reflection and exclusion bias in the estimation of endogenous peer effects - Groups

	$K = 2$			$K = 5$		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$
Panel A						
$\hat{\beta_1}^{FE}$	-0.05	0.15	0.34	-0.27	-0.04	0.18
Mean of p-value of $\hat{\beta_1}^{FE}$	0.32	0.03	0.00	0.11	0.35	0.08
Proportion of p-value $\leq$ 0.05	18.4%	87.8%	100.0%	56.9%	6.2%	51.5%
Panel B						
$\hat{\beta_1}^{ML}$ - Drukker et al (2013)	-0.03	0.07	0.17	-0.13	-0.02	0.09
Mean of p-value of $\hat{\beta_1}^{ML}$	0.30	0.02	0.00	0.08	0.43	0.14
Proportion of p-value $\leq$ 0.05	21.2%	90.6%	100.0%	70.6%	11.0%	58.0%
Panel C						
$\hat{\beta_1}^{Ref}$ - 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 $\leq$ 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 number of selection pools constant at N=50 and L=20, but we vary β_1 and group size K. Pool fixed effects are included throughout. Row 1 and row 2 in Panel A report, respectively, the naive $\hat{\beta_1}^{FE}$ 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 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. Panel B presents the equivalent results of a ML estimation following the method described in Drukker et al (2013). The first row in Panel C 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 second 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 for $\hat{\beta_1}^{Corr}$ obtained using the permutation method, as well as a statistic on the distribution of simulated p- values computed in the same way as in Panel A.

Figure 4: Power curves



Notes: This Figure illustrates the power of the MM estimator relative to the OLS estimator. The simulated model is of the form $y_{il} = \beta G_{il} Y_l + \delta_l + \epsilon_{il}$ where δ_l is a pool fixed effect. There are no other regressors. We choose a moderate total sample size of 1000 observations, divided into 50 pools of 20 observations each. Within each pool there are three groups of size 2, three groups of size 3, and one group of size 5. We generate 100 samples of 1000 observations for values of β_1 ranging from -0.3 to 0.3 in increments of 0.1, and we estimate $\hat{\beta}_1^{Corr}$ for each of them. We then use permutation inference and compare each estimate to the distribution of $\hat{\beta}_1^{Corr}$ under the null of $\beta_1 = 0$. This yields a p-value for each $\hat{\beta}_1^{Corr}$ estimate. Finally we calculate, for each value of the true β_1 separately, the proportion of the corresponding 100 simulated $\hat{\beta}_1^{Corr}$ estimates that tests different from 0 with a p-value of 5% or better. This provides an approximation of the power of rejecting $\beta_1 = 0$ when the true $\beta_1 \neq 0$. As benchmark, we present the power of an OLS univariate regression with the same sample size and the same standard deviation of the dependent variable and the regressor.

Table 6: Empirical application: Golfer data (Guryan et al, 2009)

	FE	ML (Drukker et al., 2013)	MM - Correction reflection + exclusion bias
	(1)	(2)	(3)
Endogenous peer effect β_1	0.022	0.011	0.055***
	(0.439)	(0.575)	(0.008)
Lagged own effect (β_2)	0.480***	0.480***	0.481***
	(0.000)	(0.000)	(0.000)
Exogenous peer effect (β_3)	-0.061	-0.056	-0.077
	(0.598)	(0.620)	(0.559)
Sample size	2517	2517	2517
Number of selection pools	100	100	100
Group size	3	3	3

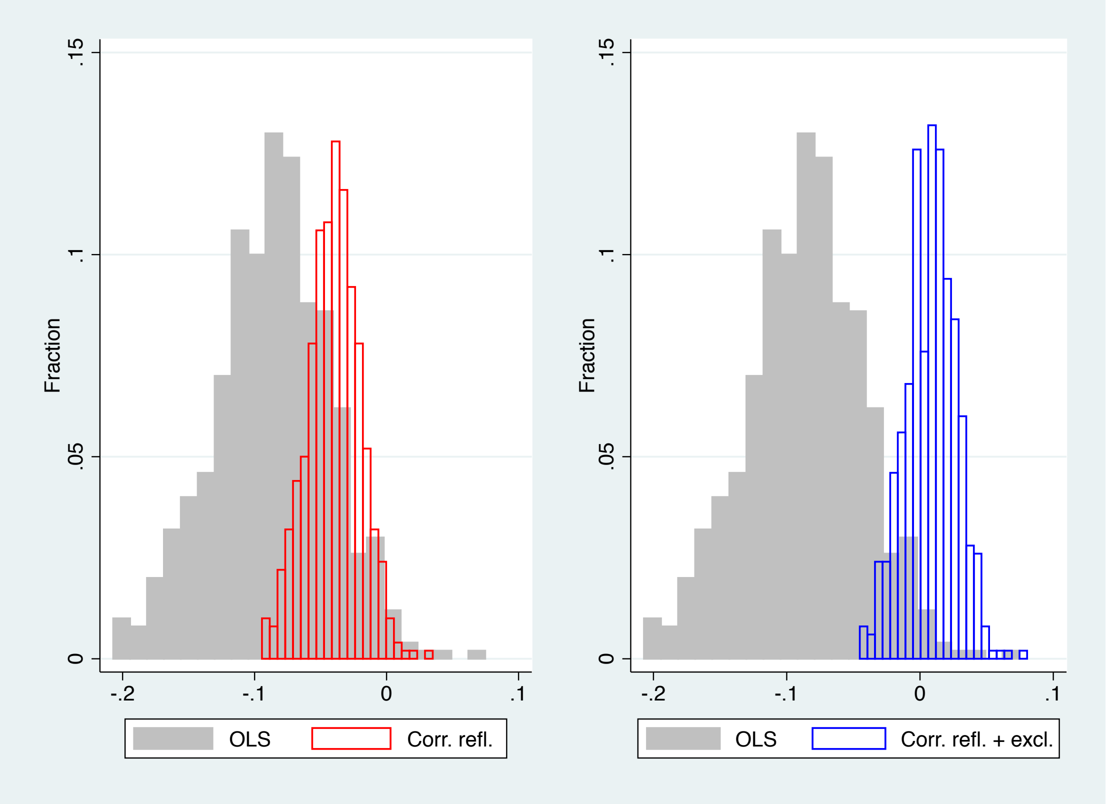
Notes: The golfer data are from Guryan et al (2009). For demonstration purpose, we restrict the golfer sample to the first tournament round, and to a random sub-set of N=100 out of 302 pools, making the overall sample size more comparable to the student application. We also focus on groups of size 3 ($K = 3$) 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 dependent variable is the golf player's score and the lagged own effect is a measure of past performance based on lagged test scores. The pool is the qualification category to which each player is assigned within each tournament. All regressions include pool fixed effects. p-values are shown in pharentheses. The p-value for the MM estimates presented in columns (3) is obtained using randomization inference with 500 iterations. The MM estimator that only corrects for reflection bias and not exclusion bias is not shown here because it essentially identical to the ML estimator reported in column (2).

Table 7: Empirical application: Student data (Fafchamps and Mo, 2018)

	FE	ML (Drukker et al., 2013)	MM - Correction reflection + exclusion bias
	(1)	(2)	(3)
Endogenous peer effect β_1	-0.113***	-0.056***	-0.023**
	(0.000)	(0.000)	(0.020)
Lagged own effect (β_2)	0.478***	0.477***	0.477***
	(0.000)	(0.000)	(0.000)
Exogenous peer effect (β_3)	0.069***	0.042**	0.027*
	(0.000)	(0.013)	(0.081)
Sample size	2960	2960	2960
Number of selection pools	155	155	155
Group size	2	2	2

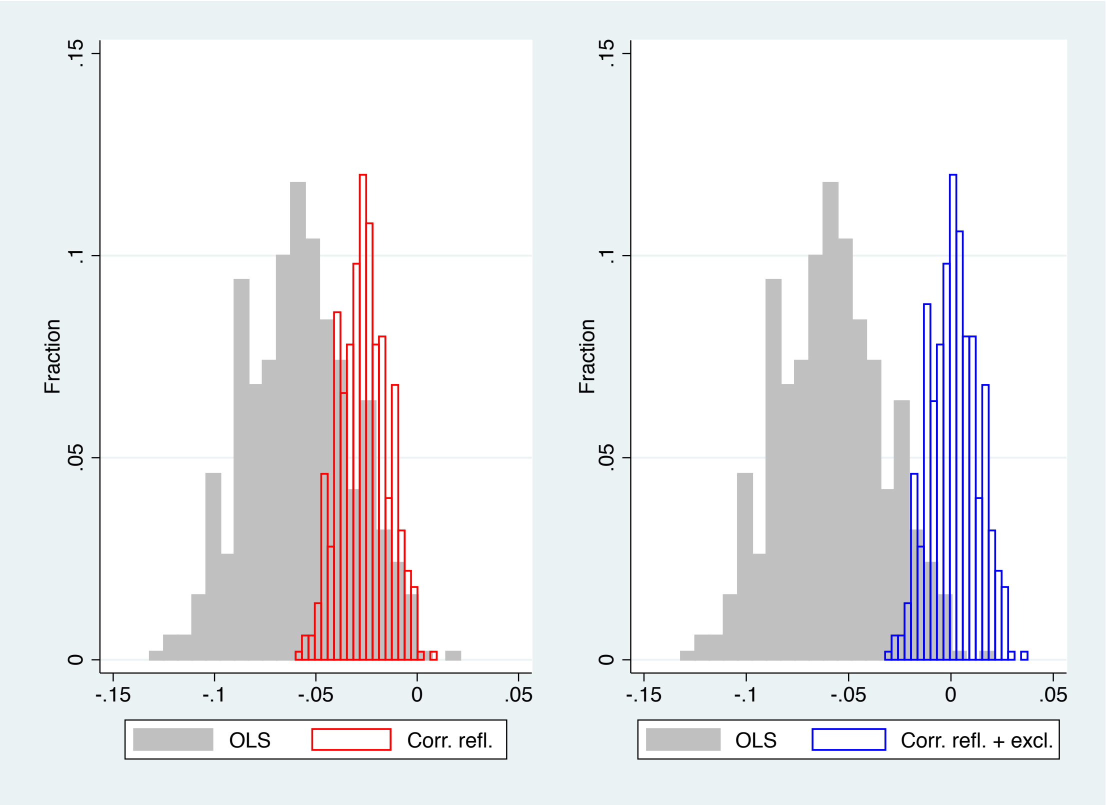
Notes: The student data are from Fafchamps and Mo (2018). We drop a few observations for which we observed inconsistencies in the indication of peers within a pair (16%). The dependent variable of interest is the math score of the students. The pool is the classroom. All regressions include pool fixed effects. p-values are shown in pharentheses. The p-value for the MM estimates presented in columns (3) is obtained using randomization inference with 500 iterations. The MM estimator that only corrects for reflection bias and not exclusion bias is not shown here because it essentially identical to the ML estimator reported in column (2).

Figure 5: Simulated $\hat{\beta}_1^{FE}$, $\hat{\beta}_1^{Ref}$, 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 $\hat{\beta}_1^{FE}$ 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 $\hat{\beta}_1^{FE}$, $\hat{\beta}_1^{Ref}$, 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 $\hat{\beta}_1^{FE}$ 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.

ONLINE APPENDIX

A Intuition behind the methodology used in the paper

A.1 What is the source of exclusion bias?

Exclusion bias is a general phenomenon that is present in all data. For most estimation, it does not matter. But in autocorrelated regression models, it does. We illustrate the intuition with a series of simple examples.

We posit that an i.i.d. data generation process y with mean μ and variance s^2 produces samples of y_i observations of size N. We denote the mean of y_i in sample n as y_n and the variance as s_y^2. We are interested in the sample correlation between any two observations y_i and y_j in sample n. We claim that, on average across sample realizations, the *sample* correlation between y_i and y_j is not 0 even though they are independently distributed. This arises from the definition of sample correlation. It is similar in nature to the Nickel (1982) bias identified in time-series data.

To illustrate with the simplest example, let $N = 2$. In this case $y_n = \frac{y_1+y_2}{2}$, the sample variance $s_y^2 = \frac{(y_1-y_n)^2+(y_2-y_n)^2}{2}$ and the sample autocorrelation $r_n = \frac{2(y_1-y_n)(y_2-y_n)}{(y_1-y_n)^2+(y_2-y_n)^2}$. By the definition of y_n we have $y_1 - y_n = -(y_2 - y_n)$. Let $d = y_1 - y_n$. Then:

$$r_n = \frac{-2d^2}{2d^2} = -1$$

In other words, y_i and y_j have a non-zero *sample* correlation even though they are two realizations of an i.i.d. process. This result generalizes to samples of any size that are divided into pools of size $L = 2$ and pool-level fixed effects are included. This is because, with pool fixed effects, each pool has a distinct mean y_n and the formula above applies within each pool.

The idea can be generalized to selection pools of any size. To see this, we first note that the **sample** correlation between any observation y_i and the average of the remaining pool observations $\overline{y}_{-i}$ is negative. This results derives from the definition of the sample average of the pool y_n: if $y_i > y_n$, by construction $\overline{y}_{-i} < y_n$ – and vice versa. Hence if we select one observation $y_{j\neq i}$ from sample n, then the *expected* sample correlation between y_i and y_j will be negative. This is because, on average, $y_j < y_n$ if $y_i > y_n$ and vice versa.

This can be shown formally as follows. By the definition of a sample mean, we have:

$$y_n = \frac{(L-1)\overline{y}_{-i} + y_i}{L}$$

Let $d = y_i - y_n$. Simple algebra yields $\overline{y}_{-i} = y_n - \frac{d}{L-1}$. It follows that the covariance between y_i and $\overline{y}_{-i}$ is simply the covariance between d and $-\frac{d}{L-1}$, and the correlation between them is:

$$r_n | y_n = \frac{cov(y_i, \overline{y}_{-i})}{sd(y_i) sd(\overline{y}_{-i})} = \frac{-s_y^2/(L-1)}{s_y^2/(L-1)} = -1$$

where covariance and variance are measured relative to pool mean y_n. This intuition generalizes to samples of any size that are divided into pools of size L and pool-level fixed effects are included. The above algebra also demonstrates that the covariance between y_i and $\overline{y}_{-i}$ falls with pool size L or, more generally, with the size of the sample if pool fixed effects are not included.

As the above examples illustrate, the negative sample correlation between sample observations within a selection pool arises mechanically because observation y_i is omitted or 'excluded' from the sample mean of the remaining observations in the pool. If it were not, this negative sample correlation would disappear. It is for this reason that we call this negative correlation an exclusion bias.

In the paper we generalize these examples to situations in which groups are formed within each selection pool and we calculate the *plim* of the within-group covariance between observations.

A.2 What is the source of reflection bias?

To illustrate the nature of reflection bias, we use a simple example with the size of the group $K = 2$. In this setting, it is straightforward to obtain an algebraic formula for the reflection bias. We start by assuming away exclusion bias to conceptually distinguish the reflection bias from exclusion bias later on. For simplicity, we assume that errors are homoskedastic and independently distributed. The latter assumption is far from innocuous since it assumes away the presence of what Manski (1993) calls correlated effects, that is, correlated errors between individuals belonging to the same peer group.[19] With this assumption, correlation in outcomes between members of the same peer

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

group constitutes evidence of endogenous peer effects.

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 $E[\epsilon^2] = \sigma^2_\epsilon$. We estimate:

$$y_1 = a + by_2 + v_1 \tag{A.1}$$

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 C.5]: If* $E[\epsilon_1 \epsilon_2] = 0$ *(i.e., there are no correlated effects), the bias in model (A.1) is given by:*

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

An immediate corollary is that $plim_{N\to\infty}[\widehat{b}^{OLS}] = 0$ iff $\beta_1 = 0$, implying that the existence of peer effects can be investigated by testing whether $b = 0$. Moreover, formula (A.2) can be solved to recover an estimate of β_1 from the naive $\widehat{b}$, yielding:[20]

$$\widehat{\beta_1}^{Ref} = \frac{1-\sqrt{1-\widehat{b}^2}}{\widehat{b}} \tag{A.3}$$

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

[20] 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.

A.3 How do reflection bias and exclusion bias combine?

Exclusion bias arises when selection pool fixed effects are added to model *(A.1)* 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 rewrite in deviation from the pool mean to eliminate the fixed effect δ_l:

$$\ddot{y}_1 = a + b\ddot{y}_2 + \ddot{\epsilon}_1 \tag{A.4}$$

where the notation $\ddot{z}_{ikl} \equiv z - \bar{z}_l$ where $\bar{z}_l$ is the sample mean of z in pool l. By applying Proposition 1, we have:

$$\rho \equiv plim_{N\to\infty} SampleCorr(\ddot{\epsilon}_{ikl}\ddot{\epsilon}_{jkl}) = -\frac{1}{L-1} \tag{A.5}$$

Using this result, we can show that the size of the combined reflection and exclusion bias is as follows:

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

$$plim_{N\to\infty}[\widehat{b}^{FE}] = \frac{2\beta_1 + (1+\beta_1^2)\rho}{1+\beta_1^2+2\beta_1\rho} \tag{A.6}$$

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

We can take roots of formula (A.6) to obtain a consistent estimate $\widehat{\beta_1}^{Corr}$ as: [21]

$$\widehat{\beta_1}^{Corr} = \frac{1 - \widehat{b}\rho - \sqrt{1 + \widehat{b}^2\rho^2 - \widehat{b}^2 - \rho^2}}{\widehat{b} - \rho} \tag{A.7}$$

[21] 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.

Table A.1: Bias in the estimation of endogenous peer effects - $K = 2$

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

(1)	(2)	(3)	(4)
True β_1	Predicted	Monte Carlo average	Monte Carlo averrage
	plim($\hat{b}^{FE}$)	of $\hat{b}^{FE}$	of $\hat{b}^{Corr}$
0.00	-0.06	-0.06	0.00
0.01	-0.04	-0.04	0.01
0.02	-0.02	-0.02	0.02
0.03	0.01	0.01	0.03
0.04	0.03	0.03	0.04
0.05	0.05	0.05	0.05
0.06	0.07	0.07	0.06
0.07	0.09	0.09	0.07
0.08	0.11	0.11	0.08
0.09	0.12	0.12	0.09
0.10	0.14	0.14	0.10

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. Column (4) shows the average of the corrected $\hat{b}$ over the same Monte Carlo replications.

We present in Table A.1 calculations based on formulas (A.6) and (A.7) 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 $K = 2$.[22] Column 1 presents the true β_1 in the data generation process. Column 2 shows the plim of $\widehat{b}^{FE}$ as predicted using our formula (A.6) and column 3 shows the simulated value of the same. Column 4 presents the consistent estimate obtained using formula (A.7). Comparison of columns 2 and 3 in the Table shows clearly that formula (A.6) works very well in predicting the magnitude of the estimation bias. Moreover, we observe that, when the true 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:

[22] 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 $N \times L = 10,000$.

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 10.

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.

Finally, column 4 in Table A.1 illustrates how in this simple case where formula $K = 2$ the estimator derived using formula (A.7) correctly estimates β_1.

A.4 Empirical example

Given that $K = 2$ in the student data in Fafchamps and Mo (2018) - described in Section 2 in the paper - we can use formulas (A.3), (A.6) and (A.7) to obtain exact predictions about the *plim* of $\hat{b}_1^{FE}$, $\hat{b}_1^{Ref}$ and $\hat{b}_1^{Corr}$ under the null in this empirical application. These predictions are shown in Table (A.2) and compared to the means of the simulated distributions of $\hat{\beta}_1^{FE}$, $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{Corr}$ shown in Figure 6 in the main body of this paper. As predicted by (A.6) $\hat{\beta}_1^{FE}$ is centered around -0.059 (considering an average pool size of 18 in this dataset) instead of being centered around the true $\beta_1 = 0$. Under the null, formula (A.3), predicts $\hat{\beta}_1^{Ref}$ to be centered around -0.029, which is close to the average of -0.026 obtained by the simulations shown in Figure 6 of the paper. Similarly, by applying formula (A.7), we expect $\hat{\beta}_1^{Corr}$ to be centered on zero. The simulation average of $\hat{\beta}_1^{Corr}$ is 0.001. Notwithstanding small differences due to Monte Carlo approximation error, the simulation results are strikingly similar to the values predicted by our formulas.

Table A.2: Mean $\hat{\beta}_1^{FE}$, $\hat{\beta}_1^{Ref}$, and $\hat{\beta}_1^{Corr}$ under $H_0 : \beta_1 = 0$ - Student data

	$\hat{\beta}_1^{FE}$		$\hat{\beta}_1^{Ref}$		$\hat{\beta}_1^{Corr}$	
	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 $\hat{\beta}_1^{FE}$, $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{Corr}$, to the exact predictions made by formulas (A.3), (A.6) and (A.7)about the plim of $\hat{\beta}_1^{FE}$, $\hat{\beta}_1^{Ref}$ and $\hat{\beta}_1^{Corr}$ under the null of no endogenous peer effects.

A.5 Why can't we allow group-level correlated effects in our model? Or can we?

Drukker and Prucha (2013) have developed a Stata command spreg that allows for spatial aucorrelation and correlated effects shared by nearby observations. The reason why the two are separately identified is because spatial autocorrelation spreads through the entire data while correlated effects are only shared locally between a group of observations and do not, by themselves, spread outside that group.

To illustrate with a simple example, imagine that the data are placed at regular intervals on a line, and calculate the sample autocorrelogram. This graph shows the sample correlation between all pairs of observations that are distance 1 from each other, then the sample correlation between all pairs that are distance 2 from each other, and so on. If the underlying data generation process only includes spatial autocorrelation, the spatial autocorrelogram has the usual declining exponential shape. In contrast, if the DGP only includes local correlated effects, the spatial autocorrelogram has a spike at distance 1 and zero otherwise. It is this difference in spatial correlation that allows spreg to estimate both effects. This logic extends to network data, in which case distance is the network distance between two observations. Autocorrelated effects spread through each network component while correlated effects remain local.

When the network data takes the form of non-overlapping groups, peer effects remain confined within that group – which forms its own component. This means that the network autocorrelogram can only be estimated for network distance 1, i.e., members of the same group. It follows that, in this case, network autocorrelation (i.e., endogenous peer effects) and correlated effects (i.e., group-level random effects) cannot be distinguished from each other since they both generate a distance 1 correlation and thus are observationally equivalent.

This reasoning also applies to IV approaches to network autocorrelation that rely on friends-of-friends for identification (e.g., Bramoulle et al. 2009; Lee et al. 2021): when peer groups are non-overlapping, there are no friends-of-friends and thus no instruments. The estimation approach we propose in the paper can, however, be extended to network data with overlapping peer groups, in which case both network autocorrelation and correlated effects can, in principle, be separately identified, even without instruments. We believe that the spreg command of Drukker and Prucha could similarly be modified to incorporate exclusion bias, a point on which we will be communicating shortly with the authors.

B Extensions

B.1 A variable transformation to address exclusion bias in tests of random peer assignment

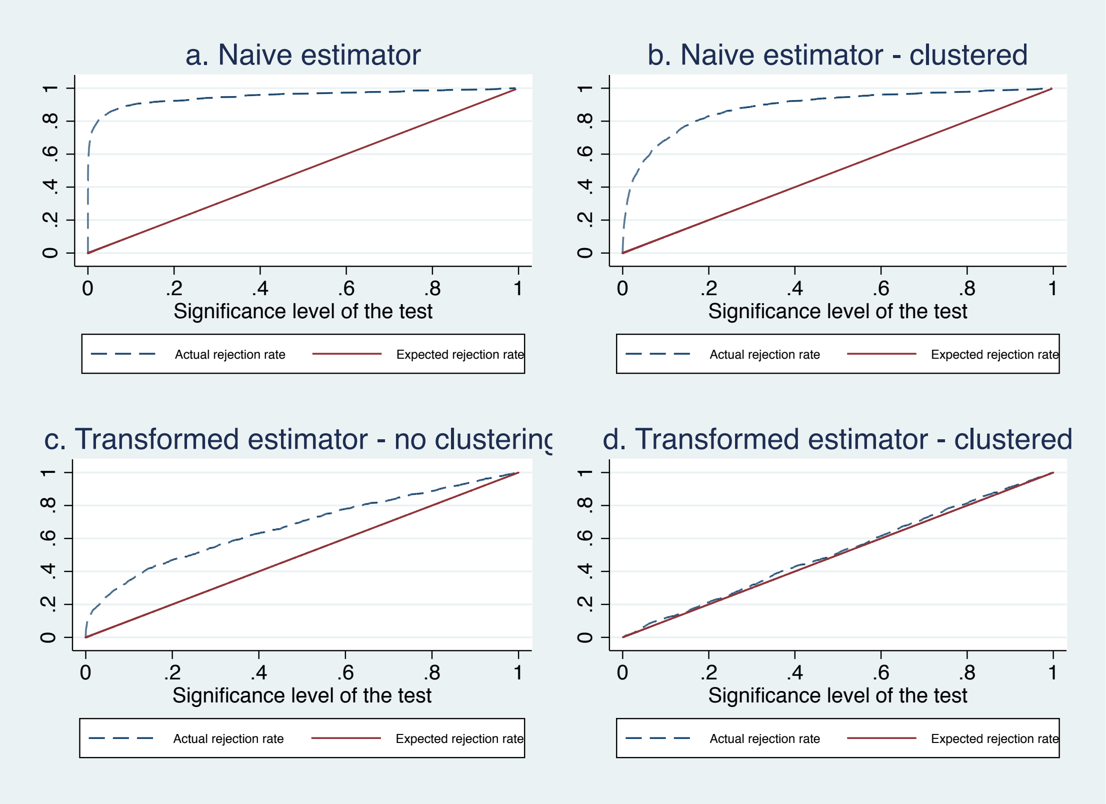
One alternative way to circumvent exclusion bias in standard tests of random peer assignment is to net out the asymptotic exclusion bias using the results from Proposition 1. Specifically, we can use formula (4.1) – or its extension to cases of varying group and pool sizes that is provided in Proposition 2 – to transform the dependent variable in model (3.1) so as to obtain a consistent point estimate of the true β_1 under the null. To this effect, we apply OLS to estimate:

$$\widetilde{y}_{iklt} = \alpha_1 \bar{y}_{-iklt} + \delta_l + \epsilon_{iklt} \tag{B.1}$$

where $\widetilde{y}_{iklt} \equiv y_{iklt} - \rho\bar{y}_{-iklt}$ with $\rho \equiv plim_{N\to\infty}[\hat{\alpha}_1^{FE}]$ given by formula (4.1).[23] Random peer assignment is verified by testing whether $\widehat{\alpha}_1^{FE} = 0$ in model (B.1) using OLS standard errors clustered at the pool level. As illustrated by simulation results presented in the bottom right panel of Figure 1, only when standard errors are clustered by selection pool does the method yield correct inference. We should point out that regression model (B.1) does not yield a consistent estimate of α_1 when the true $\alpha_1 \neq 0$ – more about this in Section 5.

[23] Under the null of $\alpha_1 = 0$, this transformed model is obtained as follows: $x_{iklt} = \left(\alpha_1 + plim_{N\to\infty}[\hat{\alpha}_1^{FE}]\right)\bar{x}_{-iklt} + \delta_l + \epsilon_{iklt} \Leftrightarrow x_{iklt} - plim_{N\to\infty}[\hat{\alpha}_1^{FE}]\bar{x}_{iklt} = \alpha_1\bar{x}_{-iklt} + \delta_l + \epsilon_{iklt}$. It immediately follows that $plim_{N\to\infty}[\tilde{\alpha}_1^{FE}] = \alpha_1$ where $\tilde{\alpha}_1^{FE}$ denotes the estimate obtained from estimating (B.1).

Figure A.1: 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 $\alpha_1 = 0$ 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.

If the model contains regressors w_{iklt} other than those shown in equation (B.1), these regressors first need to be partialled out. In practice, this means doing the following. First, express y_{iklt} and $\bar{y}_{-iklt}$ in deviation from their selection pool mean, i.e., let $\check{y}_{iklt} \equiv y_{iklt} - \frac{1}{L_k} \sum_{jk \in l} y_{jklt}$ and $\check{\bar{y}}_{-iklt} \equiv \bar{y}_{-iklt} - \frac{1}{L_k} \sum_{jk \in l} \bar{y}_{-jklt}$. Do the same for the other regressors, i.e., let $\check{w}_{iklt} = w_{iklt} - \frac{1}{L_k} \sum_{jk \in l} w_{jklt}$. Second, regress the demeaned $\check{y}_{iklt}$ on $\check{w}_{iklt}$ and keep the residuals, which we denote as $\hat{u}_{iklt}$. Similarly regress $\check{\bar{y}}_{-iklt}$ on $\check{w}_{iklt}$ and keep the residuals, which we denote as $\hat{v}_{-iklt}$. We then apply model (B.1) to the residuals, i.e., we construct $\tilde{\hat{u}}_{iklt} \equiv \hat{u}_{iklt} - \rho \hat{v}_{-iklt}$ as above and we regress $\tilde{\hat{u}}_{iklt}$ on $\hat{v}_{-iklt}$. This yields the correct test for random peer assignment in the presence of additional regressors.

The above method works when formula (4.1) can be calculated, that is, when peer assignment is to mutually exclusive groups. It does not apply to peer assignment to partially overlapping groups, or to a position in a network. In such cases randomization inference can be used instead (e.g., Fisher 1925).

B.2 Avoiding exclusion bias

B.2.1 Exogenous peer effects

When estimating exogenous peer effects, it is possible to eliminate the exclusion bias by using control variables. To illustrate, we use the peer structure used in 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:

$$y_{ikl,t+1} = \beta_0 + \beta_1 \bar{y}_{-iklt} + \delta_l + \epsilon_{ikl,t+1} \tag{B.2}$$

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 (B.2).

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 $\bar{x}_{-iklt}$ on i's subsequent outcome $y_{ikl,t+1}$ and includes pool fixed effects.[24] Here too the pre-existing characteristics of peers are negatively correlated with i's pre-existing characteristic x_{iklt}. Hence if the researcher fails to control for x_{iklt} and x_{iklt} is positively correlated with $y_{ikl,t+1}$, then estimating a model of the form:

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

will result in a negative exclusion bias.[25] This bias can be corrected by including x_{iklt} as control, as done for instance in Bayer et al. (2009):

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

If the researcher does not have data on y_{iklt} or x_{iklt}, 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_{iklt}, 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).

B.2.2 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

[24] 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.

[25] If x_{iklt} 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.

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:

$$\bar{y}_{-ikl} = \pi_0 + \pi_1 \bar{z}_{-ikl} + \pi_2 z_{ikl} + \delta_l + v_{ikl} \tag{B.3}$$

$$y_{ikl} = \beta_0 + \beta_1 \hat{\bar{y}}_{-ikl} + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \tag{B.4}$$

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.[26]

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). The removal of the negative exclusion bias by instrumentation 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 (B.3). 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.

[26] 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:

$$\begin{aligned} y_{ikl} &= \beta_0 + \beta_1 \hat{\bar{y}}_{-ikl} + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \\ &= \beta_0 + \beta_1(\hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{-ikl} + \hat{\pi}_2 z_{ikl} + \hat{\delta}_l) + \beta_2 z_{ikl} + \delta_l + \epsilon_{ikl} \end{aligned} \tag{B.5}$$

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 $\bar{z}_{-ikl} = \frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} z_{js}\right] - z_{ikl}}{L-1} + \tilde{u}_{ikl}$, where $\widetilde{u}_{ikl} \equiv \bar{z}_{-ikl} - \bar{z}_{-il}$. Since equation (B.5) controls for z_{ikl} directly, this mechanical relationship is prevented from generating an exclusion bias.

B.3 Application to time series autoregressive models

The methodological approach proposed in this paper can 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:

$$x_{it} = \beta_1 x_{it-1} + \delta_i + \epsilon_{it} \tag{B.6}$$

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 (B.6) satisfy:*

$$plim_{N\to\infty}(\hat{\beta_1^{FE}}) = -\frac{1}{T-1} = \rho \tag{B.7}$$

See Appendix C.7 for a proof. Interestingly, the limit given by formula (B.7) 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 (A.5). Formula (B.7) 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.[27]

The good news is that the different approaches proposed here also work for model (B.6). 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 (B.1). The MM estimation model (5.2) 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

[27] See Nickel (1981) and Arellano (2003) for simulations of the bias when $\beta_1 \neq 0$. As an aside, there seems to be a sign error in equation (13) of Nickel's paper: the last term should have a minus sign instead of a plus sign. If this error and its impact of subsequent equation (16) are corrected, the formula for the Nickel bias when $\rho = 0$ is identical to our equation (B.7), except that the number of time periods T in Nickel (1981) is equal to $T - 1$ in our notation.

applies as well. While this approach allows the estimation of β_1 in model (B.6) 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.

B.4 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 5.1 and assume that individuals in selection pool l are randomly assigned peers within that pool. The only difference with Section 5.1 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 (5.2) remains the same. Formally let $g_{ijl} = 1$ if i and j in selection pool l are peers, and 0 otherwise. We follow convention and set $g_{ii} = 0$ always. The network matrix in pool 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.[28] The average outcome of i's peers can then be written as $\widehat{G}_{il}Y_l$ where Y_l is the vector of all outcomes in selection pool l. The peer effect

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

model that we aim to estimate is:

$$Y_{il} = \beta \widehat{G}_{il} Y_l + \gamma X_{il} + \delta \widehat{G}_{il} X_l + \lambda_l + \epsilon_{il} \tag{B.8}$$

Let's 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. After expressing Y and X in deviation from their pool mean to eliminate λ_l, the linear-in-means network autoregressive model can thus be written in matrix form as:

$$\ddot{Y} = \beta \widehat{G} \ddot{Y} + \gamma \ddot{X} + \delta \widehat{G} \ddot{X} + \ddot{\epsilon} \tag{B.9}$$

As in the previous section, equation (5.2) combined with (A.5), (5.3) and (5.5) can be used to estimate structural parameters β, γ, δ and σ^2. It is intuitively clear that exclusion bias affects model (B.8) as well: individual i is still excluded from the selection pool of its own peers and, in the presence of selection pool fixed effects, 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 $(I - \beta\widehat{G})^{-1}$ in expression (5.2) 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 A.3. We generate each adjacency matrix $\widehat{G}_l$ 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 A.3 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.

Table A.3: 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	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$	$\beta_1 = 0.00$	$\beta_1 = 0.10$	$\beta_1 = 0.20$
Panel A						
$\hat{\beta}_1^{FE}$	-0.09	0.08	0.25	-0.26	-0.09	0.10
Mean of p-value of $\hat{\beta}_1^{FE}$	0.18	0.18	0.00	0.03	0.32	0.26
Proportion of p-value $\leq$ 0.05	51.1%	41.7%	99.9%	88.8%	27.9%	36.3%
Panel B						
$\hat{\beta}_1^{Corr}$ - 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 $\leq$ 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. Pool fixed effects are included throughout. Row 1 and row 2 in Panel A report, respectively, the naive $\hat{\beta}_1{}^{FE}$ 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 first row in Panel B presents the average of $\hat{\beta}_1{}^{Corr}$ correcting for reflection bias and exclusion bias. 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 Panel A).

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 A.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 and was introduced by 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 (B.8) 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 A.3. we compare the p-values obtained from the naive model and the permutation approach applied to model (B.8). We find that the performance of the estimation method in the network case is comparable to what it was in the peer group case.

C Proofs of propositions

The notation is as follows. In a sampled population $\mathit{\Omega}$, each individual $i \in \mathit{\Omega}$ is randomly assigned to a group of K_i people. Let $\mathit{\Pi}_i \subseteq \mathit{\Omega}$ be the pool of people from which i's $(K_i - 1)$ peers are drawn at random. When the pool $\mathit{\Pi}_i$ is the entire sample, $\mathit{\Pi}_i = \mathit{\Omega}$. The pool $\mathit{\Pi}_i$ can also be a subset of the sample of size L_i, with $\mathit{\Pi}_i \subset \mathit{\Omega}$. Section C.1 deals with cases with multiple peer selection pools, i.e., $\mathit{\Pi}_i \subset \mathit{\Omega}$ (Part 1 of Proposition 1). Section C.2 deals with $\mathit{\Pi}_i = \mathit{\Omega}$ (Part 2 of Proposition 1). Section C.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 C.4, C.5, C.6, and C.7 prove Propositions 2, 3, 4 and 5, respectively.

C.1 Proof of 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 y with the following data generating process:

$$y_{iklt} = \delta_l + \epsilon_{iklt} \tag{C.1}$$

where y_{iklt} is the value of y for individual i in group k of pool l at time t, δ_l is a pool fixed effect, and ϵ_{iklt} 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 (3.1), reproduced here:

$$y_{iklt} = \alpha_1 \bar{y}_{-iklt} + \delta_l + \epsilon_{iklt} \tag{C.2}$$

where $\bar{y}_{-iklt}$ is the sample mean of y_{iklt} for individuals other than i who are in the same group k as i, i.e.:

$$\bar{y}_{-iklt} = \frac{\left[\sum_{j=1}^{K} y_{jklt}\right] - y_{iklt}}{K-1}$$

In what follows we omit subscript t to improve clarity. Regression (C.2) can be expressed in deviation from the pool mean so as to eliminate the pool fixed effect δ_l:

$$y_{ikl} - \bar{y}_l = \beta_1(\bar{y}_{-ikl} - \bar{y}_l) + (\epsilon_{ikl} - \bar{\epsilon}_l) \tag{C.3}$$

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

We note that, by construction, $\bar{y}_l \equiv \delta_l + \bar{\epsilon}_l$. It follows that the demeaned regressor $\bar{y}_{-ikl} - \bar{y}_l$ is mechanically correlated with the demeaned error term $\epsilon_{ikl} - \bar{\epsilon}_l$, resulting in a bias in the estimation of α_1 using equation (C.3). 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 (C.2) and instrument y_{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 α_1 for our specific case of a constant pool and group size. This bias is present even when the true $\alpha_1 = 0$, leading to incorrect inference when using model (C.3) to test random peer assignment. We start by defining $u_{ikl} \equiv \bar{y}_{-ikl} - \bar{y}_{-il}$ where $\bar{y}_{-il}$ is the sample mean of y_{ikl} for individuals other than i who are in the same pool l as i, i.e.:

$$\bar{y}_{-il} \equiv \frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} y_{jsl}\right] - y_{ikl}}{L-1} \tag{C.4}$$

With this new notation, $\bar{y}_{-ikl} = \bar{y}_{-il} + u_{ikl}$ and equation (C.3) can be rewritten as:

$$y_{ikl} - \bar{y}_l = \alpha_1 \left[\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} y_{jsl}\right] - y_{ikl}}{L-1} + u_{ikl} - \left(\frac{\left[\sum_{s=1}^{\frac{L}{K}} \sum_{j=1}^{K} y_{jsl}\right] - \bar{y}_l}{L-1} \right) - \bar{u}_l \right] + \epsilon_{ikl} - \bar{\epsilon}_l \tag{C.5}$$

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

$$y_{ikl} - \bar{y}_l = \alpha_1 \left(\frac{\bar{y}_l - y_{ikl}}{L-1} + u_{ikl} - \bar{u}_l \right) + \epsilon_{ikl} - \bar{\epsilon}_l \tag{C.6}$$

If we define the notation $\ddot{z} \equiv z - \bar{z}_l$, for $z = y, \epsilon, u$, we can further simplify equation (C.3) as:

$$\ddot{y} = \alpha_1 \left(\frac{-\ddot{y}}{L-1} + \ddot{u} \right) + \ddot{\epsilon} \tag{C.7}$$

from which it is immediately apparent that the regressor used to identify α_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 α_1, which takes the following form :

$$plim_{N\to\infty}\left(\hat{\alpha}_1^{FE}\right) = \alpha_1 + \frac{cov\left(\frac{-\ddot{y}}{L-1} + \ddot{u}, \ddot{\epsilon}\right)}{var\left(\frac{-\ddot{y}}{L-1} + \ddot{u}\right)} \tag{C.8}$$

where $\hat{\alpha}_1^{FE}$ stands for the fixed effect estimator obtained using regression (C.7). Since $\alpha_1 = 0$ by construction, we can write:

$$plim_{N\to\infty}\left(\hat{\alpha}_1^{FE}\right) = \frac{cov\left(\frac{-\ddot{y}}{L-1}, \ddot{\epsilon}\right) + cov\left(\ddot{u}{,}\ddot{\epsilon}\right)}{var\left(\frac{-\ddot{y}}{L-1}\right) + 2cov\left(\frac{-\ddot{y}}{L-1}, \ddot{u}\right) + var\left(\ddot{u}\right)} \tag{C.9}$$

With some algebra, equation (C.9) will now enable us to calculate the asymptotic value of the bias in $\hat{\alpha}_1^{FE}$. We start by noting that, since $\bar{u}_l \equiv 0$ by construction, we have:

$$\begin{aligned} cov\left(\ddot{u}, \ddot{\epsilon}\right) &= E\left(\ddot{u}\ddot{\epsilon}\right) = E\left[\left(u_{ikl} - \overline{u}_l\right)\left(\epsilon_{ikl} - \overline{\epsilon}_l\right)\right] \\ &= E\left(u_{ikl}\epsilon_{ikl}\right) - E\left(u_{ikl}\overline{\epsilon}_l\right) = 0 \end{aligned} \tag{C.10}$$

by definition of the average. Similarly we can write:

$$var\left(\ddot{u}\right) = var\left(u_{ikl} - \bar{u}_l\right) = \sigma_u^2 \tag{C.11}$$

To tackle the three remaining terms in equation (C.9), we start by transforming equation (C.7) to obtain an expression for $-\frac{\ddot{y}}{L-1}$. By simple manipulation of equation (C.7), we obtain:

$$\left[\frac{L-1+\alpha_1}{L-1}\right]\ddot{y} = \alpha_1\ddot{u} + \ddot{\epsilon}$$

which leads to:

$$-\frac{\ddot{y}}{L-1} = \frac{-\alpha_1\ddot{u}}{L-1+\alpha_1} - \frac{\ddot{\epsilon}}{L-1+\alpha_1} \tag{C.12}$$

Next we note that:

$$\begin{cases} E\left(\epsilon_{ikl}\bar{\epsilon}_l\right) & = \frac{E\left(\epsilon_{ikl}^2\right)}{L} = \frac{\sigma_\epsilon^2}{L} \\ var\left(\bar{\epsilon}_l\right) & = var\left(\frac{\sum_{i=1}^{L}\epsilon_{ikl}}{L}\right) = \frac{\sum_{i=1}^{L} var(\epsilon_{ikl})}{L^2} = \frac{L\sigma_\epsilon^2}{L^2} = \frac{\sigma_\epsilon^2}{L} \end{cases} \tag{C.13}$$

from which we obtain

$$var\left(\ddot{\epsilon}\right) = \sigma_\epsilon^2 - 2\frac{\sigma_\epsilon^2}{L} + \frac{\sigma_\epsilon^2}{L} = \frac{(L-1)\sigma_\epsilon^2}{L} \tag{C.14}$$

Using the facts that $E(\ddot{\epsilon}) = E(\epsilon_{ikl} - \ddot{\epsilon}_l) = 0$ and that $\alpha_1 = 0$ by assumption, and combining these with equations (C.10), (C.14), and (C.12), we obtain:

$$\begin{aligned} cov\left(\frac{-\ddot{y}}{L-1}, \ddot{\epsilon}\right) &= E\left[\left[\frac{-\ddot{y}}{L-1} - E\left(\frac{-\ddot{y}}{L-1}\right)\right]\ddot{\epsilon}\right] \\ &= E\left[\frac{-\ddot{\epsilon}\ddot{\epsilon}}{L-1}\right] \\ &= \frac{-var(\ddot{\epsilon})}{L-1} = -\frac{\sigma_\epsilon^2}{L} \end{aligned} \tag{C.15}$$

This gives the value of the first term in the numerator of equation (C.9).

Next, we use equation (C.10) and (C.12) to get the value of the middle term in the denominator of (C.9):

$$2cov\left(\frac{-\ddot{y}}{L-1}, \ddot{u}\right) = -2\frac{E(\ddot{u}\ddot{\epsilon})}{L-1} = 0 \tag{C.16}$$

For the first term in the denominator of (C.9), we again use equation (C.12) to get:

$$\begin{aligned} var\left(\frac{-\ddot{y}}{L-1}\right) &= var\left(-\frac{\ddot{\epsilon}}{L-1}\right) \\ &= \frac{\sigma_\epsilon^2}{L(L-1)} \end{aligned} \tag{C.17}$$

Summarizing these different results, we can write the numerator and denominator of (C.8) as follows:

$$cov(\frac{-\ddot{y}}{L-1} + \ddot{u}, \ddot{\epsilon}) \quad = \quad -\frac{\sigma_\epsilon^2}{L} \tag{C.18}$$

$$var(\frac{-\ddot{y}}{L-1} + \ddot{u}) \quad = \quad \frac{\sigma_\epsilon^2}{L(L-1)} + \sigma_u^2 \tag{C.19}$$

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

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

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

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

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

$$\sigma_u^2 = \frac{(L-K)^2(K-1)}{(L-1)^2(K-1)^2}\sigma_\epsilon^2 + \frac{(L-K)}{(L-1)^2}\sigma_\epsilon^2 = \frac{(L-K)}{(L-1)(K-1)}\sigma_\epsilon^2 < \epsilon_\epsilon^2 \tag{C.20}$$

Substituting this into equation (C.19) the denominator of (C.8) can be written:

$$\begin{aligned} var(\frac{-\ddot{y}}{L-1} + \ddot{u}) &= \frac{\sigma_\epsilon^2}{L(L-1)} + \frac{(L-K)}{(L-1)(K-1)}\sigma_\epsilon^2 \\ &= \frac{(K-1) + (L-K)L}{L(L-1)(K-1)}\sigma_\epsilon^2 \end{aligned}$$

Combining these results we get:

$$\begin{aligned} plim_{N\to\infty}\left(\hat{\alpha}_1^{FE}\right) &= \frac{(-\frac{\sigma_\epsilon^2}{L})}{\frac{(K-1)+(L-K)L}{L(L-1)(K-1)}\sigma_\epsilon^2} \\ &= -\frac{(L-1)(K-1)}{(K-1)+(L-K)L} \end{aligned} \tag{C.21}$$

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

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

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 $N = 1$. In this case, the estimated regression does not include pool fixed effects δ_l.

The first part of Proposition 1 (summarized by formula (4.1) and derived in Section C.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 (4.1) it immediately follows that as L converges to infinity, the exclusion bias converges to zero. Formally, if $\mathit{\Pi}_i = \mathit{\Omega}$, then

$$plim_{L\to\infty}\left(\hat{\alpha}_1^{OLS}\right) = 0 \tag{C.22}$$

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 $\mathit{\Pi}_i = \mathit{\Omega}$.

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

Formula (C.21) only holds in the limit, that is, for large sample sizes N. The computation of $E(\hat{\alpha}_1^{FE})$ that applies in small sample sizes is not as straightforward, because $E\left[\frac{samplecov\left(\frac{-\ddot{y}}{L-1}+\ddot{u},\ddot{\epsilon}\right)}{samplevar\left(\frac{-\ddot{y}}{L-1}+\ddot{u}\right)}\right] \neq \frac{E\left[samplecov\left(\frac{-\ddot{y}}{L-1}+\ddot{u},\ddot{\epsilon}\right)\right]}{E\left[samplevar\left(\frac{-\ddot{y}}{L-1}+\ddot{u}\right)\right]}$. 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 $R = SampleCov\left(\frac{-\ddot{y}}{L-1} + \ddot{u}, \ddot{\epsilon}\right)$, $S = SampleVar\left(\frac{-\ddot{y}}{L-1} + \ddot{u}\right)$, 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 (C.21). We know from equation (C.18) and equation (C.19) 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:

$$E\left[\hat{\alpha}_1^{FE}|L\right] < plim_{N\to\infty}\left[\hat{\alpha}_1^{FE}\right] \tag{C.23}$$

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.

C.4 Proof of Proposition 2

The first part of the proof presents a simple formula for aggregating correlation coefficients across sub-samples. The second part applies the formula to the case where pool size and group size vary across pools but group size is the same within each pool. Part 3 examines the case where pool size if fixed but group sizes vary within pools. The last part concludes the proof by combining all cases within a single formula.

An elegant formula for aggregating correlation coefficients can be found in an early paper by Dunlap (1937), which we reproduce here. The author posits that the researcher has calculated correlation coefficients between z and c – and other simple statistics like their mean and variance – separately for two samples of sizes m and n from the same data generating process. Not having a computer at his disposal, the researcher wishes to calculate the correlation coefficient of the combined sample from these already calculated statistics. The solution is the following formula:

$$r_{zy} = \frac{m s_{z_m} s_{c_m} r_{z_m c_m} + m\delta_m \Delta_m + n s_{z_n} s_{c_n} r_{z_n c_n} + n\delta_n \Delta_n}{\sqrt{m(s^2_{z_m} + \delta^2_m) + n(s^2_{z_n} + \delta^2_n)}\sqrt{m(s^2_{c_m} + \Delta^2_m) + n(s^2_{c_n} + \Delta^2_n)}} \tag{C.24}$$

where: the two subsamples are indexed by m and n, respectively; r_{ab} denotes the correlation coefficient between a and b; s_a denotes the standard deviation of a; δ_s denotes the difference between the sample means of z_s and z; and Δ_s denotes the difference between the sample means of c_s and c. The formula naturally generalizes to more than two sub-samples. We also note that, in univariate regressions of c on z, the following relationship holds:

$$\alpha_1 = r_{zc}\frac{s_c}{s_z}$$

To apply the formula to our setting, imagine that we have two sub-samples m and n from the same data generating process (3.1). Within each sub-sample, pool and group sizes are constant. But they vary across the two sub-samples. From Proposition 1 we know the *plim* of $\hat{\alpha}_1$ for each of the two sub-samples with pool fixed effects is:

$$plim_{N\to\infty}[\hat{\alpha}_{1m}] = -\frac{(L_m - 1)(K_m - 1)}{(L_m - K_m)L_m + (K_m - 1)} \tag{C.25}$$

$$plim_{N\to\infty}[\hat{\alpha}_{1n}] = -\frac{(L_n-1)(K_n-1)}{(L_n-K_n)L_n+(K_n-1)} \tag{C.26}$$

We wish to know the *plim* of $\hat{\alpha}_1$ for the combined sample. To achieve this, we apply the formula (C.24). To remove the pool fixed effects, we start by transforming the regression model (3.1) into its pool de-meaned version (C.7) from the proof of part 1 of Proposition 1, which we reproduce here for convenience:

$$\ddot{y}_s = \alpha_{1s}\left(\frac{-\ddot{y}_s}{L-1}+\ddot{u}_s\right)+\ddot{\epsilon}_s \tag{C.27}$$

where $s=\{m,n\}$. For notational simplicity, let us define $x_s \equiv \ddot{y}_s$ and let $z_s \equiv \frac{-\ddot{y}_s}{L-1}+\ddot{u}_s$. Further let r_s stand for the correlation between c_s and z_s. Since (C.27) is a univariate regression, it follows that:

$$plim\hat{\alpha}_{1s} = r_s\frac{s_{c_s}}{s_{z_s}}$$

which establishes a formal link with formula (C.24). By construction, the means of $\ddot{y}_s$ and $\ddot{u}_s$ are 0, and thus the means c_s and z_s are 0 in each pool, implying that $\delta_m = 0 = \delta_n$ and $\Delta_m = 0 = \Delta_n$. We thus have:

$$\sqrt{m(s_{z_m}^2+\delta_m^2)+n(s_{z_n}^2+\delta_n^2)} = \sqrt{ms_{z_m}^2+ns_{z_n}^2}$$

$$= \sqrt{\sum_m z_m^2+\sum_n z_n^2} = (m+n)^{1/2}s_z$$

and similarly:

$$\sqrt{m(s_{c_m}^2+\Delta_m^2)+n(s_{c_n}^2+\Delta_n^2)} = (m+n)^{1/2}s_c$$

where s_z and s_c are the standard deviations of z and c in the full sample.

Since $r_{z_mc_m} = \frac{s_{zm}}{s_{cm}}plim\hat{\alpha}_{1m}$ and $r_{z_nc_n} = \frac{s_{zn}}{s_{cn}}plim\hat{\alpha}_{1n}$, we can now rewrite formula (C.24) as follows:

$$plim\hat{\alpha}_1 = \frac{s_c}{s_z}\frac{ms_{z_m}^2plim\hat{\alpha}_{1m}+ns_{z_n}^2plim\hat{\alpha}_{1n}}{(m+n)s_zs_c}$$

$$= \frac{m}{m+n}\frac{s_{z_m}^2}{s_z^2}plim\hat{\alpha}_{1m}+\frac{n}{m+n}\frac{s_{z_n}^2}{s_z^2}plim\hat{\alpha}_{1n} \tag{C.28}$$

Equations (C.25) and (C.26) provide values for $plim\hat{\alpha}_{1m}$ and $plim\hat{\alpha}_{1n}$. A formula for $s_{z_m}^2$ was derived in Proposition 1, part 1:

$$s^2_{z_m} \equiv Var(\frac{-\ddot{y}}{L_m - 1} + \ddot{u}) = \frac{(K_m - 1) + (L_m - K_m)L_m}{L_m(L_m - 1)(K_m - 1)} \sigma^2_\epsilon$$

A similar formula holds for $s^2_{z_n}$:

$$s^2_{z_n} = \frac{(K_n - 1) + (L_n - K_n)L_n}{L_n(L_n - 1)(K_n - 1)} \sigma^2_\epsilon$$

Furthermore we have, by the definition of the variance:

$$s^2_z = \frac{m}{m+n} s^2_{z_m} + \frac{n}{m+n} s^2_{z_n}$$

Since the unknown variance term σ^2_ϵ cancels out from the $\frac{s^2_{z_m}}{s^2_z}$ and $\frac{s^2_{z_n}}{s^2_z}$ ratios, we do not need it in order to calculate $plim\hat{\alpha}_1$. As in (C.24), the above reasoning naturally generalizes to multiple sub-samples. This completes the second part of the proof.

We now turn to the case when group size varies within pools. We start by assuming all pools have the same mix of group sizes. As in part 2, we regard each set of groups of a given size k as a sub-sample of the whole pool. Let p and q be the number of individual observations in each sub-sample. Under the null hypothesis of $\alpha_1 = 0$ and the maintained assumption of random assignment of peers, each sub-sample can be regarded as a representative random sample. Hence the *plim* formula (C.8) of Proposition 1 part 1 applies to each of them independently. It follows that the *plim*'s of $\hat{\alpha}_1$ are given by the formula from Proposition 1 part 1:

$$plim_{N\to\infty}[\hat{\alpha}_{1p}] = -\frac{(L-1)(K_p - 1)}{(L - K_p)L + (K_p - 1)} \tag{C.29}$$

$$plim_{N\to\infty}[\hat{\alpha}_{1q}] = -\frac{(L-1)(K_q - 1)}{(L - K_q)L + (K_q - 1)} \tag{C.30}$$

We now apply (C.24) to derive the *plim* of the regression coefficient obtained from pooling the two sub-samples p and q. As in part 2, $\delta_p = 0 = \delta_q$ and $\Delta_p = 0 = \Delta_q$. Hence equation (C.28) applies as well:

$$plim\hat{\alpha}_1 = \frac{p}{p+q} \frac{s^2_{z_p}}{s^2_z} plim\hat{\alpha}_{1p} + \frac{q}{p+q} \frac{s^2_{z_q}}{s^2_z} plim\hat{\alpha}_{1q}$$

where:

$$s^2_{z_p} = \frac{(K_p - 1) + (L - K_p)L}{L(L-1)(K_p - 1)}\sigma^2_\epsilon$$

$$s^2_{z_q} = \frac{(K_q - 1) + (L - K_q)L}{L(L-1)(K_q - 1)}\sigma^2_\epsilon$$

$$s^2_z = \frac{p}{p+q}s^2_{z_p} + \frac{p}{p+q}s^2_{z_q}$$

This formula holds within each pool.

We can now combine variation in group size within pools with variation in pool sizes to obtain the following over-arching formula for an arbitrary combination of group and pool sizes. Each group k of size n_k and pool size L_k is regarded as a distinct subsample with its own $plim\hat{\alpha}_1$ and $s^2_{z_k}$ defined as before as:

$$plim_{N\to\infty}[\hat{\alpha}_{1k}] = -\frac{(L_k - 1)(K_k - 1)}{(L_k - K_k)L_k + (K_k - 1)}$$

$$s^2_{z_k} = \frac{(K_k - 1) + (L_k - K_k)L_k}{L_k(L_k - 1)(K_k - 1)}$$

where, for simplicity, we have dropped σ^2_ϵ from the definition of $s^2_{z_k}$ since it cancels out in the final formula for $plim(\hat{\alpha}_1)$. The definition of s^2_z generalizes to:

$$s^2_z = \sum_k \frac{n_k}{M}s^2_{z_k}$$

where $M \equiv \sum_k n_k$ stands for the total number of observations in the estimation sample. The generalized formula for the $plim$ of the pooled $\hat{\alpha}_1$ can be written

$$plim\hat{\alpha}_1 = \sum_k \frac{n_k}{M}\frac{s^2_{z_k}}{s^2_z}plim\hat{\alpha}_{1k}$$

This concludes the proof.

Table A.4 and Table A.5 confirm the accuracy of the formula in Proposition 2 through a set of simulations, particularly for large sample sizes (as expected).

Table A.4: Simulated exclusion bias with random peer assignment: Varying peer group sizes

	Small sample			Large sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Simulation parameters:						
Number of pools (N)	20	20	20	100	100	100
Group size 1 (K1)	2	2	5	2	2	5
Number of groups of size K1	10	10	6	10	10	6
Group size 2 (K2)	5	10	10	5	10	10
Number of groups of size K2	6	3	2	6	3	2
Pool size	50	50	50	50	50	50
Total sample size	1000	1000	1000	5000	5000	5000
Plim of $\hat{\alpha}_1$ from Proposition 2	-0.038	-0.045	-0.115	-0.038	-0.045	-0.115
Mean of $\hat{\alpha}_1^s$ over 100 simulations	-0.040	-0.045	-0.154	-0.038	-0.046	-0.115

Notes: The Table reports simulation results from 100 Monte Carlo replications for varying peer group compositions. For example, column (1) considers pools with 10 peer groups of size 2 and 6 peer groups of size 5. Each simulation considers pools of fixed size $L = 50$ and considers observations generated with a true $\alpha_1 = 0$. In each simulated sample s, coefficient $\hat{\alpha}_1^s$ is estimated using fixed effects at the level of the selection pool. Columns (1)-(3) present results for simulations considering 20 selection pools (1000 observations). Columns (4)-(6) present results for simulations considering 100 selection pools (5000 observations).

Table A.5: Simulated exclusion bias with random peer assignment: Varying pool sizes

	Small sample			Large sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Simulation parameters:						
Pool size 1 (L1)	20	20	50	20	20	50
Number of pools of size L1	10	10	10	40	60	60
Peer group size K1	2	2	10	2	2	10
Pool size 2 (L2)	40	30	20	40	30	20
Number of pools of size L2	30	20	20	120	120	120
Peer group size K2	10	5	2	10	5	2
Total sample size	1400	800	900	5600	4800	5400
Plim of $\hat{\alpha}_1$ from Proposition 2	-0.136	-0.094	-0.070	-0.136	-0.094	-0.071
Mean of $\hat{\alpha}_1^s$ over 100 simulations	-0.132	-0.103	-0.063	-0.135	-0.094	-0.073

Notes: The Table reports simulation results from 100 Monte Carlo replications for varying peer selection pool sizes. For example, column (1) considers samples with 10 pools of 10 observations and 30 pools of 40 observations. The first set of pools contains peer groups all of size 2 and the latter set of pools contains peer groups all of size 10. Each simulation considers observations generated with a true $\alpha_1 = 0$. In each simulated sample s, coefficient $\hat{\alpha}_1^s$ is estimated using fixed effects at the level of the selection pool. Columns (1)-(3) present results for simulations considering a relatively small number of observations. Columns (4)-(6) present results for simulations considering a relatively large number of observations.

C.5 Proof of Proposition 3

To recall, we have, in each group:

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

where $0 < \beta_1 < 1, E[\epsilon_1] = E[\epsilon_2] = 0$ and $E[\epsilon^2] = \sigma_\epsilon^2$. Solving this system of simultaneous linear equations yields the following reduced forms:

$$\begin{aligned} y_1 &= \frac{\beta_0(1+\beta_1)}{1-\beta_1^2} + \frac{\epsilon_1 + \beta_1\epsilon_2}{1-\beta_1^2} \\ y_2 &= \frac{\beta_0(1+\beta_1)}{1-\beta_1^2} + \frac{\epsilon_2 + \beta_1\epsilon_1}{1-\beta_1^2} \end{aligned}$$

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 $E[y_1] = E[y_2] = \frac{\beta_0(1+\beta_1)}{1-\beta_1^2} \equiv \overline{y}$. If ϵ_1 and ϵ_2 are independent from each other, $E[\epsilon_1\epsilon_2] = 0$ and we can write:

$$E[(y_1 - \overline{y})^2] = E\left[\left(\frac{\epsilon_1 + \beta_1\epsilon_2}{1-\beta_1^2}\right)^2\right] = \sigma_\epsilon^2 \frac{1+\beta_1^2}{(1-\beta_1^2)^2}$$

The covariance between y_1 and y_2 is given by:

$$E[(y_1 - \overline{y})(y_2 - \overline{y})] = E\left[\left(\frac{\epsilon_1 + \beta_1\epsilon}{1-\beta_1^2}\right)\left(\frac{\epsilon_2 + \beta_1\epsilon_1}{1-\beta_1^2}\right)\right] = \frac{2\beta_1\sigma_\epsilon^2}{(1-\beta_1^2)^2}$$

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:

$$r = \frac{E[(y_1 - \overline{y})(y_2 - \overline{y})]}{E[(y_1 - \overline{y})^2]} = \frac{2\beta_1}{1+\beta_1^2}$$

We estimate a model of the form:

$$y_1 = a + by_2 + v_1 \tag{C.31}$$

Since equation (C.31) 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:

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

C.6 Proof of Proposition 4

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

$$\rho \equiv plim_{N\to\infty} SampleCorr(\ddot{\epsilon}_{ikl}\ddot{\epsilon}_{jkl}) = -\frac{1}{L-1} \tag{C.32}$$

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:

$$\begin{aligned} plim_{N\to\infty}[(\ddot{y}_1 - \overline{\ddot{y}})^2] &= \frac{\sigma_\epsilon^2(1+\beta_1^2+2\beta_1\rho)}{(1-\beta^2)^2} \\ plim_{N\to\infty}[(\ddot{y}_1 - \overline{\ddot{y}})(\ddot{y}_2 - \overline{\ddot{y}})] &= \frac{\sigma_\epsilon^2(2\beta_1+(1+\beta_1^2)\rho)}{(1-\beta_1^2)^2} \end{aligned}$$

Equipped with the above results, we can now derive an expression for the combined reflection and exclusion bias in model *(A.1)*. As before, we use the fact that $\widehat{b}^{FE} = \frac{SampleCov[(\ddot{y}_1-\overline{\ddot{y}})(\ddot{y}_2-\overline{\ddot{y}})]}{SampleVar[(\ddot{y}_1-\overline{\ddot{y}})^2]}$. Simple algebra yields:

$$plim_{N\to\infty}[\widehat{b}^{FE}] = \frac{2\beta_1+(1+\beta_1^2)\rho}{1+\beta_1^2+2\beta_1\rho} \tag{C.33}$$

C.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:

$$x_{it} = \delta_i + \epsilon_{it} \tag{C.34}$$

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_ϵ^2. 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:

$$x_{it} = \beta_1 x_{it-1} + \delta_i + \epsilon_{it} \tag{C.35}$$

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 ((C.35)) can be expressed in deviation from the individual mean so as to eliminate the individual fixed effect δ_l:

$$x_{it} - \bar{x}_i = \beta_1(x_{it-1} - \bar{x}'_i) + (\epsilon_{it} - \bar{\epsilon}_i) \tag{C.36}$$

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

$$\ddot{x}_{it} = \beta_1 \ddot{x}'_{it} + \ddot{\epsilon}_{it} \tag{C.37}$$

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

$$plim_{N\to\infty}\left(\hat{\beta}_1^{FE}\right) = \beta_1 + \frac{cov\left(\ddot{x}'_{it}, \ddot{\epsilon}_{it}\right)}{var\left(\ddot{x}'_{it}\right)} \tag{C.38}$$

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 C.35 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:

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

Online Appendix References

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.

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.

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.

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.

Drukker, David M., Ingmar R. Prucha, and Rafal Raciborski (2013). “Maximum likelihood and generalized spatial two-state least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances”, *The Stata Journal,* 13(2): 221-41

Dunlap, Jack W. (1937). “Combinative Properties of Correlation Coefficients”, *Journal of Experimental Education*, 5(3): 286-88

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.

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

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.

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.

Lee, Lung-Fei, Xiaodong Liu, Eleonora Patacchini, and Yves Zenou (2021). “Who is the Key Player? A Network Analysis of Juvenile Delinquency”, *Journal of Business and Economic Statistics*, 39(3): 849-57

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

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.

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

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