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 $s_y^2 = \frac{(y_1-y_n)^2+(y_2-y_n)^2}{2-1}$ and the sample autocorrelation $r_n = \frac{(y_1-y_n)(y_2-y_n)}{(y_1-y_n)^2+(y_2-y_n)^2}$. 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.[20] With this assumption, correlation in outcomes between members of the same peer

[20] 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 $E[\epsilon^2] = \sigma_\epsilon^2$. 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:[21]

$$\widehat{\beta}_1^{\,Ref} = \frac{1 - \sqrt{1-\widehat{b}^2}}{\widehat{b}} \tag{A.3}$$

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

[21] 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: [22]

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

[22] 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 plim($\hat{b}^{FE}$)	Monte Carlo average of $\hat{b}^{FE}$	Monte Carlo averrage 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 $K = 2$.[23] 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:

[23] 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 0.002. 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	-0.056	-0.029	-0.026	0.000	0.002

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 included in a Stata command spreg an ML estimator 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 developed by Drukker et al. (2013) 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

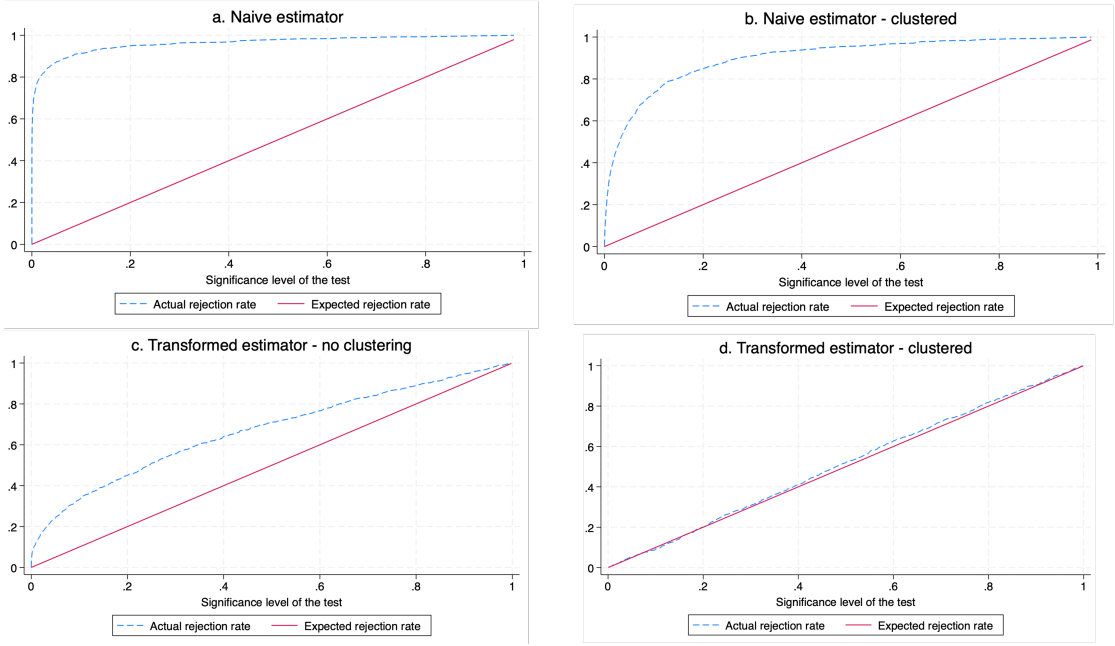
One alternative way to circumvent exclusion bias in standard tests of random peer assignment is to net out the asymptotic exclusion bias using the results from Proposition 1. Specifically, we can use formula (4.1) – or its extension to cases of varying group and pool sizes that is provided in Proposition 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).[24] 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 A.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.

[24] 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.[25]

[25] As discussed in Proposition 1 Part 3, even if the researcher does not include pool fixed effects, there is still an

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.[26] 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 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

exclusion bias if the pool size L is small enough.

[26] 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.

this 2SLS estimation strategy can be written as follows:

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

$$y_{ikl} = \beta_0 + \beta_1 \hat{\bar{y}}_{-ikl} + \beta_2 z_{ikl} + \gamma_l + \epsilon_{ikl} \quad \text{(B.4)}$$

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.[27]

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.

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

[27] 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:

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

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

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.[28]

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 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.

[28] 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.

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.[29] 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 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

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

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	0.26	-0.26	-0.08	0.10
Mean of p-value of $\hat{\beta}_1^{FE}$	0.24	0.29	0.00	0.06	0.40	0.36
Proportion of p-value ≤ 0.05	32.2%	26.0%	98.7%	76.2%	14.3%	16.3%
Panel B						
$\hat{\beta}_1^{Corr}$ - correction for reflection bias + exclusion bias	0.00	0.09	0.19	0.00	0.09	0.19
Mean of p-value of $\hat{\beta}_1^{Corr}$ (using permutation method)	0.49	0.04	0.00	0.49	0.18	0.01
Proportion of p-value ≤ 0.05	5.9%	83.9%	100.0%	6.7%	45.4%	95.8%

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 $G_i Y$ 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 Ω, each individual $i \in \Omega$ is randomly assigned to a group of K_i people. Let $\Pi_i \subseteq \Omega$ be the pool of people from which i's $(K_i - 1)$ peers are drawn at random. When the pool Π_i is the entire sample, $\Pi_i = \Omega$. The pool Π_i can also be a subset of the sample of size L_i, with $\Pi_i \subset \Omega$. Section C.1 deals with cases with multiple peer selection pools, i.e., $\Pi_i \subset \Omega$ (Part 1 of Proposition 1). Section C.2 deals with $\Pi_i = \Omega$ (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:

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

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:

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

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

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 $\Pi_i = \Omega$, 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 $\Pi_i = \Omega$.

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:

$$r_{zy} = \frac{ms_{z_m}s_{c_m}r_{z_mc_m} + m\delta_m\Delta_m + ns_{z_n}s_{c_n}r_{z_nc_n} + n\delta_n\Delta_n}{\sqrt{m(s_{z_m}^2 + \delta_m^2) + n(s_{z_n}^2 + \delta_n^2)}\sqrt{m(s_{c_m}^2 + \Delta_m^2) + n(s_{c_n}^2 + \Delta_n^2)}} \tag{C.24}$$

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 $r_{z_mc_m} = \frac{s_{z_m}}{s_{c_m}}plim\hat{\alpha}_{1m}$ and $r_{z_nc_n} = \frac{s_{z_n}}{s_{c_n}}plim\hat{\alpha}_{1n}$, 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 K_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:

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

where $M \equiv \sum_k K_k$ 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

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

This concludes the proof. Since M cancels out, it can be ignored from the Proposition.

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 1000 simulations	-0.040	-0.047	-0.121	-0.037	-0.044	-0.119

Notes: The Table reports simulation results from 1000 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 1000 simulations	-0.136	-0.096	-0.068	-0.137	-0.094	-0.071

Notes: The Table reports simulation results from 1000 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.