CENTRE FOR THE STUDY OF AFRICAN ECONOMIES

CSAE Working Paper WPS/2014-05

Exclusion bias in empirical social interaction models: causes, consequences and solutions

January 2014

Bet Caeyers[1]

Abstract

This paper formalises an unproven source of ordinary least squares estimation bias in standard linear-in-means peer effects models. I derive a formula for the magnitude of the bias and discuss its underlying parameters. I show the conditions under which the bias is aggravated in models adding cluster fixed effects and demonstrate how it affects inference and interpretation of estimation results. Further, I reveal that two-stage least squares (2SLS) estimation strategies eliminate the bias and provide illustrative simulations. The results may explain some counter-intuitive findings in the social interaction literature, such as the observation of OLS estimates of endogenous peer effects that are larger than their 2SLS counterparts.

[1] University of Oxford, UK; email: betcaeyers@gmail.com; I gratefully acknowledge the feedback I received from Oriana Bandiera, Matthew Collin, Marcel Fafchamps and Simon Quinn for their constructive feedback on previous drafts of this paper.

1. Introduction

In the social interaction literature, the standard model for estimating endogenous peer effects is the following linear-in-means specification (for a review, see Durlauf and Ioannides, 2010):[2]

$$y_{li} = \beta_0 + \beta_1 \bar{y}_{k_{li}} + \varepsilon_i \tag{1}$$

where y_{li} denotes the outcome of individual i in group l and $\bar{y}_{k_{li}}$ is the average outcome of individual i's k peers, which exclude individual i herself. The identification of the peer effect β_1 typically encounters various empirical challenges extensively discussed in the literature, such as the reflection problem (Manski, 1993), endogenous peer group sorting and correlated effects (Brock and Durlauf, 2001, 2007; Moffitt, 2001; Graham and Hahn, 2005; Soetevent, 2006; Graham, 2008). Various strategies have been suggested to deal with these identification issues (e.g. Laschever, 2005; Lee, 2007; Lin, 2007; Graham, 2008; Bramouille et al., 2009; De Giorgi et al., 2009).

One such strategy is to use randomisation methods to assign peers to individuals. Until recently, it was assumed that such a randomisation strategy ensures that no *ex ante* systematic relationship exists between individuals and their peers. As a result, any observed correlation between outcomes *ex post*, conditional on common shocks, was attributed to peer effects. For instance, this strategy has been used to study peer effects on educational outcomes, where y_{li} in equation (1) indicates the test score of student i in classroom l and $\bar{y}_{k_{li}}$ denotes the average test score of that student's k peers (e.g. Sacerdote, 2001).

However, Guryan–Kroft–Notowidigdo (2009), henceforth GKN (2009), have argued that, even with random peer assignment, a negative relationship exists between people's characteristics and those of their peers. The intuition for this result builds on the fact that, in

[2] Although my basic model considers a simple bivariate regression, this paper also briefly discusses the implications for the results of the inclusion of control variables to the model.

equation (1), individuals cannot be their own peers, so that the pool of potential peers from which an individual's peers are drawn systematically excludes the individual herself. Consequently, the pool of potential peers differs for each individual, and its expected value will be negatively correlated with the characteristics of the individual. Using Monte Carlo simulations, GKN (2009) show how this correlation yields a downward bias in the ordinary least squares (OLS) estimate of the endogenous peer effect, β_1, in typical tests of random peer assignments (where the true $\beta_1 = 0$ in a regression on baseline characteristics). Henceforth, I refer to this type of bias as the *exclusion* bias, reflecting the fact that the bias is caused by an exclusion of individuals from their own pool of potential peers. To my knowledge, the only attempt to formalise the intuitive findings of GKN (2009) has been made by Wang (2010). His study shows, both analytically and using simulations, that standard tests of random assignment of peers to students are biased, for the specific case where a student's peers are defined as all his/her classmates.

This paper contributes to the social interaction literature in a number of ways. First, it provides a more general formalisation of the proposition that OLS estimates of the endogenous peer effect are biased downwards in standard peer effects models. I generalise the linear-in-means model in GKF (2009) and Wang (2010) in two ways: (i) I allow the number of peers k to differ from the total size of the group l in equation (1) (i.e. peer groups do not necessarily have to consist of *all* a student's classmates); (ii) I allow the true endogenous peer effect to be greater than zero $(\beta_1 \geq 0)$, rather than focusing on a model that tests for random peer assignment $(\beta_1 = 0)$.

Next, I derive a clear analytical expression for the magnitude of the bias, and discuss the underlying parameters that determine it. When deriving this expression in a model that allows for cluster sampling, I make a careful distinction between models that add cluster fixed effects (e.g. classroom dummies) and models that do not. My results show that, in some common empirical settings, the exclusion bias becomes significantly more severe in models adding cluster fixed effects. Moreover, I show that in the latter type of models, the bias does not disappear as the sample size tends to infinity.

Furthermore, I suggest a new solution to the bias. I list various limitations to the method suggested by GKF (2009) and I show that two-stage least squares (2SLS) procedures are generally more useful in eliminating the bias. With the latter result, this paper further contributes to the literature by explaining a counter-intuitive yet common finding in peer effects studies. Many studies on social interactions obtain 2SLS estimates of endogenous peer effects that are significantly larger than OLS estimates (e.g. Goux and Maurin, 2007; De Giorgi et al., 2010; Brown and Laschever, 2012; de Melo, 2011; Helmers and Patnam, 2012; Krishnan and Patnam, 2012; Naguib, 2012; Collin, 2013).[3] This is counter-intuitive, as one would expect the OLS estimate to be biased *upwards* as a result of other sources of bias such as the reflection bias (Manski, 1993), endogenous peer group formation and/or unobserved correlated effects (Brock and Durlauf, 2001; Moffitt, 2001). This finding has mainly been ignored in the literature to date, except by authors such as Goux and Maurin (2007), de Melo (2011), Helmers and Patnam (2012) and Collin (2013), who attribute it to classical measurement error and/or to the local average treatment interpretation of 2SLS. This paper provides an alternative explanation for this finding, by showing that the negative exclusion bias that is present in OLS estimation is eliminated in 2SLS estimation strategies.

Although the exclusion bias is also present in models in which peer selection is non-random, the focus of this paper is on random peer assignment, which allows me to evaluate the magnitude of the bias without conflating it with the bias caused by endogenous peer group

---

[3] These studies are examples in which both OLS and 2SLS estimates are reported. Many other studies present only 2SLS results, and comparisons with OLS cannot be made (e.g. Lalive and Cattaneo, 2009; Fletcher, 2012).

formation. Moreover, this paper shows that the exclusion bias is only a concern in models in which (i) each individual is excluded from her own peer group, but is included in the peer group of other sampled individuals, and (ii) peer effects specifications fail to account for the difference in expected peer group outcomes that such self-exclusion generates. In the concluding section of this paper, I discuss examples of empirical specifications that are expected to be free of the exclusion bias (e.g. Lyle, 2007, 2009; Duflo et al., 2011; Fafchamps and Quinn, 2012).

In section 2, I begin by constructing a basic model to illustrate that the OLS estimate of the endogenous social interaction effect is biased downwards in specifications where individuals are excluded from their own peer groups. I proceed by deriving an analytical expression for the magnitude of the bias, and obtain results on the parameters that determine it. In section 3, I show that in some common empirical specifications the exclusion bias is aggravated in models adding cluster fixed effects. In section 4, I evaluate the method suggested by GKF (2009) to eliminate the bias, and argue that 2SLS procedures are generally more effective in doing so. In section 5, I present simulation results, which confirm the theoretical findings of the paper. Section 6 discusses extensions of the basic model and considers how these would alter the main results. Finally, in section 7, I discuss the implications of my results for the empirical peer effects literature.

2. Basic model

Suppose we have a sampled network Ω of N individuals. For instance, we have a school Ω with N students. Let subset $\Pi_i \subset \Omega$ be the pool of N_P potential peers for individual i, from which individual i is randomly assigned k peers. In my basic model, I ignore potential clustering and I assume that peers are drawn at the level of the sampled network Ω (e.g. school), that is, $\Pi_i = \Omega$ for all $i \in \Omega$. In the next section, I will consider a model in which peers are drawn from a subset $\Pi_i = l$ of size $L < N$ (e.g. a classroom within the school). Denote the person-specific group of randomly selected *k* peers by P_i. Throughout this paper, I assume that individual *i* is excluded from her own reference group, that is, $i \notin P_i$. Note that in this basic model, we assume that peer groups are person-specific and not necessarily overlapping; that is, the random selection of peers for one person does not depend on the random selection of peers for another person. In the next section, we will consider a more general model that allows for grouping of friends. Nevertheless, the current basic model is interesting. For instance, it forms the basis of specifications in which students are asked to identify their individual friends from the school roster (e.g. Fletcher, 2012).

Let y_i be an outcome variable of interest of individual $i \in \Omega$ (e.g. a student's test score). Consider a simple bivariate linear-in-means social interaction model in which an individual *i*'s outcome y_i is regressed on the average outcome of her *k* (randomly assigned) peers $\bar{y}_{k_i}$:[4]

$$y_i = \beta_0 + \beta_1 \bar{y}_{k_i} + \varepsilon_i \qquad (2)$$

where parameter β_1 captures the peer effect. The exposition in this paper assumes either positive or zero true peer effects, that is, $\beta_1 \geq 0$, which is natural in this setting. The error term ε_i reflects unobservable characteristics associated with *i*. To focus on the effects of exclusion bias, I assume zero correlated effects. The main result of this section is to formally show that, despite the random peer assignment, the explanatory variable $\bar{y}_{k_i}$ will be correlated with the error term ε_i, leading to bias in OLS estimates of the social interaction effect β_1.

To see this, we expand $\bar{y}_{k_i}$ in equation (2). In our basic model, the person-specific peer group P_i consists of a random set of *k* individuals from the pool of $(N-1)$ individuals in

[4] In section.6, I discuss how the addition of control variables affects the results of the model.

network Ω (excluding individual i herself). Let $\bar{y}_{-i}$ be the average outcome of this pool of $(N-1)$ individuals. Given random peer selection, the expected value of the average outcome of the k selected peers will be equal to $\bar{y}_{-i}$, that is, $E(\bar{y}_{k_i}) = \bar{y}_{-i}$. The actual draw, $\bar{y}_{k_i}$, will deviate from $\bar{y}_{-i}$ by a random component u_i:

$$\bar{y}_{k_i} = \bar{y}_{-i} + u_i \tag{3}$$

where $E(u_i) = 0$.

Inserting equation (3) into equation (2), we obtain:

$$y_i = \beta_0 + \beta_1(\bar{y}_{-i} + u_i) + \varepsilon_i \tag{4}$$

where we assume that $E(u) = E(\varepsilon) = 0$, $\mathrm{cov}(u, \varepsilon) = 0$, $\mathrm{var}(u) = \sigma_u^2$ and $\mathrm{var}(\varepsilon) = \sigma_\varepsilon^2$.

It is important to note here that, in a model where peers form part of the study sample, σ_u^2 will be correlated with σ_ε^2. Specifically, in Appendix A.1 I show that:

$$\sigma_u^2 = \frac{(N-1-k)}{(N-1)k}\sigma_\varepsilon^2 < \sigma_\varepsilon^2 \tag{5}$$

Equation (5) shows that, in this basic model in which peers are randomly drawn from within the sampled network Ω of size N, σ_u^2 is decreasing in the peer group size k and increasing in the sample size N. These findings will be important when interpreting the parameters affecting the magnitude of the exclusion bias below.

Returning to equation (4), note that $\bar{y}_{-i}$ can be written as:

$$\bar{y}_{-i} = \frac{\left(\sum_{i=1}^{N} y_i\right) - y_i}{N-1} \tag{6}$$

As a result, equation (4) can be rewritten as follows:

$$y_i = \beta_0 + \beta_1 \left(\frac{\left(\sum_{i=1}^{N} y_i \right) - y_i}{N-1} + u_i \right) + \varepsilon_i \tag{7}$$

Note that it is the presence of the dependent variable y_i on the right-hand side of equation (7) that leads the OLS estimate of β_1 to be biased downwards. To derive this correlation more formally, inserting equation (4) into equation (6) results in the reduced form of $\bar{y}_{-i}$:

$$\bar{y}_{-i} = \frac{\sum_{j=1}^{N} y_j - \beta_0}{N-1+\beta_1} - \frac{\beta_1 u_i}{N-1+\beta_1} - \frac{\varepsilon_i}{N-1+\beta_1} \tag{8}$$

Using equation (8) and the assumption that $\text{cov}(u, \varepsilon) = 0$, it is now straightforward to see that $\text{cov}(\bar{y}_{k_i}, \varepsilon_i) \neq 0$ and, therefore, that OLS estimation of equation (2) leads to a biased estimate of the social interaction effect β_1:

$$\text{cov}(\bar{y}_{k_i}, \varepsilon_i) = \text{cov}(\bar{y}_{-i} + u_i, \varepsilon_i) = \text{cov}(\bar{y}_{-i}, \varepsilon_i) = \frac{-\sigma_\varepsilon^2}{N-1+\beta_1} < 0 \tag{9}$$

Using equation (9), together with the expression for $\text{var}(\bar{y}_{k_i})$ derived in Appendix A.2, we can now determine the magnitude of the bias in OLS:

$$E\left(\hat{\beta}_1^{OLS}\right) = \beta_1 + \frac{\text{cov}(\bar{y}_{k_i}, \varepsilon_i)}{\text{var}(\bar{y}_{k_i})} = \beta_1 + \frac{\dfrac{-\sigma_\varepsilon^2}{(N-1+\beta_1)}}{\dfrac{(N-1)^2 \sigma_u^2 + \sigma_\varepsilon^2}{(N-1+\beta_1)^2}} = \beta_1 - \frac{\sigma_\varepsilon^2 (N-1+\beta_1)}{(N-1)^2 \sigma_u^2 + \sigma_\varepsilon^2}$$

Inserting the expression for σ_u^2 presented in equation (5), I obtain the first result of this paper, summarised by proposition 1.

Proposition 1. Suppose each individual i in a sampled network Ω of N individuals is randomly assigned k peers $\neq i$ from within the sampled network Ω. In a linear-in-means social interaction model, specified as in equation (2), the OLS estimate of the peer effect, $\hat{\beta}_1^{OLS}$, will be biased downwards according to the following expression:

$$E\left(\hat{\beta}_1^{OLS}\right) = \beta_1 - \frac{(N-1+\beta_1)k}{(N-1)(N-k-1)+k} < \beta_1$$

Proposition 1 tells us how, in a basic model in which peers are randomly selected from the total sample of N individuals, the magnitude of the exclusion bias is driven by three important parameters. Specifically, denoting the bias by

$$|bias| = \left|\frac{(N-1+\beta_1)k}{(N-1)(N-1-k)+k}\right|$$

we have:

1 $\frac{\Delta|bias|}{\Delta N} < 0$, *ceteris paribus*, the larger the size of the sampled network, from within which peers are drawn, the smaller the bias. This result, which is consistent with the simulation results in GKN (2009), is intuitive. The mean of the pool will be less sensitive to an exclusion of one of the outcome variables as the pool of potential peers becomes larger. Formally, as shown in equation (5), the larger the N, the larger the variance of u, σ_u^2. From equation (3) we know that, as σ_u^2 becomes larger, more variation in $\bar{y}_{k_i}$ will be explained by the random component u_i rather than being governed by the mean of the pool of potential peers, $\bar{y}_{-i}$. This implies that, *ceteris paribus*, studies that consider a larger pool of potential peers will have less severe exclusion bias. In the basic model, in which the pool from which peers are selected coincides with the sampled network Ω, this result implies that the bias disappears as the sample size tends to infinity. In the next section, we will see that this is not necessarily the case in models in which peers are selected from a sub-set of the sampled network.

2 $\frac{\Delta|bias|}{\Delta k} > 0$, *ceteris paribus*, the larger the peer group size *k*, the larger the bias. Note that the magnitude of the bias is not linear in *k*. If it were, then $\frac{\Delta|bias|}{\Delta k} > 0$ could simply reflect the fact that, although the absolute bias increases with the number of peers, the bias *per additional peer* remains constant. Proposition 1 indicates, however, that the bias per peer also increases with the total number of peers considered.[5] Intuitively, the larger the number of people who are considered in the peer group, the more closely the average peer group outcome $\bar{y}_{k_i}$ will follow the average outcome of the pool from which those peers are drawn, $\bar{y}_{-i}$. This means that $\bar{y}_{k_i}$ will be more sensitive to changes to the mean, $\bar{y}_{-i}$, such as the exclusion of an individual's outcome. Formally, as shown above in equation (3), as *k* becomes larger, the variance σ_u^2 will become smaller. From equation (4) we know that, as σ_u^2 becomes smaller, less of the variation in $\bar{y}_{k_i}$ will be explained by the random component, u_i, rather than being governed by the mean of the pool of potential peers, $\bar{y}_{-i}$. This indicates that, for a given pool size, studies that consider larger peer groups will have a larger exclusion bias.

3 $\frac{\Delta|bias|}{\Delta \beta_1} > 0$, that is, the greater the true social interaction effect β_1, the larger the bias. Also, this result is intuitive, as the greater the true peer effect β_1, the more similar the

[5] From proposition 1 it is straightforward to see that the bias per peer will be equal to $\left|\frac{(N-1+\beta_1)}{(N-1)(N-1-k)+k}\right|$.

outcomes y_i, and therefore the smaller their variance, σ_ε^2. From equation (3) we know that, if σ_ε^2 is small, σ_u^2 will be relatively small as well. Again, from equation (4), we know that, as σ_u^2 becomes smaller, less of the variation in $\bar{y}_{k_i}$ will be explained by the random component u_i and the more sensitive $\bar{y}_{k_i}$ will be to the exclusion of an individual's outcome.

A detailed discussion of the implications of these results for the interpretation of estimation results in the social interaction literature, in particular the effect of the pool size and peer group size, will follow in section 7.

3. Model allowing for cluster sampling

3.1. Motivation

In the basic model described in the previous section, we assumed that peers are randomly drawn from the pool of $(N-1)$ observations within that network. In practice, however, sampled observations are often clustered into groups within a network, and peers are often restricted to those clusters. Moreover, the researcher often relies on the addition of cluster fixed effects to control for unobserved correlated effects. For example, whereas a basic OLS estimator would consider the variation across all N individuals in a school, a model adding classroom dummies would allow us to consider the variation across students *within* different classes of size $L < N$. This is done, for instance, to control for common characteristics of students within groups, such as a teacher's contribution to the student's test score.

In this section, I show how the exclusion bias differs between models with and without cluster fixed effects, while introducing the possibility that peers are selected from within a pool $l \prec \Omega$ (e.g. a classroom within the school). I show that, in some common empirical specifications, the downward exclusion bias will be more severe in OLS models that add cluster fixed effects and the bias will not decrease as the sample size tends to infinity. Specifically, the aggravation of the bias will occur whenever peer group formation is correlated with cluster formation. The intuition for this condition will be conveyed in the next subsection, in which I set out two different empirical scenarios. The analytical discussion in section 3.3 will be structured around these two different cases.

3.2. Two empirical scenarios

Case 1: Peers assigned at the network level Ω $(N_P = N)$

All N students in a school Ω are randomly assigned k other students in the school as reference group, and classes indexed by l of size $L < N$ are formed independently from this peer allocation process.[6] That is, peer group formation operates independently from cluster formation. Thus, peers are not restricted to the class level, as the pool of potential peers consists of all students in the school (excluding oneself). In the next section, I will show that, in this case, the exclusion bias will be the same in models with and without classroom dummies.

To my knowledge, there are no studies that randomly assign peers to students at the school level. However, several studies consider non-randomly selected reference groups that contain peers from outside the classroom. For instance, studies in which peer groups are defined by individuals identified as friends from the school roster (Fletcher and Ross, 2012; Halliday and Kwak, 2012) or by all other students in the same year group and school (e.g. Hoxby, 2000; Hanushek et al., 2003; Angrist and Lang, 2004; Fletcher, 2012). To the extent that peers are selected independently from the classroom in these studies, case 1 applies.

Case 2: Peers assigned at the cluster level l $(N_P = L)$

Now consider an alternative case in which all N students in a school are, first, randomly assigned to classes indexed by l of size $L < N$, and k peers are selected from *within* the class l. I

[6] This case still assumes person-specific peer groups.

will show that, in this case, in which peer group formation is correlated with cluster formation, the exclusion bias will be more severe in a model adding classroom fixed effects than in the basic OLS model. I will also show that the bias in OLS models excluding cluster effects will be larger than their equivalents in case 1 described above.

Note that case 2 allows for both (i) person-specific peer groups within the class, such that peer groups do not necessarily overlap and (ii) grouping of friends at the cluster level l. Whereas in (i) one would set $k < L-1$, in (ii) one would set $k = L - 1$.[7] An example of (i) is de Melo (2011), in which each individual was asked to select friends from the class roster.[8] An example of (ii) is the study by Duflo et al. (2011), who randomly assigned students in non-tracking schools to one of two sections and who then defined peers by all students in the section excluding the student himself/herself.[9] Similar examples include the studies by Sacerdote (2001), Glaeser et al. (2002) and Zimmerman (2003), who randomly assigned students to rooms in dormitories and defined peers by roommates.[10]

3.3. Model

In this section, we extend the basic model by allowing for clustering of sampled observations into $\frac{N}{L}$ groups of equal size L, within the network Ω. For instance, this model allows for clustering of N students into $\frac{N}{L}$ classrooms within a school. In section 6 I will discuss the results of a specification in which I allow the size of groups to vary within the network. For simplicity, we assume that all individuals within the same group l use the same pool $\Pi_l \subset \Omega$,

[7] If grouping occurs at a level that is lower than the classroom, we would re-specify the model and denote l to be that lower level group (instead of the classroom). The same model would apply.

[8] Note that such person-specific peer groups may still overlap, but, for the purpose of simplicity, we assume here that this overlap occurs randomly.

[9] Although their randomisation strategy is in line with case 2, the study by Duflo et al. (2011) is not affected by the exclusion bias because of the empirical model they use (see section 7).

[10] Although one would not add classroom fixed effects in (ii), one could add fixed effects at a level that is higher than the classroom, but lower than the school (such as grade fixed effects). It can easily be shown that such models form special cases of the model we consider in this section.

from which to (randomly) draw peers. Suppose that the size of the pool Π_l is the same for all groups l, that is, N_P. Following the two scenarios described in the previous section, this pool is either the entire network Ω (e.g. school), as in case 1, or a subset $l \subset \Omega$ (e.g. classroom), as in case 2.

The equivalent equations of (2) and (4) become:

$$y_{li} = \beta_0 + \beta_1 \bar{y}_{k_{li}} + \varepsilon_{li} \qquad (10)$$

and

$$y_{li} = \beta_0 + \beta_1 \left(\bar{y}_{\Pi_l,-i} + u_{li} \right) + \varepsilon_{li} \qquad (11)$$

where y_{li} denotes outcome y of individual i in group l; variable $\bar{y}_{\Pi_l,-i}$ denotes the average outcome of the pool of potential peers of all individuals in group l, excluding individual i herself. As before, assume that $E(u) = E(\varepsilon) = 0$, $\mathrm{cov}(u,\varepsilon) = 0$, $\mathrm{var}(u) = \sigma_u^2$ and $\mathrm{var}(\varepsilon) = \sigma_\varepsilon^2$.[11]

Whereas, in the analytical discussion that follows, the models without cluster fixed effects will be based on equation (10), models including cluster fixed effects will be based on a transformation of this equation. Specifically, the observations of interest to the fixed effect estimator are the individual specific deviations of outcomes from their respective cluster average. Define $\ddot{x} = x_{li} - \bar{x}_l$ for $x = y, \bar{y}_k, \varepsilon$. Then, differencing equation (10) gives:

$$\ddot{y} = \beta_1 \ddot{y}_k + \ddot{\varepsilon} \qquad (12)$$

3.4. Exclusion bias in models adding cluster fixed effects

Using a similar procedure as before, in Appendix A.3 I derive the cluster fixed effect equivalent of proposition 1, for a cluster sampling model. Proposition 2 summarises the results.

[11] Note that in reality, in the context of cluster sampling, the assumption $E(\varepsilon) = 0$ will usually not be satisfied. Although the results are expected to hold in models allowing for correlated effects, in this paper I assume zero correlated effects, to focus on the partial effect of the exclusion bias.

Proposition 2. Suppose each individual i in cluster l in a clustered network Ω of size N is randomly assigned k peers $\neq i$ from within a subset $\Pi_l \subset \Omega$ of size $N_P \leq N$. In a linear-in-means social interaction model adding cluster fixed effects, specified as in equation (12), the OLS estimate of the peer effect, $\hat{\beta}_1^{FE}$, will be downward biased according to the following expression:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(N_P - 1 + \beta_1)k}{(N_P - 1)(N_P - 1 - k) + k} < \beta_1$$

Note that proposition 2 is similar to proposition 1, except that now the magnitude of the bias depends on the size of the potential pool $N_P \leq N$, rather than on the sampled network size N. Importantly, proposition 2 indicates that, for models adding cluster fixed effects, the exclusion bias is not necessarily only a small sample property. With peer pool sizes fixed at $N_P < N$, the magnitude of the bias will not decrease as N tends to infinity.

In section 2 I described two different empirical scenarios: one in which peers are selected at the network level Ω ($N_P = N$) and one in which peers are drawn from within the cluster l of size $L < N$ ($N_P = L$). In what follows, I discuss how proposition 2 applies to each of these two special cases.

Case 1: Peers assigned at the network level Ω $(N_P = N)$

This case refers to the scenario in which all N students in a school are randomly assigned to other students in the school to become friends, and classes are formed independently from this peer allocation process. In this case, the pool of potential peers equals all students in the school (excluding oneself), with size N_P equal to N for all observations in the sample. Applying the result in proposition 2, we find that, in this case, the exclusion bias in the model adding cluster fixed effects will be identical to the basic case considered in section 2:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(N-1+\beta_1)k}{(N-1)(N-1-k)+k} < \beta_1 \tag{13}$$

As before in the basic model, since the pool of potential peers coincides with the sampled network Ω, the bias will disappear as the sample size *N* tends to infinity.

Case 2: Peers assigned at the cluster level l $(N_P = L)$

This case reflects the scenario in which all *N* students in a school are, first, randomly allocated to $\left(\frac{N}{L}\right)$ classes of size *L*. Within each class, students are then randomly assigned to other students in the class to become friends. In this case, peers are restricted to the class. This means that N_P equals *L* for all observations in the sample. Again using proposition 2, we find that, in this case, the exclusion bias in the model adding cluster fixed effects becomes:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(L-1+\beta_1)k}{(L-1)(L-1-k)+k} < \beta_1 \tag{14}$$

In this case, in which the pool of potential peers is a sub-set l of the sampled network Ω, the bias depends on the size *L* of the sub-set *l*. With fixed *L*, the bias will not decrease as *N* tends to infinity.

3.5. Exclusion bias in models omitting cluster fixed effects

Case 1: Peers assigned at the network level Ω $(N_P = N)$

When the pool from which an individual's peers are drawn equals the total network Ω of *N* individuals (excluding the individual herself), and when one ignores cluster fixed effects, then it is clear that the results reduce to the basic model we considered in section 2. We obtain proposition 3, which is very similar to proposition 1, except that we now consider a clustered network.

Proposition 3. Suppose each individual i in cluster l in a clustered network Ω of size N is randomly assigned k peers $\neq i$ from within the network Ω. In a linear-in-means social interaction model omitting cluster fixed effects, specified as in equation (10), the OLS estimate of the peer effect, $\hat{\beta}_1^{OLS}$, will be downward biased according to the following expression:

$$E\left(\hat{\beta}_1^{OLS}\right) = \beta_1 - \frac{(N-1+\beta_1)k}{(N-1)(N-k-1)+k} < \beta_1$$

Case 2: Peers assigned at the cluster level l $(N_P = L)$

The results are very different, however, when peers are assigned at the level $l \prec \Omega^N$. Within the framework of hierarchical models, in which data are clustered at a sublevel smaller than the total sample size N, it has been shown that the basic OLS estimator $\hat{\beta}_1^{OLS}$ (i.e. OLS in a model excluding cluster fixed effects) is a weighted combination of the cluster fixed effects estimator $\hat{\beta}_1^{FE}$ (or *within* estimator) and the *between* estimator $\hat{\beta}_1^{BE}$ (Raudenbush and Bryk, 2002):

$$\hat{\beta}_1^{OLS} = \eta^2 \hat{\beta}_1^{BE} + (1-\eta^2)\hat{\beta}_1^{FE} \tag{15}$$

where η^2 is the ratio of the between sum of squares of the independent variable of interest, $\bar{y}_{k_i}$, to its total sum of squares. The within estimator $\hat{\beta}_1^{FE}$ is that presented in equation (14). The between estimator $\hat{\beta}_1^{BE}$ is the OLS estimator from a regression of $\bar{y}_l$ on an intercept and $\bar{\bar{y}}_{k_l}$, where $\bar{y}_l$ denotes the average outcome y_{li} over the individuals in the group and $\bar{\bar{y}}_{k_l}$ denotes the average peer group outcome of the group:

$$\bar{y}_l = \beta_0 + \beta_1 \bar{\bar{y}}_{k_l} + \bar{\varepsilon}_l \tag{16}$$

In Appendix A.4 I show that, in the limit, the between estimator $\hat{\beta}_1^{BE}$ will approach:

$$p\lim\left(\hat{\beta}_1^{BE}\right) = \beta_1 + \frac{(L-1)k}{(L-1)k + (L-1-k)} > \beta_1 \tag{17}$$

which, in contrast to $\hat{\beta}_1^{BE}$, is *larger* than the true value β_1.

Combining equations (14), (15) and (17) and the results on η^2 derived in Appendix A.5, we arrive at proposition 4.

Proposition 4. Suppose each individual *i* in a clustered network Ω is randomly assigned *k* peers $\neq$ *i* from within their cluster *l* of size *L*. In a linear-in-means social interaction model ignoring cluster fixed effects, as specified in equation (10), the OLS estimate of the peer effect, $\hat{\beta}_1^{OLS}$, will be downward biased and inconsistent according to the following expression:

$$p\lim\left(\hat{\beta}_1^{OLS}\right) = p\lim\left(\eta^2\right) p\lim\left(\hat{\beta}_1^{BE}\right) + \left[1 - p\lim\left(\eta^2\right)\right] p\lim\left(\hat{\beta}_1^{FE}\right)$$

where:

$$p\lim\left(\hat{\beta}_1^{BE}\right) = \beta_1 + \frac{(L-1)k}{(L-1)k + (L-1-k)} > \beta_1$$

$$p\lim\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(L-1+\beta_1)k}{(L-1)(L-1-k)+k} < \beta_1$$

$$0 < p\lim\left(\eta^2\right) = \frac{\dfrac{(Lk-2k+L-1)}{k(L-1)}}{\dfrac{(Lk-2k+L-1)}{k(L-1)} + \left[\dfrac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]} < 1$$

Given that $0 < p\lim\left(\eta^2\right) < 1$, we know that:

$$p\lim\left(\hat{\beta}_1^{FE}\right) < p\lim\left(\hat{\beta}_1^{OLS}\right) < \beta_1 < p\lim\left(\hat{\beta}_1^{BE}\right)$$

Note that the result summarised in proposition 4 is only valid in the limit, that is, as the sample size *N* tends to infinity. In Appendix A.5 I explain why the downward bias in $\hat{\beta}_1^{OLS}$ is expected to be more severe the smaller the sample of observations that is available. An analytical derivation of the magnitude of the exclusion bias, when using small samples, in OLS models omitting cluster fixed effects, is beyond the scope of this thesis (see Appendix A.5).

However, in section 5 I will use simulations to show that, in small samples, we also have $\hat{\beta}_1^{FE} < \hat{\beta}_1^{OLS} < \beta_1$.

3.6. Comparison results of models with and without cluster fixed effects

Comparing the results provided by propositions 1–4, I conclude this section with a fifth proposition, in which $\left|bias^{FE}\right|$ denotes the absolute value of the magnitude of the exclusion bias in a model adding cluster fixed effects, and $\left|bias^{OLS}\right|$ denotes the absolute value of the magnitude of the exclusion bias in a model ignoring fixed effects.

> **Proposition 5.**
>
> When peer group formation occurs at the total network level Ω (e.g. case 1):
>
> $$\left|bias^{FE}\right| = \left|bias^{OLS}\right|$$
>
> When peer group formation occurs at the cluster level $l \prec \Omega$ (e.g. case 2):
>
> $$\left|bias^{FE}\right| > \left|bias^{OLS}\right|$$

The intuition behind proposition 5 is the following. Recall that the exclusion bias arises if (i) an individual is excluded from her own potential peer group and (ii) if that individual is a potential peer of other individuals in the set within which one analyses the variation in outcomes. The second part of this condition is important: the exclusion bias arises because of the presence of observations for which individual i is included in her pool of potential peers. The smaller the share of such 'problematic' observations in the set of individuals across which one analyses the variation in outcomes, the less severe the exclusion bias will be. Providing peers are selected only from *within* an individual i's cluster (case 2), then the potential peers of the individuals *outside* i's cluster will never include i. Given that, in models adding cluster fixed effects, the researcher is interested only in the variation of outcomes *within* clusters, the presence of such 'unproblematic' observations outside the cluster will not reduce the exclusion bias. In contrast, in models without cluster fixed effects, the researcher is interested in the variation of outcomes across *all* individuals in the sample, which means that the exclusion bias will become relatively less severe.

An important remark is in order when comparing proposition 2 with proposition 4. Whereas in models adding cluster fixed effects the exclusion bias does *not* decrease with sample size N (see proposition 2), proposition 4 shows that, in models without cluster fixed effects, the exclusion bias *does* decrease as N tends to infinity.[12] Intuitively, with fixed cluster sizes and with peers selected at the cluster, an increase in the sample size N implies a larger number of 'unproblematic observations' in the sample. I have discussed above why this leads to a reduction in the magnitude of the exclusion bias in models without cluster fixed effects, whereas models adding cluster fixed effects remain unaffected by this.

The results summarised in proposition 5 are important. They caution researchers against naive comparisons of OLS peer effects estimation results between models with and those without cluster fixed effects. A more detailed discussion of the implications of these results for the social interaction literature follows in section 7.

4. Solutions to the bias

4.1. Method suggested by GKN (2009)

As discussed above, the exclusion bias is driven by the fact that an individual cannot be her own peer and therefore is excluded from the pool of individuals from which her peers are drawn. This empirical construction results in a correlation between the mean of the pool of potential peers, $\bar{y}_{-i}$, and the characteristics of the individual. I have shown above how this

[12] Formally, this result can be derived by considering the result in Appendix A.5 that, when N tends to infinity, more weight is given to the between estimator in the computation of $\hat{\beta}_1^{OLS}$, and, therefore, $\hat{\beta}_1^{OLS}$ becomes less negative.

causes the OLS coefficient estimate of the endogenous peer effect to be biased downwards in empirical social interaction models.

In a simple model of randomised peer group formation, GKN (2009) deals with the exclusion bias by controlling for the average outcome of the pool of potential peers, $\bar{y}_{-i}$, in the above regression (equation (2)). To highlight how, under specific conditions, this effectively deals with the exclusion bias, consider a model similar to equation (2), but now adding the term $\bar{y}_{-i}$. Substituting in for equation (3) and rearranging, I obtain:

$$y_i = \beta_0 + \beta_1 \bar{y}_{k_i} + \beta_2 \bar{y}_{-i} + \varepsilon_i \tag{18}$$

$$= \beta_0 + \beta_1 (\bar{y}_{-i} + u_i) + \beta_2 \bar{y}_{-i} + \varepsilon_i$$

$$= \beta_0 + (\beta_1 + \beta_2) \bar{y}_{-i} + \beta_1 u_i + \varepsilon_i \tag{19}$$

The inclusion of $\bar{y}_{-i}$ in the model acts as a proxy, soaking up the effect of the non-random component of $\bar{y}_{k_i}$. As a result, the coefficient estimate $\hat{\beta}_1$ measures the partial effect of the random component u_i. Since $E(u_i \varepsilon_i) = 0$ under the assumption of random peer selection, $p\lim(\hat{\beta}_1) = \beta_1$, and OLS will yield consistent estimates of the social interaction effect β_1.

However, there are a few limitations to the method suggested by GKN (2009). If the pool of potential peers is identical for each sampled individual, excluding i (such as in case 1 described above), then all variation in $\bar{y}_{-i}$ will be driven by y_i. Thus, β_1 will not be identified. If the pool of potential peers is identical for individuals sharing the same cluster, a similar problem occurs whenever cluster fixed effects are added to the model and β_1 will not be identified. To formalise the latter result, reconsider the model in equation (12) but now we add the difference between $\bar{y}_{-i}$ and its average over the cluster group, $\bar{\bar{y}}_{-i,l}$:

$$\left(y_{li}-\bar{y}_{l}\right)=\beta_1\left(\bar{y}_{k_i}-\bar{\bar{y}}_{k_l}\right)+\beta_2\left(\bar{y}_{-i}-\bar{\bar{y}}_{-i,l}\right)+\left(\varepsilon_{li}-\bar{\varepsilon}_{l}\right)$$

$$=\beta_1\left(\bar{y}_{k_i}-\bar{\bar{y}}_{k_l}\right)+\beta_2\left(\left(\frac{\sum_{j=1}^{L}y_{lj}-y_{li}}{L-1}\right)-\frac{1}{L}\sum_{i=1}^{L}\left(\frac{\sum_{j=1}^{L}y_{lj}-y_{li}}{L-1}\right)\right)+\left(\varepsilon_{li}-\bar{\varepsilon}_{l}\right)$$

$$=\beta_1\left(\bar{y}_{k_i}-\bar{\bar{y}}_{k_l}\right)+\beta_2\left(\left(\frac{\sum_{j=1}^{L}y_{lj}-y_{li}}{L-1}\right)-\frac{\sum_{j=1}^{L}y_{lj}-\bar{y}_{l}}{L-1}\right)+\left(\varepsilon_{li}-\bar{\varepsilon}_{l}\right)$$

$$=\beta_1\left(\bar{y}_{k_i}-\bar{\bar{y}}_{k_l}\right)+\beta_2\left(\left(\frac{\sum_{j=1}^{L}y_{lj}-y_{li}}{L-1}\right)-\frac{\sum_{j=1}^{L}y_{lj}}{L}\right)+\left(\varepsilon_{li}-\bar{\varepsilon}_{l}\right)$$

$$=\beta_1\left(\bar{y}_{k_i}-\bar{\bar{y}}_{k_l}\right)+\beta_2\left(\frac{y_{li}-\bar{y}_{l}}{1-L}\right)+\left(\varepsilon_{li}-\bar{\varepsilon}_{l}\right)$$

These derivations show that the term $\left(\bar{y}_{-i}-\bar{\bar{y}}_{-i,l}\right)$ (the within-model equivalent of the term suggested by GKN (2009)) equals $\left(\frac{y_{li}-\bar{y}_{l}}{1-L}\right)$, when peers are selected from within the cluster. Thus, all the variation in $\left(y_{li}-\bar{y}_{l}\right)$ will be captured by $\left(\frac{y_{li}-\bar{y}_{l}}{1-L}\right)$ and β_1 will not be identified.

Lastly, even if these issues are not a concern, the GKN (2009) method requires perfect knowledge of the pool from which peers are selected. However, as I will discuss in section 6, in practice, this is not always the case.

4.2. 2SLS estimation strategies

The social interaction literature frequently relies on 2SLS strategies to address other empirical challenges in the identification of peer effects, such as the reflection problem and correlated effects (for a discussion of these issues, see Manski 1993; Brock and Durlauf, 2001; Moffitt, 2001). For instance, various studies have instrumented the average peer group outcome by the proportion of peers that are randomly treated by an intervention that affects their outcomes (e.g. Duflo and Saez, 2002; Lalive and Cattaneo, 2009). Other studies have used a set of average exogenous characteristics of the peers of a person's peers as identifying instruments, while controlling for the individual's own exogenous characteristics (Bramouille et al., 2009; De Giorgi et al., 2010; de Melo, 2011; Helmers and Patnam, 2012). Although these studies use 2SLS to address traditional challenges to identification, I will show that it also overcomes the exclusion bias.

To see this, let $\bar{z}_{k_i}$ be the peer group average of a characteristic *z*, and assume that it is considered as an instrument for the average peer group outcome $\bar{y}_{k_i}$. If $\bar{z}_{k_i}$ is informative in the first stage of a 2SLS estimation strategy, then z_i will usually also be informative about y_i, and therefore it is often included in the second stage of 2SLS.[13] The first and second stages of such SLS estimation strategies can be modelled as follows:

First stage: $\bar{y}_{k_i} = \pi_0 + \pi_1 \bar{z}_{k_i} + \pi_2 z_i + v_i$

Second stage: $y_i = \beta_0 + \beta_1 \hat{\bar{y}}_{k_i} + \beta_2 z_i + \varepsilon_i$ (20)

where $E(z_i \varepsilon_i) = 0$, $E(\varepsilon_i) = 0$ and $\hat{\bar{y}}_{k_i} = \hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{k_i} + \hat{\pi}_2 z_i$ is the fitted value of the first-stage regression.

Expanding the second-stage 2SLS equation (equation (20)), using the expression for the fitted values, it is straightforward to see that $\text{cov}(\hat{\bar{y}}_{k_i}, \varepsilon_i \mid z_i) = 0$ and therefore $\hat{\beta}_1^{2SLS}$ does not suffer from the exclusion bias:

$$y_i = \beta_0 + \beta_1 \hat{\bar{y}}_{k_i} + \beta_2 z_i + \varepsilon_i$$

[13] Note that it is not necessary to control for *z* whenever *z* is purely random. This is the case, for instance, when an intervention is randomised only for peers and not for oneself. If all individuals in the sample are subject to the randomisation, however, it is essential that *z* is controlled for.

$$= \beta_0 + \beta_1\left(\hat{\pi}_0 + \hat{\pi}_1\bar{z}_{k_i} + \hat{\pi}_2 z_i\right) + \beta_2 z_i + \varepsilon_i$$

$$= \beta_0 + \beta_1\left[\hat{\pi}_0 + \hat{\pi}_1\left(\frac{\sum_{j=1}^{N_P} z_j - z_i}{N_P - 1} + \widetilde{u}_i\right) + \hat{\pi}_2 z_i\right] + \beta_2 z_i + \varepsilon_i$$

$$= \beta_0 + \beta_1\left(\hat{\pi}_0 + \frac{\hat{\pi}_1\sum_{j=1}^{N_P} z_j}{N_P - 1} + \hat{\pi}_1\widetilde{u}_i\right) + \left(\beta_1\hat{\pi}_2 + \beta_2\frac{-\hat{\pi}_1}{N_P - 1}\right)z_i + \varepsilon_i \qquad (21)$$

where $\widetilde{u}$ is the equivalent of u but now for z instead of y (see (3)). Note that, if y_i and z_i are correlated (i.e. if $\beta_2 \neq 0$), we would expect $\bar{z}_{k_i}$ to also be mechanically correlated with y_i, through $\bar{z}_{k_i} = \frac{\sum_{j=1}^{N_P} z_j - z_i}{N_P - 1} + \widetilde{u}_i$. However, the expression derived in equation (21) shows how controlling for z_i in the regression prevents this mechanical relationship from existing. The presence of z_i in the model captures the effect that the exclusion of z_i has on the computation of the average peer group characteristic $\bar{z}_{k_i}$, allowing $\hat{\beta}_1^{2SLS}$ to be an unbiased estimate of the peer effect.

5. Simulation results

To illustrate the presence and the consequences of the exclusion bias, this section reports on the results of various Monte Carlo simulations. The outcomes of the simulations confirm each of the theoretical predictions derived earlier in this paper. Specifically, they confirm the predicted magnitude of the bias for various values of the network size *N*, the cluster size *L* and the peer group size *k*. They confirm the difference in results between models with and without cluster fixed effects whenever peers are assigned at the cluster level. Finally, they confirm the ability of 2SLS strategies to eliminate the bias.

Although the results are also confirmed for simulations with $\beta_1 > 0$, for simplicity I focus here on the case of zero true peer effects, that is, $\beta_1 = 0$. In each simulation, I generate a dataset of *N* observations and randomly assign a normally distributed characteristic *y* to each of them. The underlying data generation process (DGP) for the first set of Monte Carlo simulations is the following:

$$y_{li} = 10 + \varepsilon_{li}, \; \varepsilon_{li} \sim N(0,1) \tag{22}$$

For the simulations based on the cluster sampling models, I then randomly assign the *N* individuals to $\frac{N}{L}$ clusters of *L* persons each. For each individual in the sampled network, I randomly select *k* reference persons out of a specified pool $\Pi_l \subset \Omega$ of potential peers. Each simulation differs in the value assigned to the network size *N*, the cluster size *L* and the peer group size *k*. The simulations also differ in the definition of the pool Π_l. This is either the total network Ω $(N_P = N)$ or the cluster $l \prec \Omega$ $(N_P = L)$.

Once the dataset is compiled, I run regressions, both with and without cluster fixed effects, of the outcome y_{li} on $\bar{y}_{k_i}$:

$$y_{li} = \beta_0 + \beta_1 \bar{y}_{k_i} + \varepsilon_{li} \tag{23}$$

where y_{li} is the outcome of individual *i* in cluster *l* and $\bar{y}_{k_i}$ is the average outcome of that individual's peer group. As noted above, the DGP is based on a true $\beta_1 = 0$, thus an unbiased estimator would be expected to yield $\hat{\beta}_1 = 0$. In each simulation, I store the coefficient estimate $\hat{\beta}_1$ and repeat this procedure 10,000 times. Tables 1–4 show the simulation results for the estimate of the endogenous peer effect, $\hat{\beta}_1$, where (S) represents the simulation result and

(T) indicates what the theory outlined above would predict given the respective values of *k*, *N* and *L*.

Table 1 Simulation results $\hat{\beta}_1$: basic model

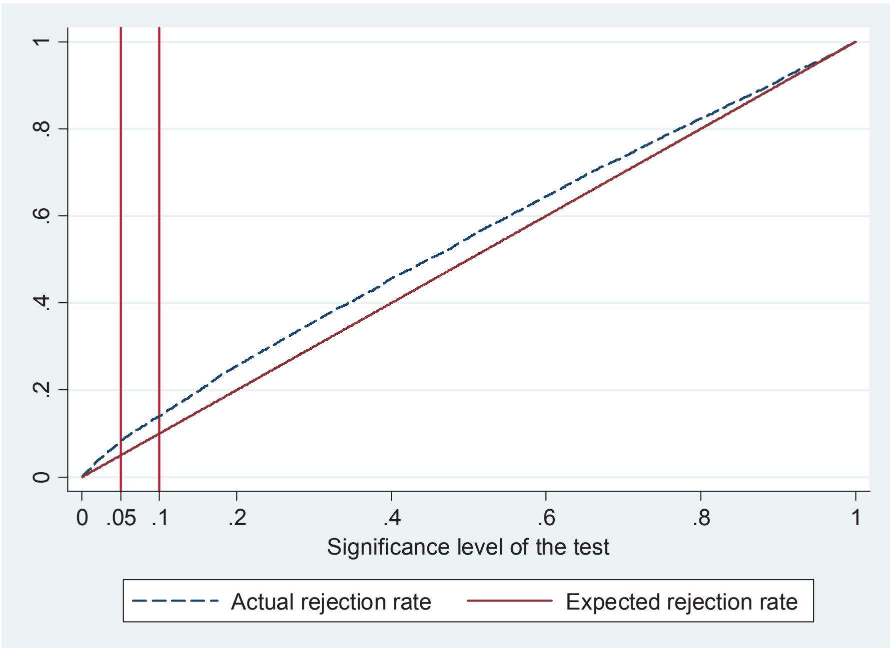
		$N_P = N$		
		$N = 50$	$N = 100$	$N = 200$
		OLS (1)	OLS (2)	OLS (3)
$k = 1$	T	–0.021	–0.010	–0.005
	S	–0.019	–0.009	–0.006
		(0.001)	(0.001)	(0.001)
	% rejection 0.05	4.7	4.7	4.9
	% rejection 0.10	9.6	10.1	9.7
$k = 5$	T	–0.113	–0.053	–0.026
	S	–0.113	–0.049	–0.028
		(0.003)	(0.002)	(0.002)
	% rejection 0.05	6.1	5.5	5.5
	% rejection 0.10	11.8	10.5	10.7
$k = 10$	T	–0.255	–0.112	–0.053
	S	–0.257	–0.105	–0.055
		(0.005)	(0.003)	(0.002)
	% rejection 0.05	8.2	6.1	5.7
	% rejection 0.10	0.140	0.116	0.111
$k = 15$	T	–0.437	–0.178	–0.082
	S	–0.442	–0.171	–0.083
		(0.007)	(0.004)	(0.003)
	% rejection 0.05	9.7	6.8	5.7
	% rejection 0.10	16.6	12.4	11.2

Notes: No cluster sampling – peer selection at the level Ω. Data generation process, $y_{li} = 10 + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. OLS, ordinary least squares; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; T, coefficient estimate predicted by proposition 1. % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

The tables also indicate, for each simulation, the proportion of repeated simulation rounds (10,000 in total) for which we reject the hypothesis that $\beta_1 = 0$, at the 5% and 10% significance levels.

Table 1 presents the results of simulations predicated on the basic model presented in section 2, with an underlying DGP that ignores clustering at a level lower than Ω. That is, all *N* sampled students belong to the same school but are not grouped into classes. The *k* peers are randomly assigned at the school level, $N_p = N$, where we vary the values of k and N. The analytical expression of the exclusion bias that is relevant here is provided in proposition 1. Note (by comparing (T) with (S)) that each of the simulations yields results for the magnitude of the bias that are very similar to what proposition 1 would predict. As expected in this model, for a given peer group size, the exclusion bias becomes less severe as the sample size grows. Moreover, with the sample size fixed, the magnitude of the bias increases as the peer group grows in size.

Figure 1 Rejection rates $H_0 : \beta_1 = 0$ (basic model OLS: $N = 50$; $k = 10$)



Note, however, that the rejection rates reported in Table 1 are relatively low. To interpret these numbers, consider Figure 1, which plots the rates at which we reject the test $H_0 : \beta_1 = 0$ in simulation $N = 50$ and $k = 10$, for various significance levels. Given that the significance level states the tolerated probability of rejecting H_0 when it is in fact true, the 45° line shows us the rejection rates that we would expect if the estimator was unbiased. Actual rejection rates that are higher than the 45° line are an indication that the estimator is biased. Figure 1 indicates that, for $N = 50$ and $k = 10$, the estimate is slightly biased: whereas, at the 5% significance level, we would expect to reject H_0 5% of the time, in our simulations we reject it approximately 8% of the time (out of 10,000 repetitions). However, although the magnitude of the bias appears to be quite significant in most of the specifications in Table 1 (e.g. for $N = 50$ and $k = 15$, we obtain a coefficient estimate of the endogenous peer effect of –0.442, which is equivalent to a bias of 3 percentage points per peer), the rejection rates are relatively low. This is not surprising, given the small sample sizes and the associated high standard errors of the estimates.

Table 2 Simulation results $\hat{\beta}_1$: case 1 ($N_P = N = 200$)

		$L = 20$		$L = 50$	
		FE	OLS	FE	OLS
		(1)	(2)	(3)	(4)
$k = 1$	T	–0.005	–0.005	–0.005	–0.005
	S	–0.004	–0.004	–0.004	–0.004
		(0.001)	(0.001)	(0.001)	(0.001)
	% rejection 0.05	5.6	5.1	5.1	5.1
	% rejection 0.10	10.8	9.9	10.1	9.9
$k = 5$	T	–0.026	–0.026	–0.026	–0.026
	S	–0.023	–0.024	–0.023	–0.024
		(0.006)	(0.006)	(0.002)	(0.002)
	% rejection 0.05	5.4	4.9	5.0	4.9
	% rejection 0.10	10.8	10.0	10.5	10.0
$k = 10$	T	–0.053	–0.053	–0.053	–0.053
	S	–0.049	–0.050	–0.050	–0.050
		(0.002)	(0.002)	(0.002)	(0.002)
	% rejection 0.05	6.0	5.6	5.7	5.6
	% rejection 0.10	11.2	10.4	10.7	10.4
$k = 15$	T	–0.082	–0.082	–0.082	–0.082
	S	–0.078	–0.079	–0.078	–0.079
		(0.003)	(0.003)	(0.003)	(0.003)
	% rejection 0.05	6.3	5.8	6.0	5.8
	% rejection 0.10	11.7	11.2	11.5	11.2

Notes: cluster sampling level $l \prec \Omega$ – peer selection level Ω. Data generation process, $y_{li} = 10 + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. FE, OLS regressions adding cluster fixed effects; OLS, OLS regressions ignoring cluster fixed effects; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; T, coefficient estimate predicted by proposition 2 and proposition 3. % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

Tables 2 and 3 report on the results of simulations in which N is fixed at 200, but in which now the N individuals are randomly allocated to groups of size L. We present the results for $L = 20$ and $L = 50$. Table 2 is based on a DGP reflecting case 1, in which k peers are randomly assigned at the school level, independently from the classroom formation process. Again, the simulation results for both models with cluster fixed effects (FE) and without cluster fixed effects (OLS) are consistent with the values predicted by the analytical expressions in proposition 2 and proposition 3, respectively. Specifically, the magnitude of the bias depends only on the network size N and not on the cluster size L; hence, the similarities between the results presented in Table 2 and the results presented in column (3) in Table 1.

Table 3 shows the results of simulations with an underlying DGP mimicking case 2, in which the k peers are randomly assigned from within the cluster l. From section 3.5 we know that, in the context of cluster sampling, the OLS estimate of the peer effect in a model omitting cluster fixed effects will be a weighted average of the between estimator (BE) and the cluster fixed effect estimator (FE). The analytical expressions for the between estimator and the fixed effects estimator were presented in equation (14) and equation (17), respectively. Recall from the discussion of proposition 4 that for small samples, such as those we use in our simulations, it is computationally difficult to analytically derive predicted values for the coefficient estimate in OLS models omitting cluster fixed effects, that is, $\hat{\beta}_1^{OLS}$. Hence, those predicted values are indicated as 'not applicable' (N/A) in Table 3. However, the simulation results confirm the prediction made in section 6 that $\hat{\beta}_1^{OLS}$ always lies somewhere in between the between estimator and the fixed effect estimator results.

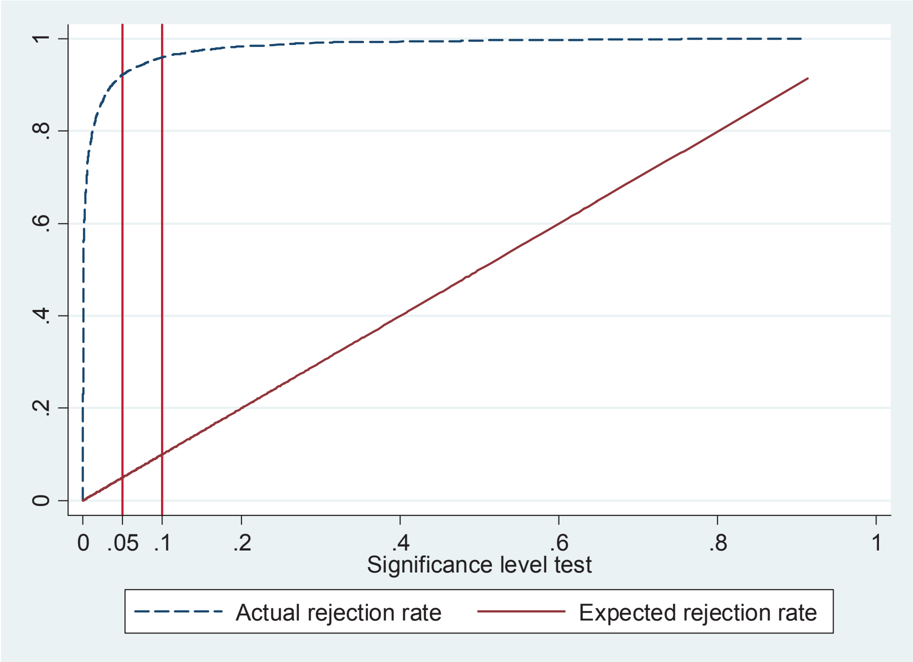
Table 3 Simulation results $\hat{\beta}_1$: case 2 ($N = 200$; $N_P = L$)

		$L = 20$			$L = 50$		
		BE (1)	FE (2)	OLS (3)	BE (4)	FE (5)	OLS (6)
$k = 1$	T	0.514	–0.055	N/A	0.505	–0.021	N/A
	S	0.515	–0.056	–0.006	0.505	–0.021	–0.005
		(0.002)	(0.001)	(0.001)	(0.005)	(0.001)	(0.001)
	% rejection 0.05		13.0	6.2		6.0	5.1
	% rejection 0.10		20.6	11.7		11.3	10.2
$k = 5$	T	0.872	–0.351	N/A	0.848	–0.113	N/A
	S	0.870	–0.351	–0.036	0.852	–0.110	–0.026
		(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
	% rejection 0.05		49.0	8.4		10.0	5.9
	% rejection 0.10		61.2	14.8		17.0	11.2
$k = 10$	T	0.955	–1.050	N/A	0.926	–0.255	N/A
	S	0.955	–1.049	–0.099	0.925	–0.251	–0.065
		(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	% rejection 0.05		92.3	12.3		17.8	7.2
	% rejection 0.10		0.961	0.193		0.271	0.134
$k = 15$	T	0.986	–3.132	N/A	0.986	–3.132	N/A
	S	0.986	–3.145	–0.212	0.955	–0.432	–0.114
		(0.001)	(0.004)	(0.005)	(0.002)	(0.003)	(0.003)
	% rejection 0.05		100	17.8		26.5	8.8
	% rejection 0.10		100	25.8		37.9	15.5

Notes: cluster sampling level $l \prec \Omega$ – peer selection level *l*. Data generation process, $y_{li} = 10 + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. BE, between-group estimation model; FE, OLS regressions adding cluster fixed effects (within group estimation model); N/A: not applicable; OLS, OLS regressions ignoring cluster fixed effects; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; T, coefficient estimate predicted by equation (14), equation (17) and proposition 4. % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

Importantly, note the high rates at which we reject the null hypothesis that $\beta_1 = 0$ at the 0.05 and 0.10 significance levels, especially in the models adding cluster fixed effects (FE). Figure 2 shows the rejection rates for the cluster fixed effects estimation results for $N = 200$, $L = 20$ and $k = 10$.

Figure 2 Rejection rates $H_0 : \beta_1 = 0$ (case 2 – FE: $N = 200$; $L = 20$; $k = 10$)



Finally, in order to test the hypothesis that 2SLS is effective in eliminating the exclusion bias, I generate an alternative DGP that is guided by the discussion in section 4.2:

$$y_{li} = 10 + 5x_{li} + \varepsilon_{li}, \ \varepsilon_{li} \sim N(0,1) \tag{24}$$

where x_{li} is a random uniformly distributed variable. Based on this DGP, I carry out simulations (again with 10,000 repetitions) in which each time I run 2SLS regressions on:

$$y_{li} = \beta_1 + \beta_2 \bar{y}_{k_i} + x_{li} + \varepsilon_{li}$$

instrumenting $\bar{y}_{k_i}$ by the average x characteristic of the peer group, $\bar{x}_{k_i}$.

Table 4 displays the 2SLS equivalent estimation results for each of the cases considered in Tables 1–3. Since the results are similar for all values of k, the table only reports on the results for $k = 10$. The results are consistent with our analytical results in section 5.2, that is, the exclusion bias is eliminated in 2SLS.

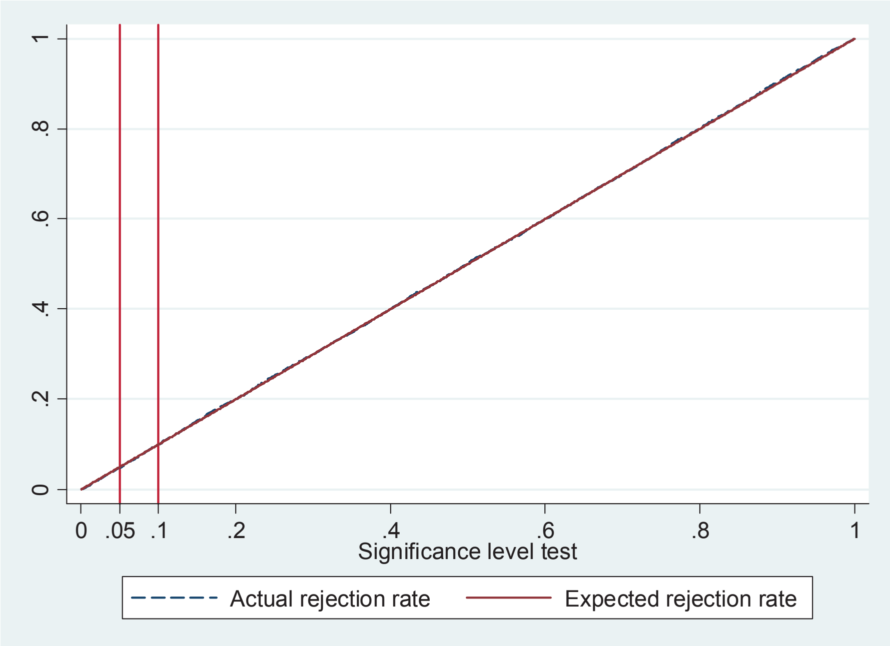
Table 4 Simulation results on $\hat{\beta}_1$: 2SLS

		Base model	Case 1: $N = 200$; $N_P = N$				Case 2: $N = 200$; $N_P = L$			
		$N = 200$	$L = 20$		$L = 50$		$L = 20$		$L = 50$	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$k = 10$	T	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	S	0.002	0.001	0.001	0.001	0.002	0.002	–0.011	0.003	–0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	% rejection 0.05	4.9	5.6	4.9	5.1	4.9	5.9	4.7	5.5	4.8
	% rejection 0.10	9.5	10.9	10.0	10.4	10.0	11.0	9.5	10.4	10.1
	Cluster FE	No	Yes	No	Yes	No	Yes	No	Yes	No

Notes: Data generation process, $y_{li} = 10 + 5x_{li} + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. FE, fixed effects; T, coefficient estimate predicted by theory outlined in section 4.2; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

Figure 3 shows the test rejection rates in case 2 ($N = 200$; $L = 20$; $k = 10$), now based on the 2SLS estimate of the endogenous peer effect β_1. The actual rejection rates now closely follow the expected rejection rates (i.e. 45 degree line), suggesting that the 2SLS estimate of the endogenous peer effect is, indeed, unbiased.

Figure 3 Rejection rates $H_0 : \beta_1 = 0$ (case 2 – 2SLS: $N = 200$; $L = 20$; $k = 10$)



6. Extensions of the basic model

In practice, determining the magnitude of the exclusion bias will not always be as straightforward as in the basic models considered above. Until now, we have assumed that all peers are randomly drawn from within a well-specified pool of individuals, of fixed size N_P. In this section, I briefly discuss some possible extensions of the standard model and anticipate how these would alter the main results.

First, in the basic cluster sampling models considered in section 3, I assumed all clusters (e.g. classrooms) were of equal size L. In reality, however, groups often vary in size. Allowing for varying group sizes L_l, in Appendix A.5, I show how the analytical expression for the magnitude of the bias in a cluster fixed effects model (with peers selected at the cluster level) becomes:

$$|Bias| = \frac{\left[\sum_{l=1}^{\frac{N}{L}} \frac{L_l - 1}{(L_l - 1 + \beta_1)L_l} \times L_l\right]}{\left[\sum_{l=1}^{\frac{N}{L}} \frac{(L_l - 1)^2 \frac{(L_l - 1 - k)}{(L_l - 1)k} + 1}{(L_l - 1 + \beta_1)^2} \times L_l\right]} \tag{25}$$

Intuitively, expression (25) is the result of assigning different weights to the sampled observations, in which the weight is determined by the share of the individuals' group l in the total sample size, $\frac{L_l}{N}$. To test the validity of expression (25), I run a simulation with 200 individuals ($N = 200$) allocated to five different groups that each differ in size: $l_1 = 20$, $l_2 = 30$, $l_3 = 40$, $l_4 = 50$, $l_5 = 60$. I set each individual to randomly select five peers from within their clusters ($k = 5$). The result of a simulation with a DGP structure identical to that used in Table 3 (with 10,000 repetitions) indicates that the exclusion bias as estimated by this simulation, –0.147, is very similar to the predicted exclusion bias of –0.146 that is obtained by inserting the values for k and the various values for L_l into equation (25).

Second, the pool of potential peers might be correlated with, but not necessarily perfectly overlap, the cluster. For instance, although students tend to select many of their friends from within their classrooms, they usually also have peers other than classmates. Such empirical specifications would be characterised by properties of both case 1 and case 2 considered above: peer group formation is likely to be correlated with cluster formation (case 1), but some peers are selected from outside the cluster (case 2). Formally, consider a specification in which a proportion, θ, of peers is selected from within the classroom and a proportion, $(1-\theta)$, is selected from outside the classroom. Instead of $E(\bar{y}_{k_{li}}) = \bar{y}_{l,-i}$ that we considered in the previous sections, we would now model $E(\bar{y}_{k_{li}})$ as follows:

$$E(\bar{y}_{k_{li}}) = \theta\bar{y}_{l,-i} + (1-\theta)\bar{y}_{\Omega,-l} \tag{26}$$

where $0 \leq \theta \leq 1$ and $\bar{y}_{\Omega,-l}$ is the average outcome of the entire network Ω, excluding individual i's group l. Although a further treatment of this extension is beyond the scope of this thesis, it is expected that, in a classroom fixed effects model, the exclusion bias would be smaller when at least one peer is selected from outside the cluster ($\theta < 1$) than when all peers are drawn from within the classroom ($\theta = 1$). Intuitively, in a model adding cluster fixed effects (which looks only at the variation in outcomes within the cluster), the exclusion bias is driven by a dependency of the expected peer group outcome $E(\bar{y}_{k_{li}})$ on $\bar{y}_{l,-i}$, and this dependency will be smaller whenever $\theta < 1$.

It is important to note here a related property of models that add cluster fixed effects: whereas an increase in peer group size unambiguously increases the magnitude of the exclusion bias as long as all peers are drawn from within the cluster (i.e. $\frac{\Delta\left|bias^{FE}\right|}{\Delta k^{inside}} > 0$), the bias becomes insensitive to additional peers who are drawn from outside the cluster (i.e. $\frac{\Delta\left|bias^{FE}\right|}{\Delta k^{outside}} = 0$).

A third complication is that, in practice, the pool of potential peers is not always well defined. Empirical studies differ in the value they – implicitly – attach to θ in equation (26). In some studies on peer effects in educational achievement, survey respondents are asked to identify peers from the entire school roster (e.g. Halliday and Kwak, 2012; Fletcher, 2012), thus $\theta < 1$.[14] In other studies, people are restricted to a list of all students in the classroom in the identification of their peers (e.g. de Melo, 2011), thus $\theta = 1$. Fletcher and Ross (2012), in an attempt to control for correlated effects, construct 'clusters of observationally equivalent individuals who face the same friendship opportunity set and make the same type of

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

friendship choices' within the school. To the extent that students select peers from within these clusters, it would be the size of these groups that would be relevant for the derivation of the magnitude of the exclusion bias. In reality, however, the boundaries of the pools from which peers are selected are usually difficult to draw.

Fourth, even if the exact pool of potential peers is perfectly known, in reality peers are usually not randomly drawn from within those pools. Therefore, the expected value of the outcomes of the selected peers might differ from the average of those pools (excluding individual i). In such cases, although we would expect the bias to be similarly present, the derivation of its magnitude would become considerably more complex.

Lastly, in this paper we have only considered simple bivariate models. However, the addition of statistically significant control variables to the model is expected to reduce the magnitude of the exclusion bias. This is because, to the extent that these covariates explain some variation in the outcome variable y, they will partially control for the difference between i and the other individuals in the sample. Hence, they will partially capture the difference in expected value of the pools from which peers are drawn. Simulations indicate (not reported) that the reduction in the magnitude of the exclusion bias in models that add control variables with high coefficient parameters (e.g. 0.8) can be very significant. The reduction is negligible, however, when the coefficient parameters are relatively small (e.g. 0.20).

In general, when it is difficult to rely on theory for reasons outlined in this section, simulations of the type we ran in the previous section might be more useful in estimating the magnitude of the exclusion bias.

7. Conclusion

This paper has formalised a cited (Guryan et al., 2009) but so far unproven source of downward estimation bias of OLS in standard empirical peer effects models. The main results derived in this paper have various implications for the empirical social interaction literature.

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

Furthermore, although I certainly do not argue that one should prefer models ignoring cluster fixed effects, I do caution against naive comparisons of these models with those that do control for such effects, in terms of their respective estimation results. As shown above, the exclusion bias is aggravated in the latter type of models, whenever peer group formation is correlated with cluster formation. For example, studies adding classroom effects (e.g. de

Melo, 2011), dormitory effects (e.g. Sacerdote, 2001) or school effects (e.g. Fletcher, 2012) will be more severely affected by the exclusion bias if peers are selected – partially or completely – from within the respective clusters. The literature tends to interpret a drop in the estimate of the endogenous peer effect in models adding cluster fixed effects as evidence of unobserved covariates at the cluster level. Although such correlates will often matter, the results in this paper suggest that such interpretations of the results may be confounded by the presence of the exclusion bias.

The finding that the downward exclusion bias, present in OLS estimation, is eliminated by 2SLS estimation strategies is equally important. It provides an alternative explanation for the common but counter-intuitive tendency of peer effects studies to obtain 2SLS estimates of endogenous peer effects that are larger than their OLS counterparts (e.g. Goux and Maurin, 2007; De Giorgi et al., 2010; de Melo, 2011; Brown and Laschever, 2012; Helmers and Patnam, 2012; Krishnan and Patnam, 2012; Naguib, 2012; Collin, 2013).

Finally, note that not all peer effects studies are affected by the exclusion bias. For example, there would be no problem if the peer group consists of *all* the individuals in student i's classroom, including student i herself. Including the individual in her own peer group average outcome computations is rarely done in practice, however, because of the associated reflection problem (Manski, 1993). Nor would there be a problem if the pool of potential peers consists entirely of people who do not form part of the study sample. For example, a study on peer effects between male students from an all-boys school and female students of a neighbouring all-girls school would not be affected by this bias if the units of interest are restricted to either male or female students. In most studies, however, peers are included in the study sample. The greater the overlap between the set of peers and the set of observations captured in the study sample, the greater the exclusion bias. Lastly, studies regressing an individual's current outcome levels on peers' baseline outcomes also prevent the exclusion bias from manifesting itself, as long as they control for the individual's own baseline outcome (e.g. Zimmerman, 2003; Lyle, 2007, 2009; Duflo et al., 2011; Fafchamps and Quinn, 2012).

If not properly addressed, however, the exclusion bias may significantly affect the results of social interaction studies. Its consequences may be especially problematic for empirical analyses that consider cluster fixed effects, for which the magnitude of the bias does not decrease as the sample size tends to infinity. This paper has provided more insights into the cause, consequences and solutions of this bias, which has largely been ignored to date. Exploring other means of addressing this concern would be desirable in the future.

References

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

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

Collin, M., 2013. Peer effects in the adoption of property rights: experimental evidence from urban Tanzania. Unpublished paper presented at a conference held at the Centre for the Study of African Economies, University of Oxford, UK.

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

de Melo, G., 2011. Peer effects identified through social networks. Evidence from Uruguayan schools. Working Paper, Department of Economics, University of Siena, 627, Department of Economics, University of Siena, Italy. Available at http://www.econ-pol.unisi.it/dipartimento/it/node/1606.

Duflo, E. and Saez, E., 2002. Participation and investment decisions in a retirement plan: the influence of colleagues' choices. *Journal of Public Economics* 85 (1), 121–148.

Duflo, E., Dupas, P. and Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. *American Economic Review* 101 (5), 1739–1774.

Durlauf, S.N. and Ioannides, Y.M., 2010. Social interactions. *Annual Review of Economics* 2, 451–478.

Fafchamps, M. and Quinn, S., 2012. Networks and manufacturing firms in Africa: initial results from a randomised experiment. Working Paper, Centre for the Study of African Economies, University of Oxford, UK.
Available at http://users.ox.ac.uk/~econ0087/networks.pdf.

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

Fletcher, J.M. and Ross, S.L., 2012. Estimating the effects of friendship networks on health behaviors of adolescents. Working Paper no. 18253, National Bureau of Economic Research, Cambridge, MA.

Glaeser, E.L., Sacerdote, B.I. and Scheinkman, J.A., 2002. The social multiplier. Discussion Paper no. 1968, Harvard Institute of Economic Research, Cambridge, MA. Available at http://post.economics.harvard.edu/hier/2002papers/2002list.html.

Goux, D. and Maurin, E., 2007. Close neighbours matter: neighbourhood effects on early performance at school. *The Economic Journal* 117 (523), 1193–1215.

Graham, B.S., 2008. Identifying social interactions through conditional variance restrictions. *Econometrica* 76 (3), 643–660.

Graham, B.S. and Hahn, J.Y., 2005. Identification and estimation of the linear-in-means model of social interactions. *Economics Letters* 88 (1), 1–6.

Guryan, J., Kroft, D. and Notowidigdo, N.J., 2009. Peer effects in the workplace: evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics* 1 (4), 34–68.

Halliday, T.J. and Kwak, S., 2012. What is a peer? The role of network definitions in estimation of endogenous peer effects. *Applied Economics* 44 (3), 289–302.

Hanushek, E.A., Kain, J.F., Markman, J.M. and Rivkin, S.G., 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* 18 (5), 527–544.

Helmers, C. and Patnam, M. 2012. Does the rotten child spoil his companion? Spatial peer effects among children in rural India. Working Paper WPS/2010-13, Centre for the Study of African Economies, University of Oxford, UK.

Hoxby, C., 2000. Peer Effects in the classroom: learning from gender and race variation. Working Paper no. 7867, National Bureau of Economic Research, Cambridge, MA.

Lee, L.F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140 (2), 333-374.

Lin, X., 2007. Peer effects and student academic achievement: An application of spatial autoregressive model with group unobservables, Mimeo, Department of Economics, Tsinghyua University, Beijing.

Lyle, D., 2007. Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics* 89 (2), 289–299.

Lyle, D., 2009. The effects of peer group heterogeneity on the production of human capital at West Point. *American Economic Journal: Applied Economics* 1 (4), 69–84.

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

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

Moffitt, R.A., 2001. Policy interventions, low level equilibria, and social interactions. In: Durlauf, S.N. and Young, H.P. (eds), *Social Dynamics*. MIT Press, Cambridge, MA, pp. 45–82.

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

Wang, L.C., 2010. Peer effects in the classroom: evidence from a natural experiment in Malaysia. Working Paper, Department of Economics Seminar Series, University of California, San Diego, CA.

Available at http://economics.missouri.edu/seminars/files/2010/100129.pdf.

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

Appendices

A.1. Relationship between σ_u^2 and σ_ε^2

We can rewrite equation (3) as follows:

$$u_i = \bar{y}_{k_i} - \bar{y}_{-i} = \frac{\sum_{j=1}^{k} y_{j_i}}{k} - \left(\frac{\sum_{j=1}^{N} y_{j_i} - y_i}{N-1} \right)$$

$$= \frac{\sum_{j=1}^{k} y_{j_i}}{k} - \left(\frac{\sum_{j=1}^{k} y_{j_i} + \sum_{j=k+1}^{N} y_{j_i} - y_i}{N-1} \right)$$

$$= \frac{(N-1)\sum_{j=1}^{k} y_{j_i}}{(N-1)k} - \frac{k\sum_{j=1}^{k} y_{j_i}}{(N-1)k} - \frac{\sum_{j=k+1}^{N} y_{j_i}}{(N-1)} + \frac{y_i}{(N-1)}$$

$$= \frac{(N-1-k)\sum_{j=1}^{k} y_{j_i}}{(N-1)k} - \frac{\left(\sum_{j=k+1}^{N} y_{j_i} - y_i\right)}{(N-1)}$$

Using this and $\mathrm{var}(y_1) = \ldots = \mathrm{var}(y_n) = \sigma_\varepsilon^2$, we can now derive $\mathrm{var}[u_i]$:

$$\mathrm{var}(u_i) = \sigma_u^2 = \frac{(N-1-k)^2 k\sigma_\varepsilon^2}{(N-1)^2 k^2} + \frac{(N-1-k)\sigma_\varepsilon^2}{(N-1)^2}$$

$$= \frac{(N-1-k)^2 k + k^2(N-1-k)}{(N-1)^2 k^2}\sigma_\varepsilon^2$$

$$= \frac{(N-1-k)k(N-1-k+k)}{(N-1)^2 k^2}\sigma_\varepsilon^2 \qquad \Rightarrow \mathrm{var}(u_i) = \sigma_u^2 = \frac{(N-1-k)}{(N-1)k}\sigma_\varepsilon^2 < \sigma_\varepsilon^2$$

$$= \frac{(N-1-k)}{(N-1)k}\sigma_\varepsilon^2$$

(27)

A.2. Deriving an expression for $\mathrm{var}(\bar{y}_{k,-i})$

Using the reduced form of $\bar{y}_{-i}$ provided in equation (8), we obtain:

$$\mathrm{var}(\bar{y}_{k,-i}) = \mathrm{var}(\bar{y}_{-i} + u_i) = \mathrm{var}(\bar{y}_{-i}) + 2\,\mathrm{cov}(\bar{y}_{-i}, u_i) + \mathrm{var}(u_i)$$

$$= \frac{\beta_1^2 \sigma_u^2}{\left(N-1+\beta_1\right)^2} + \frac{\sigma_\varepsilon^2}{\left(N-1+\beta_1\right)^2} + 2\frac{-\beta_1 \sigma_u^2}{\left(N-1+\beta_1\right)} + \sigma_u^2$$

$$= \frac{(N-1)^2 \sigma_u^2 + \sigma_\varepsilon^2}{\left(N-1+\beta_1\right)^2} \tag{28}$$

A.3. Exclusion bias in models adding cluster fixed effects (equal cluster sizes)

Starting from equation (11), we substitute for $\bar{y}_{\Pi,-i}$ by the cluster sampling equivalent of equation (6):

$$y_{li} = \beta_0 + \beta_1 \left(\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - y_{li}}{N_P - 1} + u_{li} \right) + \varepsilon_{li} \tag{29}$$

Averaging equation (29) over all L observations in group l, we obtain the group average outcome $\bar{y}_l$:

$$\bar{y}_l = \beta_0 + \beta_1 \left[\frac{\sum_{i=1}^{L} \frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - y_{li}}{N_P - 1}}{L} + \bar{u}_l \right] + \bar{\varepsilon}_l = \beta_0 + \beta_1 \left[\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - \bar{y}_l}{N_P - 1} + \bar{u}_l \right] + \bar{\varepsilon}_l$$

$$\Rightarrow \bar{y}_l = \beta_0 + \beta_1 \left[\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - \bar{y}_l}{N_P - 1} + \bar{u}_l \right] + \bar{\varepsilon}_l \tag{30}$$

Subtracting equation (30) from equation (29), we derive the following expression for the cluster fixed effects model:

$$y_{li}-\bar{y}_l=\beta_1\left[\frac{\left(\sum_{j=1}^{N_P}y_{lj}\right)-y_{li}}{N_P-1}+u_{li}-\left[\frac{\left(\sum_{j=1}^{N_P}y_{lj}\right)-\bar{y}_l}{N_P-1}\right]-\bar{u}_l\right]+\varepsilon_{li}-\bar{\varepsilon}_l$$

$$\Leftrightarrow y_{li}-\bar{y}_l=\beta_1\left(\frac{\bar{y}_l-y_{li}}{N_P-1}+u_{li}-\bar{u}_l\right)+\varepsilon_{li}-\bar{\varepsilon}_l \tag{31}$$

Denoting $\ddot{x}=x_{li}-\bar{x}_l$, for $x=y,u,\varepsilon$, we have:

$$\ddot{y}=\beta_1\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u}\right)+\ddot{\varepsilon} \tag{32}$$

Using the properties of the covariance and variance operators, I obtain the following expression for the cluster fixed effects estimate of β_1:

$$E\left(\hat{\beta}_1^{FE}\right)=\beta_1+\frac{\operatorname{cov}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u},\ddot{\varepsilon}\right)}{\operatorname{var}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u}\right)}$$

$$=\beta_1+\frac{\operatorname{cov}\left(\frac{-\ddot{y}}{N_P-1},\ddot{\varepsilon}\right)+\operatorname{cov}(\ddot{u},\ddot{\varepsilon})}{\operatorname{var}\left(\frac{-\ddot{y}}{N_P-1}\right)+2\operatorname{cov}\left(\frac{-\ddot{y}}{N_P-1},\ddot{u}\right)+\operatorname{var}(\ddot{u})} \tag{33}$$

In order to expand equation (33), we consider:

$$\operatorname{cov}(\ddot{u},\ddot{\varepsilon})=E(\ddot{u}\ddot{\varepsilon})=E\left[(u_{li}-\bar{u}_l)(\varepsilon_{li}-\bar{\varepsilon}_l)\right]$$

$$=E(u_{li}\varepsilon_{li})-E(\bar{u}_l\varepsilon_{li})+E(\bar{u}_l\bar{\varepsilon}_l)-E(u_{li}\bar{\varepsilon}_l)=0 \tag{34}$$

and

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

Furthermore, since $\bar{u}_l=\frac{\sum_{i=1}^{L}u_{li}}{L}$ and u is assumed independent across individuals:

$$\begin{cases} E(u_{li}\bar{u}_l) = \dfrac{E(u_{li}^2)}{L} = \dfrac{\sigma_u^2}{L} \\ \operatorname{var}(\bar{u}_l) = \operatorname{var}\left(\dfrac{\sum_{i=1}^{L} u_{li}}{L}\right) = \dfrac{\sum_{i=1}^{L} \operatorname{var}(u_{li})}{L^2} = \dfrac{L\sigma_u^2}{L^2} = \dfrac{\sigma_u^2}{L} \end{cases}$$

$$\Rightarrow \operatorname{var}(\ddot{u}) = \sigma_u^2 - 2\frac{\sigma_u^2}{L} + \frac{\sigma_u^2}{L} = \frac{(L-1)\sigma_u^2}{L} \tag{35}$$

Similarly:

$$\operatorname{var}(\ddot{\varepsilon}) = \frac{(L-1)\sigma_\varepsilon^2}{L} \tag{36}$$

Using equation (32), we derive the reduced form of $\left(-\dfrac{\ddot{y}}{N_P - 1}\right)$:

$$\ddot{y} = \beta_1\left(\frac{-\ddot{y}}{N_P - 1} + \ddot{u}\right) + \ddot{\varepsilon} \Leftrightarrow \left[\frac{N_P - 1 + \beta_1}{N_P - 1}\right]\ddot{y} = \beta_1\ddot{u} + \ddot{\varepsilon}$$

$$\Leftrightarrow -\frac{\ddot{y}}{N_P - 1} = \frac{-\beta_1\ddot{u}}{N_P - 1 + \beta_1} - \frac{\ddot{\varepsilon}}{N_P - 1 + \beta_1} \tag{37}$$

Using $E(\ddot{\varepsilon}) = E(\varepsilon_{li} - \bar{\varepsilon}_l) = 0$, equations (34), (36) and (37), we derive:

$$\begin{aligned} \operatorname{cov}\left(\frac{-\ddot{y}}{N_P - 1}, \ddot{\varepsilon}\right) &= E\left[\left[\frac{-\ddot{y}}{N_P - 1} - E\left(\frac{-\ddot{y}}{N_P - 1}\right)\right]\ddot{\varepsilon}\right] \\ &= E\left[-\frac{\beta_1\ddot{u}\ddot{\varepsilon}}{N_P - 1 + \beta_1}\right] + E\left[\frac{-\ddot{\varepsilon}\ddot{\varepsilon}}{N_P - 1 + \beta_1}\right] \\ &= \frac{-\beta_1 E(\ddot{u}\ddot{\varepsilon}) - \operatorname{var}(\ddot{\varepsilon})}{N_P - 1 + \beta_1} \\ &= \frac{-\operatorname{var}(\ddot{\varepsilon})}{N_P - 1 + \beta_1} \\ &= -\frac{1}{N_P - 1 + \beta_1} \cdot \frac{(L-1)\sigma_\varepsilon^2}{L} \end{aligned} \tag{38}$$

Similarly,

$$2\operatorname{cov}\left(\frac{-\ddot{y}}{N_P-1},\ddot{u}\right)=-2\frac{\beta_1\operatorname{var}(\ddot{u})}{N_P-1+\beta_1}-2\frac{E(\ddot{u}\ddot{\varepsilon})}{N_P-1+\beta_1}$$

$$=-2\frac{\beta_1\operatorname{var}(\ddot{u})}{N_P-1+\beta_1}=-\frac{2\beta_1}{N_P-1+\beta_1}.\frac{(L-1)\sigma_u^2}{L} \tag{39}$$

Again using equation (37), we have:

$$\operatorname{var}\left(\frac{-\ddot{y}}{N_P-1}\right)=\operatorname{var}\left[\frac{-\beta_1\ddot{u}}{N_P-1+\beta_1}-\frac{\ddot{\varepsilon}}{N_P-1+\beta_1}\right]$$

$$=\operatorname{var}\left(\frac{-\beta_1\ddot{u}}{N_P-1+\beta_1}\right)+\frac{2\beta_1\operatorname{cov}(\ddot{u},\ddot{\varepsilon})}{(N_P-1+\beta_1)^2}+\operatorname{var}\left(\frac{-\ddot{\varepsilon}}{N_P-1+\beta_1}\right)$$

$$=\frac{\beta_1^2\frac{(L-1)\sigma_u^2}{L}}{(N_P-1+\beta_1)^2}+\frac{\frac{(L-1)\sigma_u^2}{L}}{(N_P-1+\beta_1)^2} \tag{40}$$

Using equations (34)–(40), we obtain:

$$\operatorname{cov}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u},\ddot{\varepsilon}\right)=\left(-\frac{1}{N_P-1+\beta_1}.\frac{(L-1)\sigma_\varepsilon^2}{L}\right) \tag{41}$$

and

$$\operatorname{var}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u}\right)=\left[\frac{(N_P-1)^2\sigma_u^2+\sigma_\varepsilon^2}{(N_P-1+\beta_1)^2}\right]\frac{(L-1)}{L} \tag{42}$$

Hence, we can expand equation (33) as follows:

$$E\left(\hat{\beta}_1^{FE}\right)=\beta_1+\frac{\operatorname{cov}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u},\ddot{\varepsilon}\right)}{\operatorname{var}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u}\right)}=\beta_1+\frac{\left(-\frac{1}{N_P-1+\beta_1}\frac{(L-1)\sigma_\varepsilon^2}{L}\right)}{\left[\frac{(N_P-1)^2\sigma_u^2+\sigma_\varepsilon^2}{(N_P-1+\beta_1)^2}\right]\frac{(L-1)}{L}}$$

$$\Rightarrow E\left(\hat{\beta}_1^{FE}\right)=\beta_1-\frac{\sigma_\varepsilon^2(N_P-1+\beta_1)}{(N_P-1)^2\sigma_u^2+\sigma_\varepsilon^2}$$

Inserting the expression for σ_u^2 given by equation (27), we finally have:

$$E\left(\hat{\beta}_1^{FE}\right)=\beta_1-\frac{\left(N_P-1+\beta_1\right)k}{\left(N_P-1\right)\left(N_P-1-k\right)+k} \tag{43}$$

A.4. Exclusion bias in the between estimation model

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

$$\begin{aligned}\bar{y}_{l,-i} &= \frac{\sum_{j=1}^{L} y_j-\beta_0}{L-1+\beta_1}-\frac{\beta_1\bar{u}_l}{L-1+\beta_1}-\frac{\bar{\varepsilon}_l}{L-1+\beta_1} \\ &= \frac{L\bar{y}_l-\beta_0}{L-1+\beta_1}-\frac{\beta_1\bar{u}_l}{L-1+\beta_1}-\frac{\bar{\varepsilon}_l}{L-1+\beta_1}\end{aligned} \tag{44}$$

where $\bar{y}_{l,-i}$ is the average outcome over the individuals in the group, excluding individual i, and $\bar{y}_l$, $\bar{u}_l$ and $\bar{\varepsilon}_l$ denote the group averages of y, u and ε, respectively.

Using equation (44), we find $\operatorname{cov}\left(\bar{\bar{y}}_{k_l},\bar{\varepsilon}_l\right)$ and $\operatorname{var}\left(\bar{\bar{y}}_{k_l}\right)$, which we will need to find $p\lim\left(\hat{\beta}_1^{BE}\right)$ below:

$$\begin{aligned}\operatorname{cov}\left(\bar{\bar{y}}_{k_l},\bar{\varepsilon}_l\right) &= \operatorname{cov}\left(\bar{y}_{l,-i}+\bar{u}_l,\bar{\varepsilon}_l\right)=\operatorname{cov}\left(\bar{y}_{l,-i},\bar{\varepsilon}_l\right) \\ &= L\frac{E\left(\bar{\varepsilon}_l^2\right)}{L-1+\beta_1}-\frac{E\left(\bar{\varepsilon}_l^2\right)}{L-1+\beta_1}=\frac{L-1}{L-1+\beta_1}\operatorname{var}\left(\bar{\varepsilon}_l\right) \\ &= \frac{\left(L-1\right)\sigma_\varepsilon^2}{\left(L-1+\beta_1\right)L}=\frac{L-1}{L-1+\beta_1}\operatorname{var}\left(\frac{\sum_{j=1}^{L}\varepsilon_j}{L}\right)\end{aligned} \tag{45}$$

$$\mathrm{var}\left(\bar{\bar{y}}_{k_l}\right)=\mathrm{var}\left(\frac{\sum_{i=1}^{L}\bar{y}_{k_i}}{L}\right)=\frac{1}{L^2}\mathrm{var}\left(\sum_{i=1}^{L}\left(\frac{\sum_{j=1}^{L}y_{lj}-y_{li}}{L-1}\right)+\sum_{i=1}^{L}u_{li}\right)$$

$$=\frac{1}{L^2}\mathrm{var}\left(\frac{L\sum_{i=1}^{L}y_{li}}{L-1}-\frac{\sum_{i=1}^{L}y_{li}}{L-1}+\sum_{i=1}^{L}u_{li}\right)$$

$$=\frac{1}{L^2}\mathrm{var}\left(\sum_{i=1}^{L}y_{li}+\sum_{i=1}^{L}u_{li}\right)=\frac{\sigma_\varepsilon^2+\sigma_u^2}{L}$$

$$=\frac{\left[1+\frac{(L-1-k)}{(L-1)k}\right]\sigma_\varepsilon^2}{L} \tag{46}$$

where in the last step we used equation (27).

Using equations (45) and (46), we finally obtain:

$$p\lim\left(\hat{\beta}_1^{BE}\right)=\beta_1+\frac{\mathrm{cov}\left(\bar{\bar{y}}_{k_l},\bar{\varepsilon}_l\right)}{\mathrm{var}\left(\bar{\bar{y}}_{k_l}\right)}=\frac{\frac{(L-1)\sigma_\varepsilon^2}{(L-1+\beta_1)L}}{\frac{\left[1+\frac{(L-1-k)}{(L-1)k}\right]\sigma_\varepsilon^2}{L}}$$

$$\Rightarrow p\lim\left(\hat{\beta}_1^{BE}\right)=\beta_1+\frac{(L-1)^2k}{(L-1+\beta_1)[(L-1)k+(L-1-k)]} \tag{47}$$

A.5. Derivation of $p\lim\left(\eta^2\right)$

Weight parameter η^2 in equation (15) is the ratio of the between-group sum of squares of the independent variable of interest, $\bar{y}_{k_i}$, to its total sum of squares:

$$\eta^2=\frac{SS_{\bar{y}_k}^{BG}}{SS_{\bar{y}_k}^{Total}}=\frac{SS_{\bar{y}_k}^{BG}}{SS_{\bar{y}_k}^{BG}+SS_{\bar{y}_k}^{Within}} \tag{48}$$

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

$$SS_{\bar{y}_k}^{BG} = SS_{\bar{y}_k}^{BE} \times L \tag{49}$$

where $SS_{\bar{y}_k}^{BE}$ is the sum of squares of $\bar{\bar{y}}_{k_i}$ in the between estimation regression (35) in section 3.5. Furthermore, using the definition of the variance operator, we know that:

$$\operatorname{var}\left(\bar{\bar{y}}_{k_l}\right) = \frac{SS_{\bar{y}_k}^{BE}}{\left(\frac{N}{L}-1\right)} \Rightarrow SS_{\bar{y}_k}^{BE} = \operatorname{var}\left(\bar{\bar{y}}_{k_l}\right) \times \left(\frac{N}{L}-1\right) \tag{50}$$

Using equations (48)–(50), we obtain:

$$SS_{\bar{y}_k}^{BG} = \operatorname{var}\left(\bar{\bar{y}}_{k_l}\right) \times \left(\frac{N}{L}-1\right) \times L$$

Substituting in for the expression of $\operatorname{var}\left(\bar{\bar{y}}_{k_l}\right)$ given by equation (46), we finally have:

$$SS_{\bar{y}_k}^{BG} = \frac{(Lk-2k+L-1)}{k(L-1)}\left(\frac{N}{L}-1\right)\sigma_{\varepsilon}^2 \tag{51}$$

Next, $SS_{\bar{y}_k}^{Within}$ is the sum of the squared differences between each individual's average peer group outcome, $\bar{y}_{k_i}$, and its average for the individual's group $\bar{\bar{y}}_{k_l}$. Similarly to equation (50), we have:

$$\operatorname{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right) = \frac{SS_{\bar{y}_k}^{Within}}{N-1} \Rightarrow SS_{\bar{y}_k}^{Within} = \operatorname{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right) \times (N-1)$$

From the above, we know that $\operatorname{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right) = \operatorname{var}\left(\frac{-\ddot{y}}{N_P-1} + \ddot{u}\right)$. Therefore, we can substitute in for the expression of $\operatorname{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right)$ by using equations (42) and (27).

We have:

$$SS_{\bar{y}_k}^{Within} = \left[\frac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]\frac{1}{L}(N-1) \tag{52}$$

Combining equations (48), (51) and (52), we obtain:

$$\eta^2 = \frac{SS_{\bar{y}_k}^{BG}}{SS_{\bar{y}_k}^{BG} + SS_{\bar{y}_k}^{Within}} \tag{53}$$

where:

$$\begin{cases} SS_{\bar{y}_k}^{BG} = \dfrac{(Lk-2k+L-1)}{k(L-1)}\left(\dfrac{N}{L}-1\right)\sigma_\varepsilon^2 \\ SS_{\bar{y}_k}^{Within} = \left[\dfrac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]\dfrac{(N-1)}{L}\sigma_\varepsilon^2 \end{cases}$$

Finally, taking probability limits, we obtain the following expression for $p\lim(\eta^2)$:

$$p\lim(\eta^2) = \frac{p\lim(SS_{\bar{y}_k}^{BG})}{p\lim(SS_{\bar{y}_k}^{BG}) + p\lim(SS_{\bar{y}_k}^{Within})}$$

$$= \frac{\dfrac{(Lk-2k+L-1)}{k(L-1)}}{\dfrac{(Lk-2k+L-1)}{k(L-1)} + \left[\dfrac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]} \tag{54}$$

Note that this result only holds in the limit, that is, when the sample size N tends to infinity. For small samples, however, $E(\eta^2) < p\lim(\eta^2)$. To see this, consider:

$$E(\eta^2) = E\left[\frac{(SS_{\bar{y}_k}^{BG})}{(SS_{\bar{y}_k}^{BG}) + (SS_{\bar{y}_k}^{Within})}\right]$$

$$= E(SS_{\bar{y}_k}^{BG})E\left(\frac{1}{(SS_{\bar{y}_k}^{BG}) + (SS_{\bar{y}_k}^{Within})}\right) + \operatorname{cov}\left(SS_{\bar{y}_k}^{BG}, \frac{1}{(SS_{\bar{y}_k}^{BG}) + (SS_{\bar{y}_k}^{Within})}\right)$$

Although the derivation of $\operatorname{cov}\left(SS_{\bar{y}_k}^{BG}, \frac{1}{\left(SS_{\bar{y}_k}^{BG}\right)+\left(SS_{\bar{y}_k}^{Within}\right)} \right)$ is beyond the scope of this thesis, we know that $\operatorname{cov}\left(SS_{\bar{y}_k}^{BG}, \frac{1}{\left(SS_{\bar{y}_k}^{BG}\right)+\left(SS_{\bar{y}_k}^{Within}\right)} \right) < 0$ (this result is also confirmed by simulations). Using this result, it can easily be shown that:

$$E\left(\eta^2\right) < p\lim\left(\eta^2\right) \tag{55}$$

The smaller the sample, the smaller $E\left(\eta^2\right)$, the more weight is given to the cluster fixed effect estimator, and, therefore, the more severe the exclusion bias.

A.6. Exclusion bias in models adding cluster fixed effects, with varying cluster sizes

As before, the cluster fixed effects estimate for β_1 in $y_{li} - \bar{y}_l = \beta_1\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}\right) + \varepsilon_{li} - \bar{\varepsilon}_l$ is approached by:

$$p\lim\left(\hat{\beta}_1^{FE}\right) = \beta_1 + \frac{\operatorname{cov}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}, \varepsilon_{li} - \bar{\varepsilon}_l\right)}{\operatorname{var}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}\right)}$$

However, now the within covariance and within variance will be weighted averages of the respective within covariances and within variances of groups l in the sample. Using equations (41) and (42), we obtain:

$$|Bias| = -\frac{\sum_{l=1}^{\frac{N}{L}} \operatorname{cov}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}, \varepsilon_{li} - \varepsilon_l\right) \times \frac{L_l}{N}}{\sum_{l=1}^{\frac{N}{L}} \operatorname{var}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}\right) \times \frac{L_l}{N}}$$

$$= \frac{\left[\sum_{l=1}^{\frac{N}{L}} \frac{L_l - 1}{\left(L_l - 1 + \beta_1\right)L_l} \times L_l \right]}{\left[\sum_{l=1}^{\frac{N}{L}} \frac{\left(L_l - 1\right)^2 \frac{\left(L_l - 1 - k\right)}{\left(L_l - 1\right)k} + 1}{\left(L_l - 1 + \beta_1\right)^2} \times L_l \right]} \tag{56}$$