OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 81, 5 (2019) 0305–9049
doi: 10.1111/obes.12299

The Use of Instrumental Variables in Peer Effects Models*

Stephanie von Hinke,†,‡,§ George Leckie¶ and Cheti Nicoletti††,‡‡

†*Department of Economics, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK*
‡*Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands*
§*Institute for Fiscal Studies, London, UK (e-mail: s.vonhinke@bristol.ac.uk)*
¶*Centre for Multilevel Modelling, University of Bristol, Bristol, UK (e-mail: g.leckie @bristol.ac.uk)*
††*Department of Economics and Related Studies, University of York, Heslington, York YO10 5DD, UK*
‡‡*ISER, University of Essex, Colchester, UK (e-mail: cheti.nicoletti@york.ac.uk)*

Abstract

Instrumental variables are often used to identify peer effects. This paper shows that instrumenting the 'peer average outcome' with 'peer average characteristics' requires the researcher to include the instrument at the individual level as an explanatory variable. We highlight the bias that occurs when failing to do this.

I. Introduction

Many papers in economics provide empirical evidence on the causal effect of peers on individual outcomes using an instrumental variable (IV) approach. They usually consider linear in mean regressions of an individual outcome on the corresponding average outcome of peers and a set of individual explanatory variables. They may then instrument the average outcome of peers with the peer average of certain characteristics.[1]

JEL Classification numbers: C31, C36, D01, I12, I20.

*The authors thank the editor, three anonymous referees, Michèle Belot, Peter Burridge, Fernanda Leite Lopez de Leon, Anita Ratcliffe, Kim Scharf, Stefania Sitzia and Frank Windmeijer for helpful suggestions. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 092731) and the University of Bristol provided core support for ALSPAC. We gratefully acknowledge financial support from the UK Medical Research Council (G1002345) and the UK Economic and Social Research Council (RES-576-25-0032).

[1] Different types of instruments have been used, including, (i) the average price of peers' decisions which is exogenously shifted by the introduction of policy or programme affecting only some of the people (see the 'partial-population' identification approach defined by Moffitt (2001), and the application in Dahl, Loken and Mogstad (2014)); (ii) peer averages of predetermined variables that affect peers but only influence the individual outcome

As in any other standard linear regression, the IV estimator consistently estimates the causal peer effect if the instruments are as good as randomly assigned (independence), irrelevant in explaining the individual outcome except through the average peers' outcome (exclusion), and relevant in explaining the endogenous outcome averaged across peers (relevance).[2]

The contribution of this paper is to highlight a subtle, but important implication of the relevance assumption, something not explicitly recognized in this literature: the individual variable, say x, whose peer average, say $\bar{x}$, is used to instrument the peer average outcome $\bar{y}$ must be included as an individual explanatory variable of the dependent variable y. The idea is simple: if $\bar{x}$ is a valid instrument for $\bar{y}$, then x must also be related to y at the individual level. We show that failing to include the individual variable leads to inconsistent estimates. The only case when consistency holds is if peers are randomly allocated across individuals. However, even if peers are randomly allocated *within* clusters (e.g. schools) but not *across* clusters, the inclusion of cluster fixed effects – a necessity as randomization takes place within clusters – renders the estimates inconsistent.[3]

While most applications of peer effects that use IV do include the instrument at the individual level and therefore avoid the inconsistency and bias described here, a number of papers have not done so. More generally, we have found no discussion of this issue in the literature. Given the widespread use of IV in peer effects models, we argue that it is important to raise awareness of this among both econometricians and applied researchers.

II. The peer effects model

As the consistency of the instrumental variable estimation of a peer effect depends on whether cluster fixed effects are controlled for, we discuss both cases separately, and end with a formal proof of the asymptotic bias. To better clarify what we mean by peers and clusters, consider the case where the peer group is defined by the classmates within schools, then the peer effect is the effect of the classmates, while the cluster fixed effect is the school fixed effect.

The case without cluster fixed effects

We follow the existing literature that almost exclusively specifies a linear-in-mean peer effects model and consider the following specification

$$\mathbf{y} = \mathbf{W}\mathbf{y}\rho + \mathbf{u}, \tag{1}$$

through the peers' outcome (e.g. O'Malley *et al.*, 2014), (iii) average characteristics of peers, who are not direct peers (see Bramoullé, Djebbari and Fortin, 2009; De Giorgi, Pelizzari and Redaelli, 2010; Nicoletti and Rabe, 2016; Nicoletti, Salvanes and Tominey, 2016). Other approaches to identify peer effects include (natural) experiments (e.g. Hoxby, 2000; Duflo and Saez, 2003; Gould and Winter, 2009), random allocation of peers (e.g. Sacerdote, 2001; Kremer and Levy, 2008), and fixed effects, value-added approaches (e.g. Neidell and Waldfogel, 2010).

[2] Boozer and Cacciola (2001) and Angrist (2014) additionally show that the individual variable, say x, whose peer average, $\bar{x}$, is used to instrument the peer average outcome must have some variation *within* as well as *between* peer groups.

[3] To avoid confusion with 'peer groups', we refer to these (often larger) groupings such as schools or neighbourhoods as 'clusters'.

where $\mathbf{y}$ is the $N \times 1$ vector of the individual outcome, $\mathbf{W}$ is an $N \times N$ row-standardized weight matrix describing the social ties between individuals, ρ is the scalar peer effect parameter and $\mathbf{u}$ is the residual error vector.[4] Model (1) does not include the intercept but there is no loss of generality as long as all variables are expressed as deviations from their means. Furthermore, as we discuss below, the model can easily be adjusted to account for additional explanatory variables.

The instruments for $\mathbf{Wy}$ are defined as the peer average of characteristics $\mathbf{X}$, i.e. $\mathbf{WX}$. These must satisfy *independence*, *exclusion* and *relevance*. *Exclusion* assumes that the instruments $\mathbf{WX}$ only affect $\mathbf{y}$ through $\mathbf{Wy}$, i.e. that there is zero correlation between the error term in model (1) and $\mathbf{WX}$, or $corr(\mathbf{WX}, \mathbf{u}) = 0$; *relevance* requires the instruments to explain variation in $\mathbf{Wy}$, i.e. that $corr(\mathbf{Wy}, \mathbf{WX}) \neq 0$. The IV estimation of the peer effect ρ, which we refer to as $\hat{\rho}_{IV0}$ is then given by:

$$\hat{\rho}_{IV0} = [(\mathbf{Wy})'\mathbf{P_{WX}}(\mathbf{Wy})]^{-1}(\mathbf{Wy})'\mathbf{P_{WX}}\mathbf{y}, \tag{2}$$

where $\mathbf{P_{WX}}$ is the projection matrix $[(\mathbf{WX})[(\mathbf{WX})'(\mathbf{WX})]^{-1}(\mathbf{WX})']$. The IV estimator $\hat{\rho}_{IV0}$ is equivalent to a 2-stage least squares (2SLS) estimator where the first stage is the ordinary least squares (OLS) regression of $\mathbf{Wy}$ on $\mathbf{WX}$, and the second stage is the OLS regression of $\mathbf{y}$ on the prediction of $\mathbf{Wy}$ obtained from the first stage, i.e. $[\mathbf{P_{WX}}(\mathbf{Wy})]$ (see e.g. Cameron and Trivedi, 2005).

The peer effects literature that adopts this IV approach assumes that the individual outcome $\mathbf{y}$ is not directly affected by peers' average characteristics $\mathbf{WX}$, but they generally do not make any explicit assumption on whether the individuals' characteristics $\mathbf{X}$ directly affect $\mathbf{y}$. Appendix A shows that under the relevance and exclusion assumptions, it follows that $\mathbf{X}$ directly affects $\mathbf{y}$, and hence model (1) is misspecified because it omits $\mathbf{X}$ from the explanatory variables.[5] In other words, $\mathbf{X}$ should be included as explanatory variables in model (1):

$$\mathbf{y} = \mathbf{Wy}\rho + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{3}$$

where we still omit the constant and assume that all variables, including $\mathbf{X}$, are expressed in deviation from their mean. We therefore refer to equation (3) as the true model.[6] By replacing $\mathbf{y}$ in equation (2) with the right-hand side of equation (3), we can show that the estimator $\hat{\rho}_{IV0}$ in equation (2) is inconsistent:

$$\hat{\rho}_{IV0} = \rho + [(\mathbf{Wy})'\mathbf{P_{WX}}(\mathbf{Wy})]^{-1}(\mathbf{Wy})'\mathbf{P_{WX}}(\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}). \tag{4}$$

[4] $\mathbf{W}$ is generally constructed to have zero elements on the leading diagonal, ensuring that $\mathbf{Wy}$ excludes the individuals themselves. We also assume that the peer relationships be symmetric, so that $\mathbf{W}$ is symmetric.

[5] Appendix A shows this is true under plausible assumptions.

[6] Here, we follow the existing literature that almost exclusively considers specifications in which all covariates enter additively and linearly (including the literature that *does* account for the instrument at the individual level; section III discusses some of the relevant literature). We use this specification when deriving the asymptotic bias below. However, we note that these derivations do not generalize to situations where the true model includes some other function of the instrument at the individual level (e.g. $\mathbf{X^2}$ or $\mathbf{ln(X)}$). Hence, in such cases, the asymptotic bias is also likely to be different. Nevertheless, because the majority of studies specify the model as in equation (3), we derive the bias for this specification.

Denoting $[\mathbf{P_{WX}(Wy)}]$ with $\mathbf{(WX)}\hat{\boldsymbol{\lambda}}$, where $\hat{\boldsymbol{\lambda}}$ is the OLS estimator of the coefficients of $\mathbf{WX}$ in the first stage regression of $\mathbf{Wy}$ on $\mathbf{WX}$, and taking the probability limit, we obtain

$$p-\lim \hat{\rho}_{IV0}=\rho+(\boldsymbol{\lambda}'E((\mathbf{WX})'(\mathbf{WX}))\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'E((\mathbf{WX})'(\mathbf{X}\boldsymbol{\gamma}+\boldsymbol{\epsilon})), \tag{5}$$

where $\boldsymbol{\lambda}=p-\lim \hat{\boldsymbol{\lambda}}$, which is the vector of the true slope coefficients of $\mathbf{WX}$ in the linear regression of $\mathbf{Wy}$ on $\mathbf{WX}$. This shows that the IV estimation is consistent if and only if $E((\mathbf{WX})'(\mathbf{X}\boldsymbol{\gamma}+\boldsymbol{\epsilon}))=0$. We discuss this separately as $E((\mathbf{WX})'\mathbf{X})=0$ and $E((\mathbf{WX})'\boldsymbol{\epsilon})=0$. The latter is the main assumption imposed by empirical studies that estimate peer effects by instrumenting the peer average $\mathbf{Wy}$ with $\mathbf{WX}$. The condition $E((\mathbf{WX})'\mathbf{X})=0$ is satisfied when peers are randomly allocated across individuals. If, instead, peers are randomly allocated *within* clusters, but not *across* clusters, $\mathbf{X}$ may have a different distribution across these clusters, leading to $E((\mathbf{WX})'\mathbf{X})\neq 0$ and potentially biasing the estimation. For example, university classmates can be randomly chosen from the students enrolled in a specific degree but not from other degrees, or university roommates can be randomly chosen within a college but not across colleges (see e.g. the review by *Sacerdote 2001*). Because students do not randomly select into different colleges or degrees, peers (i.e. class or roommates) are not necessarily randomly allocated across such clusters.

Nevertheless, this potential inconsistency can be solved by controlling for the individual variables $\mathbf{X}$ as in model (3), and adopting the following IV estimation

$$\hat{\rho}_{IV1}=[(\mathbf{M_X(Wy)})'\mathbf{P_{M_X(WX)}}(\mathbf{M_X(Wy)})]^{-1}(\mathbf{M_X(Wy)})'\mathbf{P_{M_X(WX)}}(\mathbf{M_X y}), \tag{6}$$

where $\mathbf{M_X}=\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$, $\mathbf{I}$ is the identity matrix, and $\mathbf{P_{M_X(WX)}}$ is the projection matrix $(\mathbf{M_X(WX)})[(\mathbf{M_X(WX)})'(\mathbf{M_X(WX)})]^{-1}(\mathbf{M_X(WX)})'$. The estimator $\hat{\rho}_{IV1}$ is a standard two-stage least squares estimation applied to model (3) transformed by premultiplying all variables by $\mathbf{M_X}$:

$$\mathbf{M_{Xy}}=\mathbf{M_{XWy}}\rho+\mathbf{M_X}\boldsymbol{\epsilon}, \tag{7}$$

with instruments $\mathbf{M_{X(WX)}}$, i.e. the original instruments $\mathbf{(WX)}$ premultiplied by $\mathbf{M_X}$. Note that transforming model (3) by premultiplying each variable by $\mathbf{M_X}$ is equivalent to replacing each variable with the residual from the regression of the variable itself on the explanatory variables $\mathbf{X}$. By applying the Frisch–Waugh theorem, we can prove that the above transformation does not affect the estimation of the peer effect ρ.

We refer to the estimation of $\hat{\rho}_{IV1}$ as *IV approach 1*, i.e. the approach that *includes* the instrument at the individual level; we refer to the estimation of $\hat{\rho}_{IV0}$ as *IV approach 0*, i.e. the estimation approach that *omits* the instrument at the individual level.

The case with cluster fixed effects

In applied work, peers are sometimes randomized *within* but not *across* clusters. For example, class peers are often randomly chosen from the set of children enrolled in a school, but because children do not randomly sort into schools, the distribution of individual characteristics $\mathbf{X}$ is likely to differ between schools, leading to $E((\mathbf{WX})'\mathbf{X})\neq 0$ and potentially biasing the instrumental variable estimation $\hat{\rho}_{IV0}$. Because randomization in such cases is within schools, analyses of these experiments necessarily include school (or cluster) fixed

effects. We now show that failing to include the instrument at the individual level leads to inconsistent estimation of the peer effect in models with cluster fixed effects, even in cases where peers are randomized.

Consider the following fixed effects model:

$$\mathbf{y} = \mathbf{W}\mathbf{y}\rho + \mathbf{D}\boldsymbol{\delta} + \boldsymbol{v}, \tag{8}$$

where, $\mathbf{D}$ is the $N \times J$ matrix of binary cluster indicators, J is the number of clusters, $\boldsymbol{\delta}$ is the corresponding vector of fixed effects and $\boldsymbol{v} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{e}$. Applying cluster–mean deviations, we can rewrite equation (8) as follows:

$$\mathbf{y}_* = (\mathbf{W}\mathbf{y})_*\rho + \boldsymbol{v}_*, \tag{9}$$

where the subscript * indicates that the variable is premultiplied by the orthogonal projection matrix $\mathbf{M_D} = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$: $\mathbf{y}_* = \mathbf{M_D}\mathbf{y}$, $(\mathbf{W}\mathbf{y})_* = \mathbf{M_D}(\mathbf{W}\mathbf{y})$, and $\boldsymbol{v}_* = \mathbf{X}_*\boldsymbol{\gamma} + \mathbf{e}_* = \mathbf{M_D}\boldsymbol{v}$. In other words, model (9) is equal to model (8) with the variables transformed to indicate deviations from their cluster means (i.e. a within-cluster transformation).

Using the instrument $(\mathbf{W}\mathbf{X})_*$, the IV estimator that fails to control for the individual variables $\mathbf{X}_*$, i.e. *IV approach 0*, can be written as

$$\hat{\rho}_{*IV0} = [(\mathbf{W}\mathbf{y})'_*\mathbf{P}_{\mathbf{W}\mathbf{X}_*}(\mathbf{W}\mathbf{y})_*]^{-1}(\mathbf{W}\mathbf{y})'_*\mathbf{P}_{\mathbf{W}\mathbf{X}_*}\mathbf{y}_*, \tag{10}$$

where $\mathbf{P}_{\mathbf{W}\mathbf{X}_*}$ is the projection matrix $[(\mathbf{W}\mathbf{X})_*((\mathbf{W}\mathbf{X})'_*(\mathbf{W}\mathbf{X})_*)^{-1}(\mathbf{W}\mathbf{X})'_*]$.[7] With $\boldsymbol{v}_* = \mathbf{X}_*\boldsymbol{\gamma} + \mathbf{e}_*$, this converges in probability to

$$p - \lim \hat{\rho}_{*IV0} = \rho + [\boldsymbol{\lambda}'_*E((\mathbf{W}\mathbf{X})'_*(\mathbf{W}\mathbf{X})_*)\boldsymbol{\lambda}_*]^{-1}\boldsymbol{\lambda}'_*E((\mathbf{W}\mathbf{X})'_*(\mathbf{X}_*\boldsymbol{\gamma} + \mathbf{e}_*)), \tag{11}$$

where $\boldsymbol{\lambda}_* = p - \lim((\mathbf{W}\mathbf{X})'_*(\mathbf{W}\mathbf{X})_*)^{-1}(\mathbf{W}\mathbf{X})'_*(\mathbf{W}\mathbf{y})_*$ is the effect of the instruments $(\mathbf{W}\mathbf{X})_*$ on the peer average outcome $(\mathbf{W}\mathbf{y})_*$. Hence, consistency of equation (11) requires that $E((\mathbf{W}\mathbf{X})'_*(\mathbf{X}_*\boldsymbol{\gamma} + \mathbf{e}_*)) = \mathbf{0}$. Under random assignment of peers across individuals, the individual vector of characteristics $\mathbf{X}$ is uncorrelated with $\mathbf{W}\mathbf{X}$. This is because $\mathbf{W}\mathbf{X}$ is the peer average excluding the individual herself, and the random assignment of peers implies that $\mathbf{X}$ is identically and independently distributed (i.i.d.) across individuals. Nevertheless, random assignment within clusters does not imply a zero correlation between the *transformed* variables $\mathbf{X}_*$ and $(\mathbf{W}\mathbf{X})_*$, i.e. between the within-cluster deviations of $\mathbf{X}$ and $\mathbf{W}\mathbf{X}$, and therefore $E((\mathbf{W}\mathbf{X})'_*\mathbf{X}_*) = \mathbf{0}$ does not necessarily hold.

To prove this and without loss of generality, we consider a scalar exogenous variable x_i and the corresponding scalar instrumental variable $\bar{x}^p_{-i}$, which is the usual peer average of x excluding individual i. Then the within-cluster deviations of x_i and $\bar{x}^p_{-i}$ are equal to $(x_i - \bar{x}^c_i)$ and $(\bar{x}^p_{-i} - \bar{\bar{x}}^{pc}_i)$, respectively, where $\bar{x}^c_i$ is the cluster average of x_i including the individual i, $\bar{\bar{x}}^{pc}_i = \sum_{j=1}^{n_{c,i}} \bar{x}^p_{-j}/n_{c,i}$ is the cluster average of the peer average of all members in the cluster of individual i, and $n_{c,i}$ is the number of members in this cluster including individual i. By excluding the very unlikely case where individuals interact exclusively with peers who do not belong to their cluster, we can prove that $(x_i - \bar{x}^c_i)$ and $(\bar{x}^p_{-i} - \bar{\bar{x}}^{pc}_i)$ are correlated. Let us consider an individual k who is a (randomly assigned) peer of individual i belonging to the same cluster; then her observed characteristic x_k will contribute to both the cluster and the peer averages of individual i, $\bar{x}^c_i$ and $\bar{x}^p_{-i}$ respectively. Hence both $(x_i - \bar{x}^c_i)$ and

[7] Note the difference with $\hat{\rho}_{IV0}$ in equation (2), i.e. the cluster–mean deviation, indicated by *.

$(\bar{x}^{p}_{-i} - \bar{\bar{x}}^{pc}_{i})$ will be correlated with x_k and therefore $corr((x_i - \bar{x}^{c}_{i})(\bar{x}^{p}_{-i} - \bar{\bar{x}}^{pc}_{i})) \neq 0$, *despite* random assignment of peers.

Generalizing of the above proof to multivariate instruments, we can see that random assignment within clusters does not imply a zero correlation between $\mathbf{X}_*$ and $(\mathbf{WX})_*$. Ultimately, this implies that the instrumental variables $(\mathbf{WX})_*$ will be correlated with $\boldsymbol{\nu}_* = \mathbf{X}_*\boldsymbol{\gamma} + \mathbf{e}_*$, i.e. the error term in equation (9), biasing the instrumental variable estimation. Note that the bias is induced by the within transformation: it exists even if the *untransformed* instrumental variable $\mathbf{WX}$ is unrelated to the *untransformed* errors $\boldsymbol{\nu}$.[8] Avoiding this bias is possible by including the instruments at the individual level, $\mathbf{X}_*$, in the peer effects model, as in *IV Approach 1*,[9] considering the following model

$$\mathbf{y}_* = (\mathbf{Wy})_*\rho + \mathbf{X}_*\boldsymbol{\gamma} + \mathbf{e}_*. \tag{12}$$

The IV estimator for the peer effect can then be written as

$$\hat{\rho}_{*IV1} = [(\mathbf{M}_{\mathbf{X}_*}(\mathbf{Wy})_*)'\mathbf{P}_{\mathbf{M}_{\mathbf{X}_*}(\mathbf{WX})_*}(\mathbf{M}_{\mathbf{X}_*}(\mathbf{Wy})_*)]^{-1}(\mathbf{M}_{\mathbf{X}_*}(\mathbf{Wy})_*)'\mathbf{P}_{\mathbf{M}_{\mathbf{X}_*}(\mathbf{WX})_*}\mathbf{M}_{\mathbf{X}_*}\mathbf{y}_*, \tag{13}$$

where $\mathbf{M}_{\mathbf{X}_*} = \mathbf{I} - \mathbf{X}_*(\mathbf{X}'_*\mathbf{X}_*)^{-1}\mathbf{X}'_*$, and $\mathbf{P}_{\mathbf{M}_{\mathbf{X}_*}(\mathbf{WX})_*}$ is the projection matrix on the space generated by the columns of $\mathbf{M}_{\mathbf{X}_*}(\mathbf{WX})_*$.[10] By replacing $\mathbf{y}_*$ in equation (13) with the right-hand side of equation (12), we can show that $\hat{\rho}_{*IV1}$ converges in probability to ρ if $E((\mathbf{WX})'_*\mathbf{e}_*) = \mathbf{0}$.

Asymptotic bias

We next characterize the asymptotic bias. For this, we assume that equation (12) represents the true model (or equation (3) for the case without cluster fixed effects). However, if the true model specifies $\mathbf{y}$ as some other function of the instrument at the individual level (e.g. $\mathbf{X^2}$ or $\mathbf{ln(X)}$), the asymptotic bias will be different and hence, our derivations only refer to the case where $\mathbf{X}$ enters the specification in an additively separable way.

Assuming $E((\mathbf{WX})'_*\mathbf{e}_*) = 0$, the asymptotic bias of the estimator $\hat{\rho}_{*IV0}$ is given by

$$[\boldsymbol{\lambda}'_*E((\mathbf{WX})'_*(\mathbf{WX})_*)\boldsymbol{\lambda}_*]^{-1}\boldsymbol{\lambda}'_*E((\mathbf{WX})'_*\mathbf{X}_*)\boldsymbol{\gamma};$$

as shown by equation (11) above. Nevertheless, it is difficult to predict its sign and magnitude because it depends on (i) the effect of the instrument at the individual level on the individual outcome, i.e. $\boldsymbol{\gamma}$, (ii) the effect of the instruments on the peer average outcome $\boldsymbol{\lambda}_*$, (iii) $E((\mathbf{WX})'_*\mathbf{X}_*)$, and (iv) on $E((\mathbf{WX})'_*(\mathbf{WX})_*)$. Nevertheless, we can characterize the asymptotic bias in the case with one instrument as shown in the following Proposition.

Proposition 1. Let us assume that the following conditions hold.

[8] The idea is similar to the 'Nickell bias' (Nickell, 1981) in dynamic models that include individual fixed effects, leading to a correlation between the lagged-dependent variable and the mean deviation of the error term. However, the Nickell bias reduces as the number of time periods increases, the bias of $\hat{\rho}_{*IV0}$ reduces as the cluster size increases relative to the peer group, since the contribution of each peer to the cluster means becomes negligible.

[9] Although the instrument at the individual level has to be included as an additional explanatory variable, the form in which it enters matters for the bias. As the existing literature mainly considers additively separable specifications, we characterize the bias for this case only in section 'Asymptotic bias'.

[10] In addition to avoiding the bias discussed here, it also corrects for the 'exclusion bias' defined by Caeyers and Fafchamps (2016).

A1. **Correct model specification:** The true model for y_i is given by

$$y_i = \bar{y}^p_{-i}\rho + x_i\gamma + \mathbf{d}_i\boldsymbol{\delta} + e_i, \tag{14}$$

where the subscript $i = 1,\ldots,N$ denotes individuals; y_i and x_i are demeaned; $\bar{y}^p_{-i}$ is the peer average of y excluding individual i; x_i is a scalar exogenous variable; $\mathbf{d}_i$ is the $1 \times J$ vector of cluster indicators; J is the number of clusters; e_i is an idiosyncratic error uncorrelated with the explanatory variables except for the endogenous variable $\bar{y}^p_{-i}$; and (y_i, x_i, e_i) are i.i.d. with means zero and variances σ^2_x, σ^2_y and σ^2_e.

A2. **Three-level hierarchical balanced data structure:** Individuals (level 1) are nested within peer groups (level 2), which are nested within clusters (level 3). The data are balanced in the sense that all peer groups and all clusters have the same number of individuals, which we denote with n_p and n_c respectively.

A3. **Random assignment:** Peers are randomly assigned across individuals within clusters.

A4. **Exogeneity of the instrument:** There is no correlation between the deviation from the cluster mean of the error term, $e_{i,*} = e_i - \bar{e}^c_i$, and of the instrument, $\bar{x}^p_{-i,*} = \bar{x}^p_{-i} - \sum_{j=1}^{n_c} \bar{x}^p_{-j}/n_c$, where the sum is over all individuals belonging to the same cluster as individual i.

Then the asymptotic bias in the IV estimation that uses $\bar{x}^p_{-i}$ to instrument for $\bar{y}^p_{-i}$ but omits to include x_i among the explanatory variables is

$$-\frac{n_p}{n_c - n_p}\frac{\gamma}{\lambda_*}. \tag{15}$$

where γ is the effect of x_i on y_i, and λ_* is the coefficient on $\bar{x}^p_{-i}$ from an OLS regression of $\bar{y}^p_{-i}$ on $\bar{x}^p_{-i}$ and the dummy variables for each of the clusters $\mathbf{d}_i$; i.e. the first stage in a two-stage least squares procedure.

The proof is given in Appendix B. The above proposition shows that the asymptotic bias is inversely related to the effect of the instrument on the peers' average outcome, λ_*, and converges to zero if n_c tends to infinite as long as n_p remains bounded.[11] Similarly, the larger the peer group, n_p, the larger the bias. Notice that Assumption A2 implies that the size of peer groups is smaller than the size of the clusters and this ensures that the bias does not explode. In the case where there is just one cluster i.e. $n_c = N$, we have random allocation of peers across individuals and the asymptotic bias goes to zero for N which tends to ∞.

Note that *IV approach 0* and *1* can easily be adjusted to account for additional explanatory variables, by extending model (12) to include covariates, say, $\mathbf{Q}_*$. The asymptotic results can be extended to this case by applying the Frisch–Waugh–Lovell theorem which implies replacing $\mathbf{y}_*$ with the residual of the regression of $\mathbf{y}_*$ on $\mathbf{Q}_*$, i.e. $\mathbf{M}_{\mathbf{Q}_*}\mathbf{y} = [\mathbf{I} - \mathbf{Q}_*(\mathbf{Q}'_*\mathbf{Q}_*)^{-1}\mathbf{Q}'_*]\mathbf{y}_*$ and similarly replacing $(\mathbf{Wy})_*$ with $(\mathbf{M}_{\mathbf{Q}_*}\mathbf{Wy}_*)$ and $\mathbf{X}_*$ with $(\mathbf{M}_{\mathbf{Q}_*}\mathbf{X}_*)$. The conclusions remain unchanged, i.e. *IV approach 1* provides a consistent estimation for the peer effect ρ, while *IV approach 0* is inconsistent.

[11] The latter also holds for the 'exclusion bias', which Caeyers and Fafchamps (2016) show converges to zero when n_c tends to infinite while n_p remains bounded.

III. A brief discussion of the literature

Although we recognize that most empirical peer effects estimations include the instrument at the individual level, some papers have not. For example, Kang (2007) examines peer effects in students' maths attainment, estimating a school fixed effects model that uses peers' average science scores to instrument for peers' average maths scores, but excludes the individual's science score from the structural equation. Hence, despite students being quasi-randomly allocated from elementary to middle schools, not including the instrument at the individual level, combined with the inclusion of school fixed effects, leads to biased peer effects estimates. Similarly, Figlio (2007) investigates peer effects in students' disruptive behaviour, using the proportion of classroom boys with girls' names to instrument for peers' average behaviour, while adjusting for individual and grade fixed effects, but not including an indicator whether the individuals themselves have a girls' name. Lundborg (2006) investigates peer effects in adolescent substance use, estimating school-grade fixed effects models that use various peer-level instruments, several of which are excluded at the individual-level from the structural equation. For example, one of the instruments for peer average illicit drug use is the proportion of peers who indicate they know someone who could give or sell them drugs; and one of the instruments for peer average binge drinking is the proportion of peers who indicate their parents would provide beer if asked. These variables, however, are not included at the individual level.

As we discuss above, it is difficult to predict the sign and magnitude of the asymptotic bias as it depends on different factors. Nevertheless, we can comment on this to an extent. Equation (15) shows that the asymptotic bias has the same sign as $-\frac{\gamma}{\lambda_*}$. Because it is generally true that the relationship between x and y at the peer group level also holds at the individual level, γ and λ_* are of the same sign, implying the bias is negative. Furthermore, the magnitude of the asymptotic bias depends on the ratio $\frac{n_p}{n_c - n_p}$. This suggests that in primary school settings, which tend to be smaller than secondary schools but with similar class sizes, one would expect to see larger biases if classes are defined as the peer group, all else equal.

As an example, consider the study by Kang (2007). Their data include 4,813 students in 248 classes and 124 schools, suggesting that the average peer group (i.e. class) and school include 19 and 39 pupils respectively. The estimated λ_* (i.e. the effect of the instrument in the first stage) is 0.64. If we assume that $\lambda_* \approx \gamma$ (i.e. the effect of the instrument at the individual level on the individual outcome is similar to the first stage), the asymptotic bias approximates $-\frac{n_p}{n_c - n_p}\frac{\gamma}{\lambda_*} = -0.95 \times 1 = -0.95$.[12] This suggests that the bias may be relatively large, indicating that it does matter whether the instrument at the individual level is included as a covariate or not. Their peer effect is estimated to be around 0.3. Our back-of-the envelope calculations suggest that this is an underestimate, with our estimate closer to 1.25. Although this is a large difference, we cannot comment on its significance.

[12] We do not know the true value of γ, as this is precisely the parameter that is not estimated. In our illustrative application, presented in the Web Appendix, the ratio $\frac{\gamma}{\lambda_*} = \frac{0.332}{0.290} = 1.145$. Hence, although this is tentative as this estimate is obtained from a different data set, it suggests that assuming $\gamma = \lambda_*$ is a reasonable approximation. It is difficult to characterize the likely bias in Figlio (2007) and Lundborg (2006); their data contain approximately 76,000 and 3,000 students respectively, but they do not mention how many schools and classrooms they observe, and Lundborg (2006) does not report the first stage estimates.

Furthermore, we note that the bias also depends on the extent to which our assumptions, listed in the proposition above, hold. Indeed, it relies on the true model being defined by equation (12), in the sense that x_i enters the equation in an additively separable way, which may not be the case. Similarly, we assume that each individual has the same number of peers and the same number of cluster members, which is unlikely to be the case. The true data structure will therefore also impact on the estimate of the bias.

IV. Conclusion

A popular approach to estimating peer effects in the economics literature is to fit linear in mean regressions of individuals' outcomes on the corresponding average outcomes of their peers. A common approach to deal with the simultaneity of the peer effect is to use IV, instrumenting the average outcome of peers with the peer average of certain characteristics. We show that the validity of the relevance assumption in this setting has a subtle, but important implication: the instrument at the individual level must be included as an additional explanatory variable. We show that failing to do so leads to biased and inconsistent peer effect estimates. We demonstrate that the only case when consistency holds, is if peers are randomly allocated across individuals. However, even then, the IV estimation remains inconsistent if the model includes cluster fixed effects in addition to the peer effect. Examples are those where randomization takes place within, but not across, schools or neighbourhoods, where the inclusion of school or neighbourhood fixed effects (a necessity as randomization takes place within these clusters) renders the estimates inconsistent. In that case, the bias is induced by the inclusion of cluster fixed effects and its within-cluster transformation; something that has hitherto not been discussed in this literature. We present a simple solution: the instrument at the individual level must be included in the peer effects model. This leads to consistent peer effect parameter estimates under the assumptions required for IV.

Appendix A: Proof by contradiction

In the following, we prove that, if the instrumental variables **WX** satisfy the relevance and exclusion conditions for the estimation of the peer effect in model (1), then **X** directly affects **y**, and hence model (1) is misspecified because it omits **X** from the explanatory variables. The proof does not rely on any specific type of peer assignment.

As used in the spatial statistics and econometrics literature on peer effect (see e.g. Lee, 2007; Bramoullé *et al.*, 2009), we can derive the reduced form of model (1),

$$\mathbf{y} = (\mathbf{I} - \mathbf{W}\rho)^{-1}\mathbf{u}, \tag{A1}$$

where **I** is the identity matrix of size N and we assume that $|\rho| < 1$ and $\rho > 0$ so that the matrix $(\mathbf{I} - \mathbf{W}\rho)$ is invertible and the peer effect is positive. By using the series expansion $(\mathbf{I} - \mathbf{W}\rho)^{-1} = \sum_{s=1}^{\infty} \rho^s \mathbf{W}^s$ we can then rewrite the reduced form model as

$$\mathbf{y} = \sum_{s=1}^{\infty} \rho^s \mathbf{W}^s \mathbf{u}. \tag{A2}$$

Given equation (A2), the symmetry of the matrix **W** (because of the symmetry of peer relationships), and the fact that all variables are demeaned, we can prove that the covariance between **Wy** and **WX** is

$$Cov(\mathbf{Wy}', \mathbf{WX}) = E(\sum_{s=1}^{\infty} \rho^s \mathbf{u}' \mathbf{W}^{s+2} \mathbf{X}). \tag{A3}$$

This implies that **WX** are relevant instruments for **Wy** only if the right-hand side of the above equation is different from zero. We can rewrite this as a sum of expectations, with weights given by ρ^s:

$$\sum_{s=1}^{\infty} \rho^s E(\mathbf{u}' \mathbf{W}^{s+2} \mathbf{X}). \tag{A4}$$

Because $\rho^s > 0$, the above expression is different from zero if at least one of the following conditions hold: (i) **u** depends linearly on **WX**; (ii) **u** depends linearly on $\mathbf{W}^h\mathbf{X}$ for some $h > 1$ but does not depend linearly on **WX**; (iii) **u** depends linearly on **X**. Condition (i) would invalidate the instrumental variable because the exclusion restriction would not be satisfied. Condition (ii) would imply that the outcome **y** depends on the average of **X** for peers separated by h interactions[13] but not on the average of **X** for direct peers (i.e. peers separated by 1 interaction). This is unlikely, as it is implausible that peers separated by more than one interaction have a larger influence on the outcome **y** than direct peers. This implies that condition (iii) must hold to guarantee that the right-hand side of equation (A3) be non-zero. In other words, **X** and **u** are correlated, implying that **X** are omitted variables. The only situation when omitting **X** would not bias the estimation of the peer effect is when there is no correlation between the instruments **WX** and **X**.

Appendix B: Proof of Proposition 1

Proof. While the true model is given by model (14) (see Assumption A1), the estimation model omits the explanatory variable x_i and is given by

$$y_i = \bar{y}^p_{-i}\rho + \mathbf{d}_i\boldsymbol{\delta} + v_i, \tag{B1}$$

where $i = 1,\ldots,N$ and the error term $v_i = x_i\gamma + e_i$. Notice that model (B1) is identical to model (8), but it is expressed as a set of N individual equations rather than in matrix notation.

To control for the cluster effect, we can transform all variables in model (B1) using within-cluster deviations:

$$y_i - \bar{y}^c_i = (\bar{y}^p_{-i} - \bar{\bar{y}}^{pc}_i)\rho + v_i - \bar{v}^c_i, \tag{B2}$$

where $\bar{y}^c_i$ and $\bar{v}^c_i$ are the averages of y_i and v_i across all members belonging to the same cluster as individual i and, similarly, $\bar{\bar{y}}^{pc}_i = \sum_{j=1}^{n_c} \bar{y}^p_{-j}/n_c$ is the cluster average of the peer average of all members belonging to the same cluster as individual i.

[13] A peer is separated by her direct peers by one interaction, a peer is separated by her peers of peers by two interactions and so on.

Note that $(\mathbf{Wy})_*$, $\mathbf{y}_*$ and $(\mathbf{WX})_*$ defined in section 'The case with cluster fixed effects' are equivalent to the vectors of the individual within-cluster deviations $(\bar{y}^p_{-i} - \bar{\bar{y}}^{pc}_i)$, $(y_i - \bar{y}^c_i)$ and $(\bar{x}^p_{-i} - \bar{\bar{x}}^{pc}_i)$ respectively. Note also that the IV estimator of the peer effect ρ based on the misspecified model (B2), which instruments $(\bar{y}^p_{-i} - \bar{\bar{y}}^{pc}_i)$ with $(\bar{x}^p_{-i} - \bar{\bar{x}}^{pc}_i)$, is equivalent under Assumption A3/A4 to that defined in (11):

$$p - \lim \hat{\rho}_{*IV0} = \rho + [\boldsymbol{\lambda}'_* E((\mathbf{WX})'_*(\mathbf{WX})_*)\boldsymbol{\lambda}_*]^{-1}\boldsymbol{\lambda}'_* E((\mathbf{WX})'_*(\mathbf{X}_*\boldsymbol{\gamma} + \mathbf{e}_*)), \tag{B3}$$

where $\boldsymbol{\lambda}_* = p - \lim((\mathbf{WX})'_*(\mathbf{WX})_*)^{-1}(\mathbf{WX})'_*(\mathbf{Wy})_*$ is the coefficient on $\bar{x}^p_{-i}$ in the first stage regression of $\bar{y}^p_{-i}$ on $\bar{x}^p_{-i}$ and the cluster dummy variables, $\mathbf{d}_i$, and $\boldsymbol{\gamma}$ is the effect of x_i in the true model (14). Notice that because the explanatory variable x_i and the instrument $\bar{x}^p_{-i}$ are univariate variables, the coefficients $\boldsymbol{\lambda}_*$ and $\boldsymbol{\gamma}$ are actually scalars, which we denote as λ_* and γ. Under the assumption of exogeneity of the instrument (Assumption A4), $E((\mathbf{WX})'_*\mathbf{e}_*) = 0$ so that the asymptotic bias becomes:

$$p - \lim \hat{\rho}_{*IV0} - \rho = [\boldsymbol{\lambda}'_* E((\mathbf{WX})'_*(\mathbf{WX})_*)\boldsymbol{\lambda}_*]^{-1}\boldsymbol{\lambda}'_* E((\mathbf{WX})'_*(\mathbf{X}_*\boldsymbol{\gamma})). \tag{B4}$$

Because we assume that each individual has the same number of peers n_p and all his/her peers belong to the same cluster (see Assumption A2), $\bar{\bar{x}}^{pc}_i = (\sum_{j=1}^{n_c}\sum_{s=1,s\neq j}^{n_p} x_s)/(n_c n_p) = \bar{x}^c_i$. The intuition here is that the characteristic x_k of individual k belonging to the same cluster as individual i appears n_p times in the sum of the numerator of $[(\sum_{j=1}^{n_c}\sum_{s=1,s\neq j}^{n_p} x_s)/(n_c n_p)]$ as a peer of her n_p peers. This implies that

$$\left(\sum_{j=1}^{n_c}\sum_{s=1,s\neq j}^{n_p} x_s\right)/(n_c n_p) = \left(\sum_{j=1}^{n_c} x_j n_p\right)/(n_c n_p) = \sum_{j=1}^{n_c} \bar{x}_j/n_c = \bar{x}^c_i.$$

Because all variables are demeaned, x_i is i.i.d. across individuals (see Assumption A1) and peers are randomly allocated across individuals within clusters (Assumption A3), $E((\mathbf{WX})'_*\mathbf{X}_*)$ is the covariance between $(\bar{x}^p_{-i} - \bar{x}^c_i)$, and $(x_i - \bar{x}^c_i)$ and $E((\mathbf{WX})'_*(\mathbf{WX})_*)$ is the variance of $(\bar{x}^p_{-i} - \bar{x}^c_i)$. Hence, equation (B3) can be rewritten as

$$p - \lim \hat{\rho}_{*IV0} - \rho = Cov(\bar{x}^p_{-i} - \bar{x}^c_i, x_i - \bar{x}^c_i) Var(\bar{x}^p_{-i} - \bar{x}^c_i)^{-1} \frac{\gamma}{\lambda_*}. \tag{B5}$$

We can prove that

$$\begin{aligned} Cov((\bar{x}^p_{-i} - \bar{x}^c_i), (x_i - \bar{x}^c_i)) &= Cov(\bar{x}^p_{-i}, x_i) - Cov(\bar{x}^p_{-i}, \bar{x}^c_i) - Cov(\bar{x}^c_i, x_i) + Var(\bar{x}^c_i) \\ &= 0 - \frac{\sigma_x^2}{n_c} - \frac{\sigma_x^2}{n_c} + \frac{\sigma_x^2}{n_c} = -\frac{\sigma_x^2}{n_c}. \end{aligned} \tag{B6}$$

by using the following conditions

(i) x_i is i.i.d. across individuals with mean zero and variance σ_x^2 (see Assumption A1);
(ii) peers are randomly allocated across individuals within clusters (see Assumption A3);
(iii) all peers of members of a cluster belong to the same cluster (see Assumption A2).

- Conditions (i) and (ii) implies that x_i is uncorrelated with $\bar{x}^p_{-i}$ so that $Cov(\bar{x}^p_{-i}, x_i) = 0$.

- Using assumptions (i) and (iii),

$$Cov(\bar{x}^p_{-i}, \bar{x}^c_i) = Cov(\sum_{j=1, j\neq i}^{n_p} x_j, \sum_{s=1}^{n_c} x_s)/(n_c n_p) = E(\sum_{j=1, j\neq i}^{n_p} x_j^2)/(n_c n_p) = \sigma_x^2/n_c$$

.

- Because x_i is included in the cluster average, $\bar{x}^c_i$,

$$Cov(\bar{x}^c_i, x_i) = Cov(\sum_{s=1}^{n_c} x_s, x_i)/n_c = \sigma_x^2/n_c.$$

- Finally, using condition (i), $Var(\bar{x}^c_i) = \frac{\sigma_x^2}{n_c}$.

Using the same reasoning, we can show that

$$Var(\bar{x}^p_{-i} - \bar{x}^c_i) = Var(\bar{x}^p_{-i}) + Var(\bar{x}^c_i) - 2Cov(\bar{x}^p_{-i}, \bar{x}^c_i) = \frac{\sigma_x^2}{n_p} + \frac{\sigma_x^2}{n_c} - 2\frac{\sigma_x^2}{n_c} = \sigma_x^2 \frac{n_c - n_p}{n_c n_p}. \quad \text{(B7)}$$

Replacing $Cov((\bar{x}^p_{-i} - \bar{x}^c_i), (x_i - \bar{x}^c_i))$ and $Var(\bar{x}^p_{-i} - \bar{x}^c_i)$ in equation (B5) with the last right hand side terms in equations (B6) and (B7), we get

$$p - \lim \hat{\rho}_{*IV0} - \rho = -\frac{n_p}{n_c - n_p} \frac{\gamma}{\lambda_*}. \quad \text{(B8)}$$

Final Manuscript Received: December 2018

References

Angrist, J. (2014). 'The perils of peer effects', *Labour Economics*, Vol. 30, pp. 98–108.

Boozer, M. and Cacciola, S. (2001). *Inside the 'Black Box' of Project Star: Estimation of Peer Effects Using Experimental Data*, Yale Economic Growth Center No. DP832.

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

Caeyers, B. and Fafchamps, M. (2016). *Exclusion Bias in the Estimation of Peer Effects*, NBER Working Paper No. 22565.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

Dahl, G., Loken, K. and Mogstad, M. (2014). 'Peer effects in program participation', *American Economic Review*, Vol. 104, pp. 2049–2074.

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

Duflo, E. and Saez, fnmE. (2003). 'The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment', *Quarterly Journal of Economics*, Vol. 118, pp. 815–842.

Figlio, D. (2007). 'Boys named sue: Disruptive children and their peers', *Education, Finance and Policy*, Vol. 2, pp. 376–394.

Gould, E. and Winter, E. (2009). 'Interactions between workers and the technology of production: Evidence from professional baseball', *The Review of Economics and Statistics*, Vol. 91, pp. 188–200.

Hoxby, C. (2000). 'The effects of class size on student achievement: New evidence from population variation', *Quarterly Journal of Economics*, Vol. 115, pp. 1239–1285.

Kang, C. (2007). 'Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea', *Journal of Urban Economics*, Vol. 61, pp. 458–495.

Kremer, M. and Levy, D. (2008). 'Peer effects and alcohol use among college students', *Journal of Economic Perspectives*, Vol. 22, pp. 189–206.
Lee, L. F. (2007). 'Identification and estimation of econometric models with group interactions, contextual factors and fixed effects', *Journal of Econometrics*, Vol. 140, pp. 333–374.
Lundborg, P. (2006). 'Having the wrong friends? Peer effects in adolescent substance use', *Journal of Health Economics*, Vol. 25, pp. 214–233.
Moffitt, R. (2001). 'Policy interventions, low-level equilibria, and social interactions', in Durlauf S. and Young H. (eds.), *Social Dynamics*, Cambridge: MIT Press, pp. 6–17.
Neidell, M. and Waldfogel, J. (2010). 'Cognitive and noncognitive peer effects in early education', *The Review of Economics and Statistics*, Vol. 92, pp. 562–576.
Nickell, S. (1981). 'Biases in dynamic models with fixed effects', *Econometrica*, Vol. 49, pp. 1417–1426.
Nicoletti, C. and Rabe, B. (2016). *Sibling Spillover Effects in School Test Scores*, IZA Discussion Paper No. 8615.
Nicoletti, C., Salvanes K. and Tominey, E. (2016). *The Family Peer Effect on Mothers Labour Supply*, University of York Discussion Paper No. 16-4.
O'Malley, A. J., Elwert, F., Rosenquist, J. N., Zaslavsky, A. M., Christakis, N. A. (2014). 'Estimating peer effects in longitudinal dyadic data using instrumental variables', *Biometrics*, Vol. 70, pp. 506–515.
Sacerdote, B. (2001). 'Peer effects with random assignment: Results for dartmouth roommates', *Quarterly Journal of Economics*, Vol. 116, pp. 681–704.