Exclusion Bias in the Estimation of Peer Effects

Bet Caeyers (Institute for Fiscal Studies, London)
Marcel Fafchamps (Stanford University)

Workshop Social Identity and Social Interactions in Economics - Laval University

29 April 2016

Motivation

- Well-known sources of bias in peer effect estimation: reflection bias and correlated effects (Manski, 1993; Moffitt, 2001; Blume and Durlauf, 2005)

Motivation

- Well-known sources of bias in peer effect estimation: reflection bias and correlated effects (Manski, 1993; Moffitt, 2001; Blume and Durlauf, 2005)
- OLS expected to be upward biased $\rightarrow$ 2SLS often used to obtain consistent peer effects estimates

Motivation

- Well-known sources of bias in peer effect estimation: reflection bias and correlated effects (Manski, 1993; Moffitt, 2001; Blume and Durlauf, 2005)
- OLS expected to be upward biased $\rightarrow$ 2SLS often used to obtain consistent peer effects estimates
- Counter-intuitive but common finding in the literature: $\hat{\beta}^{2SLS} > \hat{\beta}^{OLS}$

Motivation

- Well-known sources of bias in peer effect estimation: reflection bias and correlated effects (Manski, 1993; Moffitt, 2001; Blume and Durlauf, 2005)
- OLS expected to be upward biased $\rightarrow$ 2SLS often used to obtain consistent peer effects estimates
- Counter-intuitive but common finding in the literature: $\hat{\beta}^{2SLS} > \hat{\beta}^{OLS}$
- OLS maybe affected by yet another bias?

Intuition (Guryan, Kroft and Notowidogo, 2009)

Intuition (Guryan, Kroft and Notowidogo, 2009)

- Typical test of random peer assignment (e.g. Sacerdote, 2001):

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

  - x_{ikl} is pre-determined characteristic of individual i in peer group k in cluster l
  - $\bar{x}_{-i,k,l}$ is average x of peers in individual i 's peer group, excluding individual i
  - δ_l is set of cluster dummies

Intuition (Guryan, Kroft and Notowidogo, 2009)

- Typical test of random peer assignment (e.g. Sacerdote, 2001):

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

  - x_{ikl} is pre-determined characteristic of individual i in peer group k in cluster l
  - $\bar{x}_{-i,k,l}$ is average x of peers in individual i 's peer group, excluding individual i
  - δ_l is set of cluster dummies

- Researchers testing random peer assignment assume ex-ante $\beta_1 = 0$

Intuition (Guryan, Kroft and Notowidogo, 2009)

- Typical test of random peer assignment (e.g. Sacerdote, 2001):

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

  - x_{ikl} is pre-determined characteristic of individual i in peer group k in cluster l
  - $\bar{x}_{-i,k,l}$ is average x of peers in individual i 's peer group, excluding individual i
  - δ_l is set of cluster dummies

- Researchers testing random peer assignment assume ex-ante $\beta_1 = 0$
- Guryan et al (2009) provide intuition for why this is not true

Intuition (Guryan, Kroft and Notowidogo, 2009)

- Typical test of random peer assignment (e.g. Sacerdote, 2001):

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

  - x_{ikl} is pre-determined characteristic of individual i in peer group k in cluster l
  - $\bar{x}_{-i,k,l}$ is average x of peers in individual i 's peer group, excluding individual i
  - δ_l is set of cluster dummies

- Researchers testing random peer assignment assume ex-ante $\beta_1 = 0$
- Guryan et al (2009) provide intuition for why this is not true
- We refer to this bias as the 'exclusion bias'

Intuition (Guryan et al, 2009)

- Guryan et al (2009):

Intuition (Guryan et al, 2009)

- Guryan et al (2009):
  - Provide intuition

Intuition (Guryan et al, 2009)

- Guryan et al (2009):
  - Provide intuition
  - Provide Monte Carlo simulation results that confirm this intuition

Intuition (Guryan et al, 2009)

- Guryan et al (2009):
  - Provide intuition
  - Provide Monte Carlo simulation results that confirm this intuition
  - Suggest a simple correction method but not not one that is generally applicable

Intuition (Guryan et al, 2009)

- Guryan et al (2009):
  - Provide intuition
  - Provide Monte Carlo simulation results that confirm this intuition
  - Suggest a simple correction method but not not one that is generally applicable
  - No formal proof

Intuition (Guryan et al, 2009)

- Guryan et al (2009):
  - Provide intuition
  - Provide Monte Carlo simulation results that confirm this intuition
  - Suggest a simple correction method but not not one that is generally applicable
  - No formal proof
  - Limited to test of random peer assignment; Do not consider exclusion bias in the estimation of endogenous peer effects

Main contributions of our paper

Main contributions of our paper

1. We formalize the intuitive results in Guryan et al (2009) for typical **tests of random peer assignment** ($\beta_1 = 0$ and non-overlapping peer groups of size $K \geq 2$)

Main contributions of our paper

1. We formalize the intuitive results in Guryan et al (2009) for typical **tests of random peer assignment** ($\beta_1 = 0$ and non-overlapping peer groups of size $K \geq 2$)
   - We derive an exact formula for the OLS inconsistency caused by exclusion bias and discuss underlying parameters (peer group size and pool size)

Main contributions of our paper

2. We generalize results to case $\beta_1 \geq 0$ & non-overlapping peer groups of size $K = 2$
   $\Rightarrow$ **Estimation of endogenous peer effects** (basic model)

$$y_{ikl} = \beta_0 + \beta_1 \bar{y}_{-i,k,l} + \delta_l + \epsilon_{ikl}$$

Main contributions of our paper

2. We generalize results to case $\beta_1 \geq 0$ & non-overlapping peer groups of size $K = 2$
   $\Rightarrow$ **Estimation of endogenous peer effects** (basic model)

$$y_{ikl} = \beta_0 + \beta_1 \bar{y}_{-i,k,l} + \delta_l + \epsilon_{ikl}$$

   - We derive exact formulas for both exclusion bias and reflection bias

Main contributions of our paper

2. We generalize results to case $\beta_1 \geq 0$ & non-overlapping peer groups of size $K = 2$
   $\Rightarrow$ **Estimation of endogenous peer effects** (basic model)

$$y_{ikl} = \beta_0 + \beta_1 \bar{y}_{-i,k,l} + \delta_l + \epsilon_{ikl}$$

- We derive exact formulas for both exclusion bias and reflection bias
- We determine conditions under which exclusion bias dominates reflection bias, changing the sign of peer effect estimates

Main contributions of our paper

2. We generalize results to case $\beta_1 \geq 0$ & non-overlapping peer groups of size $K = 2$
   $\Rightarrow$ **Estimation of endogenous peer effects** (basic model)

$$y_{ikl} = \beta_0 + \beta_1 \bar{y}_{-i,k,l} + \delta_l + \epsilon_{ikl}$$

   - We derive exact formulas for both exclusion bias and reflection bias
   - We determine conditions under which exclusion bias dominates reflection bias, changing the sign of peer effect estimates
   - We show that exclusion bias is significantly stronger when cluster FEs are added at the level of the selection pool (e.g. classroom dummies)

Main contributions of our paper

3. We **generalize results to allow for** $K \geq 2$ **and arbitrary network data**

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i$$

Main contributions of our paper

3. We **generalize results to allow for** $K \geq 2$ **and arbitrary network data**

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i$$

4. We propose specific solutions to the problem caused by exclusion bias and reflection bias

Main contributions of our paper

3. We **generalize results to allow for $K \geq 2$ and arbitrary network data**

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i$$

4. We propose specific solutions to the problem caused by exclusion bias and reflection bias
5. We show when 2SLS procedures do not suffer from exclusion bias $\Rightarrow$ explains counter-intuitive finding $\hat{\beta}^{2SLS} > \hat{\beta}^{OLS}$

Main contributions of our paper

3. We **generalize results to allow for $K \geq 2$ and arbitrary network data**

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i$$

4. We propose specific solutions to the problem caused by exclusion bias and reflection bias
5. We show when 2SLS procedures do not suffer from exclusion bias $\Rightarrow$ explains counter-intuitive finding $\hat{\beta}^{2SLS} > \hat{\beta}^{OLS}$
6. Simulation results confirm all theoretical predictions of the paper

Main contributions of our paper

3. We **generalize results to allow for $K \geq 2$ and arbitrary network data**

$$Y_i = \beta G_i Y + \gamma X_i + \delta G_i X + \epsilon_i$$

4. We propose specific solutions to the problem caused by exclusion bias and reflection bias
5. We show when 2SLS procedures do not suffer from exclusion bias $\Rightarrow$ explains counter-intuitive finding $\hat{\beta}^{2SLS} > \hat{\beta}^{OLS}$
6. Simulation results confirm all theoretical predictions of the paper
7. We (will) review the literature and discuss the type of peer effects studies that are likely to be/not to be affected by exclusion bias, and how

Exclusion bias in the test of random peer assignment ($\beta_1 = 0$)

- Typical test of random peer assignment:

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

Exclusion bias in the test of random peer assignment ($\beta_1 = 0$)

- Typical test of random peer assignment:

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

- Exclusion bias:

$$plim(\hat{\beta}_1) = -\frac{K-1}{N_P - K + 1}$$

where $K=$ size of peer group and N_P = size of pool from which peers are drawn

Exclusion bias in the test of random peer assignment ($\beta_1 = 0$)

- Typical test of random peer assignment:

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

- Exclusion bias:

$$plim(\hat{\beta}_1) = -\frac{K-1}{N_P - K + 1}$$

where $K=$ size of peer group and $N_P =$ size of pool from which peers are drawn

1. $\frac{\triangle|bias|}{\triangle N_P} < 0$: *Ceteris paribus*, exclusion bias is less severe in datasets with a larger pool of potential peers.

Exclusion bias in the test of random peer assignment ($\beta_1 = 0$)

- Typical test of random peer assignment:

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

- Exclusion bias:

$$plim(\hat{\beta}_1) = -\frac{K-1}{N_P - K + 1}$$

where K= size of peer group and N_P = size of pool from which peers are drawn

1. $\frac{\triangle |bias|}{\triangle N_P} < 0$: *Ceteris paribus*, exclusion bias is less severe in datasets with a larger pool of potential peers.
2. $\frac{\triangle |bias|}{\triangle K} > 0$: *Ceteris paribus*, exclusion bias is more severe in datasets with larger peer groups.

Cluster FEs versus Pooled OLS

- Another important result:
  1. When peers are selected at the level of the entire population Ω (e.g. school) and clusters are formed independently from peer group formation:
  $$E(\hat{\beta}_1^{FE}) = E(\hat{\beta}_1^{POLS})$$
  2. When peers are selected within clusters indexed by $l \subset \Omega$ (e.g. classroom within a school):
  $$E(\hat{\beta}_1^{FE}) < E(\hat{\beta}_1^{POLS})$$

Cluster FEs versus Pooled OLS

- Another important result:
  1. When peers are selected at the level of the entire population Ω (e.g. school) and clusters are formed independently from peer group formation:
  $$E(\hat{\beta}_1^{FE}) = E(\hat{\beta}_1^{POLS})$$
  2. When peers are selected within clusters indexed by $l \subset \Omega$ (e.g. classroom within a school):
  $$E(\hat{\beta}_1^{FE}) < E(\hat{\beta}_1^{POLS})$$
- **Intuitive?**

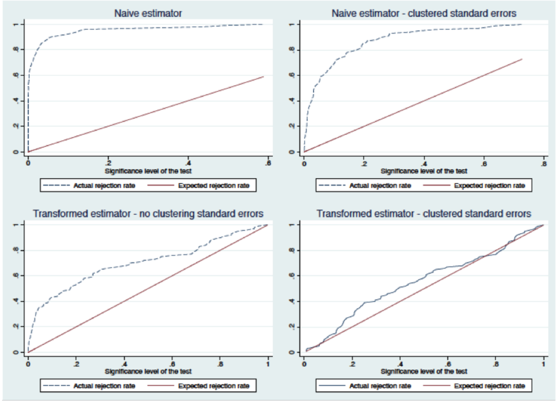
Illustration of the magnitude of the exclusion bias

Table: Exclusion bias when true $\beta_1 = 0$

		$N_P = 20$	$N_P = 50$	$N_P = 100$
K = 2	plim($\hat{\beta}_1$)	-0.053	-0.020	-0.010
K = 5	plim($\hat{\beta}_1$)	-0.250	-0.087	-0.042
K = 10	plim($\hat{\beta}_1$)	-0.818	-0.220	-0.099

Note: N= 1000 and cluster fixed effects added to all models

Figure: Expected versus actual rejection rate of $H_0 : \beta_1 = 0$; $N = 1000$; $N_P = 20$; K = 5; Cluster FE model



Note: Monte Carlo simulation results based only 100 repetitions

Exclusion bias in the estimation of endogenous peer effects ($\beta_1 \geq 0$)

Model setup

- The peer effects model we seek to estimate has the following form:

$$y_{ik} = \beta_0 + \beta_1 \bar{y}_{-i,k} + \epsilon_{ik}$$

Exclusion bias in the estimation of endogenous peer effects ($\beta_1 \geq 0$)

Model setup

- The peer effects model we seek to estimate has the following form:

$$y_{ik} = \beta_0 + \beta_1 \bar{y}_{-i,k} + \epsilon_{ik}$$

- To make this illustration as clear as possible, to start:
  - We assume peer groups of size $K = 2$
  - We abstract from exogenous peer effects and contextual effects
  - We abstract from unobserved common shocks and other correlated effects

Simple model (K = 2) - Reflection bias ($\beta_1 \geq 0$)

- We consider a system of equations similar to that of Moffit (2001):

$$\begin{aligned} y_1 &= \alpha + \beta_1 y_2 + \epsilon_1 \\ y_2 &= \alpha + \beta_1 y_1 + \epsilon_2 \end{aligned} \tag{2}$$

$0 < \beta < 1$, $E[\epsilon_1] = E[\epsilon_2] = 0$ and $E[\epsilon^2] = \sigma_\epsilon^2$

Simple model (K = 2) - Reflection bias ($\beta_1 \geq 0$)

- We consider a system of equations similar to that of Moffit (2001):

$$
\begin{aligned}
y_1 &= \alpha + \beta_1 y_2 + \epsilon_1 \\
y_2 &= \alpha + \beta_1 y_1 + \epsilon_2
\end{aligned}
\tag{2}
$$

$0 < \beta < 1$, $E[\epsilon_1] = E[\epsilon_2] = 0$ and $E[\epsilon^2] = \sigma_\epsilon^2$

- We start by ignoring exclusion bias to derive a precise formula of the reflection bias in our model.
    - That is, we start by assuming $E(\epsilon_1 \epsilon_2) = 0$

Simple model (K = 2) - Reflection bias ($\beta_1 \geq 0$)

- We consider a system of equations similar to that of Moffit (2001):

$$\begin{aligned} y_1 &= \alpha + \beta_1 y_2 + \epsilon_1 \\ y_2 &= \alpha + \beta_1 y_1 + \epsilon_2 \end{aligned} \tag{2}$$

$0 < \beta < 1$, $E[\epsilon_1] = E[\epsilon_2] = 0$ and $E[\epsilon^2] = \sigma_\epsilon^2$

- We start by ignoring exclusion bias to derive a precise formula of the reflection bias in our model.
  - That is, we start by assuming $E(\epsilon_1 \epsilon_2) = 0$
- We obtain the following expression for reflection bias:

$$E[\widehat{\beta_1}] = \frac{2\beta_1}{1 + \beta_1^2} \neq \beta_1$$

Simple model (K = 2) - Exclusion bias ($\beta_1 \geq 0$)

- So far we have assumed that $E\left[\epsilon_1 \epsilon_2\right] = 0 \Rightarrow$ not true because of the presence of exclusion bias.

Simple model (K = 2) - Exclusion bias ($\beta_1 \geq 0$)

- So far we have assumed that $E\left[\epsilon_1 \epsilon_2\right] = 0 \Rightarrow$ not true because of the presence of exclusion bias.
- From Proposition 1 we know that if we regress $\epsilon_1 = \alpha_0 + \alpha_1 \epsilon_2 + u$:

$$E[\widehat{\alpha_1}] = -\frac{K-1}{N_P - K + 1} = -\frac{1}{N_P - 1} \equiv \rho \tag{3}$$

Simple model (K = 2) - Exclusion bias ($\beta_1 \geq 0$)

- So far we have assumed that $E\left[\epsilon_1\epsilon_2\right] = 0 \Rightarrow$ not true because of the presence of exclusion bias.
- From Proposition 1 we know that if we regress $\epsilon_1 = \alpha_0 + \alpha_1\epsilon_2 + u$:

$$E[\widehat{\alpha_1}] = -\frac{K-1}{N_P - K + 1} = -\frac{1}{N_P - 1} \equiv \rho \tag{3}$$

- The sample covariance between ϵ_1 and ϵ_2 is thus:

$$Cov[\epsilon_1, \epsilon_2] = E[\epsilon_1\epsilon_2] = \rho\sigma_\epsilon^2 < 0 \tag{4}$$

Simple model (K = 2) - Exclusion bias ($\beta_1 \geq 0$)

- So far we have assumed that $E\left[\epsilon_1 \epsilon_2\right] = 0 \Rightarrow$ not true because of the presence of exclusion bias.
- From Proposition 1 we know that if we regress $\epsilon_1 = \alpha_0 + \alpha_1 \epsilon_2 + u$:

$$E[\widehat{\alpha_1}] = -\frac{K-1}{N_P - K + 1} = -\frac{1}{N_P - 1} \equiv \rho \tag{3}$$

- The sample covariance between ϵ_1 and ϵ_2 is thus:

$$Cov[\epsilon_1, \epsilon_2] = E[\epsilon_1 \epsilon_2] = \rho \sigma_\epsilon^2 < 0 \tag{4}$$

- We recalculate everything as before but now we assume $E[\epsilon_1 \epsilon_2] = \rho \sigma_\epsilon^2$ instead of $E[\epsilon_1 \epsilon_2] = 0$

Simple model (K = 2) - Exclusion bias ($\beta_1 \geq 0$)

- So far we have assumed that $E\left[\epsilon_1 \epsilon_2\right] = 0 \Rightarrow$ not true because of the presence of exclusion bias.
- From Proposition 1 we know that if we regress $\epsilon_1 = \alpha_0 + \alpha_1 \epsilon_2 + u$:

$$E[\widehat{\alpha_1}] = -\frac{K-1}{N_P - K + 1} = -\frac{1}{N_P - 1} \equiv \rho \tag{3}$$

- The sample covariance between ϵ_1 and ϵ_2 is thus:

$$Cov[\epsilon_1, \epsilon_2] = E[\epsilon_1 \epsilon_2] = \rho \sigma_\epsilon^2 < 0 \tag{4}$$

- We recalculate everything as before but now we assume $E[\epsilon_1 \epsilon_2] = \rho \sigma_\epsilon^2$ instead of $E[\epsilon_1 \epsilon_2] = 0$
- We obtain:

$$E[\hat{\beta}_1] = \frac{2\beta_1 + (1 + \beta_1^2)\rho}{1 + \beta_1^2 + 2\beta_1 \rho} < \frac{2\beta_1}{1 + \beta_1^2} \neq \beta_1 \tag{5}$$

Illustration: Reflection bias vs. exclusion bias ($\beta_1 \geq 0$)

Table: Simulation results - Exclusion bias versus reflection bias in the estimation of endogenous peer effects

$K = 2$; $N_P = 10$; $N = 500$					
True β_1	Predicted reflection bias	Prediction exclusion bias	Total predicted bias	Predicted E($\hat{\beta}_1$)	Simulated E($\hat{\beta}_1$)
0.00	0.000	-0.111	-0.111	-0.111	-0.117
0.02	0.020	-0.111	-0.091	-0.071	-0.077
0.04	0.040	-0.111	-0.072	-0.032	-0.038
0.06	0.060	-0.111	-0.051	0.009	0.002
0.08	0.079	-0.110	-0.031	0.049	0.042
0.10	0.098	-0.109	-0.011	0.098	0.082
0.12	0.117	-0.108	0.009	0.129	0.122
0.14	0.135	-0.106	0.029	0.169	0.162
0.16	0.152	-0.104	0.048	0.208	0.201
0.18	0.169	-0.102	0.067	0.247	0.240
0.20	0.185	-0.099	0.086	0.286	0.279

Correction methods - Guryan et al (2009)

- Control for differences in mean outcome across selection pools by adding to the estimation equation the mean outcome $\bar{y}_{-i,l}$ of individuals other than i in selection cluster l:

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \theta \bar{x}_{-i,l} + \epsilon_{ikl} \tag{6}$$

Correction methods - Guryan et al (2009)

- Control for differences in mean outcome across selection pools by adding to the estimation equation the mean outcome $\bar{y}_{-i,l}$ of individuals other than i in selection cluster l:

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \theta \bar{x}_{-i,l} + \epsilon_{ikl} \quad (6)$$

- Limitations:
  - Requires knowledge of the selection pool
  - Parameters β_1 and θ are separately identified only if there is variation in pool sizes N_P
  - Variation in N_P may be insufficient, resulting in multicollinearity and quasi-underidentification
  - Does not correct for reflection bias (mainly useful for test of random peer assignment)

Correction methods - An alternative ($K = 2$)

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?
  - To conduct inference we suggest using **randomization inference** to derive exact p-values (Fisher, 1925; Rosenbaum, 1984)

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?
  - To conduct inference we suggest using **randomization inference** to derive exact p-values (Fisher, 1925; Rosenbaum, 1984)
  - BUT in (importantly) different way than standard randomization inference applications (e.g. Athey, Eckles and Imbens, 2015)

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?
  - To conduct inference we suggest using **randomization inference** to derive exact p-values (Fisher, 1925; Rosenbaum, 1984)
  - BUT in (importantly) different way than standard randomization inference applications (e.g. Athey, Eckles and Imbens, 2015)
    - Start from observational dataset and obtain $\widehat{\beta}_1^{naive}$

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?
  - To conduct inference we suggest using **randomization inference** to derive exact p-values (Fisher, 1925; Rosenbaum, 1984)
  - BUT in (importantly) different way than standard randomization inference applications (e.g. Athey, Eckles and Imbens, 2015)
    - Start from observational dataset and obtain $\widehat{\beta}_1^{naive}$
    - Re-shuffle observations by randomly assigning them to different peers

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?
  - To conduct inference we suggest using **randomization inference** to derive exact p-values (Fisher, 1925; Rosenbaum, 1984)
  - BUT in (importantly) different way than standard randomization inference applications (e.g. Athey, Eckles and Imbens, 2015)
    - Start from observational dataset and obtain $\widehat{\beta}_1^{naive}$
    - Re-shuffle observations by randomly assigning them to different peers
    - Re-estimate your regression, obtain and store $\widehat{\beta}_1^{s}$

Correction methods - An alternative ($K = 2$)

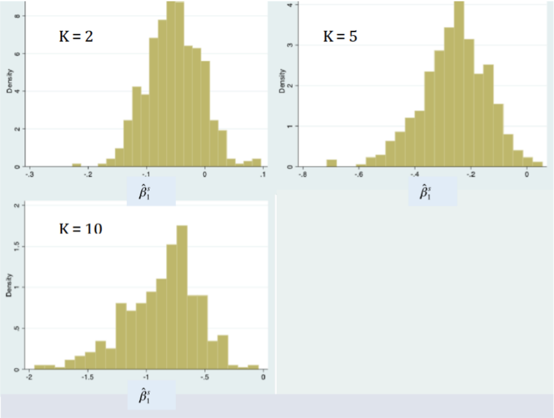
- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?
  - To conduct inference we suggest using **randomization inference** to derive exact p-values (Fisher, 1925; Rosenbaum, 1984)
  - BUT in (importantly) different way than standard randomization inference applications (e.g. Athey, Eckles and Imbens, 2015)
    - Start from observational dataset and obtain $\widehat{\beta}_1^{naive}$
    - Re-shuffle observations by randomly assigning them to different peers
    - Re-estimate your regression, obtain and store $\widehat{\beta}_1^s$
    - Repeat this process many times and trace the distribution of $\widehat{\beta}_1^s$ = distribution under the null

Correction methods - An alternative ($K = 2$)

- How to correct point estimates?
  - Our formula can be used to adjust the point estimate of $\widehat{\beta}_1$ to make it consistent
- How to correct inference?
  - To conduct inference we suggest using **randomization inference** to derive exact p-values (Fisher, 1925; Rosenbaum, 1984)
  - BUT in (importantly) different way than standard randomization inference applications (e.g. Athey, Eckles and Imbens, 2015)
    - Start from observational dataset and obtain $\widehat{\beta}_1^{naive}$
    - Re-shuffle observations by randomly assigning them to different peers
    - Re-estimate your regression, obtain and store $\widehat{\beta}_1^{s}$
    - Repeat this process many times and trace the distribution of $\widehat{\beta}_1^{s}$ = distribution under the null
    - Obtain correct p-value by taking proportion of $\widehat{\beta}_1^{s}$ that are above $\widehat{\beta}_1^{naive}$ (similar to bootstrapping procedures)

Randomization inference - Example

Figure: Histogram $\hat{\beta}_1^s$ under null hypothesis ($N = 1000$; $N_P = 20$)



Correction - generalized model (groups ≥ 2 or network data)

$$Y = \beta GY + \gamma X + \delta GX + \epsilon \tag{7}$$

Correction - generalized model (groups ≥ 2 or network data)

$$Y = \beta GY + \gamma X + \delta GX + \epsilon \tag{7}$$

- If simply interested in correcting inference for $H_0 : \beta_1 = 0$ then randomization inference method described before will do the trick

Correction - generalized model (groups ≥ 2 or network data)

$$Y = \beta GY + \gamma X + \delta GX + \epsilon \tag{7}$$

- If simply interested in correcting inference for $H_0 : \beta_1 = 0$ then randomization inference method described before will do the trick
- If interested in correcting point estimates as well as inference: No closed-form formulas available

Correction - generalized model (groups ≥ 2 or network data)

$$Y = \beta GY + \gamma X + \delta GX + \epsilon \tag{7}$$

- If simply interested in correcting inference for $H_0 : \beta_1 = 0$ then randomization inference method described before will do the trick
- If interested in correcting point estimates as well as inference: No closed-form formulas available
- We can characterize the DGP under exclusion bias and use nonlinear method of moments estimation techniques to provide consistent estimates for β_1:

$$E[YY'] = (I-\beta G)^{-1}E[(\gamma X+\delta GX)(\gamma X+\delta GX)'](I-\beta G')^{-1}+(I-\beta G)^{-1}E[\epsilon\ \epsilon'](I-\beta G')^{-1}$$

Correction - generalized model (groups ≥ 2 or network data)

$$Y = \beta GY + \gamma X + \delta GX + \epsilon \tag{7}$$

- If simply interested in correcting inference for $H_0 : \beta_1 = 0$ then randomization inference method described before will do the trick
- If interested in correcting point estimates as well as inference: No closed-form formulas available
- We can characterize the DGP under exclusion bias and use nonlinear method of moments estimation techniques to provide consistent estimates for β_1:

$$E[YY'] = (I - \beta G)^{-1} E[(\gamma X + \delta GX)(\gamma X + \delta GX)'](I - \beta G')^{-1} + (I - \beta G)^{-1} E[\epsilon\ \epsilon'](I - \beta G')^{-1}$$

- Where $E[\epsilon\epsilon'] = \sigma_\epsilon^2 \begin{bmatrix} B & 0 & ... \\ 0 & B & ... \\ ... & ... & ... \end{bmatrix}$ and where B is a $K \times K$ of the form:

$$B = \begin{bmatrix} 1 & \rho & ... \\ \rho & 1 & ... \\ ... & ... & ... \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{N_P - 1} & ... \\ -\frac{1}{N_P - 1} & 1 & ... \\ ... & ... & ... \end{bmatrix}$$

Simulation results correction method - General groups

	K = 2 ; $L = 20$; $N = 100$			K = 5 ; $L = 20$; $N = 100$		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	0.0	0.1	0.2	0.0	0.1	0.2
Naive E($\hat{\beta}_1$)	-0.05	0.14	0.33	-0.26	-0.04	0.17
E($\hat{\beta}_1$) correction reflection bias only	-0.03	0.07	0.17	-0.12	-0.01	0.10
E($\hat{\beta}_1$) correction reflection and exclusion bias	0.00	0.10	0.20	0.00	0.10	0.20

Note: Cluster fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions.

Example randomization inference on one fictional dataset - General groups

	K $= 2$; $L = 10$; $N = 500$			K $= 5$; $L = 10$; $N = 500$		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	0.0	0.1	0.2	0.0	0.1	0.2
Naive $\hat{\beta}_1$	-0.15	0.05	0.25	-0.60	-0.34	-0.08
Naive p-value	0.00	0.27	0.00	0.00	0.00	0.43
Corrected p-value	0.63	0.01	0.00	0.72	0.11	0.03

Note: Cluster fixed effects added in all regressions; Randomization over 100 Monte Carlo replications.

Simulation correction method - Network data

p = probability of link between i and j within a cluster	$p = 0.1$; $L = 20$; $N = 100$			$p = 0.3$; $L = 20$; $N = 100$		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	0.0	0.1	0.2	0.0	0.1	0.2
Naive E($\hat{\beta}_1$)	-0.09	0.08	0.24	-0.31	-0.13	0.04
E($\hat{\beta}_1$) correction reflection bias only	-0.04	0.03	0.11	-0.11	-0.05	0.01
E($\hat{\beta}_1$) correction reflection and exclusion bias	0.01	0.10	0.19	0.02	0.11	0.19

Note: Cluster fixed effects added in all regressions; Simulations $\hat{\beta}_1$ over 100 Monte Carlo repetitions.

Example randomization inference on one fictional dataset - Network data

p = probability of link between i and j within a cluster	$p = 0.1$; $L = 10$; $N = 500$			$p = 0.3$; $L = 10$; $N = 500$		
	(1)	(2)	(3)	(4)	(5)	(6)
True β_1	0.0	0.1	0.2	0.0	0.1	0.2
Naive $\hat{\beta}_1$	-0.14	0.03	0.19	-0.40	-0.24	-0.06
Naive p-value	0.02	0.60	0.00	0.00	0.01	0.52
Corrected p-value	0.65	0.09	0.00	0.40	0.01	0.00

Note: Cluster fixed effects added in all regressions; Randomization over 100 Monte Carlo replications.

Example - Sacerdote (2001)

Parameters	**Values**	**Example Sacerdote (2001)**
Sample size N	1589	Number of graduate students enrolled in 1998-1999
Number of peer selection pools	42	Number of housing preference blocks within which students are randomly allocated to dorms and dorm rooms
Pool size N_P	38	Average number of students in each housing preference block
Peer group size K	53% in double room ($K = 2$) 44% in triple rooms ($K = 3$) 3% in quad rooms ($K = 4$)	Number of students in each dorm room
Exclusion bias: $-\frac{K-1}{N_P - K + 1}$	-0.04	Weighted average of biases for different K: $0.53 * (bias \mid K = 2)$ $+0.44 * (bias \mid K = 3)$ $+0.03 * (bias \mid K = 4)$

Example - Sacerdote (2001) - test of random peer assignment

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

	SATH Math	SAT verbal	High school Academic class index	High school academic index
$\hat{\beta}_1^{Naive}$ - Sacerdote (2001)	-0.025 (0.028)	-0.009 (0.029)	0.010 (0.028)	-0.032 (0.028)
$\hat{\beta}_1^{Corrected}$ (Caveat! Not clustering s.e.)	0.015 (0.028)	0.031 (0.029)	0.05* (0.028)	0.008 (0.028)

Example - Sacerdote (2001) - estimation of peer effects

$$y_{ikl} = \beta_0 + \beta_1 \bar{y}_{-i,k,l} + \delta_l + \epsilon_{ikl}$$

	GPA test score
$\hat{\beta}_1^{Naive}$ - Sacerdote (2001)	0.07** (0.029)
$\hat{\beta}_1^{CorrectionReflectionOnly}$ Conservative correction (assuming $K = 2$ and no clustering s.e.)	0.03 (0.029)
$\hat{\beta}_1^{TotalCorrection}$ Conservative correction (assuming $K = 2$ and no clustering s.e.)	0.05* (0.029)

Which type of studies are NOT affected by exclusion bias?

Which type of studies are NOT affected by exclusion bias?

- Certain studies that use an RCT as a means to estimate peer effects (e.g. Fafchamps and Vicente, 2013)

$$y_i = b_0 + b_1 \sum_{j \in N_i} T_j + b_2 T_i + u_i$$

Which type of studies are NOT affected by exclusion bias?

- Certain studies that use an RCT as a means to estimate peer effects (e.g. Fafchamps and Vicente, 2013)

$$y_i = b_0 + b_1 \sum_{j \in N_i} T_j + b_2 T_i + u_i$$

- Studies that use lagged outcome of peers and control for lagged outcome of individal i herself (e.g. Munshi, 2004)

$$y_{i,t+1} = b_0 + b_1 \bar{y}_{-i,t} + b_2 y_{it} + u_{i,t+1}$$

Which type of studies are NOT affected by exclusion bias?

- Certain studies that use an RCT as a means to estimate peer effects (e.g. Fafchamps and Vicente, 2013)

$$y_i = b_0 + b_1 \sum_{j \in N_i} T_j + b_2 T_i + u_i$$

- Studies that use lagged outcome of peers and control for lagged outcome of individal i herself (e.g. Munshi, 2004)

$$y_{i,t+1} = b_0 + b_1 \bar{y}_{-i,t} + b_2 y_{it} + u_{i,t+1}$$

- Studies that use peers' pre-treatment characteristics and control for i 's pre-treatment characteristic (e.g. Bayer et al, 2009)

$$y_{i,t+1} = b_0 + b_1 \bar{x}_{-i} + b_2 x_i + u_i$$

Which type of studies are NOT affected by exclusion bias?

Which type of studies are NOT affected by exclusion bias?

- Certain studies that employ 2SLS:

Which type of studies are NOT affected by exclusion bias?

- Certain studies that employ 2SLS:
  - 2SLS can - under certain conditions - eliminate the exclusion bias

Which type of studies are NOT affected by exclusion bias?

- Certain studies that employ 2SLS:
  - 2SLS can - under certain conditions - eliminate the exclusion bias
  - Condition: if and only if one instruments $\bar{y}_{-i,k}$ by $\bar{z}_{-i,k}$ whilst controlling for individual i's own value z_{ik} of the instrument

Which type of studies are NOT affected by exclusion bias?

- Certain studies that employ 2SLS:
  - 2SLS can - under certain conditions - eliminate the exclusion bias
  - Condition: if and only if one instruments $\bar{y}_{-i,k}$ by $\bar{z}_{-i,k}$ whilst controlling for individual i's own value z_{ik} of the instrument
    - First stage: $\bar{y}_{-i,k} = \pi_0 + \pi_1 \bar{z}_{-i,k} + \pi_2 z_{ik} + v_{ik}$
    - Second stage: $y_{ik} = \beta_0 + \beta_1 \hat{\bar{y}}_{-i,k} + \beta_2 z_{ik} + \epsilon_{ik}$

Which type of studies are NOT affected by exclusion bias?

- Certain studies that employ 2SLS:
  - 2SLS can - under certain conditions - eliminate the exclusion bias
  - Condition: if and only if one instruments $\bar{y}_{-i,k}$ by $\bar{z}_{-i,k}$ whilst controlling for individual i's own value z_{ik} of the instrument
    - First stage: $\bar{y}_{-i,k} = \pi_0 + \pi_1 \bar{z}_{-i,k} + \pi_2 z_{ik} + v_{ik}$
    - Second stage: $y_{ik} = \beta_0 + \beta_1 \hat{\bar{y}}_{-i,k} + \beta_2 z_{ik} + \epsilon_{ik}$
  - Example Bramouille et al (2009) or De Giorgi et al (2010):

$$y_i = b_0 + b_1 \bar{y}_j + b_2 \bar{x}_j + b_3 x_i + u_i$$

using peers' peers exogenous characteristics $\bar{x}_k$ as instruments for $\bar{y}_j$ whilst controlling for x_i

Which type of studies are NOT affected by exclusion bias?

- Certain studies that employ 2SLS:
  - 2SLS can - under certain conditions - eliminate the exclusion bias
  - Condition: if and only if one instruments $\bar{y}_{-i,k}$ by $\bar{z}_{-i,k}$ whilst controlling for individual i's own value z_{ik} of the instrument
    - First stage: $\bar{y}_{-i,k} = \pi_0 + \pi_1 \bar{z}_{-i,k} + \pi_2 z_{ik} + v_{ik}$
    - Second stage: $y_{ik} = \beta_0 + \beta_1 \hat{\bar{y}}_{-i,k} + \beta_2 z_{ik} + \epsilon_{ik}$
  - Example Bramouille et al (2009) or De Giorgi et al (2010):

$$y_i = b_0 + b_1 \bar{y}_j + b_2 \bar{x}_j + b_3 x_i + u_i$$

    using peers' peers exogenous characteristics $\bar{x}_k$ as instruments for $\bar{y}_j$ whilst controlling for x_i
  - Alternative explanation for the common but counter-intuitive tendency of peer effects studies to obtain $\hat{\beta}_1^{2SLS} > \hat{\beta}_1^{OLS}$

Which type of studies are NOT affected by exclusion bias?

- BUT, limitations to 2SLS:
  - Condition of controlling for z_{ik} not always satisfied
  - Requires suitable strong instruments (Bound et al, 1995)
  - Biased in finite samples (Bound et al, 1995)

Which type of studies are NOT affected by exclusion bias?

- BUT, limitations to 2SLS:
  - Condition of controlling for z_{ik} not always satisfied
  - Requires suitable strong instruments (Bound et al, 1995)
  - Biased in finite samples (Bound et al, 1995)
- We suggest a correction method that deals wit both reflection bias and exclusion bias and which does not require any IVs and that is valid even in small finite samples.

Concluding remarks

- Theory and simulations suggest that implications of exclusion bias can indeed be significant, especially in settings with relatively large peer groups and small pools from which peers are drawn

Concluding remarks

- Theory and simulations suggest that implications of exclusion bias can indeed be significant, especially in settings with relatively large peer groups and small pools from which peers are drawn
- Caution against naive comparisons between models with peer groups that vary in size

Concluding remarks

- Theory and simulations suggest that implications of exclusion bias can indeed be significant, especially in settings with relatively large peer groups and small pools from which peers are drawn
- Caution against naive comparisons between models with peer groups that vary in size
- Caution against naive comparisons between $\hat{\beta}_1^{FE}$ and $\hat{\beta}_1^{OLS}$

Concluding remarks

- Theory and simulations suggest that implications of exclusion bias can indeed be significant, especially in settings with relatively large peer groups and small pools from which peers are drawn
- Caution against naive comparisons between models with peer groups that vary in size
- Caution against naive comparisons between $\hat{\beta}_1^{FE}$ and $\hat{\beta}_1^{OLS}$
- Caution against naive comparisons between $\hat{\beta}_1^{2SLS}$ and $\hat{\beta}_1^{OLS}$

Concluding remarks

- Theory and simulations suggest that implications of exclusion bias can indeed be significant, especially in settings with relatively large peer groups and small pools from which peers are drawn
- Caution against naive comparisons between models with peer groups that vary in size
- Caution against naive comparisons between $\hat{\beta}_1^{FE}$ and $\hat{\beta}_1^{OLS}$
- Caution against naive comparisons between $\hat{\beta}_1^{2SLS}$ and $\hat{\beta}_1^{OLS}$
- We suggest methods that can be used to correct point estimates and inference for both reflection bias and exclusion bias when no suitable instruments are available

Concluding remarks

- Theory and simulations suggest that implications of exclusion bias can indeed be significant, especially in settings with relatively large peer groups and small pools from which peers are drawn
- Caution against naive comparisons between models with peer groups that vary in size
- Caution against naive comparisons between $\hat{\beta}_1^{FE}$ and $\hat{\beta}_1^{OLS}$
- Caution against naive comparisons between $\hat{\beta}_1^{2SLS}$ and $\hat{\beta}_1^{OLS}$
- We suggest methods that can be used to correct point estimates and inference for both reflection bias and exclusion bias when no suitable instruments are available
- Next step: review the literature and demonstrate the impact of exclusion bias in practice

THANK YOU!

APPENDIX

APPENDIX SLIDES

Exclusion bias in test of random peer assignment ($\beta_1 = 0$)

$$x_{ik} = \beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik} \tag{8}$$

Exclusion bias in test of random peer assignment ($\beta_1 = 0$)

$$x_{ik} = \beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik} \quad (8)$$

- Under random peer assignment we have

$$\bar{x}_{-i,k} = \bar{x}_{-i} + u_{ik} \quad (9)$$

where $\bar{x}_{-i}$ = average outcome of pool of $(N-1)$ potential peers and u_{ik} is random term

Exclusion bias in test of random peer assignment ($\beta_1 = 0$)

$$x_{ik} = \beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik} \tag{8}$$

- Under random peer assignment we have

$$\bar{x}_{-i,k} = \bar{x}_{-i} + u_{ik} \tag{9}$$

where $\bar{x}_{-i}$ = average outcome of pool of $(N-1)$ potential peers and $u_{ik}$is random term

- Inserting equation (9) into equation (8), we obtain:

$$x_{ik} = \beta_0 + \beta_1 \left(\bar{x}_{-i} + u_{ik}\right) + \epsilon_{ik} \tag{10}$$

Exclusion bias in test of random peer assignment ($\beta_1 = 0$)

$$x_{ik} = \beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik} \tag{8}$$

- Under random peer assignment we have

$$\bar{x}_{-i,k} = \bar{x}_{-i} + u_{ik} \tag{9}$$

where $\bar{x}_{-i}$ = average outcome of pool of $(N-1)$ potential peers and $u_{ik}$is random term

- Inserting equation (9) into equation (8), we obtain:

$$x_{ik} = \beta_0 + \beta_1 \left(\bar{x}_{-i} + u_{ik}\right) + \epsilon_{ik} \tag{10}$$

- Note that:

$$\bar{x}_{-i} = \frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} x_{js}\right] - x_{ik}}{N-1} \tag{11}$$

Exclusion bias in test of random peer assignment ($\beta_1 = 0$)

$$x_{ik} = \beta_0 + \beta_1 \bar{x}_{-i,k} + \epsilon_{ik} \tag{8}$$

- Under random peer assignment we have

$$\bar{x}_{-i,k} = \bar{x}_{-i} + u_{ik} \tag{9}$$

where $\bar{x}_{-i}$ = average outcome of pool of $(N-1)$ potential peers and $u_{ik}$is random term

- Inserting equation (9) into equation (8), we obtain:

$$x_{ik} = \beta_0 + \beta_1 \left(\bar{x}_{-i} + u_{ik}\right) + \epsilon_{ik} \tag{10}$$

- Note that:

$$\bar{x}_{-i} = \frac{\left[\sum_{s=1}^{\frac{N}{K}} \sum_{j=1}^{K} x_{js}\right] - x_{ik}}{N-1} \tag{11}$$

- Through reduced form we derive formula for expected bias in $\hat{\beta}_1^{OLS}$:

$$E(\hat{\beta}_1^{OLS}) = -\frac{K(K-1)}{N + (N-K)(K-1)}$$

Formula - with clustered stratification

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \epsilon_{ikl} \tag{12}$$

where l is cluster of size $L < N$ at the level of which peers are randomised (e.g. classroom)

Formula - with clustered stratification

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \epsilon_{ikl} \tag{12}$$

where l is cluster of size $L < N$ at the level of which peers are randomised (e.g. classroom)

- Cluster FE equation can be rewritten as follows:

$$x_{ikl} - \bar{x}_l = \beta_1 \left(\bar{x}_{-i,k,l} - \bar{\bar{x}}_{-i,l} \right) + \left(\epsilon_{ikl} - \bar{\epsilon}_l \right) \tag{13}$$

Formula - with clustered stratification

$$x_{ikl} = \beta_0 + \beta_1 \bar{x}_{-i,k,l} + \delta_l + \epsilon_{ikl} \tag{12}$$

where l is cluster of size $L < N$ at the level of which peers are randomised (e.g. classroom)

- Cluster FE equation can be rewritten as follows:

$$x_{ikl} - \bar{x}_l = \beta_1 \left(\bar{x}_{-i,k,l} - \bar{\bar{x}}_{-i,l}\right) + (\epsilon_{ikl} - \bar{\epsilon}_l) \tag{13}$$

- Proof proceeds in a similar way as in the non-stratified case but now based on a model in terms of deviations of outcomes from their respective cluster averages. We obtain:

$$E(\hat{\beta}_1^{FE}) = -\frac{K(K-1)}{L + (L-K)(K-1)} \tag{14}$$

Which type of studies are affected by exclusion bias?

Which type of studies are affected by exclusion bias?

- The exclusion bias arises whenever:
  1. Individual i is excluded from her potential peer group (selection without replacement);
  2. But is a potential peer for other observed individuals j
  3. And one considers a peer effects model that regresses individual i's characteristic on an average characteristic of i's peer group without controlling for i's own characteristic

Which type of studies are NOT affected by exclusion bias?

Which type of studies are NOT affected by exclusion bias?

- Studies that use an RCT as a means to estimate peer effects (e.g. Duflo and Saez, 2003)

Which type of studies are NOT affected by exclusion bias?

- Studies that use an RCT as a means to estimate peer effects (e.g. Duflo and Saez, 2003)
- Studies that do not exclude i's own outcome when calculating the average peer group characteristic (e.g. Conley and Udry, 2010)

Which type of studies are NOT affected by exclusion bias?

- Studies that use an RCT as a means to estimate peer effects (e.g. Duflo and Saez, 2003)
- Studies that do not exclude i's own outcome when calculating the average peer group characteristic (e.g. Conley and Udry, 2010)
- Studies that use lagged outcome of peers rather than contemporeneous outcomes and then control for lagged outcome of individal i herself (e.g. Munshi, 2004)

Which type of studies are NOT affected by exclusion bias?

- Studies that use an RCT as a means to estimate peer effects (e.g. Duflo and Saez, 2003)
- Studies that do not exclude i's own outcome when calculating the average peer group characteristic (e.g. Conley and Udry, 2010)
- Studies that use lagged outcome of peers rather than contemporeneous outcomes and then control for lagged outcome of individal i herself (e.g. Munshi, 2004)
- Studies that use peers' pre-determined characteristics rather than peers' outcomes as the peer effect of interest and that control for i's own pre-treatment characteristic (e.g. Bayer et al, 2009)

Which type of studies are NOT affected by exclusion bias?

- Studies that use an RCT as a means to estimate peer effects (e.g. Duflo and Saez, 2003)
- Studies that do not exclude i's own outcome when calculating the average peer group characteristic (e.g. Conley and Udry, 2010)
- Studies that use lagged outcome of peers rather than contemporeneous outcomes and then control for lagged outcome of individal i herself (e.g. Munshi, 2004)
- Studies that use peers' pre-determined characteristics rather than peers' outcomes as the peer effect of interest and that control for i's own pre-treatment characteristic (e.g. Bayer et al, 2009)
- Studies that employ a particular type of 2SLS estimation that eliminates exclusion bias (e.g. Fletcher, 2012)