Identification of a rank-dependent peer effect model

Eyo I. Herstad [1] Myungkou Shin [2]

[1]Erasmus University Rotterdam and Tinbergen Institute

[2]University of Surrey

26th of August, 2025

Table of contents

Motivation

- Individuals interact in networks, creating spillovers/ peer effects.

Motivation

- Individuals interact in networks, creating spillovers/ peer effects.
- The most common way of modeling this is through a linear-in-means model

$$y_i = \bar{y}_i \beta_{\text{LiM}} + x_i \gamma + \epsilon_i$$

Motivation

- Individuals interact in networks, creating spillovers/ peer effects.
- The most common way of modeling this is through a linear-in-means model

$$y_i = \bar{y}_i \beta_{\text{LiM}} + x_i \gamma + \epsilon_i$$

- However,
  - Peer effects might be heterogeneous

Motivation

- Individuals interact in networks, creating spillovers/ peer effects.
- The most common way of modeling this is through a linear-in-means model

$$y_i = \bar{y}_i \beta_{\text{LiM}} + x_i \gamma + \epsilon_i$$

- However,
  - Peer effects might be heterogeneous
  - The effect of a peer might depend on the group you interact with

Motivation

- Individuals interact in networks, creating spillovers/ peer effects.
- The most common way of modeling this is through a linear-in-means model

$$y_i = \bar{y}_i \beta_{\text{LiM}} + x_i \gamma + \epsilon_i$$

- However,
  - Peer effects might be heterogeneous
  - The effect of a peer might depend on the group you interact with
- We introduce a model where the effect of each peer is endogenously determined by the distribution of the outcomes of an individual's peers

The model

The model

- Let

$$d_i = \sum_{j=1}^{n} A_{i,j} = \text{number of peers for individual } i$$

$$\tilde{Y}_{i,k} = \text{outcome of the } k\text{-th lowest performing peer of individual } i$$
$$= k\text{-th ordered statistic from } \{Y_j : A_{i,j} = 1\}.$$

The model

- Let

$$d_i = \sum_{j=1}^{n} A_{i,j} = \text{number of peers for individual } i$$

$$\begin{aligned} \tilde{Y}_{i,k} &= \text{outcome of the } k\text{-th lowest performing peer of individual } i \\ &= k\text{-th ordered statistic from } \{Y_j : A_{i,j} = 1\}\,. \end{aligned}$$

- Our model is:

$$Y_i = \sum_{k=1}^{d_i} \beta_{k,d_i} \tilde{Y}_{i,k} + x_i{}^{\intercal}\gamma + \varepsilon_i \tag{1}$$

for $i = 1, \ldots, n$. x_i are exogenous covariates that may depend on the network.

What does this mean?

- $\beta_{k,d}$ gives the effect of the k'th ranked friend of individuals with d friends. For example:

What does this mean?

- $\beta_{k,d}$ gives the effect of the k'th ranked friend of individuals with d friends. For example:
  - $\beta_{3,5}$ is the effect of the 3rd lowest outcome friend for people with 5 friends.

What does this mean?

- $\beta_{k,d}$ gives the effect of the k'th ranked friend of individuals with d friends. For example:
  - $\beta_{3,5}$ is the effect of the 3rd lowest outcome friend for people with 5 friends.
  - $\beta_{3,3}$ is the effect of the highest outcome friend for people with 3 friends.

What does this mean?

- $\beta_{k,d}$ gives the effect of the k'th ranked friend of individuals with d friends. For example:
    - $\beta_{3,5}$ is the effect of the 3rd lowest outcome friend for people with 5 friends.
    - $\beta_{3,3}$ is the effect of the highest outcome friend for people with 3 friends.
- Application-specific simplifying assumptions can give more interpretable coefficients

$$Y_i = \tilde{Y}_{i,1}\beta_{\text{min}} + \tilde{Y}_{i,d_i}\beta_{\text{max}} + \bar{Y}_{i,-1,-d_i}\beta_{\text{mean}} + x_i\gamma + \epsilon_i$$

- Assumes homogeneity across d_i and all "middle" peers to have a homogeneous effect.
- Creates a separate min/max peer effect, with the rest following "linear-in-means".

Assumptions for existence

- A sufficient condition for the existence of outcomes y_i satisfying Equation (1) is that the *total spillover for each individual* is less than one

Assumptions for existence

- A sufficient condition for the existence of outcomes y_i satisfying Equation (1) is that the *total spillover for each individual* is less than one
  - For people with one friend, this means $|\beta_{1,1}| \leq 1$

Assumptions for existence

- A sufficient condition for the existence of outcomes y_i satisfying Equation (1) is that the *total spillover for each individual* is less than one
  - For people with one friend, this means $|\beta_{1,1}| \leq 1$
  - For people with two friends, this means $|\beta_{1,2}| + |\beta_{2,2}| \leq 1$

Assumptions for existence

- A sufficient condition for the existence of outcomes y_i satisfying Equation (1) is that the *total spillover for each individual* is less than one
  - For people with one friend, this means $|\beta_{1,1}| \leq 1$
  - For people with two friends, this means $|\beta_{1,2}| + |\beta_{2,2}| \leq 1$
  - ..and so on.

Assumptions for existence

- A sufficient condition for the existence of outcomes y_i satisfying Equation (1) is that the *total spillover for each individual* is less than one
  - For people with one friend, this means $|\beta_{1,1}| \leq 1$
  - For people with two friends, this means $|\beta_{1,2}| + |\beta_{2,2}| \leq 1$
  - ..and so on.

Assumption 1 (Bounded Peer Effect).

For all $d \leq \bar{d} = \max_{i=1,\ldots,n} d_i$

$$\sum_{j=1}^{d} |\beta_{j,d}| \leq 1$$

Assumptions for existence

- A sufficient condition for the existence of outcomes y_i satisfying Equation (1) is that the *total spillover for each individual* is less than one
  - For people with one friend, this means $|\beta_{1,1}| \leq 1$
  - For people with two friends, this means $|\beta_{1,2}| + |\beta_{2,2}| \leq 1$
  - ..and so on.

Assumption 1 (Bounded Peer Effect).

For all $d \leq \bar{d} = \max_{i=1,\ldots,n} d_i$

$$\sum_{j=1}^{d} |\beta_{j,d}| \leq 1$$

- This is a weakened version of the "classic" existence assumption of the Linear-in-Mean model

Reduced form

- Assume the network and covariates are exogenous to ϵ_i.

Assumption 2 ((Exogenous Network and Covariates)).

For each $i = 1, \ldots, n$

$$\mathbf{E}_n[\varepsilon_i] = \mathbf{E}\left[\varepsilon_i \middle| \{x_i, A_{i,j}\}_{1 \leq i,j \leq n}\right] = 0.$$

Reduced form

- Assume the network and covariates are exogenous to ϵ_i.

Assumption 2 ((Exogenous Network and Covariates)).

For each $i = 1, \ldots, n$

$$\mathbf{E}_n[\varepsilon_i] = \mathbf{E}\left[\varepsilon_i \middle| \{x_i, A_{i,j}\}_{1 \le i,j \le n}\right] = 0.$$

- Then under Assumption 1 and 2 we have

$$\mathbf{E}_n\left[Y_i\right] = \sum_{j=1}^{n} \theta_{i,j} x_j^{\intercal} \gamma + \eta_i$$

Where $\theta_{i,j}, \eta_i$ are complicated functions of $A_{i,j}$ and β Definitions

Identification through Instruments

Identification through Instruments

- Assume we have instruments z_i constructed from $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$
  - In implementation, we use the covariates of peers *ranked in order of the covariate values*

Identification through Instruments

- Assume we have instruments z_i constructed from $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$
  - In implementation, we use the covariates of peers *ranked in order of the covariate values*
- Then we can identify all parameters from the standard IV moments – if our instruments are relevant

Identification through Instruments

- Assume we have instruments z_i constructed from $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$
  - In implementation, we use the covariates of peers *ranked in order of the covariate values*
- Then we can identify all parameters from the standard IV moments – if our instruments are relevant
- With many instruments and many instrumental variables, this becomes a rank-condition on

$$\mathbf{E}_n[\mathbb{Z}\mathbb{W}],$$

where $\mathbb{Z}$ is a matrix containing our instruments and $\mathbb{W}$ is a matrix containing the covariates x_i and endogenous variables $\tilde{Y}_{i,k}$.

Identification through Instruments

- Assume we have instruments z_i constructed from $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$
  - In implementation, we use the covariates of peers *ranked in order of the covariate values*
- Then we can identify all parameters from the standard IV moments – if our instruments are relevant
- With many instruments and many instrumental variables, this becomes a rank-condition on

$$\mathbf{E}_n[\mathbb{Z}\mathbb{W}],$$

  where $\mathbb{Z}$ is a matrix containing our instruments and $\mathbb{W}$ is a matrix containing the covariates x_i and endogenous variables $\tilde{Y}_{i,k}$.
- Discussion in the paper, but easily checked in the data. Conditions for instrument

Identification

Theorem 1.

Suppose that Assumptions 1-3 hold and $n \geq \frac{\bar{d}(\bar{d}+1)}{2} + l$*. Then,* β *and* γ *are identified from the moment condition below:*

$$\mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{W}\right]\begin{pmatrix}\beta \\ \gamma\end{pmatrix} = \mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{Y}\right]. \tag{2}$$

Simulation set-up

- We simulate from the DGP

$$Y_i = \gamma_0 + \gamma_1 X_i + \tilde{Y}_i^\intercal \beta + \varepsilon_i$$

with X_i and ϵ_i as independent standard normals and $n = 200$.

- We fix $\bar{d}$ to either 2 or 5 for the simulations and vary γ_1 to see the effect of the instruments strength.
- We set $\beta_{d,d} = 0.4$ and $\beta_{k,d} = \frac{0.4}{d}$ for $k < d$.

Simulation results - Bias for People with two friends

	(1)	(2)	(3)	(4)	(5)
$\beta_{1,2}$	-0.005	-0.000	-0.004	-0.001	0.001
$\beta_{2,2}$	0.002	-0.000	0.000	-0.001	-0.000
γ_1	1	1	2	3	5
$\bar{d}$	2	5	5	5	5

Simulation results - MSE for People with 2 friends

	(6)	(7)	(8)	(9)	(10)
$\beta_{1,2}$	0.049	0.260	0.023	0.010	0.003
$\beta_{2,2}$	0.025	0.109	0.012	0.006	0.002
γ_1	1	1	2	3	5
$\bar{d}$	2	5	5	5	5

Simulation results - Bias for People with 5 friends

	(1)	(2)	(3)	(4)	(5)
$\beta_{1,5}$		0.058	-0.008	-0.002	-0.002
$\beta_{2,5}$		0.044	0.021	-0.009	0.007
$\beta_{3,5}$		-0.022	-0.019	0.031	0.000
$\beta_{4,5}$		-0.001	0.022	-0.045	-0.006
$\beta_{5,5}$		-0.029	-0.016	0.018	0.004
γ_1	1	1	2	3	5
$\bar{d}$	2	5	5	5	5

Simulation results - MSE for People with 5 friends

	(6)	(7)	(8)	(9)	(10)
$\beta_{1,5}$		1.872	0.305	0.082	0.017
$\beta_{2,5}$		4.371	1.500	0.523	0.094
$\beta_{3,5}$		5.635	2.233	0.788	0.143
$\beta_{4,5}$		4.531	1.339	0.428	0.086
$\beta_{5,5}$		1.300	0.240	0.072	0.013
γ_1	1	1	2	3	5
$\bar{d}$	2	5	5	5	5

Simulation Results - Bias in Restricted model

	(1)	(2)	(3)	(4)	(5)
$\beta_{-\text{max}}$		-0.005	-0.002	0.000	0.001
β_{max}		-0.010	-0.000	-0.001	-0.001
γ_1	1	1	2	3	5
$\bar{d}$	2	5	5	5	5

Simulation results - MSE for People with 5 friends

	(6)	(7)	(8)	(9)	(10)
$\beta_{-\max}$		0.104	0.011	0.004	0.001
$\beta_{\max}$		0.082	0.008	0.003	0.001
γ_1	1	1	2	3	5
$\bar{d}$	2	5	5	5	5

Set-up

- We investigate peer effects in GPA in two Norwegian middle schools using data from Alne et al. (2025)
- Sample consists of two schools with a total of around 550 students.
- Data is merged to registry data allowing for a large set of controls
    - See paper and Alne et al. (2025) for details.
- Compare Linear-in-Means with alternative models that separate out the highest/lowest performing peers of each individual.

	(1)	(2)	(3)	(4)	(5)
$\bar{Y}$	0.483***				
	(0.131)				
$\bar{Y}_{-d_i}$			-0.003		
			(0.278)		
$\bar{Y}_{-1}$		0.678**			
		(0.295)			
$\bar{Y}_{-1,-d_i}$				0.156	
				(0.475)	
$\tilde{Y}_{d_i}$			0.43	0.303	0.407***
			(0.313)	*(0.391)*	*(0.151)*
$\tilde{Y}_1$		-0.067		0.138	0.18*
		(0.221)		*(0.202)*	*(0.106)*
Observations	529	529	529	529	529
Adjusted R^2	0.476	0.474	0.450	0.480	0.476

Conclusion

- We show how to identify and estimate rank-dependent peer effects.
- The identifying conditions are natural extensions to standard assumptions, but with stronger requirements for the instruments.
- Simulations show the estimator performs well even in small samples
- In the paper we also show how to relate the estimand of the Linear-in-Mean model to our parameters and derive the asymptotic distribution of the estimators.

References I

Alne, R., E. I. Herstad, and A. Myhre (2025). How socioeconomic and parental background shape peer networks and educational spillovers. Available on SSRN: https://www.ssrn.com/abstract=5160331.

Definitions of θ and η

Define π as an ordering on $\{1, \ldots, n\}$ in terms of $\{y_i\}_{i=1}^{n}$ and $\mathbb{B}(\pi)$ denotes a $n \times n$ peer effect coefficient matrix such that its i-th row j-th column component corresponds to the peer effect coefficient for individual j's outcome on individual i's outcome, given the ordering π.
Then the reduced form coefficients can be written as

$$\theta_{i,j} = \sum_{\pi \in \Pi} \theta_{i,j}(\pi) \mathrm{Pr}_n \{\pi\}$$

$$\eta_i = \sum_{\pi \in \Pi} \sum_{j=1}^{n} \theta_{i,j}(\pi) \mathbb{E}[\varepsilon_j | \pi] \cdot \mathrm{Pr}_n \{\pi\}$$

and $\left(\theta_{i,1}(\pi), \ldots, \theta_{i,n}(\pi)\right)$ is the i-th row of the $n \times n$ matrix $(I_n - \mathbb{B}(\pi))^{-1}$

Back

Instrument conditions

1. (Exogeneity) $\{z_{i,1}, \ldots, z_{i,\bar{d}}\}_{i=1}^{n}$ are known, predetermined functions of $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$.
2. (Relevance) The construction of the instrument z_i and the reduced-form representation of

$$\mathbb{E}\left[\tilde{Y}_{i,k}\right] = \sum_{j=1}^{n} \tilde{\theta}_{i,k,j} {x_j}^{\mathsf{T}} \gamma + \tilde{\eta}_{i,k}$$

from Corollary 1 satisfy that

$$\sum_{i=1}^{n} \begin{pmatrix} z_{i,d} \\ x_i \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} \tilde{\theta}_{i,1,j} {x_j}^{\mathsf{T}} \gamma + \tilde{\eta}_{i,1} \\ \vdots \\ \sum_{j=1}^{n} \tilde{\theta}_{i,d_i,j} {x_j}^{\mathsf{T}} \gamma + \tilde{\eta}_{i,d_i} \\ x_i \end{pmatrix}^{\mathsf{T}} \mathbf{1}\{d_i = d\}$$

has full rank, for each $d = 1, \ldots, \bar{d}$.

Back