

# Identification of a rank-dependent peer effect model

Eyo I. Herstad\* and Myungkou Shin<sup>†</sup>

February 28, 2025

## Abstract

We develop a model that captures peer effect heterogeneity by modeling the endogenous spillover to be linear in ordered peer outcomes. Unlike the canonical linear-in-means model, our approach accounts for the distribution of peer outcomes as well as the size of peer groups. Under a minimal condition, our model admits a unique equilibrium and is therefore tractable and identified. Simulations show our estimator has good finite sample performance. Finally, we apply our model to educational data from Norway, finding that higher-performing friends disproportionately drive GPA spillovers. Our framework provides new insights into the structure of peer effects beyond aggregate measures.

## 1 Introduction

Economic agents are often connected. These connections raise important questions about how linked units influence one another through spillovers, known as peer effects. It has been well established in many empirical settings that peer effects are heterogeneous (Sacerdote, 2011; Masten, 2018; Bramoullé, Djebbari and Fortin, 2020). Despite this, the canonical models of peer effects assume homogeneous effects from peers (Manski, 1993; Bramoullé, Djebbari and Fortin, 2009). This paper develops an econometric framework that allows for flexible and heterogeneous patterns of peer effects. Specifically, we will focus on heterogeneous effects based on the ranking of each peer’s outcome among the peer group: a rank-dependent peer effect model. As such, the effect of each peer of an individual is endogenously determined by the distribution of the outcomes of that individual’s peers.

This framework provides empirical researchers with a new tool to uncover patterns in peer influence, particularly in settings where relative standing plays a crucial role, such as education, workplace productivity, and social interactions. For instance, when students in a school interact, the skills of your highest-performing friend may be the key determinant of how well you perform in school. On the other hand, in a manufacturing setting where workers operate on a conveyor-belt-style assembly line, the performance of the slowest worker might

---

\*Corresponding author. Erasmus University Rotterdam and Tindbergen Institute. email: herstad@ese.eur.nl

<sup>†</sup>School of Social Sciences, University of Surrey. email: m.shin@surrey.ac.uk

be the key factor determining the spillover effect. This motivates a model of peer effects that allows for differently ranked peers to have different effects on an individual's outcome.

The current literature on peer effects has primarily focused on estimating a scalar spillover parameter. The canonical model for peer effects has, following the seminal works of Manski (1993) and Bramoullé, Djebbari and Fortin (2009), been the linear-in-means (LIM) peer effect model. This model is defined as: for  $i = 1, \dots, n$ ,

$$Y_i = \beta \bar{Y}_i + x_i^\top \gamma + \varepsilon_i. \quad (1)$$

For example, consider spillovers in exam scores. Then  $Y_i$  is the exam score for student  $i$ ,  $\bar{Y}_i$  is the mean of student  $i$ 's peers' exam scores and  $x_i$  is the vector of student-level covariates for student  $i$ .  $x_i$  could also include contextual effects: e.g., the mean of student  $i$ 's peers' covariates. In the LIM model, the total peer effect is the product of  $\bar{Y}_i$  and a scalar peer effect parameter  $\beta$ .

However, the LIM model has several caveats, limiting the patterns of spillover it can analyse. For example, the LIM model assumes that inputs from different peers are perfectly substitutable. As such, there is no compositional effect from having perfectly homogeneous peers compared to having heterogeneous peers, as long as the mean is equal. This means that the LIM model assumes that the higher moments such as standard deviation of peer outcomes does not matter.

The limitations of the LIM model are also well established in the empirical literature (Sacerdote, 2011). To overcome them, empirical papers have investigated heterogeneous patterns of peer effects by interacting average peer outcomes with exogenous variables, or by estimating non-linear functions of the average outcome of peers (Foster, 2006; Han and Li, 2009; Guryan, Kroft and Notowidigdo, 2009). More recently, there has been extensions to the LIM model by allowing for more general aggregators than the mean (Boucher et al., 2024). However, this analysis is still limited to looking at the effect of peer outcomes through a single aggregated measure. We further expand the scope of peer effects by developing a generalized econometric model that flexibly estimates heterogeneous spillovers from peers under an exchangeability assumption<sup>1</sup>. Specifically, we will let the inputs from peers affect an individual's outcome differently depending on the peer's relative position in the individual's peer group.

In our generalized model, the peer effect is a linear function of 'ordered' peer outcomes. That is, for  $i = 1, \dots, n$ ,

$$Y_i = \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} + x_i^\top \gamma + \varepsilon_i. \quad (2)$$

$d_i$  is the number of peers for individual  $i$  and  $\tilde{Y}_{i,k}$  is the outcome of the  $k$ -th lowest performing peer of individual  $i$ ;  $\tilde{Y}_{i,k}$  takes the  $k$ -th lowest value from the set

$$\{Y_j : \text{individual } j \text{ is a peer of individual } i\}.$$

---

<sup>1</sup>In this context, exchangeability means that the spillover an individual experiences does not depend on the identity of their peers. This is implicitly assumed in existing models of peer effects, but needs to be made explicit in our model due to its flexible nature.

This model identifies the general peer effect under an assumption that peers are exchangeable and that the peer effect is linear in ordered peer outcomes. Note that this does not impose linear separability of peer outcomes. To see this, consider two connected individuals,  $A$  and  $B$ . The impact of  $A$ 's outcome on  $B$ 's outcome depends on the outcomes of all the individuals that individual  $B$  is connected to. This means that both the magnitude and sign of the peer effect coefficient from individual  $A$  to  $B$  is endogenous to all the peers of  $B$ , as their outcomes decide the ranking of  $A$  in  $B$ 's peer group. In contrast, the canonical LIM model assumes this peer effect coefficient only depends on how many peers individual  $B$  has.

The generality of this model allows us to capture many economically relevant features of peer effects. Firstly, as the model does not impose perfect substitution, it can analyse whether there are complementarities or substitutability in peer effects. Secondly, as the spillover from a peer can depend on the size of the peer group, the model can detect if individuals with more connections are systemically different from other individuals. For example, consider a simple model where individuals are affected differently by their relatively higher- and lower-performing peers:

$$Y_i = \beta_{\text{low}} \sum_{k=1}^{\lceil d_i/2 \rceil} \tilde{Y}_{i,k} + \beta_{\text{high}} \sum_{k=\lceil d_i/2 \rceil+1}^{d_i} \tilde{Y}_{i,k} + x_i^\top \gamma + \varepsilon_i. \quad (3)$$

This model is a simplification of the rank-dependent peer effect model, with  $\beta_{1,d} = \dots = \beta_{\lceil d/2 \rceil, d} = \beta_{\text{low}}$  and  $\beta_{\lceil d/2 \rceil+1, d} = \dots = \beta_{d,d} = \beta_{\text{high}}$  for each  $d$ . When  $\beta_{\text{low}} = \beta_{\text{high}}$ , all peer outcomes are perfectly substitutable with each other, as in the LIM model. When  $\beta_{\text{low}} \neq \beta_{\text{high}}$ , a sufficiently large change will move a peer from the lower-performing half to higher-performing half. This will change that peer's coefficient, and therefore also change their relative rate of substitution with the other peers.

Similarly, as the models' coefficients depend on  $d_i$ , it allows for returns to scale in the number of connections. As an illustrative example, consider a model where the median of peer outcomes is the sole determining factor in the peer effect, this becomes:

$$Y_i = \beta_{d_i} \tilde{Y}_{i, \lceil d_i/2 \rceil} + x_i^\top \gamma + \varepsilon_i.$$

If  $\beta_d < \beta_{d'}$  when  $d < d'$ , we would conclude that peer effects exhibit returns to scale in the number of connections.

More generally, the rank-dependent peer effect model can be motivated as a general regression of an individual's outcome on the distribution of their peers' outcomes. As will be further discussed in Section 2, ordered peer outcomes have one-to-one relationship with the (empirical) distribution of peer outcomes. As such, our model allows for a wide variety of heterogeneous patterns in the peer effects in terms of the distribution of peer outcomes.

The primary contribution of this paper is the introduction and identification of the rank-dependent peer effect model, as shown in Equation (2). Following the literature on peer effects, our identification result uses the conditional distribution of the outcomes given the network connections and the individual characteristics, with the network assumed to be exogenous. The steps of the identification argument are as follows. First, we show that the rank-dependent peer effect model admits a unique equilibrium when the magnitude of the total peer effect is bounded by one. Then, using the unique equilibrium result, we derive

a reduced-form representation of the model. Given the reduced-form representation, we construct moment conditions to identify the peer effect coefficients  $\{\beta_{k,d}\}_{k,d}$  and  $\gamma$ . This step relies on an instrument relevance condition closely related to the equivalent condition of the LIM model.

In implementation, we use Two-stage Least-squares (TSLS) to estimate the parameters. We show, through simulation exercises, how the finite-sample performance of the TSLS estimator depends on the complexity of the model, the sample size and the strength of our instruments. We also show how ex-ante knowledge about the structure of peer effects can aid in estimation in settings where the fully saturated model is challenging to estimate. We then show how our estimator can be applied to empirical settings by estimating spillovers in learning between students at two Norwegian middle schools. Our estimates show that the homogeneous estimates of the LIM model does not capture the heterogeneity of peer effects we find in the data. Instead, the highest-GPA peers are the driving factor in GPA spillovers. This supports the empirical relevance of the rank-dependent peer effect model in real world settings.

The paper most closely related to ours is Boucher et al. (2024). They let peer outcomes be aggregated to a scalar summary measure by a broad class of aggregators, which determines the peer effect. This class of aggregators nests the mean used in the LIM model. Unlike our model, this aggregator depends on a single parameter, and does not allow for returns to centrality. In two extreme cases of their model, Boucher et al. (2024) construct the peer effect to be linear in the minimum peer outcome or the maximum peer outcome alone, both of which are allowed in our model as well. Similarly, Tao and Lee (2014) shows how to identify a model in which only the highest outcome friend matters for the peer effect. This is a special case of our model with  $\beta_{d,d} = \beta_{max}$  for all  $d$ , and zero otherwise.

The empirical study of peer effects has been a large literature in economics (Bramoullé, Djebbari and Fortin, 2020), with a seminal contribution by Sacerdote (2001). Some of the papers in this literature are close in spirit to our model, showcasing an interest in exploring these kinds of heterogeneities in the empirical literature. Existing work has, for example, let the peer effect be the mean of non-linear functions of peer outcomes, such as the proportion of peers above a cutoff (Kang, 2007; Guryan, Kroft and Notowidigdo, 2009), through quantile regressions (Kang, 2007), or looked at heterogeneity by interacting the peer effect with exogenous variables (Foster, 2006; Han and Li, 2009; Alne, Herstad and Myhre, 2025; Tincani, 2024). To the best of our knowledge, however, there is no previous work allowing for peer effects to depend on the distribution of peer outcomes, or a theoretical discussion of these methods.

The rest of this paper is structured as follows. Section 2 discusses the model and its implications. Section 3 formally identifies the parameters of the rank-dependent peer effect model. Section 4 discusses finite-sample performances of a TSLS estimator in the rank-dependent peer effect model, using simulations. Finally, Section 5 applies our methodology to educational peer effects in Norwegian schools.

## 2 Model

In our setting, a researcher observes individual-level outcomes and individual-level control covariates  $\{Y_i, x_i\}_{i=1}^n$  and the network adjacency matrix  $\{A_{i,j}\}_{1 \leq i,j \leq n}$ . We let  $A_{i,j} = 1$  indicate that individual  $i$  is linked to individual  $j$ , and  $A_{i,j} = 0$  otherwise. We assume there are no self-loops in the network, so  $A_{i,i} = 0$ , and links may be directed, i.e.,  $A_{i,j}$  and  $A_{j,i}$  are allowed to be different.

Consider a general model of nonparametric peer effects:

$$Y_i = h(\{A_{i,j}Y_j\}_{j=1}^n) + x_i^\top \gamma + \varepsilon_i. \quad (4)$$

We call  $h(\cdot)$  the peer effect function, or simply the peer effect, which takes a  $n \times 1$  vector and produces a scalar peer effect.

**Assumption 1** (Exchangeability).

$$h(\{A_{i,j}Y_j\}_{j=1}^n) = h(\{A_{i,\pi(j)}Y_{\pi(j)}\}_{j=1}^n)$$

for any permutation  $\pi$  on  $\{1, \dots, n\} \setminus \{i\}$

**Lemma 1.** Under Assumption 1, there exists some function  $\tilde{h}$  such that

$$h(\{A_{i,j}Y_j\}_{j=1}^n) = \tilde{h}(\{\tilde{Y}_{i,k}\}_{k=1}^{d_i}, d_i).$$

Furthermore, if  $\tilde{h}(\cdot)$  is linear then

$$h(\{A_{i,j}Y_j\}_{j=1}^n) = \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i}.$$

*Proof.* For any realized values of  $\{A_{i,j}, Y_j\}_{j \neq i}$  such that student  $i$  has  $d$  peers and their ordered peer outcomes are  $\{\tilde{y}_{i,k}\}_{k=1}^d$ , Assumption 1 assumes that the peer effect function  $h$  has the same value. By constructing  $\tilde{h}$  to be a function that maps the number of peers and the ordered peer outcomes to the evaluated  $h$ , Lemma 1 holds.  $\square$

Assumption 1 assumes that the peer effect function  $h$  is invariant to a permutation on its inputs. Then, we can construct a (derived) peer effect function  $\tilde{h}$  that takes the *ordered* peer outcomes and the number of peers as its inputs, with which the nonlinear peer effect model (4) holds. By additionally assuming linearity on  $\tilde{h}$ , i.e.

$$\tilde{h}(\{\tilde{Y}_{i,k}\}_{k=1}^{d_i}, d_i) = \sum_{k=1}^{d_i} \beta_{k,d_i} \tilde{Y}_{i,k},$$

we obtain the rank-dependent peer effect model of our paper. In this sense, the key motivation for the rank-dependent peer effect model is Assumption 1 where we assume that the names of individuals do not matter.

We define the rank-dependent peer effect model as:

$$Y_i = \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} + x_i^\top \gamma + \varepsilon_i \quad (5)$$

for  $i = 1, \dots, n$ , where

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

$$\tilde{Y}_{i,k} = \text{outcome of the } k\text{-th lowest performing peer of individual } i$$

$$= k\text{-th ordered statistic from } \{Y_j : A_{i,j} = 1\}.$$

Note that  $\tilde{Y}_{i,1} \leq \dots \leq \tilde{Y}_{i,d_i}$  for any  $i$  and  $\tilde{Y}_{i,k} = Y_j$  for some  $j \neq i$ . The individual-level control covariates  $x_i$  may include functions of the network adjacency matrix  $\{A_{i,j}\}_{1 \leq i,j \leq n}$  and the peer characteristics. For example, given some individual-specific characteristics  $\{w_i\}_{i=1}^n$ , the control covariate  $x_i$  could be functions of  $\{w_i\}_{i=1}^n$  and  $\{A_{i,j}\}_{1 \leq i,j \leq n}$ , allowing for contextual effects:

$$Y_i = \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} + \left( w_i \quad \frac{1}{d_i} \sum_{j \neq i} A_{i,j} w_j \right)^\top \gamma + \varepsilon_i =: \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} + x_i \gamma + \varepsilon_i.$$

The rank-dependent peer effect model (5) assumes that the outcomes of individual  $i$ 's peers affect individual  $i$  in a way that the coefficient on a peer outcome depends on the peer's rank among individual  $i$ 's peer group.  $\beta_{k,d}$  is the coefficient for the  $k$ -th lowest performing peer's outcome, when the given individual has  $d$  peers in total. Note that the model allows a individual to affect different peers differently. Suppose that individual 1 is friends with both individual 2 and 3. If individual 1 is the lowest-performing peer of individual 2 while being the highest-performing peer of individual 3, the effect of individual 1's outcome will be  $\beta_{1,d_2}$  for individual 2 and  $\beta_{d_3,d_3}$  for individual 3.

The model nests both the linear-in-means (LIM) model and the linear-in-sums (LIS) model. By letting  $\beta_{k,d} = \beta/d$ , the peer effect term in (5) becomes the LIM peer effect:

$$\sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} = \frac{1}{d_i} \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta = \beta \bar{Y}_i \quad (6)$$

with  $\bar{Y}_i = \frac{1}{d_i} \sum_{k=1}^{d_i} \tilde{Y}_{i,k} = \frac{\sum_{j=1}^n A_{i,j} Y_j}{\sum_{j=1}^n A_{i,j}}$ . By letting  $\beta_{k,d} = \beta$ , the peer effect term becomes the LIS peer effect:

$$\sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} = \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta = \beta \sum_{j=1}^n A_{i,j} Y_j. \quad (7)$$

Though widely used, the LIM model and the LIS model have a caveat that using only a mean or a sum may ignore important information contained in the second or higher moments of the peer outcomes.<sup>2</sup>

The flexibility of our peer effect model (5) allows us to address key questions in the peer effect literature. First, because the coefficients are rank-dependent, our model enables the

---

<sup>2</sup>A notable exception is where the peer effect occurs through a binary variable, as in (Gazze, Persico and Spirovska, 2024)

analysis of complementarity versus substitutability, as has been discussed in Kang (2007) amongst others. For instance, when  $\beta_{1,d} \neq 0$  and  $\beta_{k,d} = 0$  for every  $k \geq 2$  and  $d$ , the peer effect reflects perfect complementarity, meaning that only the lowest-performing peer influences the outcome. On the other hand, when  $\beta_{k,d} = \beta_{k',d}$  for all  $k, k'$ , the peer effect exhibits global perfect substitutability, where each peer contributes equally. It is important to note that the linear structure of our model imposes local perfect substitutability. Specifically, all peer outcomes are perfectly substitutable with a constant marginal rate of substitution for small changes in  $\tilde{Y}_{i,k}$  such that  $\tilde{Y}_{i,k-1} \leq \tilde{Y}_{i,k} + \Delta \leq \tilde{Y}_{i,k+1}$ . However, when the change in  $\tilde{Y}_{i,k}$  is large enough to alter its rank, the slope of the isoquant curve adjusts accordingly.

Secondly, because the coefficients also depend on  $d_i$ , the size of the peer group, our peer effect model enables the exploration of returns to centrality. For simplicity, assume  $Y_j = y$  for every  $j$  in the peer group of individual  $i$ . In the LIM model, there is zero return to centrality, meaning that the peer effect remains unchanged when an additional peer is added. The LIS model imposes a constant return to centrality. In contrast, our model we can document the return to centrality in a flexible manner, by plotting  $d \mapsto \sum_{k=1}^d \beta_{k,d} y$ . These two features of peer effect heterogeneity are directly related to several empirically important questions: does a student benefit from a high-performing peer or get negatively influenced by a low-performing one? how substitutable are peer outcomes? does the number of friends impact the spillover effect?

While the rank-dependent peer effect model (5) allows for nonlinearity in the peer effect, allowing for the peer effect of one peer to depend on the outcomes of all peers, it still admits a unique equilibrium under a natural extension of the assumption used in LIM models (Bramoullé, Djebbari and Fortin, 2009).

**Assumption 2.** (BOUNDED PEER EFFECT) *There exists some  $\bar{d}$  such that  $\max_i d_i \leq \bar{d}$ . For each  $d = 1, \dots, \bar{d}$ ,*

$$\sum_{k=1}^d |\beta_{k,d}| < 1.$$

Assumption 2 assumes that the sum of the absolute values of peer effect coefficients for each individual is less than 1. Proposition 1 states our unique equilibrium result.

**Proposition 1.** *Let Assumption 2 hold. Then, for any realization of  $\{x_i, A_{i,j}, \epsilon_i\}_{1 \leq i,j \leq n}$ , the peer effect model (5) admits a unique equilibrium.*

*Proof.* See Appendix B.1 □

Since the peer effect model admits a unique equilibrium, there exists a well-defined function from  $\{\epsilon_i\}_{i=1}^n$  to  $\{Y_i\}_{i=1}^n$  given  $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$ .

The unique equilibrium result from Proposition 1 allows us to introduce a key ingredient in the identification of the peer effect parameters: the conditional distribution of  $\{Y_i\}_{i=1}^n$  given  $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$ . Without a unique equilibrium result as in Proposition 1, we would have to assume an equilibrium selection mechanism to derive a conditional distribution of  $\{Y_i\}_{i=1}^n$  given  $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$  from a conditional distribution of  $\{\epsilon_i\}_{i=1}^n$  given

$\{x_i, A_{i,j}\}_{1 \leq i, j \leq n}$ .<sup>3</sup> However, thanks to Proposition 1, we can assume Assumption 2 and additional conditions on the distribution of  $\{\varepsilon_i\}_{i=1}^n$ , to construct moment conditions with  $\{Y_i\}_{i=1}^n$ .

### 3 Identification

To simplify notation, we use  $\mathbf{E}_n$  to denote the conditional expectation of a random variable given the exogenous variables throughout the paper. That is for any random variable  $W$ ,

$$\mathbf{E}_n[W] := \mathbf{E}[W | \{x_i, A_{i,j}\}_{1 \leq i, j \leq n}].$$

We will follow the literature by assuming that the error term  $\varepsilon_i$  is exogenous to all individual characteristics and the network.

**Assumption 3.** (EXOGENOUS NETWORK AND COVARIATES) *For each  $i = 1, \dots, n$*

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

Assumption 3 rules out endogenous network formation and treats the network structure as fixed, meaning it does not depend on the randomness in  $\{\varepsilon_i\}_{i=1}^n$ . While there has been advances in analysing peer effect models with endogenous network formation (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021; Jochmans, 2023), we instead focus on endogeneity in peer effects conditional on the network.<sup>4</sup>

Assumption 3 allows us to construct a reduced-form representation of  $\mathbf{E}_n[Y_i]$ , though unlike the LIM model this reduced form is not linear in  $\{x_i\}_{i=1}^n$ . Instead, the rank-dependence creates a nonlinear reduced-form representation as shown in Corollary 1. To discuss the coefficients in the corollary, let us introduce the following notations:  $\pi$  denotes an ordering on  $\{1, \dots, n\}$  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  $\pi$ . For any realization of  $\{\varepsilon_i\}_{i=1}^n$ , Proposition 1 gives us a unique equilibrium  $\{y_i\}_{i=1}^n$  and thus a unique  $\pi$  such that

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbb{B}(\pi) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} x_1^\top \gamma + \varepsilon_1 \\ \vdots \\ x_n^\top \gamma + \varepsilon_n \end{pmatrix}.$$

---

<sup>3</sup>Identification does not necessarily require a unique equilibrium, see for example De Paula (2013) for examples from the literature on games with multiple equilibria. For example, an alternative identification strategy in our setting based on this literature would be to limit our attention to sets of  $\{y_i\}_{i=1}^n$  that can only be an outcome of a single realization of  $\{\varepsilon_i\}_{i=1}^n$ . Or, we may focus on extreme values of  $\{x_i\}_{i=1}^n$  and discuss identification for samples with extreme values only. Though these approaches have been valuable in the contexts discussed in De Paula (2013), they are not straightforward to implement in our setting. If there were multiple equilibria in the rank-dependent peer effect model, there would be  $n!$ , a massive number, different orderings to consider to find values of  $\{y_i\}_{i=1}^n$  consistent with  $\{\varepsilon_i\}_{i=1}^n$ .

<sup>4</sup>To see this, note that we can view our model as a second step of a two-stage process where firstly, the network is formed and secondly, the peer effect is determined based on the connections established in the first step. Models of endogenous network formation address endogeneity in the first step, whereas our model addresses endogeneity in the second step by allowing peer effect coefficients to depend on  $\{\varepsilon_i\}_{i=1}^n$ .

Corollary 1 takes expectations over the set of every possible ordering,  $\Pi$ , and gives us a reduced-form representation of  $\mathbf{E}_n[Y_i]$ .

**Corollary 1.** *Let Assumptions 2 and 3 hold. Then, the peer effects model (5) admits a reduced-form relationship with residual terms  $\eta_i$ :*

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

where

$$\begin{aligned} \theta_{i,j} &= \sum_{\pi \in \Pi} \theta_{i,j}(\pi) \Pr_n\{\pi\} \\ \eta_i &= \sum_{\pi \in \Pi} \sum_{j=1}^n \theta_{i,j}(\pi) \mathbf{E}_n[\varepsilon_j | \pi] \cdot \Pr_n\{\pi\} \end{aligned}$$

and  $(\theta_{i,1}(\pi), \dots, \theta_{i,n}(\pi))$  is the  $i$ -th row of the  $n \times n$  matrix  $(I_n - \mathbb{B}(\pi))^{-1}$ . Moreover, there also exists a reduced-form relationship between the ordered peer outcome  $\tilde{Y}_{i,k}$  and  $\{x_j\}_{j=1}^n$ , with residual terms  $\{\tilde{\eta}_{j,k}\}_{j=1}^n$ :

$$\mathbf{E}_n[\tilde{Y}_{i,k}] = \sum_{j=1}^n \tilde{\theta}_{i,k,j} x_j^\top \gamma + \tilde{\eta}_{i,k}$$

where

$$\begin{aligned} \tilde{\theta}_{i,k,j} &= \sum_{\pi \in \Pi} \tilde{\theta}_{i,k,j}(\pi) \Pr_n\{\pi\} \\ \tilde{\eta}_{i,k} &= \sum_{\pi \in \Pi} \sum_{j=1}^n \tilde{\theta}_{i,k,j}(\pi) \mathbf{E}_n[\varepsilon_j | \pi] \cdot \Pr_n\{\pi\}. \end{aligned}$$

and  $(\tilde{\theta}_{i,k,1}(\pi), \dots, \tilde{\theta}_{i,k,n}(\pi))$  is one row of the  $n \times n$  matrix  $(I_n - \mathbb{B}(\pi))^{-1}$ , which corresponds to the  $k$ -th lowest performing peer of individual  $i$ .

*Proof.* See Appendix B.2 □

Note that the residual terms  $\eta_i$  and  $\tilde{\eta}_{i,k}$  may not be linear in  $\{x_i\}_{i=1}^n$ .

To compare the reduced form from Corollary 1 with more familiar models, such as the LIM model in (1) and the LIS model, observe that for these models, Assumptions 2 and 3 directly yield a reduced-form representation of  $\{Y_i\}_{i=1}^n$ :

$$\begin{pmatrix} \mathbf{E}_n[Y_1] \\ \vdots \\ \mathbf{E}_n[Y_n] \end{pmatrix} = (I_n - \mathbb{B})^{-1} \begin{pmatrix} x_1^\top \gamma \\ \vdots \\ x_n^\top \gamma \end{pmatrix} \quad (8)$$

where the  $n \times n$  matrix  $\mathbb{B}$  is  $\beta G := \beta \cdot (A_{i,j}/d_i)_{i,j}$  in the LIM model and  $\beta A := \beta \cdot (A_{i,j})_{i,j}$  in the LIS model. Importantly, this equation is linear in  $\{x_i\}_{i=1}^n$ .

However, this linearity does not hold in our model. Instead, the rank-dependent peer effect coefficients introduce nonlinearities in the conditional expectations due to the connection between ordering and  $\{x_i\}_{i=1}^n$ . Consider a simple case with three individuals where individual 1 is connected to both individuals 2 and 3, but individuals 2 and 3 are not connected to each other. From Assumption 3, we have:

$$\begin{aligned} \mathbf{E}_n[Y_1] &= \beta_{12} \mathbf{E}_n[Y_2 \mathbf{1}\{Y_2 \leq Y_3\} + Y_3 \mathbf{1}\{Y_2 > Y_3\}] \\ &\quad + \beta_{22} \mathbf{E}_n[Y_3 \mathbf{1}\{Y_2 \leq Y_3\} + Y_2 \mathbf{1}\{Y_2 > Y_3\}] + x_1^\top \gamma. \end{aligned}$$

Substituting  $Y_2$  and  $Y_3$  using their reduced forms, we get:

$$\begin{aligned} \tilde{\beta} \mathbf{E}_n[Y_1 | \{x_i, A_{i,j}\}_{1 \leq i,j \leq n}] &= \beta_{12} \mathbf{E}_n[\varepsilon_2 \mathbf{1}\{Y_2 \leq Y_3\} + \varepsilon_3 \mathbf{1}\{Y_2 > Y_3\}] \\ &\quad + \beta_{22} \mathbf{E}_n[\varepsilon_3 \mathbf{1}\{Y_2 \leq Y_3\} + \varepsilon_2 \mathbf{1}\{Y_2 > Y_3\}] + \sum_{i=1}^3 x_i^\top \tilde{\gamma}_i \end{aligned}$$

for some  $\tilde{\beta}$ ,  $\tilde{\gamma}_1$ ,  $\tilde{\gamma}_2$ , and  $\tilde{\gamma}_3$ . Without further assumptions on the distribution of  $\{\varepsilon_i\}_{i=1}^n$ , the conditional expectations involve terms like  $\mathbf{E}_n[\varepsilon_2 \mathbf{1}\{Y_2 \leq Y_3\}]$ , which may not be zero or linear in  $x_i$ . An exception occurs when  $\beta_{12} = \beta_{22}$ , in which case the conditional expectations cancel out, resulting in a form similar to the LIM or LIS model.

These complications affect the relevance conditions we use to construct valid moment conditions. In the standard LIM and LIS models, the relevance condition on a set of instruments  $\{z_i\}_{i=1}^n$  for identifying  $\beta$  is given by ensuring that the matrix

$$(z_1 \quad \cdots \quad z_n) \begin{pmatrix} \mathbb{B} (I_n - \mathbb{B})^{-1} \begin{pmatrix} x_1^\top \gamma \\ \vdots \\ x_n^\top \gamma \end{pmatrix} & \begin{pmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{pmatrix} \end{pmatrix} \quad (9)$$

has full rank. Bramoullé, Djebbari and Fortin (2009) discusses a similar condition in the LIM model when instruments are based on the mean covariates of peers or peers of peers.

Our conditions are different and will instead relate to the reduced forms presented in Corollary 1. These conditions are given in Assumption 4.

**Assumption 4.**

- a. (EXOGENEITY)  $\{z_{i,1}, \dots, z_{i,d}\}_{i=1}^n$  are known, predetermined functions of  $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$ .*
- b. (RELEVANCE) The construction of the instrument  $z_i$  and the reduced-form representation of*

$$\mathbf{E}_n[\tilde{Y}_{i,k}] = \sum_{j=1}^n \tilde{\theta}_{i,k,j} x_j^\top \gamma + \tilde{\eta}_{i,k}$$

from Corollary 1 satisfy that

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

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

Part **a.** of Assumption 4, together with Assumption 3, provides the standard instrument exogeneity assumption commonly found in the literature. This includes, for example, instruments based on the average covariates of peers and peers-of-peers, as introduced by Bramoullé, Djebbari and Fortin (2009) for estimating the effects of peers' average outcomes. Given the increased number of endogenous variables in our model compared to the LIM model, a larger set of instruments is required. Beyond the instruments discussed in Bramoullé, Djebbari and Fortin (2009), we also consider alternatives such as ordered peer covariates or higher moments of peer covariates. In the following discussion, we do not commit to a specific set of instruments, we focus on outlining the necessary conditions that any valid instruments must satisfy for identification.

Assumption 4-**b.** establishes a rank condition analogous to those found in Bramoullé, Djebbari and Fortin (2009). The matrix in Assumption 4-**b.** decomposes into two components:

$$\sum_{i:d_i=d} \begin{pmatrix} z_{i,d} \\ x_i \end{pmatrix} \begin{pmatrix} \sum_{j=1}^n \tilde{\theta}_{i,1,j} x_j^\top \gamma \\ \vdots \\ \sum_{j=1}^n \tilde{\theta}_{i,d_i,j} x_j^\top \gamma \\ x_i \end{pmatrix}^\top \quad \text{and} \quad \sum_{i:d_i=d} \begin{pmatrix} z_{i,d} \\ x_i \end{pmatrix} \begin{pmatrix} \tilde{\eta}_{i,1} \\ \vdots \\ \tilde{\eta}_{i,d} \\ \mathbf{0}_l \end{pmatrix}^\top.$$

Constructing instruments  $\{z_{i,d}\}_{i=1}^n$  such that the first matrix is full rank is straightforward as long as there is sufficient non-transitivity in the network and variation in the covariates  $\{x_i\}_{i=1}^n$ . Specifically, we need the columns of  $\mathbb{X} = (x_1, \dots, x_n)^\top$  and

$$\begin{pmatrix} \vdots \\ \sum_{j=1}^n \tilde{\theta}_{i,1,j} x_j^\top \gamma \\ \vdots \end{pmatrix} \quad \cdots \quad \begin{pmatrix} \vdots \\ \sum_{j=1}^n \tilde{\theta}_{i,d_i,j} x_j^\top \gamma \\ \vdots \end{pmatrix}$$

to be linearly independent. This condition is closely related to that found in Bramoullé, Djebbari and Fortin (2009), where there are no residual terms  $\{\tilde{\eta}_{i,k}\}_{i,k}$ . As such, when  $\{\tilde{\eta}_{i,k}\}_{i,k}$  are close to zero, the sum of the two matrices will also have full rank, satisfying our relevance condition. In cases where  $\{\tilde{\eta}_{i,k}\}_{i,k}$  are nonzero, it is challenging to interpret the rank condition, but we can empirically check whether or not it holds in a given sample.

To simplify the notation of our identification results, let

$$\beta = (\beta_{11} \quad \beta_{21} \quad \beta_{22} \quad \cdots \quad \beta_{1\bar{d}} \quad \cdots \quad \beta_{\bar{d}\bar{d}})^\top$$

$$\tilde{Y}_i = (0 \quad \cdots \quad 0 \quad \tilde{Y}_{i,1} \quad \cdots \quad \tilde{Y}_{i,d_i} \quad 0 \quad \cdots 0)^\top$$

so that the rank-dependent peer effect model (5) can be written as follows:

$$Y_i = \tilde{Y}_i^\top \beta + x_i^\top \gamma + \varepsilon_i = W_i^\top \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \varepsilon_i. \quad (10)$$

Let  $W_i = (\tilde{Y}_i^\top, x_i^\top)^\top \in \mathbb{R}^{\frac{\bar{d}(\bar{d}+1)}{2}+l}$  and  $z_i = (z_{i,1}^\top, \dots, z_{i,\bar{d}}^\top, x_i^\top)^\top$ . Lastly, let  $\mathbb{Y}$ ,  $\mathbb{W}$  and  $\mathbb{Z}$  denote the row-stacked matrices of  $Y_i$ ,  $W_i$  and  $z_i$ . Theorem 1 provides the identification result.

**Theorem 1.** *Suppose that Assumptions 2-4 hold and  $n \geq \frac{\bar{d}(\bar{d}+1)}{2} + l$ . Then,  $\beta$  and  $\gamma$  are identified from the moment condition below:*

$$\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}] \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \mathbf{E}_n [\mathbb{Z}^\top \mathbb{Y}]. \quad (11)$$

*Proof.* See Appendix B.3. □

For identification, we consider a fixed  $n$  and  $\bar{d}$  environment. The minimal condition  $n \geq \frac{\bar{d}(\bar{d}+1)}{2} + l$ , along with Assumption 4, guarantees that the matrix  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}]$  has a left inverse. However, in practice, when  $\bar{d}$  is large relative to  $n$ , this can pose challenges. In such cases, it may be beneficial to consider an asymptotic framework where  $\bar{d}$  grows with  $n$  and thus the dimension of  $\beta$  also grows with  $n$ . If the parameter  $\beta$  satisfies some sparsity or smoothness restrictions, it may be possible to take advantage of recent developments for high-dimensional models with many endogenous variables. See for example Belloni, Hansen and Newey (2022). However, this literature primarily analyses cross-sectional data, and it is unclear if these results will extend to our setting due to the dependent nature of network data.

The moment condition (11) consists of  $n$  different equations, one for each student  $i = 1, \dots, n$ . Since  $\tilde{Y}_i$  concatenates  $\bar{d}$  different vectors of peer outcomes where only one out of the  $\bar{d}$  vectors is nonzero for a given student, the  $n \times (\frac{\bar{d}(\bar{d}+1)}{2} + l)$  matrix  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}]$  can also be decomposed into  $\bar{d}$  submatrices. Part **b.** of Assumption 4 gives us each of the  $\bar{d}$  submatrices having full rank, leading to  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}]$  having full rank as a result. Given that  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}]$  has full rank, it is trivial to write  $\beta$  and  $\gamma$  as functions of the conditional moments of  $\{Y_i\}_{i=1}^n$ . In the case of the TSLS estimand, the first stage coefficient matrix from regressing  $W_i$  on  $z_i$  is used. Let  $\Gamma$  denote the first stage coefficient matrix. Then,

$$\begin{pmatrix} \beta^{TSLS} \\ \gamma^{TSLS} \end{pmatrix} := (\mathbf{E}_n [(\mathbb{Z}\Gamma)^\top \mathbb{W}])^{-1} \mathbf{E}_n [(\mathbb{Z}\Gamma)^\top \mathbb{Y}] = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}.$$

For Sections 4 and 5, we use the sample analogue of the TSLS estimand as our estimator.

### 3.1 Comparison to other peer effect estimands

A natural question, given that our model generalizes the standard LIM model, is to what extent existing estimands recover the key parameters of interest from our model. A commonly accepted minimal standard for such estimands is that they represent weighted averages of the underlying heterogeneity, such as in the analysis of instrument variables (Mogstad and Torgovitsky, 2024) or the analysis of Difference-in-Difference estimands (De Chaisemartin

and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). To explore this, we first assume instruments that satisfy the relevance condition used in the LIM model and define a LIM estimand as follows:

$$\beta^{\text{LIM}} := (\mathbf{E}_n [\mathbb{Z}^\top \mathbb{M}_X G \mathbb{Y}])^{-1} \mathbf{E}_n [\mathbb{Z}^\top \mathbb{M}_X \mathbb{Y}]$$

where  $G = (A_{i,j}/d_i)_{i,j}$  and  $\mathbb{M}_X = I_n - \mathbb{X}(\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top$  with  $\mathbb{X}$ , a row-stacked matrix constructed with  $\{x_i\}_{i=1}^n$ . Similarly, for the LIS model, we define

$$\beta^{\text{LIS}} := (\mathbf{E}_n [\mathbb{Z}^\top \mathbb{M}_X A \mathbb{Y}])^{-1} \mathbf{E}_n [\mathbb{Z}^\top \mathbb{M}_X \mathbb{Y}]$$

where  $A = (A_{i,j})_{i,j}$ .

**Assumption 5.**

- i* There exist an instrument  $\{z_i\}_{i=1}^n$ , a known, predetermined function of  $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$ , that satisfies  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{M}_X G \mathbb{Y}] \neq 0$ .
- ii* There exist an instrument  $\{z_i\}_{i=1}^n$ , a known, predetermined function of  $\{x_i, A_{i,j}\}_{1 \leq i,j \leq n}$ , that satisfies  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{M}_X A \mathbb{Y}] \neq 0$ .

These conditions are the same instrument relevance conditions used in the LIM model and the LIS model, as in Bramoullé, Djebbari and Fortin (2009).

**Proposition 2.** Under Assumption 2, 3 and 5-i,

$$\beta^{\text{LIM}} = \sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIM}} d \beta_{k,d}$$

where  $\sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIM}} = 1$  and  $w_{k,d}^{\text{LIM}} \leq 0$ . Similarly, under Assumptions 2, 3 and 5-ii,

$$\beta^{\text{LIS}} = \sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIS}} \beta_{k,d}$$

where  $\sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIS}} = 1$  and  $w_{k,d}^{\text{LIS}} \leq 0$ .

*Proof.* See Appendix B.4. □

Note that, in the case of LIM misspecification, the weights are applied not directly to  $\beta_{k,d}$  but to  $d\beta_{k,d}$ . The rank-dependent peer effect coefficient  $\beta_{k,d}$  represents the impact of a single peer outcome when there are  $d$  peers in total, so it should be rescaled by  $d$  to interpret it as a coefficient on a ‘representative’ or ‘average’ peer. For instance, when  $\beta_{1,d} = \dots = \beta_{d,d} = \beta$ , we have:

$$\sum_{k=1}^d \beta_{k,d} \tilde{y}_{i,k} = d\beta \bar{y}_i,$$

where the coefficient on the average peer outcome is  $d\beta_{k,d}$ .

Proposition 2 demonstrates that both the LIM estimand  $\beta^{\text{LIM}}$  and the LIS estimand  $\beta^{\text{LIS}}$  are weighted sums of the rank-dependent peer effect coefficients  $\{\beta_{k,d}\}_{k,d}$ . While the weights sum to one, which is reasonable, there is no guarantee that the weights have the same sign. As a result, it is possible for all  $\beta_{k,d}$  to be positive, while the LIM estimand  $\beta^{\text{LIM}}$  is negative, and vice versa. The expressions for the weights are provided in the Appendix.

## 4 Simulation

To investigate the finite sample performance of the TSLS estimator for the rank-dependent peer effects, we simulate 1000 samples from the rank-dependent peer effect model while varying the model specifications. To investigate the small sample properties of the estimator, we consider sample sizes of either  $n = 100$  or  $n = 200$ . The DGP we use is as follows: for  $i = 1, \dots, n$ ,

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

and

$$\begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} \stackrel{\text{iid}}{\sim} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

To simulate the network, we first generate the links based on a logit model, defining

$$A_{i,j} = \mathbf{1}\{-4.5 + R_{i,j} \leq V_{i,j}\}$$

with  $V_{i,j} \stackrel{\text{iid}}{\sim} \text{logit}$  and  $R_{i,j} \sim N(0, 1)$ . We then drop links from individuals with more links than  $\bar{d}$  until all individuals have at most  $\bar{d}$  links. This procedure generates a network with a roughly uniform degree distribution, with very few individuals having zero links. For the results in this section we will fix the network across simulations. Results when we do not fix the network across simulations can be seen in Appendix C. Finally, to avoid unnecessary repetition in our tables, we will only show results for a subset of  $\beta$ .

The instruments  $\{Z_i = (Z_{i,1}, \dots, Z_{i,\bar{d}})^\top\}_{i=1}^n$  used in the TSLS estimation are constructed by taking ordered statistics of peers' covariate  $\{X_j\}_{j:A_{i,j}=1}$  for each student  $i$ .  $Z_{i,d}$  is a  $d$ -dimensional vector that orders  $\{X_j\}_{j:A_{i,j}=1}$  when  $d_i = d$  and contains zeros when  $d_i \neq d$ . To avoid the simulations potentially suffering from the contamination bias discussed in Goldsmith-Pinkham, Hull and Kolesár (2024+), we stratify our estimator. We do this by splitting the dataset into  $\bar{d} + 1$  strata of  $\{i : d_i = 0\}, \dots, \{i : d_i = \bar{d}\}$  and the sets of the coefficients  $(\gamma_0, \gamma_1), \beta_{1,1}, \dots, (\beta_{1,\bar{d}}, \dots, \beta_{\bar{d},\bar{d}})$  are separately estimated from each strata. A pooled version of our estimator generally performs similarly, but can sometimes be less robust. See Appendix C for a discussion.

First, we set  $\bar{d} = 2$ . The bias and MSE of the TSLS estimator are compared to those of the OLS estimator in Table 1. In this table, two key model parameters,  $\gamma_1$  and  $n$ , vary across the columns. As expected, the OLS estimator shows a larger bias compared to the TSLS estimator. However, the TSLS estimator has a higher variance than the OLS estimator. Notably, as  $|\gamma_1|$  increases, the variance of the TSLS estimator decreases. This indicates the finite sample performance of the estimator depends on the existence of strong covariates.

Additionally, as  $|\gamma_1|$  increases, the performance of the OLS estimator also improves in terms of both bias and MSE. A larger  $|\gamma_1|$  means that the ordering of  $\{Y_i\}_{i=1}^n$  relies less on the variation in  $\{\varepsilon_i\}_{i=1}^n$  and more on the variation in  $\{\gamma_1 x_i\}_{i=1}^n$ . If the endogeneity problem is exacerbated by the randomness in the mapping  $\{Y_j\}_{j=1}^n \mapsto \tilde{Y}_{i,k}$ , then a larger  $|\gamma_1|$  may help reduce bias by decreasing this randomness.

Next, we increase the parameter space by setting  $\bar{d} = 5$ . Since the dimension of  $\beta$  is proportional to  $\bar{d}^2$ , this introduces 15 different coefficients for the rank-dependent peer effect

Table 1: Comparison between the OLS estimator and the TSLS estimator

	Bias				MSE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\gamma}_1^{OLS}$	-0.019	-0.023	-0.033	-0.031	0.012	0.012	0.007	0.007
$\hat{\beta}_{1,2}^{OLS}$	0.120	0.054	0.112	0.045	0.035	0.011	0.017	0.004
$\hat{\gamma}_1^{TSLS}$	-0.003	-0.009	-0.016	-0.002	0.043	0.042	0.262	0.256
$\hat{\beta}_{1,2}^{TSLS}$	-0.007	0.002	-0.003	-0.002	0.107	0.016	0.077	0.015
$\gamma_1$	1	2	1	2	1	2	1	2
$n$	100	100	200	200	100	100	200	200

*Notes:* The values of the coefficients used in DGP are as follows:  $\gamma_0 = 1$  and  $\beta = (0.3, 0.15, 0.4)$ . The instruments used in TSLS estimation are ordered peers' covariates. TSLS estimation is stratified in the sense that the dataset is separated into three subdata— $\{i : d_i = 0\}$ ,  $\{i : d_i = 1\}$  and  $\{i : d_i = 2\}$ —and the sets of coefficients— $(\gamma_0, \gamma_1)$ ,  $\beta_{1,1}$  and  $(\beta_{1,2}, \beta_{2,2})$ —are estimated from each subdata.

in our model. Panel A of Table 2 presents the bias and MSE of the TSLS estimator under this setting. We observe that the TSLS estimator suffers from significant bias and highly volatile variance when  $\gamma_1 = 1$  (columns 2 and 7). However, as  $|\gamma_1|$  increases, the variance of the TSLS estimator decreases, as shown in columns 8 through 10.

Interestingly, the MSE is larger for the coefficients  $\beta_{k,5}$  where  $k = 2, 3, 4$ , compared to  $k = 1, 5$ . This is related to our earlier observation that larger  $|\gamma_1|$  improves the performance of the OLS estimator. The identity of the  $k$ -th lowest-performing peer tends to fluctuate more frequently for  $k = 2, 3, 4$ , leading to a more severe endogeneity problem from the randomness in the mapping  $\{Y_j\}_{j=1}^n \mapsto \tilde{Y}_{i,k}$ . In contrast, this issue is less pronounced for  $k = 1, 5$ . Therefore, it may be reasonable to impose smoothness restrictions on  $\beta_{k,d}$  for values of  $k$  that are not close to 1 or  $d$ .

To discuss the empirical advantage of a smoothness restriction, we additionally considered a restricted peer effect model with correct smoothness restrictions: for  $d = 1, \dots, 5$ ,

$$\beta_{k,d} = \frac{\beta_{-\max}}{d-1} \quad \forall k < d \quad \text{and} \quad \beta_{d,d} = \beta_{\max}.$$

With the restricted model, we estimated only two peer effect parameters:  $\beta_{-\max}, \beta_{\max}$ . Panel B of Table 2 contains the estimation results. Both in terms of bias and MSE, correct smoothness restriction greatly improves the performance of the TSLS estimator when  $|\gamma_1|$  is small.

Finally, Table 3 reports the average first-stage  $R^2$  and the probability of the  $R^2$  exceeding 0.7 for each of the endogenous variables  $\tilde{Y}_{1,5}, \dots, \tilde{Y}_{5,5}$ . The results show that a weak first-stage regression exacerbates the performance of the TSLS estimator. This suggests that the

Table 2: The TSLS estimation when the network is more connected

	Bias					MSE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<b>Panel A: Saturated model</b>										
$\beta_{1,2}$	-0.005	-0.000	-0.004	-0.001	0.001	0.049	0.260	0.023	0.010	0.003
$\beta_{2,2}$	0.002	-0.000	0.000	-0.001	-0.000	0.025	0.109	0.012	0.006	0.002
$\beta_{1,5}$		0.058	-0.008	-0.002	-0.002		1.872	0.305	0.082	0.017
$\beta_{2,5}$		0.044	0.021	-0.009	0.007		4.371	1.500	0.523	0.094
$\beta_{3,5}$		-0.022	-0.019	0.031	0.000		5.635	2.233	0.788	0.143
$\beta_{4,5}$		-0.001	0.022	-0.045	-0.006		4.531	1.339	0.428	0.086
$\beta_{5,5}$		-0.029	-0.016	0.018	0.004		1.300	0.240	0.072	0.013
<b>Panel B: Restricted model</b>										
$\beta_{-\max}$		-0.005	-0.002	0.000	0.001		0.104	0.011	0.004	0.001
$\beta_{\max}$		-0.010	-0.000	-0.001	-0.001		0.082	0.008	0.003	0.001
$\gamma_1$	1	1	2	3	5	1	1	2	3	5
$\bar{d}$	2	5	5	5	5	2	5	5	5	5

*Notes:* The values of the coefficients used in DGP are as follows:  $n = 200$ ,  $\gamma_0 = 1$ ,  $\beta_{1,1} = 0.3$ ,  $\beta_{k,d} = \frac{\beta_{-\max}}{d} = \frac{0.4}{d}$  for  $k < d$  and  $\beta_{d,d} = \beta_{\max} = 0.3$  for  $d = 2, \dots, 5$ . The instruments used in TSLS estimation are ordered peers' covariates. The TSLS estimation is stratified in the sense that the dataset is separated into three subdata— $\{i : d_i = 0\}, \dots, \{i : d_i = 5\}$ —and the sets of coefficients— $(\gamma_0, \gamma_1), \dots, (\beta_{1,5}, \dots, \beta_{5,5})$ —are estimated from each subdata. Panel A estimates the fully saturates model with fifteen  $\beta$ s and Panel B focuses on individuals with at least two peers and estimates the correctly specific model with two  $\beta$ s:  $\beta_{-\max} = 0.4$  and  $\beta_{\max} = 0.3$ .

potential costs of allowing for a more complex parameter space in terms of peer effect coefficients can be mitigated by carefully selecting the instruments  $\{Z_i\}_{i=1}^n$  to ensure a sufficiently strong first-stage regression.

## 5 Empirical Illustration

To see how our model behaves with real data, we apply it to a dataset from Alne, Herstad and Myhre (2025), which is collected in two Norwegian middle schools and contains school grades, test scores for a nationwide test (“National Test”) taken prior to entering middle

Table 3:  $R^2$  in the first stage regression

LHS	Average first stage $R^2$				$\Pr\{R^2 \geq 0.7\}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\tilde{Y}_{1,5}$	0.389	0.611	0.704	0.771	0.002	0.201	0.580	0.856
$\tilde{Y}_{2,5}$	0.412	0.616	0.701	0.765	0.004	0.207	0.562	0.828
$\tilde{Y}_{3,5}$	0.421	0.619	0.703	0.763	0.004	0.230	0.552	0.835
$\tilde{Y}_{4,5}$	0.409	0.609	0.694	0.749	0.004	0.189	0.510	0.770
$\tilde{Y}_{5,5}$	0.381	0.587	0.678	0.745	0.004	0.140	0.423	0.772
$\gamma_1$	1	2	3	5	1	2	3	5
$\bar{d}$	5	5	5	5	5	5	5	5

*Notes:* The values of the coefficients used in DGP are as follows:  $n = 200$ ,  $\gamma_0 = 1$ ,  $\beta_{1,1} = 0.3$ ,  $\beta_{k,d} = \frac{0.4}{d}$  for  $k < d$  and  $\beta_{d,d} = 0.3$  for  $d = 2, \dots, 5$ . The instruments used in TSLS estimation are ordered peers' covariates. The TSLS estimation is stratified in the sense that the dataset is separated into three subdata— $\{i : d_i = 0\}, \dots, \{i : d_i = 5\}$ —and the sets of coefficients— $(\gamma_0, \gamma_1), \dots, (\beta_{1,5}, \dots, \beta_{5,5})$ —are estimated from each subdata.

school, parental background variables, and friendship network of the students. We follow the regression specification of Alne, Herstad and Myhre (2025); the GPA is the outcome variable with possible spillovers and the National Test score and other socioeconomic and demographic variables are the control covariates. To simplify the presentation, we focus on a subset of the control covariates used in Alne, Herstad and Myhre (2025). Similarly, we run the estimation only for individuals with more than two friends to avoid any potential contamination bias. The instruments are constructed with the National test score from the test the students took before they started in middle school, as it is arguably exogenous to shocks to GPA in middle school.

In this paper, we expand the LIM model to allow for ordered peer outcomes to affect a student differently, as in (3). Specifically, we focus on the highest-performing peer and the lowest-performing peer. Furthermore, whenever well-defined, we let

$$\begin{aligned}\bar{Y}_{i,-1-d_i} &= \frac{1}{d_i - 2} \sum_{k=2}^{d_i-1} \tilde{Y}_{i,k} \\ \bar{Y}_{i,-1} &= \frac{1}{d_i - 1} \sum_{k=2}^{d_i} \tilde{Y}_{i,k} \\ \bar{Y}_{i,-d_i} &= \frac{1}{d_i - 1} \sum_{k=1}^{d_i-1} \tilde{Y}_{i,k}.\end{aligned}$$

Table 4: Min/max models for educational peer effects

	<i>Dependent variable: GPA</i>				
	(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 (0.313)	0.303 (0.391)	0.407*** (0.151)
$\tilde{Y}_1$		-0.067 (0.221)		0.138 (0.202)	0.18* (0.106)
Parent Income	0.146** (0.066)	0.15** (0.068)	0.148** (0.069)	0.135** (0.065)	0.136** (0.066)
Born in Norway	-0.277** (0.135)	-0.26* (0.136)	-0.363** (0.148)	-0.298** (0.149)	-0.321** (0.139)
National Test	0.388*** (0.053)	0.383*** (0.053)	0.401*** (0.052)	0.381*** (0.053)	0.381*** (0.053)
Parental Education	0.053*** (0.016)	0.051*** (0.016)	0.049*** (0.017)	0.05*** (0.016)	0.05*** (0.016)
Distance to School	-0.011 (0.048)	-0.015 (0.051)	0 (0.052)	-0.022 (0.052)	-0.024 (0.05)
Female	0.468*** (0.101)	0.4*** (0.093)	0.586*** (0.103)	0.442*** (0.093)	0.466*** (0.104)
Constant	-1.377*** (0.345)	-1.565*** (0.454)	-1.83*** (0.477)	-1.434*** (0.382)	-1.488*** (0.387)
Observations	529	529	529	529	529
Adjusted R <sup>2</sup>	0.476	0.474	0.450	0.480	0.476

Notes: \* $p < 0.01$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ . The standard errors are computed from the asymptotic variance formula discussed in Section Appendix A of the appendix.

As  $\tilde{Y}_{i,k}$  is the  $k$ -th lowest GPA among student  $i$ 's peers,  $\bar{Y}_{i,-1-d_i}$  is the average GPA of student  $i$ 's peers minus the one with the lowest GPA and the one with the highest GPA, and so on. For the instruments, we construct  $z_{i,1}$ ,  $z_{i,d_i}$ ,  $\bar{z}_{i,-1-d_i}$ ,  $\bar{z}_{i,-1}$  and  $\bar{z}_{i,-d_i}$  accordingly, using the National Test score. For each specification, we use the corresponding instruments to the  $\tilde{Y}_i$  variables contained in the specification.

As controls for student background, we use the log of household income, whether the student was born in Norway, the average number of years the parents have been educated,

the distance the students live from school as well as the student’s gender.

Table 4 shows the estimation results. There are five specifications we consider. The first is the classic LIM model. The second and third specifications separate out the lowest and highest friends GPA respectively. In the fourth we separate out both the highest and lowest GPA, while in the final specification we only include the highest and lowest GPA friends, not using the average GPA of the peers inbetween.

While the estimates are noisy, there are indications that the majority of estimated peer effect in the LIM model is driven by  $\tilde{Y}_{d_i}$ , the friend with the highest grade. We see that the coefficient on  $\tilde{Y}_1$  is consistently smaller than the coefficient on  $\tilde{Y}_{d_i}$ . In our specification where we only separate out  $\tilde{Y}_{d_i}$ , the coefficient on the remaining LIM component is almost zero. However, the estimate of the coefficient on  $\tilde{Y}_{d_i}$  is not significant in this specification. In total, this exercise shows that there may be large heterogeneity in peer spillovers in education masked by the estimated effect of the LIM model, though we do not have the power in this sample to conclude anything concrete.

## 6 Conclusion

This paper introduced the rank-dependent peer effect model. In this model, how you affect your peer depends not only on your own outcome level but also on the outcomes of the other peers that they have. The construction of the peer effect in the model is flexible enough to allow us to discuss many interesting questions regarding how the composition of a peer group affects the peer effect: complementarity v. substitutability, returns to centrality, etc. Traditional peer effect models, such as the linear-in-means (LIM) model, tend to oversimplify the dynamics of peer influence by focusing on a single scalar parameter and assuming perfect substitutability among peers. This oversimplification neglects the heterogeneity within peer effects, where individuals may experience varying degrees of influence depending on the structure and nature of their interactions, across different empirical contexts. By developing the rank-dependent peer effect model, this paper allows researchers to investigate richer patterns of peer interactions, with specific peers—such as the highest or lowest performers—having disproportionately larger effects on outcomes in certain contexts.

While we have focused on identification and estimation through TSLS, there are multiple avenues to establish estimators with better finite sample properties. For example, the model can benefit from having an estimator that has data-driven smoothness or sparsity property.

Finally, this paper has only considered settings with exogenous networks. As work on incorporating endogenous networks into peer effect models continue to develop, it would similarly be interesting to see if similar approaches can be applied to the rank-dependent peer effect model.

# APPENDIX

## Appendix A Asymptotic theory for the TSLS estimator

Assumption 6 gives a set of sufficient conditions for the asymptotic normality of the TSLS estimator. Recall the peer effect equation (10).

**Assumption 6.**

*a. The two matrices*

$$\frac{1}{n} \sum_{i=1}^n z_i W_i^\top \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n z_i z_i^\top$$

*converge to full rank matrices (in probability) as  $n \rightarrow \infty$ .*

*b.  $\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \varepsilon_i \xrightarrow{d} \mathcal{N}(0, \Sigma_W)$  with some  $\Sigma_W > 0$  as  $n \rightarrow \infty$ .*

Part **a.** of Assumption 6 is the sample counterpart of the relevance conditions discussed in Section 3. Part **b.** assumes that  $\{\varepsilon_i\}_{i=1}^n$  satisfies some weak dependence condition and  $\{z_i\}_{i=1}^n$  does not explode so that we have asymptotic normality on  $\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \varepsilon_i$ .

**Theorem 2.** *Assumptions 2-4 and 6 hold. Then,*

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$$

*as  $n \rightarrow \infty$ , with some consistently estimable positive definite matrix  $\Sigma$ .*

*Proof.* The TSLS estimator is

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \left( \frac{1}{n} \sum_{i=1}^n W_i z_i^\top \left( \frac{1}{n} \sum_{i=1}^n z_i z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_i W_i^\top \right)^{-1} \cdot \frac{1}{n} \sum_{i=1}^n W_i z_i^\top \left( \frac{1}{n} \sum_{i=1}^n z_i z_i^\top \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i Y_i.$$

Then,

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} &= \left( \frac{1}{n} \sum_{i=1}^n W_i z_i^\top \left( \frac{1}{n} \sum_{i=1}^n z_i z_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_i W_i^\top \right)^{-1} \\ &\quad \cdot \frac{1}{n} \sum_{i=1}^n W_i z_i^\top \left( \frac{1}{n} \sum_{i=1}^n z_i z_i^\top \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \varepsilon_i. \end{aligned}$$

By Assumption 6  $\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \varepsilon_i \xrightarrow{d} \mathcal{N}(0, \Sigma_W)$ , and the matrices

$$C = \text{plim} \frac{1}{n} \sum_{i=1}^n z_i W_i^\top \quad \text{and} \quad B = \text{plim} \frac{1}{n} \sum_{i=1}^n z_i z_i^\top$$

exist and are full rank. Therefore by the continuous mapping theorem we get

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

With  $\Sigma = (CB^{-1}C)^{-1} CB^{-1} \Sigma_W \left( (CB^{-1}C)^{-1} CB^{-1} \right)^\top$

□

## Appendix B Proofs

### B.1 Proof for Proposition 1

We want to show that for any  $\{x_i, \epsilon_i\}_{i=1}^n$ , the peer effect model (5) admits a unique equilibrium  $\{y_i\}_{i=1}^n$  such that for  $i = 1, \dots, n$ ,

$$y_i = \sum_{k=1}^{d_i} \tilde{y}_{i,k} \beta_{kd_i} + x_i^\top \gamma + \epsilon_i.$$

$\tilde{y}_{i,k}$  is constructed in the same way as  $\tilde{Y}_{i,k}$ .

Consider a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that

$$g(y) = \begin{pmatrix} g_1(y) \\ \vdots \\ g_n(y) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{d_1} \tilde{y}_{1,k} \beta_{kd_1} + x_1^\top \gamma + \epsilon_1 \\ \vdots \\ \sum_{k=1}^{d_n} \tilde{y}_{n,k} \beta_{kd_n} + x_n^\top \gamma + \epsilon_n \end{pmatrix}$$

when  $y = (y_1, \dots, y_n)^\top$ . Note that the function  $g$  depends on  $\{x_i, A_{ij}, \epsilon_i\}_{1 \leq i, j \leq n}$  in addition to  $\{\beta_{k,d}\}_{k,d}$  and  $\gamma$ . We show that the function  $g$  is a contraction mapping with the supremum norm: for any  $y, y' \in \mathbb{R}^n$ ,

$$\|g(y) - g(y')\|_\infty \leq \bar{\beta} \|y - y'\|_\infty.$$

Firstly, suppose that  $y$  and  $y'$  have the same order. The inequality holds trivially: for each  $i$ ,

$$\begin{aligned} |g_i(y) - g_i(y')| &= \left| \sum_{k=1}^{d_i} (\tilde{y}_{i,k} - \tilde{y}'_{i,k}) \beta_{k,d_i} \right| \\ &\leq \sum_{k=1}^{d_i} |\beta_{k,d_i}| \cdot \|y - y'\|_\infty \leq \bar{\beta} \|y - y'\|_\infty \end{aligned}$$

with some  $\bar{\beta} = \max_d \sum_{k=1}^d |\beta_{k,d}| < 1$ . The first inequality holds since  $y$  and  $y'$  have the same order and therefore  $\tilde{y}_{i,k}$  and  $\tilde{y}'_{i,k}$  are the outcome of the same individual:  $|\tilde{y}_{i,k} - \tilde{y}'_{i,k}|$  is bounded by  $\|y - y'\|_\infty$ . The second equality is from Assumption 2.

Secondly, suppose that  $y$  and  $y'$  do not have the same order.<sup>5</sup> Fix  $i$  and  $k$  and consider

$$|(\tilde{y}_{i,k} - \tilde{y}'_{i,k}) \beta_{k,d_i}|.$$

When individual  $i$ 's  $k$ -th lowest performing peer under the outcome vector  $y$  is the same individual with that under the outcome vector  $y'$ , the quantity above is bounded by  $|\beta_{k,d_i}| \cdot \|y - y'\|_\infty$ . Suppose that the order change affects individual  $i$ 's  $k$ -th lowest performing peer

---

<sup>5</sup>It is assumed throughout the proof that some tie-breaking rule is applied here so that there is one-to-one relationship between an individual's peers and the ordered outcomes  $y_{i1}^p, \dots, y_{id_i}^p$ .

and WLOG let  $\tilde{y}_{i,k} = y_1$  and  $\tilde{y}'_{i,k} = y_2'$ ; individual  $i$ 's  $k$ -th lowest performing peer is individual 1 under the outcome vector  $y$  and is individual 2 under the outcome vector  $y'$ . Then,

$$|(\tilde{y}_{i,k} - \tilde{y}'_{i,k}) \beta_{k,d_i}| \leq |\beta_{k,d_i}| \cdot |y_1 - y_2'|$$

and individual 1's rank among individual  $i$ 's peers under the outcome vector  $y'$  is either higher than  $k$  or lower than  $k$ .

**Case 1.** Individual 1's rank among individual  $i$ 's peers under the outcome vector  $y'$  is higher than  $k$ : there exists some  $k' > k$  such that  $y_1' = y_{i,k'}^p$ . Then, we have that

$$y_1' = y_{i,k'}^p \geq \tilde{y}'_{i,k} = y_2'$$

and therefore

$$y_1 - y_2' \geq y_1' - \|y - y'\|_\infty - y_2' \geq -\|y - y'\|_\infty.$$

Since individual 1 is ranked higher under the outcome vector  $y'$  compared to the outcome vector  $y$  where his rank was  $k$ , there must be another individual  $j \neq 1$  whose rank is lower than or equal to  $k$  under the outcome vector  $y'$  and is higher than  $k$  under the outcome vector  $y$ :

$$y_j' \leq \tilde{y}'_{i,k} = y_2' \quad \text{and} \quad y_j \geq \tilde{y}_{i,k} = y_1$$

and therefore

$$y_1 - y_2' \leq y_j - y_2' \leq y_j' + \|y - y'\|_\infty - y_2' \leq \|y - y'\|_\infty.$$

By combining the two inequalities, we get  $|y_1 - y_2'| \leq \|y - y'\|_\infty$ .

**Case 2.** Individual 1's rank among individual  $i$ 's peers under the outcome vector  $y'$  is lower than  $k$ : there exists some  $k' < k$  such that  $y_1' = y_{i,k'}^p$ . Then, we have that

$$y_1' = y_{i,k'}^p \leq \tilde{y}'_{i,k} = y_2'$$

and therefore

$$y_1 - y_2' \leq y_1' + \|y - y'\|_\infty - y_2' \leq \|y - y'\|_\infty.$$

Since individual 1 is ranked lower under the outcome vector  $y'$  compared to the outcome vector  $y$  where his rank was  $k$ , there must be another individual  $j \neq 1$  whose rank is higher than or equal to  $k$  under the outcome vector  $y'$  and is lower than  $k$  under the outcome vector  $y$ :

$$y_j' \geq \tilde{y}'_{i,k} = y_2' \quad \text{and} \quad y_j \leq \tilde{y}_{i,k} = y_1$$

and therefore

$$y_1 - y_2' \geq y_j - y_2' \geq y_j' - \|y - y'\|_\infty - y_2' \geq -\|y - y'\|_\infty.$$

By combining the two inequalities, we get  $|y_1 - y_2'| \leq \|y - y'\|_\infty$ .

By aggregating across  $k$  for the fixed  $i$  and then aggregating across  $i$ , we get

$$\|g(y) - g(y')\|_\infty \leq \bar{\beta} \|y - y'\|_\infty.$$

Assume to the contrary that there exist  $y, y' \in \mathbb{R}^n$  such that  $y \neq y'$ ,  $y = g(y)$  and  $y' = g(y')$ . Let  $c = \|y - y'\|_\infty > 0$ . Then,

$$c = \|y - y'\|_\infty = \|g(y) - g(y')\|_\infty \leq \bar{\beta} \|y - y'\|_\infty \leq \bar{\beta} c < c$$

from Assumption 2, leading to a contradiction. □

## B.2 Proof for Corollary 1

Let  $\pi$  denote a ordering on  $\{Y_i, \dots, Y_n\}$  with some tiebreaking rule.

$$\pi : \{1, \dots, n\} \mapsto \{1, \dots, n\}$$

and satisfies that  $Y_{\pi(1)} \leq \dots \leq Y_{\pi(n)}$ . Note that  $\pi$  is a function of  $\{\epsilon_i\}_{i=1}^n$ ; from Proposition 1, we have shown that there exists a unique equilibrium and thus a unique ordering for each realization  $\{\epsilon_i\}_{i=1}^n$ . Let  $\Pi$  denote the set of all possible ordering on  $\{Y_1, \dots, Y_n\}$ .  $\Pi$  is a function of  $\{x_i, A_{ij}\}_{i,j}$  and the distribution of  $\{\epsilon_i\}_i$ .

For any realization of  $\{\epsilon_i\}_{i=1}^n$  and the corresponding ordering  $\pi$ , there is a reduced-form linear relationship between  $\{y_i\}_{i=1}^n$  and  $\{x_i\}_{i=1}^n$ : construct a  $n \times n$  matrix  $\mathbb{B}(\pi)$  such that

$$\mathbb{B}(\pi) = \left( \beta_{ij}(\pi) \right)_{i,j}$$

$$\beta_{ij}(\pi) = \begin{cases} 0 & \text{if } A_{ij} = 0 \\ \beta_{kd_i} \text{ for some } k \text{ s.t. } \sum_{j'=1}^{\pi^{-1}(j)} A_{i\pi(j')} = k & \text{if } A_{ij} = 1 \end{cases}$$

$\mathbb{B}(\pi)$  takes the ordering  $\pi$  as fixed and finds the corresponding rank-dependent coefficient  $\beta_{k,d}$  for each of the peer outcomes. Find that  $(I_n - \mathbb{B}(\pi))$  is invertible for any  $\pi$ : suppose that there is some  $x \neq \mathbf{0}$  such that  $x = \mathbb{B}(\pi)x$ ; for some  $i$  such that  $|x_i| = \|x\|_\infty$ ,

$$|x_i| = \left| \sum_{j=1}^n \beta_{ij}(\pi) x_j \right| \leq \sum_{j=1}^j |\beta_{ij}(\pi)| \cdot |x_j| \leq \bar{\beta} |x_i| < |x_i|,$$

leading to a contradiction;  $I - \mathbb{B}(\pi)$  is full rank. Then,

$$\mathbb{Y} = \mathbb{B}(\pi)\mathbb{Y} + \mathbb{X}\gamma + \mathbb{E}$$

$$\mathbb{Y} = (I - \mathbb{B}(\pi))^{-1} \mathbb{X}\gamma + (I - \mathbb{B}(\pi))^{-1} \mathbb{E}.$$

We have a reduced-form linear relationship between  $\{y_i\}_{i=1}^n$  and  $\{x_i\}_{i=1}^n$ .

Since the reduced-form relationship holds for every realization of  $\{\epsilon_i\}_{i=1}^n$  such that the ordering  $\pi$  stays the same, we can consider a conditional expectation of the linear relationship

given the event that  $\{\varepsilon_i\}_{i=1}^n$  induces the ordering  $\pi$ . Let  $\Pi$  denote the set of all possible orderings  $\pi$ , then

$$\begin{aligned}\mathbf{E}_n[Y_i] &= \sum_{\pi \in \Pi} \sum_{j=1}^n \theta_{ij}(\pi) (x_j^\top \gamma + \mathbf{E}_n[\varepsilon_j | \pi]) \cdot \Pr_n\{\pi\} \\ &= \sum_{j=1}^n \theta_{ij} x_j^\top \gamma + \eta_i\end{aligned}$$

where  $(\theta_{i1}(\pi), \dots, \theta_{in}(\pi))$  is the  $i$ -th row of the  $n \times n$  matrix  $(I - \mathbb{B}(\pi))^{-1}$  and

$$\begin{aligned}\theta_{ij} &= \sum_{\pi \in \Pi} \theta_{ij}(\pi) \Pr_n\{\pi\} \\ \eta_i &= \sum_{\pi \in \Pi} \sum_{j=1}^n \theta_{ij}(\pi) \mathbf{E}_n[\varepsilon_j | \pi] \cdot \Pr_n\{\pi\}.\end{aligned}$$

Likewise, by taking a different student while summing over the ordering  $\pi$ , we get

$$\begin{aligned}\mathbf{E}_n[\tilde{Y}_{i,k}] &= \sum_{\pi \in \Pi} \sum_{j=1}^n \tilde{\theta}_{i,k,j}(\pi) (x_j^\top \gamma + \mathbf{E}_n[\varepsilon_j | \pi]) \cdot \Pr_n\{\pi\} \\ &= \sum_{j=1}^n \tilde{\theta}_{i,k,j} x_j^\top \gamma + \tilde{\eta}_{i,k}\end{aligned}$$

where  $(\tilde{\theta}_{i,k,1}(\pi), \dots, \tilde{\theta}_{i,k,n}(\pi))$  is one row of the  $n \times n$  matrix  $(I_n - \mathbb{B}(\pi))^{-1}$ , which corresponds to the  $k$ -th lowest performing peer of student  $i$ , and

$$\begin{aligned}\tilde{\theta}_{i,k,j} &= \sum_{\pi \in \Pi} \tilde{\theta}_{i,k,j}(\pi) \Pr_n\{\pi\} \\ \tilde{\eta}_{i,k} &= \sum_{\pi \in \Pi} \sum_{j=1}^n \tilde{\theta}_{i,k,j}(\pi) \mathbf{E}_n[\varepsilon_j | \pi] \cdot \Pr_n\{\pi\}.\end{aligned}$$

□

### B.3 Proof for Theorem 1

Construct a  $\frac{\bar{d}(\bar{d}+1)}{2} \times 1$  vector  $\tilde{Y}_i$  such that

$$\tilde{Y}_i = (0 \quad \dots \quad Y_{i1}^p \quad \dots \quad Y_{id_i}^p \quad \dots \quad \mathbf{0}_{\bar{d}}^\top).$$

$\tilde{Y}_i$  stacks up (hypothetical) vectors of peer outcomes with varying peer numbers  $d = 1, \dots, \bar{d}$ ; the peer outcome vector of peer number  $d$  is nonzero only when  $d = d_i$ . Let  $W_i = (\tilde{Y}_i^\top, x_i^\top)^\top$ . Then, the rank-dependent peer effect model can be written as follows:

$$Y_i = \tilde{Y}_i^\top \beta + x_i^\top \gamma + \varepsilon_i = W_i^\top \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \varepsilon_i.$$

Likewise, construct a vector of instrument  $z_i$  such that  $z_i$  stacks up vectors of instruments with varying peer numbers  $d = 1, \dots, \bar{d}$ .

$$z_i = (z_{i1}^\top \ \cdots \ z_{i\bar{d}}^\top \ x_i^\top)^\top.$$

$z_{id}$ , the instrument vector of peer number  $d$ , is nonzero only when  $d = d_i$ .

Consider a stacked up matrix of the peer effect model (5):

$$\begin{aligned} \mathbb{Y} &= \mathbb{W} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \mathbb{E} \\ \mathbb{Z}^\top \mathbb{Y} &= \mathbb{Z}^\top \mathbb{W} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \mathbb{Z}^\top \mathbb{E}. \end{aligned}$$

From Assumption 4, by taking expectation conditional upon  $\{x_i, A_{ij}\}_{i,j}$ ,

$$\mathbf{E}_n [\mathbb{Z}^\top \mathbb{Y}] = \mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}] \begin{pmatrix} \beta \\ \gamma \end{pmatrix}.$$

The parameter of interest  $\beta$  and  $\gamma$  are identified from the conditional distribution of  $\{Y_i\}_{i=1}^n$  given  $\{x_i, A_{ij}\}_{1 \leq i, j \leq n}$  when  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}]$  is full rank.

From Corollary 1, we have the tools to write out the matrix  $\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}]$ :

$$\mathbf{E}_n [\mathbb{Z}^\top \mathbb{W}] = \begin{pmatrix} Q_{ZY^p,1} & \mathbf{O}_{l \times 2} & \cdots & \mathbf{O}_{l \times \bar{d}} & Q_{ZX,1} \\ \mathbf{O}_{2l \times 1} & Q_{ZY^p,2} & \cdots & \mathbf{O}_{2l \times \bar{d}} & Q_{ZX,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{O}_{\bar{d}l \times 1} & \mathbf{O}_{\bar{d}l \times 2} & \cdots & Q_{ZY^p,\bar{d}} & Q_{ZX,\bar{d}} \\ Q_{XY^p,1} & Q_{XY^p,2} & \cdots & Q_{XY^p,\bar{d}} & Q_{XX} \end{pmatrix}$$

where

$$\begin{aligned} Q_{Z\tilde{Y},d} &= \sum_{i=1}^n z_{id} \left( \mathbf{E}_n [\tilde{Y}_{i,1}] \ \cdots \ \mathbf{E}_n [\tilde{Y}_{i,d_i}] \right) \mathbf{1}\{d_i = d\}, \\ Q_{ZX,d} &= \sum_{i=1}^n z_{id} x_i^\top \mathbf{1}\{d_i = d\}, \\ Q_{X\tilde{Y},d} &= \sum_{i=1}^n x_i \left( \mathbf{E}_n [\tilde{Y}_{i,1}] \ \cdots \ \mathbf{E}_n [\tilde{Y}_{i,d_i}] \right) \mathbf{1}\{d_i = d\}, \\ Q_{XX} &= \sum_{i=1}^n x_i x_i^\top. \end{aligned}$$

A slight abuse of notation is applied here: the row vectors in the summation of  $Q_{Z\tilde{Y},d}$  and  $Q_{X\tilde{Y},d}$  do not have the same dimension, but the indicator  $\mathbf{1}\{d_i = d\}$  only selects the ones with  $d$  components.

Thanks to the diagonal structure of  $Q_{Z\tilde{Y},1}, \dots, Q_{Z\tilde{Y},\bar{d}}$ , each submatrix

$$Q_d := \begin{pmatrix} Q_{Z\tilde{Y},d} & Q_{ZX,d} \\ Q_{X\tilde{Y},d} & Q_{XX} \end{pmatrix}$$

being full rank for  $d = 1, \dots, \bar{d}$  implies that  $\mathbb{Z}^\top \mathbf{E}_n [\mathbb{W}]$  is full rank. From the reduced-form representation of  $\mathbf{E}_n [\tilde{Y}_{i,k}]$  with  $\theta$  and  $\eta$ , we get

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

From Assumption 4,  $Q_d$  has full rank. □

## B.4 Proof for Proposition 2

The LIM model is:

$$Y_i = \beta^{\text{LIM}} \bar{Y}_i + x_i^\top \gamma + \varepsilon_i,$$

With some  $z_i$  that is a function of  $\{x_i, A_{i,j}\}_{i,j}$  is used as instruments, the moment condition used in the estimation of the LIM estimator is:

$$\begin{aligned} \mathbf{E}_n [\mathbb{Y}] &= \beta^{\text{LIM}} G \mathbf{E}_n [\mathbb{Y}] + \mathbb{X} \gamma \\ \mathbb{M}_X \mathbf{E}_n [\mathbb{Y}] &= \beta^{\text{LIM}} \mathbb{M}_X G \mathbf{E}_n [\mathbb{Y}] + \mathbb{M}_X \mathbb{X} \gamma \\ &= \beta^{\text{LIM}} \mathbb{M}_X G \mathbf{E}_n [\mathbb{Y}] \\ \mathbb{Z}^\top \mathbb{M}_X \mathbf{E}_n [\mathbb{Y}] &= \beta^{\text{LIM}} \mathbb{Z}^\top \mathbb{M}_X G \mathbf{E}_n [\mathbb{Y}] \\ \beta^{\text{LIM}} &= (\mathbb{Z}^\top \mathbb{M}_X G \mathbf{E}_n [\mathbb{Y}])^{-1} \mathbb{Z}^\top \mathbb{M}_X \mathbf{E}_n [\mathbb{Y}] \end{aligned}$$

where  $G$  is the row-normalized network adjacency matrix and

$$\mathbb{M}_X = I_n - \mathbb{X} (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top.$$

We know that

$$\mathbf{E}_n [Y_i | \{x_i, A_{i,j}\}_{i,j}] = \sum_{k=1}^{d_i} \beta_{k,d} \mathbf{E}_n [\tilde{Y}_{i,k}] + x_i^\top \gamma$$

from Assumption 4. Then, the misspecified parameter  $\beta^{\text{LIM}}$  becomes

$$\beta^{\text{LIM}} = (\mathbb{Z}^\top \mathbb{M}_X G \mathbf{E}_n [\mathbb{Y}])^{-1} \mathbb{Z}^\top \mathbb{M}_X \begin{pmatrix} \sum_{k=1}^{d_1} \beta_{k,d} \mathbf{E}_n [\tilde{Y}_{1,k}] \\ \vdots \\ \sum_{k=1}^{d_n} \beta_{k,d} \mathbf{E}_n [\tilde{Y}_{n,k}] \end{pmatrix}$$

By rewriting the matrix multiplications as a summation with

$$\hat{\mathcal{A}} = \sum_{i=1}^n z_i x_i^\top \left( \sum_{i=1}^n x_i x_i^\top \right)^{-1},$$

we get

$$\begin{aligned}\beta^{\text{LIM}} &= \frac{\sum_{i=1}^n (z_i - \hat{\mathcal{A}}x_i) \sum_{k=1}^{d_i} \mathbf{E}_n [\tilde{Y}_{i,k}] \beta_{k,d}}{\sum_{i=1}^n (z_i - \hat{\mathcal{A}}x_i) \frac{1}{d_i} \sum_{k=1}^{d_i} \mathbf{E}_n [\tilde{Y}_{i,k}]} \\ &= \frac{\sum_{d=1}^{\bar{d}} \sum_{k=1}^d \sum_{i=1}^n \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right] \beta_{k,d}}{\sum_{d=1}^{\bar{d}} \sum_{i=1}^n \frac{1}{d} \sum_{k=1}^d \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]}.\end{aligned}$$

By letting

$$w_{k,d}^{\text{LIM}} = \frac{\sum_{i=1}^n \frac{1}{d} \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]}{\sum_{d=1}^{\bar{d}} \sum_{k=1}^d \sum_{i=1}^n \frac{1}{d} \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]},$$

$\beta^{\text{LIM}}$  becomes

$$\beta^{\text{LIM}} = \sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIM}} d \beta_{k,d}.$$

The weights satisfy the sum-to-one constraints:  $\sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIM}} = 1$ .

Likewise, for the LIS peer effect parameter, we get

$$\begin{aligned}\beta^{\text{LIM}} &= \frac{\sum_{d=1}^{\bar{d}} \sum_{k=1}^d \sum_{i=1}^n \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right] \beta_{k,d}}{\sum_{d=1}^{\bar{d}} \sum_{i=1}^n \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \sum_{k=1}^d \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]} \\ &= \sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIS}} \beta_{k,d}.\end{aligned}$$

where

$$w_{k,d}^{\text{LIS}} = \frac{\sum_{i=1}^n \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]}{\sum_{d=1}^{\bar{d}} \sum_{k=1}^d \sum_{i=1}^n \mathbf{E}_n \left[ (z_i - \hat{\mathcal{A}}x_i) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]}.$$

Again, the weights sum to one:  $\sum_{d=1}^{\bar{d}} \sum_{k=1}^d w_{k,d}^{\text{LIS}} = 1$ .

When  $z_i$  is not a scalar, we can modify the argument above and get similar expressions, by replacing  $z_i - \hat{\mathcal{A}}x_i$  with  $c^\top (z_i - \hat{\mathcal{A}}x_i)$  with some weighting  $c$ . In the canonical TSLS case, the weighting vector is

$$c^\top = \mathbf{E}_n [\mathbf{Y}^\top G^\top \mathbf{M}_X \mathbf{Z} (\mathbf{Z}^\top \mathbf{Z})^{-1}];$$

misspecification weights depend on the weighting matrix when the LIM or LIS is overidentified.

□

## Appendix C Supplementary Simulation Results

This appendix presents additional simulation results that are complementary to those in the main text. We first show a simulation set-up similar to our empirical design, and then give versions of Tables 1,2,3 when we re-draw the network in each iteration.

### C.1 Simulation of empirical set-up

To make these simulations closer to the empirical application, we no longer drop links to make the maximum number of links equal  $\bar{d}$ . This results in a network where the average number of friends is around 5, and the maximum number around 12.

We will investigate the performance of the estimator for different levels of instrument strength, sample size and for two different estimators. Our sample size will be either one or two networks with 300 individuals. When the second network is included, it has an identical network structure to the first network. The first of our two estimators is a pooled estimator that estimates all our parameters simultaneously. The second stratifies the sample by excluding individuals with less than three links.

As we can see from Table 5, the pooled estimator struggles to estimate the parameter for  $\tilde{Y}_{1,1}$  in a smaller sample with a weaker instrument. This is not too surprising, as there are only six individuals with only one link in the network we simulate. Interestingly, the poor estimation of this parameter also spreads to the rest of the variables in the model. However, this contamination disappears as we get either a larger sample or a stronger instruments. This indicates the pooling estimator is most vulnerable when either there are very few individuals with a given number of friends, or if the instruments are weak. The stratified estimator is also robust to the badly estimated variable, reinforcing how stratification makes our estimators more robust.

Table 5: Simulating our empirical specification

	Bias				MSE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A: Pooled Estimator</b>								
$\beta_{1,1}$	-0.180	0.049	-0.002	-0.062	46.034	34.758	3.291	3.057
$\beta_{\min}$	-0.002	-0.009	-0.006	0.004	1.460	0.096	0.048	0.013
$\beta_{\max}$	-0.065	0.012	-0.004	-0.003	10.088	0.111	0.076	0.010
$\beta_{\text{mid}}$	0.053	-0.001	0.007	-0.001	15.705	0.012	0.062	0.006
<b>Panel B: Stratified Estimator</b>								
$\beta_{1,1}$								
$\beta_{\min}$	-0.014	-0.001	-0.005	0.001	0.131	0.005	0.026	0.002
$\beta_{\max}$	-0.012	0.003	-0.004	-0.001	0.150	0.004	0.026	0.002
$\beta_{\text{mid}}$	0.023	-0.002	0.006	-0.001	0.569	0.014	0.094	0.007
$\gamma_1$	1	2	1	2	1	2	1	2
$n_{\text{network}}$	1	1	2	2	1	1	2	2

*Notes:* The values of the coefficients used in DGP are as follows:  $n = 200$ ,  $\gamma_0 = 1$ ,  $\beta_{1,1} = 0.3$ ,  $\beta_{k,d} = \frac{\beta_{\text{mid}}}{d} = \frac{0.15}{d}$  for  $1 < k < d$  for all  $d > 1$ ,  $\beta_{1,d} = \beta_{\min} = 0.15$  and  $\beta_{d,d} = \beta_{\max} = 0.3$  for all  $d > 1$ . The instruments used in TSLS estimation are ordered peers' covariates.

## C.2 Simulation results with networks re-drawn

Table 6: Comparison between the OLS estimator and the TSLS estimator

	Bias				MSE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\gamma}_1^{OLS}$	-0.028	-0.020	-0.030	-0.030	0.013	0.011	0.006	0.007
$\hat{\beta}_{1,2}^{OLS}$	0.122	0.052	0.126	0.049	0.037	0.011	0.022	0.005
$\hat{\gamma}_1^{TSLS}$	0.001	0.009	-0.002	-0.006	0.051	0.047	0.046	0.049
$\hat{\beta}_{1,2}^{TSLS}$	-0.016	-0.001	-0.001	0.004	0.113	0.016	0.039	0.007
$\gamma_1$	1	2	1	2	1	2	1	2
$n$	100	100	200	200	100	100	200	200

*Notes:* The values of the coefficients used in DGP are as follows:  $\gamma_0 = 1$  and  $\beta = (0.3, 0.15, 0.4)$ . The instruments used in TSLS estimation are ordered peers' covariates. TSLS estimation is stratified in the sense that the dataset is separated into three subdata— $\{i : d_i = 0\}$ ,  $\{i : d_i = 1\}$  and  $\{i : d_i = 2\}$ —and the sets of coefficients— $(\gamma_0, \gamma_1)$ ,  $\beta_{1,1}$  and  $(\beta_{1,2}, \beta_{2,2})$ —are estimated from each subdata.

Table 7: The TSLS estimation when the network is more connected

	Bias					MSE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<b>Panel A: Saturated model</b>										
$\beta_{1,2}$	0.002	0.009	-0.001	-0.002	-0.001	0.039	0.339	0.030	0.013	0.004
$\beta_{2,2}$	-0.002	-0.004	-0.005	0.003	0.000	0.018	0.142	0.017	0.008	0.003
$\beta_{1,5}$		0.018	0.030	0.018	0.004		1.855	0.352	0.129	0.030
$\beta_{2,5}$		0.168	-0.069	0.007	0.007		6.665	1.952	0.899	0.248
$\beta_{3,5}$		-0.236	0.065	-0.046	-0.022		10.501	2.895	1.368	0.404
$\beta_{4,5}$		0.130	-0.038	0.009	0.020		6.664	1.961	0.688	0.162
$\beta_{5,5}$		-0.044	0.009	0.015	-0.005		1.842	0.300	0.087	0.020
<b>Panel B: Restricted model</b>										
$\beta_{-\max}$		-0.001	-0.002	0.003	0.001		0.109	0.010	0.004	0.001
$\beta_{\max}$		-0.012	-0.003	-0.004	-0.001		0.101	0.007	0.003	0.001
$\gamma_1$	1	1	2	3	5	1	1	2	3	5
$\bar{d}$	2	5	5	5	5	2	5	5	5	5

*Notes:* The values of the coefficients used in DGP are as follows:  $n = 200$ ,  $\gamma_0 = 1$ ,  $\beta_{1,1} = 0.3$ ,  $\beta_{k,d} = \frac{\beta_{-\max}}{d} = \frac{0.4}{d}$  for  $k < d$  and  $\beta_{d,d} = \beta_{\max} = 0.3$  for  $d = 2, \dots, 5$ . The instruments used in TSLS estimation are ordered peers' covariates. The TSLS estimation is stratified in the sense that the dataset is separated into three subdata— $\{i : d_i = 0\}, \dots, \{i : d_i = 5\}$ —and the sets of coefficients— $(\gamma_0, \gamma_1), \dots, (\beta_{1,5}, \dots, \beta_{5,5})$ —are estimated from each subdata. Panel A estimates the fully saturates model with fifteen  $\beta$ s and Panel B focuses on individuals with at least two peers and estimates the correctly specified model with two  $\beta$ s:  $\beta_{-\max} = 0.4$  and  $\beta_{\max} = 0.3$ .

Table 8:  $R^2$  in the first stage regression

LHS	Average first stage $R^2$				$\Pr \{R^2 \geq 0.7\}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\tilde{Y}_{1,5}$	0.396	0.614	0.715	0.775	0.004	0.240	0.603	0.849
$\tilde{Y}_{2,5}$	0.423	0.625	0.709	0.766	0.009	0.245	0.569	0.839
$\tilde{Y}_{3,5}$	0.429	0.622	0.708	0.768	0.003	0.228	0.574	0.865
$\tilde{Y}_{4,5}$	0.418	0.613	0.696	0.761	0.008	0.211	0.524	0.821
$\tilde{Y}_{5,5}$	0.382	0.587	0.683	0.746	0.004	0.168	0.470	0.757
$\gamma_1$	1	2	3	5	1	2	3	5
$\bar{d}$	5	5	5	5	5	5	5	5

*Notes:* The values of the coefficients used in DGP are as follows:  $n = 200$ ,  $\gamma_0 = 1$ ,  $\beta_{1,1} = 0.3$ ,  $\beta_{k,d} = \frac{0.4}{d}$  for  $k < d$  and  $\beta_{d,d} = 0.3$  for  $d = 2, \dots, 5$ . The instruments used in TSLS estimation are ordered peers' covariates. The TSLS estimation is stratified in the sense that the dataset is separated into three subdata— $\{i : d_i = 0\}, \dots, \{i : d_i = 5\}$ —and the sets of coefficients — $(\gamma_0, \gamma_1), \dots, (\beta_{1,5}, \dots, \beta_{5,5})$ —are estimated from each subdata.

## References

- Alne, Ragnar, Eyo I. Herstad, and Andreas Myhre.** 2025. “How Socioeconomic and Parental Background Shape Peer Networks and Educational Spillovers.” Available on SSRN: <https://www.ssrn.com/abstract=5160331>.
- Belloni, Alexandre, Christian Hansen, and Whitney Newey.** 2022. “High-dimensional linear models with many endogenous variables.” Journal of Econometrics, 228(1): 4–26.
- Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou.** 2024. “Toward a general theory of peer effects.” Econometrica, 92(2): 543–565.
- Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin.** 2009. “Identification of peer effects through social networks.” Journal of Econometrics, 150(1): 41–55.
- Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin.** 2020. “Peer effects in networks: A survey.” Annual Review of Economics, 12(1): 603–629.
- De Chaisemartin, Clément, and Xavier d’Haultfoeuille.** 2020. “Two-way fixed effects estimators with heterogeneous treatment effects.” American Economic Review, 110(9): 2964–2996.
- De Paula, Aureo.** 2013. “Econometric analysis of games with multiple equilibria.” Annu. Rev. Econ., 5(1): 107–131.
- Foster, Gigi.** 2006. “It’s not your peers, and it’s not your friends: Some progress toward understanding the educational peer effect mechanism.” Journal of Public Economics, 90(8–9): 1455–1475.
- Gazze, Ludovica, Claudia Persico, and Sandra Spirovska.** 2024. “The long-run spillover effects of pollution: How exposure to lead affects everyone in the classroom.” Journal of Labor Economics, 42(2): 357–394.
- Goldsmith-Pinkham, Paul, and Guido W Imbens.** 2013. “Social networks and the identification of peer effects.” Journal of Business & Economic Statistics, 31(3): 253–264.
- Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár.** 2024+. “Contamination bias in linear regressions.” The American Economic Review.
- Goodman-Bacon, Andrew.** 2021. “Difference-in-differences with variation in treatment timing.” Journal of Econometrics, 225(2): 254–277.
- Guryan, Jonathan, Kory Kroft, and Matthew J Notowidigdo.** 2009. “Peer effects in the workplace: Evidence from random groupings in professional golf tournaments.” American Economic Journal: Applied Economics, 1(4): 34–68.
- Han, Li, and Tao Li.** 2009. “The gender difference of peer influence in higher education.” Economics of Education review, 28(1): 129–134.

- Hsieh, Chih-Sheng, and Lung Fei Lee.** 2016. "A social interactions model with endogenous friendship formation and selectivity." Journal of Applied Econometrics, 31(2): 301–319.
- Jochmans, Koen.** 2023. "Peer effects and endogenous social interactions." Journal of Econometrics, 235(2): 1203–1214.
- Johnsson, Ida, and Hyungsik Roger Moon.** 2021. "Estimation of peer effects in endogenous social networks: Control function approach." Review of Economics and Statistics, 103(2): 328–345.
- Kang, Changhui.** 2007. "Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea." Journal of Urban Economics, 61(3): 458–495.
- Manski, Charles F.** 1993. "Identification of endogenous social effects: The reflection problem." The Review of Economic Studies, 60(3): 531–542.
- Masten, Matthew A.** 2018. "Random coefficients on endogenous variables in simultaneous equations models." The Review of Economic Studies, 85(2): 1193–1250.
- Mogstad, Magne, and Alexander Torgovitsky.** 2024. "Instrumental Variables with Unobserved Heterogeneity in Treatment Effects." National Bureau of Economic Research.
- Sacerdote, Bruce.** 2001. "Peer effects with random assignment: Results for Dartmouth roommates." The Quarterly journal of economics, 116(2): 681–704.
- Sacerdote, Bruce.** 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In Handbook of the Economics of Education. Vol. 3, 249–277. Elsevier.
- Tao, Ji, and Lung-fei Lee.** 2014. "A social interaction model with an extreme order statistic." The Econometrics Journal, 17(3): 197–240.
- Tincani, Michela M.** 2024. "Heterogeneous Peer Effects and Rank Concerns in the Classroom." CEPR Discussion Paper No. 18921. CEPR Press, Paris & London. <https://cepr.org/publications/dp18921>.