DISCUSSION PAPER SERIES

Identifying Peer Effects in Networks with Unobserved Effort and Isolated Students

Aristide Houndetoungan
Cristelle Kouame
Michael Vlassopoulos

Identifying Peer Effects in Networks with Unobserved Effort and Isolated Students

Aristide Houndetoungan
CY Cergy Paris University and Thema

Cristelle Kouame
World Bank Group

Michael Vlassopoulos
University of Southampton and IZA

ABSTRACT

Identifying Peer Effects in Networks with Unobserved Effort and Isolated Students*

Peer influence on effort devoted to some activity is often studied using proxy variables when actual effort is unobserved. For instance, in education, academic effort is often proxied by GPA. We propose an alternative approach that circumvents this approximation. Our framework distinguishes unobserved shocks to GPA that do not affect effort from preference shocks that do affect effort levels. We show that peer effects estimates obtained using our approach can differ significantly from classical estimates (where effort is approximated) if the network includes isolated students. Applying our approach to data on high school students in the United States, we find that peer effect estimates relying on GPA as a proxy for effort are 40% lower than those obtained using our approach.

Corresponding author:
Michael Vlassopoulos
Department of Economics
University of Southampton
Southampton, SO17 1BJ
United Kingdom

E-mail: m.vlassopoulos@soton.ac.uk

1 Introduction

In recent years, there has been a growing interest in the impact of peers on educational outcomes (Sacerdote, 2011; Epple and Romano, 2011). As peer effects on students' academic effort may involve a social multiplier effect, understanding whether students are influenced by their friends and the size of this influence is crucial for evaluating policies aimed at improving academic achievement (Manski, 1993). However, estimating peer effects based on academic effort presents challenges, since effort is generally unobserved. Consequently, many empirical studies resort to using grade point average (GPA) as a proxy variable for effort. While microfounded models exploring the impact of peer interactions on academic achievement are often focused on effort (e.g., see Calvó-Armengol et al., 2009; Arcidiacono et al., 2012; Fruehwirth, 2013, 2014; Hong and Lee, 2017), GPA is typically employed in the empirical analysis because measures of student effort are unavailable. Yet, this approximation overlooks that GPA is not solely influenced by effort but also by many other factors, including unobserved student and school characteristics.

In this paper, we investigate the implications of using GPA as a proxy for academic effort on the estimation and interpretation of peer effects. We develop a structural model of educational effort with social interactions, in which students decide on their academic effort while taking into account that of their peers (friends), assuming complementarity between students' and peers' efforts. We also explicitly model the production function of academic achievement (GPA), which includes effort as a key input. Without using a proxy for student academic effort, we show that the model allows for identifying peer effects on academic effort. Furthermore, we demonstrate that the size of the effect may differ from that estimated using GPA as a proxy for effort, because the proxy-based approach suffers from an issue akin to an omitted variables problem. We derive a reduced-form equation for GPA, which differs from the classical linear-in-means peer effects specification, that highlights the need to control for two types of unobserved GPA shocks to disentangle peer effects on academic effort from other common effects captured by GPA. First, one needs to account for common shocks that directly influence GPA, such as improvements in teaching quality, irrespective of the effort level. These shocks result in a GPA increase for the same level of effort and do not involve a social multiplier effect. Second, one needs to account for common shocks affecting students' preferences, such as increasing motivation to value academic achievement through information, which influence both academic effort and GPA, and may have social multiplier effects among students with friends.

We demonstrate that approximating student effort with GPA may result in biased estimates of peer effects when some students in the network do not have friends. This is because students who have friends are affected differently by the two types of GPA shocks mentioned above compared to those without friends. Standard approaches using GPA as a proxy fail to differentiate between these

two types of shocks, leading to biased estimates of peer effects when the network includes students without friends. However, we show that in networks without isolated students, the standard model does not produce biased estimates of peer effects. The key difference between the reduced-form of the standard model and our framework is that while the standard model has a single intercept for each school, our approach accounts for unobserved school-level heterogeneity based on whether a student has friends or not. This amounts to introducing two types of fixed effects per school, one for students with friends and one for isolated students.

We illustrate the importance of this distinction through both a Monte Carlo simulation study and an empirical application using the National Longitudinal Study of Adolescent to Adult Health (Add Health) data.[1] Estimating peer effects through our proposed method suggests that increasing the average GPA of peers by one point leads to a 0.856 point increase in students' GPA. In contrast, the standard linear-in-means model using GPA as a proxy for student academic effort estimates this effect at 0.507. This substantial difference highlights that failing to account for two types of shocks at the school level may yield substantially biased estimates. The intuition behind this bias is that the standard approach misidentifies direct shocks to GPA that do not influence effort as effort shocks, thereby overestimating their impact through a social multiplier effect. Importantly, we also find that other plausible approaches, such as estimating a model that accounts for only one type of school-level shock and incorporating a dummy variable for isolated students, fail to rectify the bias, as does excluding isolated students from the sample.

Our econometric model incorporates two types of school-level unobserved shocks to GPA, which renders the well-known *reflection problem* more complex.[2] In the case of the standard linear-in-means model, Bramoullé et al. (2009) provide straightforward conditions relating to the network structure under which the reflection problem is resolved. We extend their identification analysis to our framework; our main condition for identification requires that the network must include at least two students separated by a path of distance three, which is slightly stronger than the condition in Bramoullé et al. (2009).

We also extend our analysis to the case of endogenous networks. Network endogeneity can occur because we do not observe certain student characteristics, such as intellectual quotient (IQ), that may influence both students' likelihood to form links with others and their GPA. We control for these unobserved characteristics using a two-stage estimation approach. Our method is nonparametric as in Johnsson and Moon (2021). We do not impose a specific parametric restriction in the relationship between the unobserved characteristics and GPA. Our approach is similar to generalized additive

---

[1]The Add Health dataset comprises 22% of students without friends, including 11% who are not fully isolated, in the sense that they are nominated as friends of others. Other studies reporting social network data from educational settings with isolated students include those by Alan et al. (2021), Conti et al. (2013), and Boucher et al. (2021).

[2]The reflection problem arises when one cannot disentangle endogenous peer effects from exogenous contextual peer effects (Manski, 1993)

models (GAM), which are widely employed in the nonparametric regression literature (see Hastie, 2017).

Our structural model and econometric approach can also be applied to study peer influence on other outcomes that depend on exerted effort. An example is the body mass index (BMI), which cannot be directly chosen (e.g., Fortin and Yazbeck, 2015). People need to exert effort, such as developing healthy diet habits or engaging in physical exercise to improve their BMI. Peer influences are more related to effort than BMI. Another example is peer effects on a worker's effort (e.g., Mas and Moretti, 2009; Cornelissen et al., 2017). The observed outcome is generally worker's productivity, whereas peer effects take source in effort.

Related Literature

There is a large literature studying social interactions both theoretically and empirically (Durlauf and Ioannides, 2010; Blume et al., 2011). We follow the games in networks approach to the analysis of social interactions (see Jackson and Zenou, 2015, for a comprehensive overview of this literature). Ballester et al. (2006) analyze a noncooperative game with linear-quadratic utilities and strategic complementarities, in which each player decides how much effort they exert. Applying a similar framework to an education application, Calvó-Armengol et al. (2009) use GPA as a proxy for the exerted effort. We contribute to this strand of the literature by explicitly modelling the production function that captures how effort, along with other factors, translates into GPA, while distinguishing between unobserved shocks that directly impact GPA without affecting effort, from shocks that impact both academic effort and GPA. This leads to a reduced-form equation for GPA that differs from the standard linear-in-means peer effects specification.

This paper makes a methodological contribution to the extensive empirical literature on peer effects on educational outcomes (Sacerdote, 2011; Epple and Romano, 2011). We show both analytically and through an empirical application using AddHealth data that approximating student effort with GPA may result in biased estimates of peer effects, when some students in the network lack friends. Since isolated students are a common feature in many social network datasets, this finding highlights the limitations of using proxy variables, such as GPA, for estimating peer effects reliably.

Our paper also contributes to the econometric literature on peer effects (De Paula, 2017; Kline and Tamer, 2020) in two ways. First, a key challenge in this field is the reflection problem (Manski, 1993). A recent wave of papers have addressed this issue by imposing conditions on the network structure (Bramoullé et al., 2009; De Giorgi et al., 2010).[3] Our contribution is to study the reflection problem in a setting where the GPA is influenced by various types of common shocks at the school

---

[3]Other studies address the reflection problem using group size variation (Davezies et al., 2009; Lee, 2007) or imposing restrictions on the error terms (Graham, 2008; Rose, 2017). For an overview of this literature, see Bramoullé et al. (2020).

level, and where students may have no peers. As in Bramoullé et al. (2009), our main identification condition involves the network structure and can be readily tested in empirical applications. Our second contribution lies in addressing the network endogeneity issue. We allow for unobserved attributes to influence students' likelihood of forming links with others and their GPA (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021). Unlike many studies that impose a strong parametric restriction between unobserved attributes and the peer effect model, we adopt a nonparametric approach to connect GPA to these attributes. Our approach is similar to the control function method used by Johnsson and Moon (2021), but it is adaptable to complex models. Specifically, we employ cubic B-splines, which are commonly used in Generalized Additive Models (GAM), to establish a smooth function linking the unobserved attributes and GPA (Hastie, 2017).

The remainder of the paper is organized as follows. Section 2 presents the microeconomic foundations of the model using a network game in which students decide their academic effort. Section 3 describes the econometric model and addresses the identification and estimation of the parameters. Section 4 presents our empirical analysis using Add Health data. Section 5 provides an extension of our framework to endogenous networks. Section 6 concludes this paper.

2 The Structural Model

In this section, we introduce a structural model based on a game of complete information, where students choose their effort level, which impacts their educational achievements (GPA). Students' effort levels are also influenced by the effort exerted by their peers. This model can also be applied to investigating peer effects on other inputs that are not typically directly observed by the econometrician. For instance, it can be used to examine peer effects on (unobserved) effort in a workplace setting where observed productivity serves as the outcome. Another application is peer effects on effort directed toward improving BMI, such as adopting healthy diet habits or engaging in physical exercise, which is more readily observable than the effort itself.

We consider S independent schools and denote by $n_s \geqslant 3$ the number of students in the s-th school, $s \in \{1, \ldots, S\}$. Let n be the total number of students, that is, $n = \sum_{s=1}^{S} n_s$. Each student i in school s has a GPA denoted by $y_{s,i}$ and observable characteristics represented by a K-vector $\mathbf{x}_{s,i}$. Students interact in their school through a directed network that can be represented by an $n_s \times n_s$ adjacency matrix $\mathbf{A}_s = [a_{s,ij}]_{ij}$, where $a_{s,ij} = 1$ if student j is $i's$ friend and $a_{s,ij} = 0$ otherwise. We assume that $a_{s,ii} = 0$ for all i and s so that students cannot interact with themselves. In addition, we only consider within-school interactions; students do not interact with peers in other schools (see Calvó-Armengol et al., 2009). We define the social interaction matrix $\mathbf{G}_s = [g_{s,ij}]_{ij}$ as the row-normalized adjacent matrix $\mathbf{A}_s$, that is, $g_{s,ij} = 1/n_{s,i}$ if j is a $i's$ friend and $g_{s,ij} = 0$ otherwise, where

$n_{s,i}$ is the number of friends of student i within school s.

Student i chooses their effort $e_{s,i}$, which in turn affects their GPA. More precisely, GPA is a function of student effort $e_{s,i}$, observable characteristics $\mathbf{x}_{s,i}$, and a random term $\eta_{s,i}$ that captures unobservable characteristics.[4] Following Fruehwirth (2013) and Boucher and Fortin (2016), we posit that this relationship is given by the production function:

$$y_{s,i} = \alpha_s + \delta e_{s,i} + \mathbf{x}'_{s,i}\boldsymbol{\theta} + \eta_{s,i}, \tag{1}$$

where $\delta > 0$ and α_s and $\boldsymbol{\theta}$ are unknown parameters. The parameter α_s captures unobserved school-level GPA shifters, such as teacher quality, operating as fixed effects. The linear production function is somewhat restrictive and can be generalized to a nonlinear or nonparametric production function (e.g., see Fruehwirth, 2013). However, identification of the resulting econometric model can be intractable, especially when the model allows for unobserved school heterogeneity.

The effort exerted and the GPA obtained provide students with a benefit that is captured by a payoff function, which, as in Calvó-Armengol et al. (2009) and Boucher and Fortin (2016), takes a linear-quadratic form:

$$u_{s,i}(e_{s,i}, \mathbf{e}_{s,-i}, y_{s,i}) = \underbrace{(c_s + \mathbf{x}'_{s,i}\boldsymbol{\beta} + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma} + \varepsilon_{s,i})y_{s,i} - \frac{e_{s,i}^2}{2}}_{\text{private sub-payoff}} + \underbrace{\lambda e_{s,i}\mathbf{g}_{s,i}\mathbf{e}_s}_{\text{social sub-payoff}}, \tag{2}$$

where $\mathbf{X}_s = (\mathbf{x}_{s,1}, \ldots, \mathbf{x}_{s,n_s})'$, $\mathbf{g}_{s,i}$ is the i-th row of $\mathbf{G}_s$, $\mathbf{e}_{s,-i} = (e_{s,1}, \ldots, e_{s,i-1}, e_{s,i}, \ldots, e_{s,n})'$, $\mathbf{e}_s = (e_{s,1}, \ldots, e_{s,n_s})'$, the term $\mathbf{g}_{s,i}\mathbf{e}_s$ is the average effort of peers, and c_s, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are unknown parameters. The parameter λ captures endogenous peer effects. The payoff function (2) encompasses two components: a private sub-payoff and a social sub-payoff. The term $(c_s + \mathbf{x}'_{s,i}\boldsymbol{\beta} + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma} + \varepsilon_{s,i})$ represents the benefit enjoyed per unit of GPA achieved, where $\varepsilon_{s,i}$ is the student type (observable by all students). This benefit accounts for student observed heterogeneity, as it depends on $\mathbf{x}_{s,i}$ and peer group average characteristics $\mathbf{g}_{s,i}\mathbf{X}_s$ termed *contextual variables* (see Manski, 1993). The benefit also accounts for school unobserved heterogeneity through the parameter c_s. The second term of the private sub-payoff, $e_{s,i}^2/2$ reflects the cost of exerting effort. The social sub-payoff $\lambda e_{s,i}\mathbf{g}_{s,i}\mathbf{e}_s$ implies that an increase in the average peer group's effort $\mathbf{g}_{s,i}\mathbf{e}_s$ influences student i's marginal payoff if $\lambda \neq 0$. When $\lambda > 0$, the payoff function (2) implies complementarity between students' and peers' efforts, whereas $\lambda < 0$ indicates substitutability in efforts.[5]

The parameters α_s and c_s capture different unobserved shocks at the school level, and are conceptually different in terms of their policy implications. α_s captures unobserved shocks on GPA that do

---

[4] Whether or not the random term $\eta_{s,i}$ is observed by student i or their peers is inconsequential for the analysis of the game.

[5] An alternative specification considers the social-payoff as $-\frac{\lambda}{2}(e_{s,i} - \mathbf{g}_{s,i}\mathbf{e}_s)^2$, which represents a social cost depending on the gap between the student's effort level and the average peer effort. This specification leads to conformist preferences when $\lambda > 0$. Our approach can also be extended to this alternative specification.

not affect student effort. These shocks, such as variation in teaching quality and school management, directly impact GPA irrespective of student effort.[6] On the other hand, c_s captures shocks on student preferences, particularly on the marginal payoff. For instance, interventions aimed at making students more aware of the returns to academic achievement could influence the marginal payoff. Such a shock can be captured by c_s and would influence effort and consequently GPA through equation (1). We show that α_s and c_s do not impact GPA in the same way (see Section 3.1).

By substituting $y_{s,i}$ from Equation (1) into Equation (2), we obtain a payoff function that does not depend on GPA (see Appendix A.1). This new payoff function defines a static game with complete information, in which students simultaneously choose their effort levels to maximize their payoff. The best response function for the students is given by:

$$e_{s,i} = \delta c_s + \lambda \mathbf{g}_{s,i} \mathbf{e}_s + \delta \mathbf{x}'_{s,i} \boldsymbol{\beta} + \delta \mathbf{g}_{s,i} \mathbf{X}_s \boldsymbol{\gamma} + \delta \varepsilon_{s,i}. \tag{3}$$

In Equation (3), students' levels of effort are expressed as a function of the average effort of their peers $\mathbf{g}_{s,i} \mathbf{e}_s$, observed students' characteristics $\mathbf{x}_{s,i}$, and the average characteristics of peers $\mathbf{g}_{s,i} \mathbf{X}_s$ (contextual variables). The parameter λ represents the impact of peers on a student's effort level. A positive value of λ indicates that a student's effort level increases if their peers put in more effort. Furthermore, Equation (3) shows that the optimal effort level (and thus the resulting GPA) is influenced by shocks at the school level (on c_s) that affect student preferences. However, shocks directly affecting GPA through the parameter α_s do not impact effort.

The best response function (3) in matrix form can be expressed as $\mathbf{e}_s = \delta c_s \mathbf{1}_{n_s} + \lambda \mathbf{G}_s \mathbf{e}_s + \delta \mathbf{X}_s \boldsymbol{\beta} + \delta \mathbf{G}_s \mathbf{X}_s \boldsymbol{\gamma} + \delta \boldsymbol{\varepsilon}_s$, where $\boldsymbol{\varepsilon}_s = (\varepsilon_{s,1}, \ldots, \varepsilon_{s,n_s})'$ and $\mathbf{1}_{n_s}$ is an n_s-vector of ones. A solution of this equation in $\mathbf{e}_s$ is a Nash equilibrium (NE) of the game. As $\mathbf{G}_s$ is row-normalized, the NE is unique under Assumption 2.1 and can be expressed as $\mathbf{e}_s = (\mathbf{I}_{n_s} - \lambda \mathbf{G}_s)^{-1} (\delta c_s \mathbf{1}_{n_s} + \delta \mathbf{X}_s \boldsymbol{\beta} + \delta \mathbf{G}_s \mathbf{X}_s \boldsymbol{\gamma} + \delta \boldsymbol{\varepsilon}_s)$, where $\mathbf{I}_{n_s}$ is the $n_s \times n_s$ identity matrix (see Appendix A.1).

Assumption 2.1. $|\lambda| < 1$.

The condition $|\lambda| < 1$ implies that students do not increase (in absolute value) their effort as much as the increase in the effort of their peers. Put differently, when the average effort of a student's friends increases by one unit, the corresponding change in the student's own effort is less than one unit in absolute value.

---

[6]This reasoning holds because GPA is unbounded. An increase in α_s necessarily results in a higher GPA, and thus a higher payoff, regardless of the effort level. If we were to consider a framework where GPA is bounded, an increase in α_s may decrease the effort for students nearing the upper bound of GPA (e.g., see Fruehwirth, 2013). However, an increase in α_s does not have the same implication as an increase in c_s, which is the key factor in our framework.

3   The Econometric Model and Identification Strategy

If effort were directly observable, we could estimate the peer effect parameter from Equation (3) following Kelejian and Prucha (1998). However, as we do not observe effort, the equation cannot be estimated directly. Instead, we derive from this equation an estimable equation based on GPA, which is influenced by effort. We show that this equation is econometrically different from the equation that we obtain if we proxy effort using GPA in Equation (2). We also present an identification strategy to identify the peer effect parameter λ in the effort equation.

3.1   Reduced-Form Equation for GPA

From Equation (1), we express effort as a function of GPA and replace this expression in Equation (3). This yields a reduced-form equation for GPA that does not directly depend on effort (see Appendix A.2). The equation is given by

$$y_{s,i} = \kappa_{s,i} + \lambda \mathbf{g}_{s,i} \mathbf{y}_s + \mathbf{x}'_{s,i} \tilde{\boldsymbol{\beta}} + \mathbf{g}_{s,i} \mathbf{X}_s \tilde{\boldsymbol{\gamma}} + (\boldsymbol{\omega}_{s,i} - \lambda \mathbf{g}_{s,i}) \boldsymbol{\eta}_s + \delta^2 \varepsilon_{s,i}, \tag{4}$$

where $\kappa_{s,i} = \delta^2 c_s + (1 - \lambda \mathbf{g}_{s,i} \mathbf{1}_{n_s}) \alpha_s$, $\tilde{\boldsymbol{\beta}} = \delta^2 \boldsymbol{\beta} + \boldsymbol{\theta}$, $\tilde{\boldsymbol{\gamma}} = \delta^2 \boldsymbol{\gamma} - \lambda \boldsymbol{\theta}$, and $\boldsymbol{\omega}_{s,i}$ is a row-vector of dimension n_s in which all the elements are zero, except the i-th element, which is one.

If instead, we proxy effort by GPA in the payoff function (2), the resulting reduced-form equation of GPA would be:

$$y_{s,i} = \bar{c}_s + \bar{\lambda} \mathbf{g}_{s,i} \mathbf{y}_s + \mathbf{x}'_{s,i} \bar{\boldsymbol{\beta}} + \delta \mathbf{g}_{s,i} \mathbf{X}_s \bar{\boldsymbol{\gamma}} + \bar{\varepsilon}_{s,i}, \tag{5}$$

where $\bar{\lambda}$ is the peer effect parameter, $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\gamma}}$ measure the effects of own and contextual characteristics, $\bar{c}_s$ controls for unobserved school heterogeneity and $\bar{\varepsilon}_{s,i}$ is an error term. We will refer to this specification as the *standard* (*classical*) *model*.

Let $\mathcal{V}_s^{NI}$ denote the subsample of students of school s who have friends (non-isolated) and $\mathcal{V}_s^I$ denote the subsample of students of school s who have no friends (isolated). As $\mathbf{G}_s$ is row-normalized, we have $\mathbf{g}_{s,i} \mathbf{1}_{n_s} = 1$ if $i \in \mathcal{V}_s^{NI}$ and $\mathbf{g}_{s,i} \mathbf{1}_{n_s} = 0$ otherwise. Thus, $\kappa_{s,i} = \kappa_s^I$ if $i \in \mathcal{V}_s^I$ and $\kappa_{s,i} = \kappa_s^{NI}$ if $i \in \mathcal{V}_s^{NI}$, where $\kappa_s^I = \delta^2 c_s + \alpha_s$ and $\kappa_s^{NI} = \delta^2 c_s + (1 - \lambda) \alpha_s$. If there are no isolated students in the network, $\kappa_{s,i}$ would be a simple school fixed effect, and our framework would be equivalent to the classical model.[7] The difference between the standard model and our framework is that the standard model has a single intercept term per school. In Equation (4), $\kappa_{s,i}$ accounts for unobserved school-level heterogeneity depending on whether the student i has friends or not. We would have $\kappa_s^{NI} = \kappa_s^I$ if and only if $\alpha_s = 0$, $\lambda = 0$, or if every student has friends. These conditions may not hold in many

---

[7]An isolated student is a student who has no friends. However, this student may be a friend of others. We later refer to a *fully* isolated student as a student who has no friends and who is not a friend of others.

settings.

To appropriately account for the unobserved factor $\kappa_{s,i}$ in Equation (4), the above discussion suggests that we need to incorporate both school-fixed effects and a school-specific variable indicating whether a student has friends or not. This involves including S school dummy variables and S dummy variables indicating whether each student has friends or not. Each of the latter dummy variables corresponds to one school and takes the value of one if the student has friends. The rationale for this lies in the distinction between α_s and c_s, which affect GPA differently. To understand why distinguishing between these two shocks requires controlling for whether a student is isolated or not, it is important to consider the implications of the two shocks. As per Equations (3) and (4), an increase in α_s (e.g., by improving teaching quality) suggests an increase in GPA without affecting effort. Importantly, this increase does not depend on whether the student is isolated and does not involve a social multiplier effect.[8] In contrast, a preference shock on c_s could result in a social multiplier effect on the effort (according to Equation (3)), and thereby on GPA. This social multiplier effect is only present among students who have friends. Therefore, the distinction between the two types of shocks is essentially captured by the student's social status: unlike for α_s, the impact of c_s on GPA is contingent upon whether the student is isolated or not.

The main distinction between the standard model and Equation (4) arises due to an omitted variables problem inherent in the standard model. These omitted variables are $-\lambda\alpha_s\mathbf{g}_{s,i}\mathbf{1}_{n_s}$, for $s = 1,\ldots,S$, and may introduce a significant discrepancy between the peer effect estimates of the two models. Following the logic of omitted variable bias (Theil, 1957), we know that the difference between the peer effect estimate from the standard model and the estimate from our specification has the same sign as the partial correlation between $-\lambda\alpha_s\mathbf{g}_{s,i}\mathbf{1}_{n_s}$ and $\mathbf{g}_{s,i}\mathbf{y}_s$. In most scenarios, it will be the sign of $-\lambda\alpha_s$ if $\mathbf{y}_s$ is nonnegative. This suggests that peer effects estimated using the standard model would be underestimated compared to the estimate from our specification if $\lambda\alpha_s > 0$.[9]

3.2   Identification and Estimation

As the effort $e_{s,i}$ is not observed, not all the parameters of the structural model can be identified. Specifically, δ cannot be identified because it only appears in Equation (4) through a product with other parameters. Similarly, $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ cannot be identified separately, as they only appear in the expression of $\tilde{\boldsymbol{\gamma}}$. Nevertheless, we can identify the composite parameters $\tilde{\boldsymbol{\beta}} = \delta^2\boldsymbol{\beta} + \boldsymbol{\theta}$ and $\tilde{\boldsymbol{\gamma}} = \delta^2\boldsymbol{\gamma} - \lambda\boldsymbol{\theta}$,

---

[8]Equation (4) implies that the variation in $\mathbf{y}_s$, denoted $\Delta^\alpha\mathbf{y}_s$, following an increase $\Delta\alpha_s$ in α_s is such that $\Delta^\alpha\mathbf{y}_s = \Delta\alpha_s(\mathbf{I}_{n_s} - \lambda\mathbf{G}_s)\mathbf{1}_{n_s} + \lambda\mathbf{G}_s\Delta^\alpha\mathbf{y}_s$. This implies that $(\mathbf{I}_{n_s} - \lambda\mathbf{G}_s)\Delta^\alpha\mathbf{y}_s = \Delta\alpha_s(\mathbf{I}_{n_s} - \lambda\mathbf{G}_s)\mathbf{1}_{n_s}$ and thus, $\Delta^\alpha\mathbf{y}_s = \Delta\alpha_s\mathbf{1}_{n_s}$. Hence, the increase in α_s results in the same increase in GPA for all students.

[9]Another difference between the standard model and Equation (4) is that the standard model does not take into account the term $(\boldsymbol{\omega}_{s,i} - \lambda\mathbf{g}_{s,i})\boldsymbol{\eta}_s$. However, it is worth pointing out that this second difference does not lead to inconsistent estimates if $\boldsymbol{\eta}_s$ is independent of $\mathbf{G}_s$. Indeed, even in the case of correlated effects, estimating the model without controlling for these effects leads to a consistent estimator (Kelejian and Prucha, 1998).

which are parameters in the reduced-form equation (4) that capture the causal effects of students' characteristics $\mathbf{x}$ and average friends' characteristics $\bar{\mathbf{x}}$ on GPA. Consequently, it is not necessary to identify δ and $\boldsymbol{\theta}$ to estimate these causal effects.[10]

Let n_s^{NI} denote the number of students in school s with peers, and n_s^I the number of students in the school s without peers. Let also $\boldsymbol{\ell}_s^{NI} = \mathbf{G}_s \mathbf{1}_{n_s}$ and $\boldsymbol{\ell}_s^I = \mathbf{1}_{n_s} - \boldsymbol{\ell}_s^{NI}$. Equation (4) can be written in matrix form at the school level as

$$\mathbf{y}_s = \kappa_s^I \boldsymbol{\ell}_s^I + \kappa_s^{NI} \boldsymbol{\ell}_s^{NI} + \lambda \mathbf{G}_s \mathbf{y}_s + \mathbf{X}_s \tilde{\boldsymbol{\beta}} + \mathbf{G}_s \mathbf{X}_s \tilde{\boldsymbol{\gamma}} + (\mathbf{I}_{n_s} - \lambda \mathbf{G}_s) \boldsymbol{\eta}_s + \delta^2 \boldsymbol{\varepsilon}_s. \tag{6}$$

Note that the number of unknown parameters to be estimated in Equation (6) grows to infinity with the number of schools (there are $2S$ dummy variables). This issue is known as an incidental parameter problem and may lead to inconsistent estimators (Lancaster, 2000). A common approach to consistently estimate the model is to eliminate the fixed effects κ_s^I and κ_s^{NI}. To do so, we define $\mathbf{J}_s := \mathbf{I}_{n_s} - \frac{1}{n_s^I} \boldsymbol{\ell}_s^I \boldsymbol{\ell}_s^{I\prime} - \frac{1}{n_s^{NI}} \boldsymbol{\ell}_s^{NI} \boldsymbol{\ell}_s^{NI\prime}$ and impose by convention that $\frac{1}{n_s^I} \boldsymbol{\ell}_s^I \boldsymbol{\ell}_s^{I\prime} = 0$ if $n_s^I = 0$, and that $\frac{1}{n_s^{NI}} \boldsymbol{\ell}_s^{NI} \boldsymbol{\ell}_s^{NI\prime} = 0$ if $n_s^{NI} = 0$. Note that $\boldsymbol{\ell}_s^{NI\prime} \boldsymbol{\ell}_s^{NI} = \hat{n}_s$, $\boldsymbol{\ell}_s^{I\prime} \boldsymbol{\ell}_s^I = n_s^I$, $\boldsymbol{\ell}_s^{NI\prime} \boldsymbol{\ell}_s^I = 0$, $\boldsymbol{\ell}_s^{I\prime} \boldsymbol{\ell}_s^{NI} = 0$. Thus, $\mathbf{J}_s \boldsymbol{\ell}_s^I = \mathbf{J}_s \boldsymbol{\ell}_s^{NI} = 0$. One can eliminate the term $\kappa_s^I \boldsymbol{\ell}_s^I + \kappa_s^{NI} \boldsymbol{\ell}_s^{NI}$ by premultiplying each term of Equation (6) by the matrix $\mathbf{J}_s$.[11] This implies that

$$\mathbf{J}_s \mathbf{y}_s = \lambda \mathbf{J}_s \mathbf{G}_s \mathbf{y}_s + \mathbf{J}_s \mathbf{X}_s \tilde{\boldsymbol{\beta}} + \mathbf{J}_s \mathbf{G}_s \mathbf{X}_s \tilde{\boldsymbol{\gamma}} + \mathbf{J}_s (\mathbf{I}_{n_s} - \lambda \mathbf{G}_s) \boldsymbol{\eta}_s + \delta^2 \mathbf{J}_s \boldsymbol{\varepsilon}_s. \tag{7}$$

The random term $\boldsymbol{v}_s := \mathbf{J}_s (\mathbf{I}_{n_s} - \lambda \mathbf{G}_s) \boldsymbol{\eta}_s + \delta^2 \mathbf{J}_s \boldsymbol{\varepsilon}_s$ is comprised of two error vectors $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$. To consistently estimate $\boldsymbol{\psi} := (\lambda, \tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\gamma}}')'$, we impose the following assumption.

Assumption 3.1. *For any $s = 1, \ldots, S$, $\mathbb{E}(\boldsymbol{\eta}_s | \mathbf{G}_s, \mathbf{X}_s) = 0$ and $\mathbb{E}(\boldsymbol{\varepsilon}_s | \mathbf{G}_s, \mathbf{X}_s) = 0$*

Assumption 3.1 implies that $\mathbf{X}_s$ and $\mathbf{G}_s$ are exogenous with respect to $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$. This suggests that there is no omission of important variables in $\mathbf{X}_s$, which are captured by $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$. We later relax this assumption in our approach to controlling for network endogeneity by allowing for $\mathbf{X}_s$ and $\mathbf{G}_s$ to depend on $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$ through unobserved factors to the econometrician (see Section 5).

Identification and Estimation of $\boldsymbol{\psi}$

Identification in peer effects models can be challenging, particularly due to the reflection problem (Manski, 1993). Bramoullé et al. (2009) address this problem and provide necessary and sufficient conditions for identification. Their main condition requires that $\mathbf{I}_{n_s}$, $\mathbf{G}_s$, $\mathbf{G}_s^2$, and $\mathbf{G}_s^3$ are linearly

---

[10]However, the non-identification of δ poses challenges for implementing certain counterfactual analyses, particularly when examining direct shocks on GPA and preference shocks on student effort. Equation (4) suggests that an increase in c_s by Δc_s (preference shock) leads to a change of $\delta^2 \Delta c_s (\mathbf{I}_{n_s} - \lambda \mathbf{G}_s)^{-1} \mathbf{1}_{n_s}$ in $\mathbf{y}_s$. The impact of this increase on GPA depends on δ.

[11]By premultiplying each term by $\mathbf{J}_s$, we consider Equation (6) in deviation to the average within the student group, that is, $\mathcal{V}_s^I$ or $\mathcal{V}_s^{NI}$. This eliminates the parameters κ_s^I and κ_s^{NI}.

independent. However, their approach does not apply to our framework because their identification results assume that there are no isolated students (see their main assumption in their Section 2.1).[12]

Given that existing identification results do not directly apply to our model, we extend the analysis in Bramoullé et al. (2009). We derive easy-to-verify conditions to address the reflection problem when the network includes students without friends.

Assumption 3.2. *(i) $\lambda\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}} \neq 0$; (ii) There are students in the network separated by a link of distance three.*

Condition (i) is equivalent to stating that GPA is influenced by at least one contextual variable.[13] With several characteristics in $\mathbf{X}_s$, this condition can be satisfied. Condition (ii) is slightly stronger than the assumption that $\mathbf{I}_{n_s}$, $\mathbf{G}_s$, $\mathbf{G}_s^2$, and $\mathbf{G}_s^3$ are linearly independent. It means that there are students who have connections to students that extend to three degrees of separation—friends of friends of friends—who are neither directly their friends nor friends of their friends (see an illustration in Figure 1 below). As we allow isolated students, Assumption 3.2 is not a necessary condition for identification, as is the case in Bramoullé et al. (2009); rather, it is a sufficient condition that holds in many cases.

Under Assumption 3.2, we show that ψ is identified. We estimate ψ using a standard GMM approach. It is well known that the regressor $\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s$ is endogenous in Equation (7). However, it can be instrumented by $\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s$, as suggested by Kelejian and Prucha (1998) and Bramoullé et al. (2009).[14] Let $\hat{\psi}$ be the GMM estimator of ψ. We have the following result.

Proposition 3.1. *Under Assumptions 2.1–3.2 and A.1 (stated in Appendix A.3), ψ is globally identified, $\hat{\psi}$ is a consistent estimator, and $\sqrt{n}(\hat{\psi} - \psi_0) \xrightarrow{d} \mathcal{N}\big(0, \lim_{n\to\infty} n\,\mathbb{V}(\hat{\psi})\big)$, where ψ_0 is the true value of ψ.*

The consistency and asymptotic normality of $\hat{\psi}$ are directly derived from Kelejian and Prucha (1998). We later discuss how the asymptotic variance of $\hat{\psi}$ can be estimated.

We show that the reflection problem arises within a network with links of distance three only when variations in friends' average characteristics have no influence on students' outcomes, which would directly contradict Condition (i) of Assumption 3.2.[15] Below, we provide an intuition behind our identification approach and a formal proof in Appendix A.3. Figure 1 presents an example of a network with a link of distance three: i_2 is at three nodes from i_1; they are a friend of a friend of a

---

[12]Bramoullé et al. (2020) also discuss the case involving isolated students. They argue that the presence of isolated students can help identify the peer effect parameter (see their Section 2.1.1). Recall that the two intercepts of our model are $\kappa_s^I = \delta^2 c_s + \alpha_s$ and $\kappa_s^{NI} = \delta^2 c_s + (1-\lambda)\alpha_s$. The fact that only one intercept depends on λ implies that λ can be identified using isolated students if $\kappa_s^I = \kappa_s^{NI}$. However, as we allow GPA shocks α_s to be different from preference shocks c_s in terms of their impact on effort, this condition is unlikely to hold. Consequently, the presence of isolated students does not help identify λ in our case.

[13]See Equation A.7, which quantifies the total effect of an increase in a contextual variable on GPA.

[14]To avoid a weak instrument issue, the pool of instruments can also be expanded to $\mathbf{J}_s[\mathbf{G}_s^2\mathbf{X}_s, \ldots, \mathbf{G}_s^p\mathbf{X}_s]$ for some integer $p > 2$.

[15]Houndetoungan (2022) applies a similar approach to the case of nonlinear models.

friend of i_1. The reflection problem arises when $\mathbb{E}(\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s|\mathbf{G}_s,\mathbf{X}_s)$ is perfectly collinear with $\mathbf{J}_s\mathbf{X}_s$ and $\mathbf{J}_s\mathbf{G}_s\mathbf{X}_s$. For a non-isolated student, this collinearity is equivalent to saying that there exist vector of parameters, $\dot{\boldsymbol{\beta}}$ and $\dot{\boldsymbol{\gamma}}$, such that $\mathbb{E}(\mathbf{g}_{s,i}\mathbf{y}_s - y_s^{NI}|\mathbf{G}_s,\mathbf{X}_s) = (\mathbf{x}_{s,i}-\hat{\mathbf{x}}_s)'\dot{\boldsymbol{\beta}}+(\mathbf{g}_{s,i}\mathbf{X}_s-\bar{\mathbf{x}}_s^{NI})'\dot{\boldsymbol{\gamma}}$, where y_s^{NI}, $\mathbf{x}_s^{NI}$, and $\bar{\mathbf{x}}_s^{NI}$ are respectively the averages of $\mathbf{g}_{s,i}\mathbf{y}_s$, $\mathbf{x}_{s,i}$, and $\mathbf{g}_{s,i}\mathbf{X}_s$ in $\mathcal{V}_s^{NI}$. The variables $\mathbf{g}_{s,i}\mathbf{y}_s$, $\mathbf{x}_{s,i}$, and $\mathbf{g}_{s,i}\mathbf{X}_s$ are taken in deviation with respect to their average in $\mathcal{V}_s^{NI}$ because of the matrix $\mathbf{J}_s$ that multiplies the terms of Equation (7). If we take the previous equation in difference between students i_1 and i_3, we obtain

$$\mathbb{E}(\mathbf{g}_{s,i_1}\mathbf{y}_s - \mathbf{g}_{s,i_3}\mathbf{y}_s|\mathbf{G}_s,\mathbf{X}_s) = (\mathbf{x}_{s,i_1} - \mathbf{x}_{s,i_3})'\dot{\boldsymbol{\beta}} + (\mathbf{g}_{s,i_1}\mathbf{X}_s - \mathbf{g}_{s,i_3}\mathbf{X}_s)'\dot{\boldsymbol{\gamma}}. \tag{8}$$

As i_3 is i_1's only friend and i_4 is i_3's only friend, we have $\mathbf{g}_{s,i_1}\mathbf{y}_s = y_{s,i_3}$, $\mathbf{g}_{s,i_3}\mathbf{y}_s = y_{s,i_4}$, $\mathbf{g}_{s,i_1}\mathbf{X}_s = \mathbf{x}_{s,i_3}$, and $\mathbf{g}_{s,i_3}\mathbf{X}_s = \mathbf{x}_{s,i_4}$. Thus Equation (8) implies $y_{i_3}^e - y_{i_4}^e = (\mathbf{x}_{s,i_1} - \mathbf{x}_{s,i_3})'\dot{\boldsymbol{\beta}} + (\mathbf{x}_{s,i_3} - \mathbf{x}_{s,i_4})'\dot{\boldsymbol{\gamma}}$, where $y_i^e = \mathbb{E}(y_{s,i}|\mathbf{G}_s,\mathbf{X}_s)$. As $\mathbf{x}_{s,i_2}$ does not appear in the previous equation, we can say that $y_{i_3}^e - y_{i_4}^e$ is independent of $\mathbf{x}_{s,i_2}$, conditional on $\mathbf{x}_{s,i_1}$, $\mathbf{x}_{s,i_3}$ and $\mathbf{x}_{s,i_4}$. Put differently, an increase in $\mathbf{x}_{s,i_2}$ (*ceteris paribus*) has no impact on $y_{i_3}^e - y_{i_4}^e$, which means that the increase has either no impact on $y_{i_3}^e$ and $y_{i_4}^e$ or the same impact on $y_{i_3}^e$ and $y_{i_4}^e$. The first implication is not possible because $\mathbf{x}_{s,i_2}$ is the vector of contextual variables for i_4 and Condition (i) of Assumption 3.2 implies that GPA is influenced by at least one contextual variable. Moreover, $\mathbf{x}_{s,i_2}$ cannot influence $y_{i_3}^e$ and $y_{i_4}^e$ in the same way because i_2 is a direct friend of i_4 and not directly linked to i_3. This is in contradiction to equation (8), therefore, $\mathbb{E}(\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s|\mathbf{G}_s,\mathbf{X}_s)$ cannot be a linear combination of $\mathbf{J}_s\mathbf{X}_s$ and $\mathbf{J}_s\mathbf{G}_s\mathbf{X}_s$.
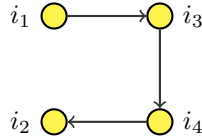
Figure 1: Solving the reflection problem
Note: → means that the node on the right side is a friend of the node on the left side.

Asymptotic variance of $\hat{\boldsymbol{\psi}}$

We now discuss how the asymptotic variance of $\hat{\boldsymbol{\psi}}$ can be consistently estimated. One simple approach is to use the robust estimator *à la* White (1980) that allows for the variance of the component of $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$ to vary across schools. Let $\tilde{\mathbf{X}}_s = \mathbf{J}_s[\mathbf{X}_s,\ \mathbf{G}_s\mathbf{X}_s]$, $\mathbf{R}_s = [\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s,\ \tilde{\mathbf{X}}_s]$, $\mathbf{Z}_s = [\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s,\ \tilde{\mathbf{X}}_s]$, $\mathbf{R}'\mathbf{Z} = \sum_{s=1}^S \mathbf{R}_s'\mathbf{Z}_s$, $\mathbf{Z}'\mathbf{Z} = \sum_{s=1}^S \mathbf{Z}_s'\mathbf{Z}_s$. Let also $\hat{\boldsymbol{v}}_s := \mathbf{J}_s\mathbf{y}_s - \mathbf{R}_s\hat{\boldsymbol{\psi}}$, that is, $\hat{\boldsymbol{v}}_s$ is the residual vector from Equation (7). We denote by diag the bloc diagonal matrix operator. The asymptotic variance of $\hat{\boldsymbol{\psi}}$ can be estimated by $\hat{\mathbf{B}}_n^{-1}\hat{\mathbf{D}}_n\hat{\mathbf{B}}_n^{-1}/n$, where $\hat{\mathbf{B}}_n = (\mathbf{R}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{R}'\mathbf{Z})'/n$ and $\hat{\mathbf{D}}_n = (\mathbf{R}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}\,\text{diag}\{\mathbf{Z}_1'\hat{\boldsymbol{v}}_1\hat{\boldsymbol{v}}_1'\mathbf{Z}_1,\dots,\mathbf{Z}_S'\hat{\boldsymbol{v}}_S\hat{\boldsymbol{v}}_S'\mathbf{Z}_S\}(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{R}'\mathbf{Z})'/n$.

However, for this estimator of the asymptotic variance, we need to impose that n_s is bounded, that is, the network is composed of many bounded and independent schools. This assumption is

important for $\mathbf{Z}'_s\hat{\boldsymbol{v}}_s$ to be $O_p(1)$ and have a second-order moment. We also present another estimator that does not require this condition. This estimator is based on the covariance structure of the error term $\boldsymbol{v}_s$. For this second approach, we set the following assumptions.

Assumption 3.3. *(i) $\mathbf{G}_s$ is uniformly bounded in column sum; (ii) $\lambda \neq 0$; (iii) $(\eta_{s,i},\ \varepsilon_{s,i})$ is independently distributed across i such that $\mathbb{E}(\eta_{s,i}^2|\mathbf{G}_s,\mathbf{X}_s) = \sigma_\eta^2 > 0$, $\mathbb{E}((\delta^2\varepsilon_{s,i})^2|\mathbf{G}_s,\mathbf{X}_s) = \sigma_\epsilon^2 > 0$, $\mathbb{E}(\delta^2\varepsilon_{s,i}\eta_{s,i}|\mathbf{G}_s,\mathbf{X}_s) = \rho\sigma_\eta\sigma_\epsilon$, and $|\rho| \leqslant 1$.*

If both S and n_s tend to infinity, Condition (i) rules out the cases where the sum of certain columns of $\mathbf{G}_s$ is unbounded. This condition is also considered by (Lee, 2004) in the case of the standard model. Condition (iii) imposes constant variances for the error terms $\eta_{s,i}$ and $\varepsilon_{s,i}$ but accounts for their potential correlation. This is important as $\eta_{s,i}$ and $\varepsilon_{s,i}$ characterize the same student i. The parameter ρ measures the correlation between $\eta_{s,i}$ and $\varepsilon_{s,i}$. We impose no restrictions on the joint distribution of $(\eta_{s,i},\ \varepsilon_{s,i})$. The restriction $\lambda \neq 0$ of Condition (ii) is necessary for identifying σ_η, σ_ϵ, and ρ. If $\lambda = 0$, the disturbance of Equation (7) would be $\mathbf{J}_s\boldsymbol{\eta}_s + \delta^2\mathbf{J}_s\boldsymbol{\varepsilon}_s$, and one cannot disentangle σ_η, σ_ϵ, and ρ.

Under Assumption 3.3, we can estimate $(\sigma_\eta^2,\ \sigma_\epsilon^2,\ \rho)$ and construct a consistent estimator for $\mathbb{V}(\hat{\boldsymbol{\psi}})$. We employ a quasi-maximum likelihood (QML) approach, where the dependent variable is the vector of the residual vector $\hat{\boldsymbol{v}}_s = \mathbf{J}_s\mathbf{y}_s - \mathbf{R}_s\hat{\boldsymbol{\psi}}$. The likelihood of $\hat{\boldsymbol{v}}_s$ is based on a multivariate normal distribution with mean and covariance matrix equal to those of the true error term $\boldsymbol{v}_s$, where λ is replaced by its estimator $\hat{\lambda}$. Importantly, we do not require $\boldsymbol{v}_s$ to be actually normally distributed. Let $(\hat{\sigma}_\eta^2,\ \hat{\sigma}_\epsilon^2,\ \hat{\rho})$ be the QML estimator of $(\sigma_\varepsilon^2,\ \sigma_\epsilon^2,\ \rho)$. We establish the following result.

Proposition 3.2. *Under Proposition 3.1 and Assumptions A.2–A.4 stated in Appendix A.4, $(\sigma_\epsilon^2,\ \sigma_\epsilon^2,\ \rho)$ is globally identified, and $(\hat{\sigma}_\eta^2,\ \hat{\sigma}_\epsilon^2,\ \hat{\rho})$ is a consistent estimator.*

We can now consistently estimate the asymptotic variance of $\hat{\boldsymbol{\psi}}$ using the estimator $(\hat{\sigma}_\eta^2,\ \hat{\sigma}_\epsilon^2,\ \hat{\rho})$. We denote by $\mathbf{B} = \text{plim}(\mathbf{R}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{R}'\mathbf{Z})'/n$, where plim stands for the limit in probability as n grows to infinity. We also define $\check{\boldsymbol{\Omega}} = \sum_{s=1}^S \mathbf{Z}'_s\big(\sigma_\epsilon^2\mathbf{J}_s + \sigma_\eta^2\mathbf{J}'_s\mathbf{W}_s\mathbf{W}'_s\mathbf{J}_s + \rho\sigma_\epsilon\sigma_\eta\mathbf{J}_s(\mathbf{W}_s + \mathbf{W}'_s)\mathbf{J}_s)\big)\mathbf{Z}_s$. We make the assumption that $\text{plim}(\mathbf{R}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}\check{\boldsymbol{\Omega}}(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{R}'\mathbf{Z})'/n$ exists and is denoted by $\mathbf{D}$. As a result $\lim_{n\to\infty} n\,\mathbb{V}(\hat{\boldsymbol{\psi}}) = \mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}$. We can obtain a consistent estimator of the variance by replacing σ_ϵ^2, σ_η^2, and ρ with their estimator, and $\mathbf{B}$ and $\mathbf{D}$ with their empirical counterparts.

3.3 Simulation Study

In this section, we conduct a simulation study to illustrate the importance of controlling for separate school fixed effects for isolated and non-isolated students. We consider a scenario with $S = 20$ schools, each school having $n_s = 50$ students. The network is defined such that student i in school

s has $n_{s,i}$ friends chosen randomly among their schoolmates. We randomly assign to $n_{s,i}$ a value in $\{0, 1, \ldots, 10\}$ according to the probability distribution given by $\mathbb{P}(n_{s,i} = k) = \dfrac{\pi}{(1+k)^{0.6}}$, where π is a constant such that $\sum_{k=0}^{10} \mathbb{P}(n_{s,i} = k) = 1$. This method for determining the number of friends results in 21% of students having no friends, with a maximum number of 10 friends per student. These characteristics of the simulated networks resemble those observed in real network data, such as the data provided by the Add Health survey (see Section 4).

We consider two exogenous variables in the matrix $\mathbf{X}_s$: $\mathbf{x}_{1s} \sim N(E_{1s},\ 16)$ and $\mathbf{x}_{2s} \sim \text{Poisson}(E_{2s})$, where E_{1s}, E_{2s} are fixed for each school and drawn from a Uniform distribution over $[0,\ 10]$. The distribution of these two control variables is specific to each school.

We investigate three data-generating processes (DGPs) denoted A, B, and C. For all DGPs, we assign the following values to the parameters: $\lambda = 0.7$, $\boldsymbol{\beta} = (1,\ 1.5)'$, $\boldsymbol{\gamma} = (5,\ -3)'$, $\delta = 1$, $\boldsymbol{\theta} = 0$, $\sigma_\eta^2 = 6$, $\sigma_\epsilon^2 = 3$, and $\rho = 0.4$, which implies $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1,\ \tilde{\beta}_2)' = (1,\ 1.5)'$ and $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1,\ \tilde{\gamma}_2)' = (5,\ -3)'$. We assume that $(\eta_{s,i},\ \varepsilon_{s,i})$ follows a bivariate normal distribution. We consider the same school preference shock for all DGPs, defined as $c_s = -1.5 q_{90}(\mathbf{x}_{2s})$, where $q_a(.)$ represents the a-th percentile. Generating c_s in this way ensures that c_s is a fixed effect because it is dependent on the control variables. To illustrate how the nature of the shock α_s can affect the results, we vary the definition of α_s for each DGP. For DGP A, we set $\alpha_s = 0$ for all s. For DGP C, we define $\alpha_s = 10 q_{90}(\mathbf{x}_{1s})$. For DGP B, we set α_s to the average of the values of $\alpha_1,\ \ldots,\ \alpha_S$ obtained for DGP C. For both DGPs A and B, α_s does not vary across schools. The only difference is that $\alpha_s = 0$ for DGP A, whereas $\alpha_s \neq 0$ for DGP B.

We consider four estimation approaches for each DGP. The most flexible estimation approach follows Equation (6), which can be expressed as follows:

$$y_{s,i} = \kappa_s^{NI} \ell_{s,i}^{NI} + \kappa_s^I (1 - \ell_{s,i}^{NI}) + \mathbf{g}_{s,i} \mathbf{y}_s + \mathbf{x}_{s,i}' \tilde{\boldsymbol{\beta}} + \mathbf{g}_{s,i} \mathbf{X}_s \tilde{\boldsymbol{\gamma}} + \tilde{v}_{s,i}, \tag{9}$$

where $\ell_{s,i}^{NI}$ is a dummy variable indicating whether the student is not isolated, $\kappa_s^I = \delta^2 c_s + \alpha_s$, $\kappa_s^{NI} = \delta^2 c_s + (1 - \lambda)\alpha_s$, and $\tilde{v}_{s,i} = (\boldsymbol{\omega}_{s,i} - \lambda \mathbf{g}_{s,i})\boldsymbol{\eta}_s + \delta^2 \varepsilon_{s,i}$. This specification, referred to as *Model 4*, is our proposed model. Additionally, we estimate several variants of this specification that impose constraints on c_s and α_s. These constrained specifications align with those commonly estimated in the literature when approximating students' efforts by their GPA:

- *Model 1*: Specification (9) without school fixed effects which assumes that $\kappa_s^I = \kappa_s^{NI} = \kappa$, where κ is a constant (c_s does not vary across schools and $\alpha_s = 0$).

- *Model 2:* Specification (9) with school fixed effects, which assumes that $\kappa_s^I = \kappa_s^{NI} = \kappa_s$, where κ_s varies across schools (c_s varies across schools, whereas $\alpha_s = 0$).

- *Model 3*: Specification (9) with school fixed effects and a dummy variable for isolated students, which assumes that κ_s^{NI} varies across schools and $\kappa_s^I = \kappa$, where κ is a constant. This means that $\alpha_s = \bar{\alpha}$ may not be zero but does not vary across schools, whereas c_s varies across schools.

Table 1 presents a summary of our Monte Carlo results across 1,000 replications. Model 1 yields biased estimates for all GPAs, which is expected due to the absence of fixed effects in this model. Although the bias of the peer effect estimate is not substantial for DGP A, estimates for other parameters, particularly $\tilde{\beta}_2$ and $\tilde{\gamma}_2$, exhibit notable biases. This is because DGP A features preference shocks c_s, generated from $\mathbf{x}_{2s}$, which Model 1 fails to account for.

Model 2 is the standard model commonly employed in the literature, especially when GPA approximates effort. This model demonstrates strong performance with DGP A, despite the presence of 21% isolated students. Interestingly, this result suggests that the dummy variable for isolated students might not be relevant, even when the network includes students without friends. As this dummy variable accounts for GPA shocks without influencing effort, its absence in DGP A (because $\alpha_s = 0$) does not compromise the validity of estimates. However, Model 2 leads to biased estimates for DGPs B and C, where $\alpha_s \neq 0$. Importantly, these biases carry substantial policy implications. For instance, while the true value of the social multiplier effect, measured by $1/(1 - \lambda)$ for non-isolated students, stands at 3.33, Model 2 underestimates it at 1.93 for DGP B and 1.73 for DGP C.

Model 3 incorporates a dummy variable for isolated students but assumes a constant coefficient associated with this variable across all schools. Adding this dummy variable means that α_s is a constant that may not be zero. This resolves the bias observed for DGP B but still results in biased estimates for DGP D. This is because DGP D assumes variability in α_s across schools.

Model 4, which is the most flexible specification, performs well across all DGPs. The simulation results also suggest that we can consistently estimate the covariance structure parameters σ_ϵ^2, σ_η^2, and ρ, as stated in Proposition 3.2. Yet, it is worth noting that the standard deviations of the estimates for DGPs A and B are higher in Model 4 compared to Model 3. This is because both models are suitable for these DGPs, whereas Model 4 is overparameterized. One can address this issue by comparing Model 3 to Model 4 using a Hausman specification test (Hausman, 1978). If the null hypothesis that the parameter ψ is the same for Models 3 and 4 is not rejected, then Model 3 would be preferred over Model 4.

4 Empirical Illustration

In this section, we provide an empirical illustration of our econometric approach to estimating peer effects using a unique and now widely-used dataset from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Specifically, our main objective in this section is to compare

Table 1: Simulation results

	Model 1		Model 2		Model 3		Model 4	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
				DGP A				
$\lambda = 0.7$	0.727	0.011	0.700	0.012	0.700	0.012	0.701	0.018
$\tilde{\beta}_1 = 1$	0.917	0.061	1.001	0.047	1.001	0.047	1.001	0.047
$\tilde{\beta}_2 = 1.5$	0.633	0.110	1.500	0.083	1.500	0.083	1.499	0.083
$\tilde{\gamma}_1 = 5$	5.018	0.099	4.999	0.091	5.000	0.096	5.003	0.098
$\tilde{\gamma}_2 = -3$	-3.542	0.109	-3.004	0.090	-3.002	0.124	-2.997	0.172
$\sigma_\epsilon^2 = 8$					9.354	5.266	9.122	5.216
$\sigma_\eta^2 = 15$					15.752	6.105	15.329	6.024
$\rho = 0.4$					0.448	0.513	0.459	0.511
				DGP B				
$\lambda = 0.7$	0.569	0.035	0.483	0.033	0.700	0.012	0.701	0.018
$\tilde{\beta}_1 = 1$	1.655	0.151	0.979	0.103	1.001	0.047	1.001	0.047
$\tilde{\beta}_2 = 1.5$	2.413	0.286	1.424	0.174	1.500	0.083	1.499	0.083
$\tilde{\gamma}_1 = 5$	3.862	0.312	3.719	0.313	5.000	0.096	5.003	0.098
$\tilde{\gamma}_2 = -3$	-8.343	0.413	-7.673	0.524	-3.002	0.124	-2.997	0.172
$\sigma_\epsilon^2 = 8$					9.354	5.266	9.122	5.216
$\sigma_\eta^2 = 15$					15.752	6.105	15.329	6.024
$\rho = 0.4$					0.448	0.513	0.459	0.511
				DGP C				
$\lambda = 0.7$	0.651	0.038	0.422	0.021	0.536	0.023	0.701	0.018
$\tilde{\beta}_1 = 1$	2.995	0.337	0.976	0.087	0.992	0.059	1.001	0.047
$\tilde{\beta}_2 = 1.5$	3.069	0.730	1.437	0.149	1.490	0.107	1.499	0.083
$\tilde{\gamma}_1 = 5$	4.026	0.374	3.464	0.310	4.308	0.163	5.003	0.098
$\tilde{\gamma}_2 = -3$	-9.189	0.504	-6.719	0.344	-3.365	0.442	-2.997	0.172
$\sigma_\epsilon^2 = 8$					9.135	8.653	9.122	5.216
$\sigma_\eta^2 = 15$					26.975	10.576	15.329	6.024
$\rho = 0.4$					0.198	0.661	0.459	0.511

Model 1 disregards school heterogeneity (i.e., c_s is constant across schools and $\alpha_s = 0$), while Model 2 accounts for it (with $\alpha_s = 0$). Model 3 includes a single fixed effect per school with a dummy variable capturing isolated students (i.e., α_s may not be zero but is constant across schools, whereas c_s varies). Model 4 is our structural model.

the estimate of peer effects obtained using our approach (equation 4) to that obtained using the classical linear-in-means peer effects specification (equation 5), in a context where isolated students are present.

4.1 Data

We use the Wave I in-school Add Health data, collected between September 1994 and April 1995. This is a dataset of a nationally representative sample of 90,118 students (7th to 12th grade) from 145 middle, junior high, and high schools across the US. It includes information on the social and demographic characteristics of students as well as their friendship links–in particular, their best friends, up

to 5 females and up to 5 males.

After removing observations with missing data, the sample used for our empirical analysis encompasses 68,430 students from 141 schools. The number of students per school ranges widely from 18 to 2,027, with an average of 485 per school. On average, each student reports having 3.4 friends (1.6 male friends and 1.9 female friends). Moreover, there are 14,900 (22%) students who have no peers (isolated students), including 7,655 (11%) who are not fully isolated, that is, they are nominated by others. Only 1% of students nominate 10 friends. This suggests that the top coding issue resulting from students' inability to nominate more than 10 friends (Griffith, 2022) is not a serious concern here.

The dependent variable, GPA, is the average grade across four subjects: mathematics, science, English/language arts, and history or social science. GPA is calculated on the basis of students' grades in these four subjects, which we recoded as follows: A = 4, B = 3, C = 2, and D = 1. In our analysis, we include controls for several potential factors that may influence GPA (Duncan et al., 2001; Lin, 2010). These factors include sex, age, Hispanic ethnicity, race, living arrangements (whether the student lives with both parents), duration of attendance at the current school, participation in school clubs, mother's education level, and mother's profession. We also control for contextual variables associated with the student's social network by including the average of friends' control variables.

Table 2 provides definitions of the variables included in the empirical analysis and presents summary statistics. The average GPA of students is 2.8, while the average GPA of their friends is 2.9. On average, students are approximately 15 years old and have attended their current school for 2.5 years. In terms of demographic composition, the sample includes 48.7% boys, 16.4% Hispanics, 64.7% whites, 16.8% Blacks, 7% Asians, and 9.5% from other racial backgrounds. Furthermore, 74.1% of the students live with both parents. Regarding maternal education, about 30.6% of students' mothers have attained a high school (HS) education level, while 16.9% have education levels beyond HS, and 41.9% have education levels below HS. In terms of mothers' occupations, 20.2% of the students' mothers work in professional occupations (such as teachers, doctors, lawyers, and executives), 20% are homemakers or do not work, and 43.5% hold other jobs.[16]

As an important feature of the proposed method is to account for whether students are isolated or not isolated, we also present summary statistics for the subsample of students who have no friends. The average GPA for this group is slightly lower than that of students who have friends. We also observe that students who have no peers are more likely to be males and are often from minority groups, such as Hispanics, Blacks, and Asians.

---

[16] Add Health provides a detailed list with more than 15 categories. We combine these occupations into four broader categories along with a missing indicator.

Table 2: Variables and summary statistics

| Variable | Definition | Own characteristics | | | | Average friend's characteristics | |
| | | All students | | Without peers | | | |
		Mean	SD	Mean	SD	Mean	SD
GPA	Average grade in mathematics, science, English or language arts, and history or social science	2.813	0.806	2.705	0.834	2.871	0.569
Gender							
(Male)	1 if male; 0 otherwise	0.487	0.500	0.582	0.493	0.459	0.318
Female	1 if female; 0 otherwise	0.513	0.500	0.418	0.493	0.541	0.318
Age	Student's age in year	15.074	1.680	15.364	1.656	15.034	1.552
Hispanic	1 if Hispanic; 0 otherwise	0.164	0.370	0.228	0.420	0.149	0.287
Race							
(White)	1 if White; 0 otherwise	0.647	0.478	0.550	0.498	0.671	0.399
Black	1 if Black; 0 otherwise	0.168	0.374	0.205	0.404	0.161	0.334
Asian	1 if Asian; 0 otherwise	0.070	0.254	0.085	0.279	0.066	0.197
Other	1 if the student's race is not in the list above; 0 otherwise	0.095	0.293	0.120	0.325	0.086	0.196
Lives with both parents	1 if the student lives with both of their parents; 0 otherwise	0.741	0.438	0.691	0.462	0.762	0.275
Year in school	Number of years spent in the current school	2.506	1.422	2.361	1.351	2.611	1.197
Member of a club	1 if the student is a member of one of the school's clubs	0.937	0.243	0.877	0.328	0.950	0.144
Mother's education							
(High)	1 if mother's education is high school; 0 otherwise	0.450	0.306	0.461	0.282	0.309	0.286
< High	1 if mother's education is less than high school; 0 otherwise	0.169	0.375	0.191	0.393	0.152	0.241
> High	1 if mother's education is greater than high school; 0 otherwise	0.484	0.419	0.493	0.376	0.446	0.330
Missing	1 if mother's education is missing; 0 otherwise	0.106	0.308	0.151	0.358	0.094	0.184
Mother's job							
Professional	1 if mother is a doctor, lawyer, scientist, teacher, executive, director, and the like; 0 otherwise	0.202	0.402	0.172	0.377	0.221	0.252
(Stay home)	1 if mother is a homemaker, retired, or does not work; 0 otherwise	0.206	0.405	0.214	0.410	0.196	0.242
Other	1 if mother's job is not in the list above; 0 otherwise	0.435	0.496	0.406	0.491	0.442	0.300
Missing	1 if mother's job is missing; 0 otherwise	0.157	0.364	0.208	0.406	0.141	0.220

This table presents the mean and the standard deviation (SD) of the variables used in the empirical analysis. The columns corresponding to "All students" show the mean and the SD in the full sample. The next two columns indicate the mean and the SD in the subsample of students having no friends. The last two columns present the contextual variables (averages within friends) in the subsample of students with friends. For the non-numeric variables, the category in parentheses is omitted from the econometric model to allow for identification.

4.2 Main Results

We consider the same specifications as in the simulation study and present our main empirical results in Table 3.[17] Estimation results for Model 4 suggest that a one-point increase in the average GPA of one's peers results in a 0.856-point increase in own GPA. This finding is aligned with the empirical literature, highlighting the importance of peers as determinants of student performance (Sacerdote, 2011; Epple and Romano, 2011). Importantly, this peer effect estimate derived from our preferred specification is 60% larger compared to the size of the coefficient in the standard models that we estimate: 0.502 in Model 1 without unobserved school heterogeneity and 0.507 in Model 2 when we allow for unobserved school heterogeneity. Note that school heterogeneity does not appear to significantly influence the endogenous peer effect parameter in the standard specification, as the estimates are similar for Models 1 and 2. As argued in Section 3.1, the classical specification leads to a biased estimate of peer effects because it does not distinguish between the common shock parameters α_s and c_s. This distinction is important, as it affects the causal interpretation of the peer effect parameters.

In Model 3, when in addition to school fixed effects we control for whether a student is isolated or not, we obtain a larger peer effects estimate of 0.751, albeit one that is still smaller than the one we obtain in Model 4. It is noteworthy that the parameter associated with the dummy variable "Has friends" in Model 3 is negative, however, this does not represent the causal effect of being a non-isolated student. This is because if we look at Equation (9), the coefficient associated with this dummy variable is $-\lambda\bar{\alpha}$, which captures the GPA shock in Equation (1). The fact that this coefficient is significant suggests the presence of shocks that directly affect GPA, irrespective of effort.

Moreover, the negative estimate indicates that $\lambda\bar{\alpha} > 0$, which is in line with the negative bias in Models 1 and 2. Indeed, as argued in Section 3.1, peer effects estimated using standard models would be biased downward if $\lambda\alpha_s > 0$. The intuition behind this bias is as follows: the standard model (Model 2) only accounts for effort shocks (c_s) and wrongly classifies direct shocks to GPA that do not influence effort as effort shocks, leading to an overestimation of their impact through a social multiplier effect. Since co-movements in students' and their peers' GPA are either peer effects or common shock effects at the school level, overestimating one type of effect reduces the second type of effect.

Note also that the weak instrument test suggests that the specifications do not suffer from a weak instrument issue. However, the Sargan-Hansen test of overidentifying restrictions suggests that in Models 1–3 not all instruments are valid, which is not the case with Model 4, indicating that the issue is addressed when we allow for additional heterogeneity. Indeed, since Models 1–3 are misspecified, the instruments used for the average friends' GPA are not exogenous. These instruments are

---

[17]Replication codes are available at https://github.com/ahoundetoungan/PeerEffectsEffort.

correlated to the omitted variables $-\lambda\alpha_s\mathbf{g}_{s,i}\mathbf{1}_{n_s}$.

We also find that several characteristics of students and contextual variables significantly influence their GPA. Female students score 0.165 grade points higher than male students. Older students tend to do worse, while students who have been in the current school for longer periods tend to do better. Regarding race and ethnicity, Black and students of other races score 0.121, and 0.026 points lower than white students, respectively, whereas Asian students score 0.194 points higher than white students. Hispanics also fare worse than non-Hispanics. Students who participate in club activities and who live with both parents score 0.138 and 0.091 points higher, respectively. Furthermore, mother's education is an important determinant of student GPA.

Finally, a number of contextual variables have significant coefficients. For example, a student's GPA increases with the mean age of their peers or when their peers are Black and Hispanic. On the other hand, the student's GPA decreases when their peers are female, Asian, participate in club activities, and when their peers' mother's job is professional.

4.3   Counterfactual analysis

We use the estimated models to simulate policy interventions, such as enhancing teacher or school quality while maintaining students' effort levels and raising students' awareness of the importance of academic achievement. Improving school quality while holding students' effort levels constant can be achieved by increasing α_s in the production equation (1). Introducing a preference shock to raise students' awareness of their academic performance involves increasing c_s. However, Models 2 and 3 are unable to isolate these shocks; instead, they confound them into a common school shock by increasing c_s (akin to an increase in $\bar{c}_s$ in Equation (5) for the standard model).

Figure 2 illustrates the implications of these shocks, represented by a one-unit increase in the fixed effects parameters. The results reveal social multiplier effects in Models 2 and 3, with the magnitude of these effects varying based on students' centrality (see Calvó-Armengol et al., 2009). For students without friends, the increase in GPA is one, indicating the absence of social multiplier effects. However, for connected students, the effects can reach 2 with Model 2 and 4 with Model 3. Note that the effects are higher in Model 3, as this model partially addresses the issue of biased peer effect estimates. The main issue regarding these results is that it is unclear what they capture, as Models 2 and 3 fail to separate GPA shocks that do not affect effort from preference shocks. By addressing this problem in Model 4, we observe that GPA shocks yield no social multiplier effects; the increase in GPA remains one for all students, irrespective of their centrality. In contrast, preference shocks can indeed lead to social multiplier effects, ranging from one for isolated students to seven for the most connected students. It is important to highlight that, in the absence of our structural model, if a

Table 3: Estimation results

	Model 1		Model 2		Model 3		Model 4	
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.502	0.025	0.507	0.028	0.751	0.041	0.856	0.044
Has friends					−2.266	0.141		
Own effects								
Female	0.183	0.006	0.176	0.006	0.165	0.006	0.165	0.006
Age	0.007	0.003	−0.015	0.003	−0.044	0.003	−0.043	0.003
Hispanic	−0.110	0.010	−0.101	0.010	−0.099	0.010	−0.091	0.010
Race								
Black	−0.144	0.012	−0.131	0.012	−0.141	0.012	−0.121	0.013
Asian	0.206	0.013	0.218	0.013	0.204	0.013	0.194	0.014
Other	−0.024	0.011	−0.026	0.011	−0.028	0.011	−0.026	0.011
Lives with both parents	0.116	0.008	0.107	0.007	0.097	0.007	0.091	0.008
Years in school	0.023	0.003	0.033	0.003	0.029	0.003	0.027	0.003
Member of a club	0.166	0.013	0.157	0.012	0.146	0.012	0.138	0.012
Mother's education								
< High	−0.076	0.009	−0.076	0.009	−0.071	0.009	−0.068	0.009
> High	0.163	0.007	0.151	0.007	0.131	0.008	0.124	0.008
Missing	0.032	0.013	0.031	0.012	0.026	0.012	0.026	0.012
Mother's job								
Professional	0.044	0.009	0.039	0.009	0.036	0.009	0.032	0.009
Other	−0.041	0.007	−0.040	0.007	−0.037	0.008	−0.037	0.008
Missing	−0.083	0.011	−0.078	0.011	−0.073	0.011	−0.070	0.011
Contextual effects								
Female	−0.138	0.012	−0.108	0.012	−0.103	0.013	−0.123	0.013
Age	−0.067	0.003	−0.073	0.004	0.023	0.005	0.024	0.006
Hispanic	0.033	0.016	0.050	0.017	0.096	0.019	0.087	0.020
Race								
Black	−0.001	0.015	−0.007	0.015	0.076	0.018	0.070	0.020
Asian	−0.051	0.019	−0.043	0.021	−0.109	0.025	−0.135	0.027
Other	−0.048	0.020	−0.046	0.020	−0.013	0.021	−0.001	0.022
Lives with both parents	−0.040	0.017	−0.040	0.016	−0.006	0.017	−0.019	0.018
Years in school	0.003	0.004	0.028	0.004	−0.011	0.005	−0.009	0.006
Member of a club	−0.175	0.029	−0.142	0.028	−0.052	0.028	−0.084	0.029
Mother's education								
< High	−0.044	0.017	−0.050	0.016	0.016	0.018	0.025	0.019
> High	0.062	0.016	0.027	0.017	−0.021	0.020	−0.032	0.021
Missing	−0.057	0.025	−0.070	0.024	−0.034	0.025	−0.031	0.026
Mother's job								
Professional	−0.058	0.018	−0.056	0.018	−0.022	0.019	−0.034	0.020
Other	−0.131	0.014	−0.105	0.014	−0.021	0.016	−0.022	0.016
Missing	−0.144	0.022	−0.116	0.021	0.003	0.024	0.008	0.024
σ_η^2					0.285		0.286	
σ_ϵ^2	0.515		0.503		0.107		0.046	
ρ					0.230		0.605	
Weak instrument F	141		202		117		120	
Sargan test prob.	0.000		0.000		0.014		0.223	
Number of schools	141		141		141		141	
Number of students	68,430		68,430		68,430		68,430	

Model 1 disregards school heterogeneity (i.e., c_s is constant across schools), while Model 2 accounts for it. Model 3 includes a single fixed effect per school with a dummy variable capturing isolated students (i.e., α_s may not be zero but is constant across schools, whereas c_s varies). Model 4 is our structural model. The columns "Coef" report the coefficient estimates, followed by their corresponding standard errors in the "Sd Err" columns.

researcher were to estimate a reduced-form specification with two types of fixed effects, as specified in Equation (9), they would then carry out policy analysis by associating the two types of shocks with changes in the fixed effects κ_s^{NI} and κ_s^I. Consequently, they would erroneously conclude that both shocks would yield social multiplier effects among students with friends. However, our structural model allows us to distinguish between the two types of shocks by clearly demonstrating how the two fixed effects are connected to the unobserved factors α_s and c_s.

To summarize, these simulations highlight that performing policy analysis using the standard model may lead to misleading conclusions. It may falsely attribute multiplier effects to shocks that do not generate them, or it may produce biased estimates of multiplier effects for shocks that indeed generate them.
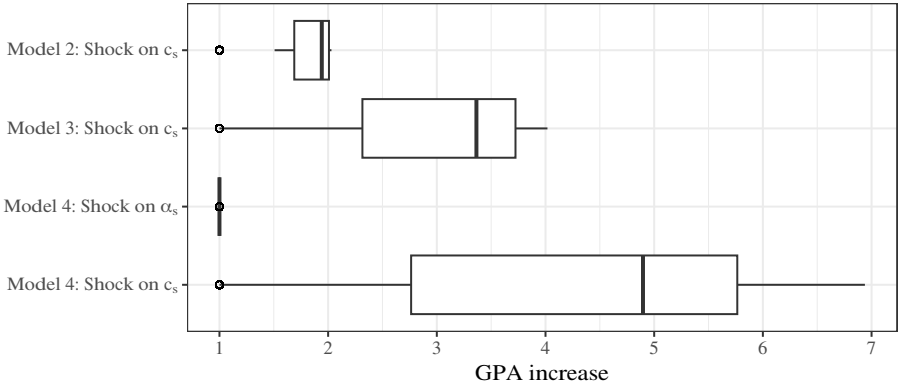

Figure 2: Effects of Shocks on the GPA

This figure presents the distribution of the increase in the GPA subsequent to a one-unit increase in α_s and $\delta^2 c_s$ for the student sample (n = 68,430). Since δ is not identified, the shock on c_s is scaled by δ^2 because the parameter c_s in our specification (9) is multiplied by δ^2. However, it is important to note that regardless of the value of δ, increasing α_s will still have the same influence on all students, whereas the implication of an increase in c_s will vary depending on centrality.

4.4 Excluding Isolated Students

We next examine the performance of the various models estimated above when we exclude isolated students from our analysis. Note that in our main Add Health sample, 22% of the students have not nominated any friends. Among them, half (11% of the full sample) are not fully isolated, meaning they are nominated as friends by others, while the other half is fully isolated, having neither nominated friends nor being nominated by others.

Excluding the 22% isolated students from the sample will result in missing values in the network dataset, because the 11% of students who are not fully isolated have been nominated as friends by others. As demonstrated by Boucher and Houndetoungan (2022), this can lead to biased estimates of peer effects. In contrast, the exclusion of the "fully isolated" students allows us to conduct a robust-

ness analysis, as it does not involve a missing network data issue. We thus define a new subsample by excluding the "fully isolated" students from our main sample, resulting in a subsample comprising 61,183 students from 139 schools.

Results are presented in Table 4. The peer effect estimate using the structural model (Model 4) is slightly larger at 0.878 after excluding fully isolated students. Similarly, the estimates using the standard linear-in-means model are also higher at 0.561 (Model 2) and 0.788 (Model 3).

This robustness analysis indicates that the bias in the estimation of peer effects when using the standard approach persists even after removing fully isolated students. This is because removal of fully isolated students still leaves in the sample some students who did not nominate any friends but were nominated by others. If we remove these students as well (Model 2′), we still find that the peer effect estimate using the standard approach is downward biased. This is likely because, as mentioned above, removal of the partially isolated students creates missing values in the network dataset.

Table 4: Estimation results after excluding isolated students

| | Excluding fully isolated | | | | | | Excluding no friends | |
| | Model 2 | | Model 3 | | Model 4 | | Model 2′ | |
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.561	0.030	0.788	0.042	0.878	0.044	0.581	0.034
Weak instrument F	160		100		105		112	
Sargan test prob.	0.000		0.095		0.493		0.000	
Number of schools	139		139		139		139	
Number of students	61,183		61,183		61,183		53,529	

Models 2, 3, and 4 are estimated using the subsample excluding fully isolated students (students who nominate no friends and who have not been nominated by others), whereas Model 2′ is estimated using the sample excluding all students who do not nominate any friends. Model 2 controls for unobserved school heterogeneity but do not include dummy variables capturing isolated students i.e., $\alpha_s = 0$ and c_s varies across schools). Model 3 includes a single fixed effect per school with a dummy variable capturing isolated students (i.e., α_s may not be zero but is constant across schools, whereas c_s varies). Model 4 is our structural model. The columns "Coef" report the coefficient estimates followed by their corresponding standard errors in the columns "Sd Err". Full results with the estimates of the coefficients associated with other variables in Table B.1.

5 Extension to Endogenous Networks

5.1 Method

The fact that certain characteristics of students that are unobserved might influence both educational outcomes and their social connections calls into question the assumption of an exogenous network. For instance, a student's IQ or level of extroversion are likely to affect both their GPA and their choice of friends. Because these student characteristics are typically not observed by the econometrician, they would not be included in $\mathbf{X}_s$ and would instead be captured by the error terms $\boldsymbol{\eta}_s$ and ε_s, giving rise to an omitted variables bias (network endogeneity).

Specifically, Equation (9) can now be written as:

$$y_{s,i} = \kappa_s^{NI}\ell_{s,i}^{NI} + \kappa_s^{I}(1 - \ell_{s,i}^{NI}) + \mathbf{g}_{s,i}\mathbf{y}_s + \mathbf{x}_{s,i}'\tilde{\boldsymbol{\beta}} + \mathbf{g}_{s,i}\mathbf{X}_s\tilde{\boldsymbol{\gamma}} + h_{s,i} + \tilde{v}_{s,i}, \tag{10}$$

where $h_{s,i}$ captures missing variables correlated with GPA that may also explain link formation in the social network. Treating the problem of network endogeneity as an omitted variables problem is a common approach in the literature (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021; Jochmans, 2023). To address this problem, we rely on a two-stage method similar to the control function approach proposed by Johnsson and Moon (2021). We first estimate the omitted factors using a network formation model. In the second stage, we add them to our model as additional explanatory variables.

We consider a network formation model with degree heterogeneity (see Graham, 2017; Dzemski, 2019; Yan et al., 2019).[18] In this model, the conditional probability of observing a link from student i to student j (that is, i declares that j is a friend) within the same school s is denoted as

$$\mathbb{P}(a_{s,ij} = 1|\ddot{\mathbf{x}}_{s,ij}, \mu_{s,i}^{out}, \mu_{s,j}^{in}) = \Phi(\ddot{\mathbf{x}}_{s,ij}'\ddot{\boldsymbol{\beta}} + \mu_{s,i}^{out} + \mu_{s,j}^{in}), \tag{11}$$

where Φ is either the normal or logistic distribution function, depending on whether the model follows a probit or logit specification; $\ddot{\mathbf{x}}_{s,ij}$ is a vector of observed dyad-specific variables, such as the distance between the characteristics of students i and j, which influence the probability of forming a friendship link; $\mu_{s,i}^{out}$ and $\mu_{s,j}^{in}$ account for unobserved heterogeneity affecting student i's likelihood of initiating friendships (outdegree) and student j's likelihood of receiving friendship nominations (indegree), respectively. For instance, a student with a high IQ may be nominated as a friend by many classmates, resulting in a large indegree $\mu_{s,i}^{in}$. However, because the network is directed, this student's outdegree can be low, if $\mu_{s,i}^{out}$ is low.

In Equation (11), there are more than $2n_s$ parameters to be estimated per school. However, Yan et al. (2019) show that the standard logit estimators of $\mu_{s,i}^{out}$ and $\mu_{s,j}^{in}$ are consistent if the network is dense.[19] In this logit model, $\mu_{s,i}^{out}$ and $\mu_{s,j}^{in}$ are treated as fixed effects, that is, they can be correlated to the observed dyad-specific variables. We refer the interested reader to Yan et al. (2019) for a formal discussion of the model, including its identification and consistent estimation. Alternatively, a Bayesian probit model based on the data augmentation technique can be used to simulate the posterior distributions of $\mu_{s,i}^{out}$ and $\mu_{s,j}^{in}$ (see Albert and Chib, 1993). However, this approach treats $\mu_{s,i}^{out}$ and $\mu_{s,j}^{in}$ as random effects.

---

[18]See De Paula (2020) for a recent review.

[19]Assuming that the network is dense requires each school's size to increase to infinity with the number of schools. Using simulations, Yan et al. (2019) also claims that the logit model performs quite well even if the network is sparse.

As in Johnsson and Moon (2021) and Houndetoungan (2022), we use a nonparametric approach to connect $h_{s,i}$ in Equation (10) to the unobserved factors $\mu_{s,i}^{out}$ and $\mu_{s,i}^{in}$. We impose that $h_{s,i} = h^{out}(\mu_{s,i}^{out}) + h^{in}(\mu_{s,i}^{in})$, where h^{out} and h^{in} are continuous functions. This specification is more flexible than the assumption that $h_{s,i}$ is a linear function of $\mu_{s,i}^{out}$ and $\mu_{s,i}^{in}$, as imposed by Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016). We approximate h^{out} and h^{in} using cubic B-splines as in generalized additive models (Hastie, 2017). The key idea behind this approach stems from the Weierstrass theorem, which states that any continuous function defined on a compact interval can be well approximated using polynomials. Specifically, we approximate $h^{out}(\mu_{s,i}^{out})$ by cubic polynomials on ten different intervals covering the range of $\mu_{s,i}^{out}$. The intervals are defined so that each comprises approximately the same share of observations. We also apply this approach to $h^{in}(\mu_{s,i}^{in})$. Given the number of intervals and the degree of the polynomials, this approach results in approximating $h_{s,i}$ by a combination of 26 variables, called bases, that are computed from the estimates of $\mu_{s,i}^{out}$ and $\mu_{s,i}^{in}$.[20]

In the second stage, we use these new 26 variables as additional explanatory variables in the GPA model. Note that our identification analysis outlined in Section 3.2 is also valid here, under the assumption that $\mu_{s,i}^{out}$ and $\mu_{s,i}^{in}$ are well identified in the first stage. In contrast, achieving asymptotic normality in this two-stage estimation process becomes more complicated and cannot be generalized without new restrictions. In general, one needs the first-stage estimator to converge as fast as possible so that its approximation error does not influence the second-stage estimator asymptotically (see OA C.4). Alternatively, a bootstrap approach can be used to approximate the asymptotic distribution of the second-stage estimator.

5.2 Empirical Results

Table 5 presents the estimation results of peer effects, accounting for network endogeneity. Full results, including coefficients associated with other variables, can be found in Appendix B. We explore two approaches for estimating the unobserved factors μ_i^{out} and μ_i^{in}: a fixed effects logit model (Models 2a, 3a, and 4a) and a random effects Bayesian model (Models 2b, 3b, and 4b). In all specifications, the new 26 regressors are globally significant, indicating the endogeneity of the network (we do not present the estimates for these new regressors).

In our preferred specifications (models 4a and 4b), the endogeneity does not significantly affect the estimate of peer effects. This result aligns with many other findings on the Add Health data arguing that the endogeneity of the network does not involve a substantial bias in the peer effects (e.g., Hsieh and Lee, 2016; Hsieh and Lin, 2017). Similar to those studies, we observe a slight decrease

---

[20]Normally, applying a cubic polynomial with ten intervals would result in 40 bases for each $\mu_{s,i}^{out}$ and $\mu_{s,i}^{in}$. However, constraints are imposed on the coefficients of many variables for the approximation to be a continuous function and to avoid a problem of multicollinearity. Given the constraints, each approximation is ultimately a combination of 13 variables. The detailed method for the construction of these bases can be found in Hastie (2017).

in the peer effect estimate (from 0.856 to 0.828). This decline occurs because of unobserved factors such as IQ that are positively correlated with GPA. For instance, an exogenous shock to IQ would simultaneously influence both students' and peers' GPAs. Accounting for the endogeneity helps disentangle true peer effects from these co-movements in GPA.

In the standard models (Models 2a and 2b), controlling for network endogeneity through the fixed effect approach significantly increases the peer effect estimate from 0.507 to 0.672. This control also mitigates the overidentification problem. The Sargan test probability increases from 0.000 to 0.039. Overidentification can occur due to the endogeneity of the network, which is likely to be the case in the standard model because the error term includes dummy variables for isolated students. By accounting for network endogeneity, the bias stemming from the omission of these variables is reduced. However, the Bayesian random effect approach still yields biased estimates. This is because this method only captures unobserved factors μ_i^{out} and μ_i^{in} that are independent of the regressions in $\mathbf{x}_{s,i}$. For example, this approach cannot control for the omission of the IQ because the latter would be corrected with $\mathbf{x}_{s,i}$. For the standard models with a dummy variable for isolated students (Model 3a, 3b), controlling for network endogeneity does not significantly influence the peer effects. Nonetheless, it effectively resolves the issue of overidentification.

Table 5: Estimation results controlling for network endogeneity

| | Model 2a | | Model 2b | | Model 3a | | Model 3B | | Model 4a | | Model 4b | |
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.672	0.036	0.478	0.029	0.729	0.042	0.717	0.042	0.828	0.044	0.826	0.044
Weak instrument F	131		190		113		113		114		114	
Sargan test prob.	0.039		0.000		0.076		0.090		0.447		0.453	
Number of schools	141		141		141		141		141		141	
Number of students	68,430		68,430		68,430		68,430		68,430		68,430	

Models 2a, 3a, and 4a use logit fixed effects estimations for $\mu_{s,i}^{out}$ and $\mu_{s,i}^{in}$ in the first stage, whereas Models 2b, 3b, and 4c consider a Bayesian probit random effects estimate. In Models 2a and 2b, unobserved school heterogeneity is controlled for, but dummy variables capturing isolated students are not included ($\alpha_s = 0$ and c_s varies across schools). Models 3a and 3b include a single fixed effect per school with a dummy variable capturing isolated students (i.e., α_s may not be zero but is constant across schools, whereas c_s varies). Models 4a and 4b are specified according to our structural model. The columns "Coef" report the coefficient estimates followed by their corresponding standard errors in the "Sd Err" columns. Full results with the estimates of the coefficients associated with other variables are presented in Appendix Table B.2.

6 Conclusion

This paper proposes a peer effect model in which students choose their level of academic effort, which in turn impacts their academic achievement (GPA). Unlike standard models used in the literature to estimate peer effects on GPA, our structural model accounts for two types of common shocks at the school level and allows for identifying peer effects on effort itself, even though effort is unobserved. We introduce common shocks that directly influence GPA, irrespective of effort levels, and common shocks affecting both students' effort and their GPA.

We show that these two types of shocks have different impacts on GPA. Shocks exerted directly on GPA without influencing academic effort do not involve a social multiplier, whereas preference shocks that affect both academic effort and GPA may involve a social multiplier effect. We also show that failure to differentiate the two types of shocks results in a biased estimate of peer effects, when there are isolated students. Practically, accounting for the difference between the shocks amounts to controlling for student heterogeneity on the basis of whether they have friends or not.

Our model leads to an econometric specification that poses identification challenges. This occurs in particular because of the presence of unobserved school heterogeneity and students with no peers in the network. We derive conditions for identification and propose a multi-stage estimation strategy that combines the GMM and QML approaches. Our approach yields a consistent estimator, and we establish asymptotic normality. We also extend the estimation strategy and examine the case of endogenous networks.

We present an empirical illustration using Add Health data. We find that increasing the average GPA of peers by one point results in a 0.856 point increase in a student's GPA. The peer effect estimate obtained using standard models is 40% lower than that obtained from our proposed approach. Controlling for network endogeneity in the standard models reduces the bias.

More generally, our framework can be used to study peer effects on activities that cannot be directly observed. An example is body mass index (BMI), which cannot be directly chosen. People need to exert effort, such as developing healthy diet habits, engaging in physical exercise, and avoiding fast food, to improve their BMI. Peer influence is more related to effort than BMI. Another example is peer effects on workers' effort. The observed outcome is generally worker's productivity, whereas peer effects stem from effort.

References

ALAN, S., C. BAYSAN, M. GUMREN, AND E. KUBILAY (2021): "Building social cohesion in ethnically mixed schools: An intervention on perspective taking," *The Quarterly Journal of Economics*, 136, 2147–2194.

ALBERT, J. H. AND S. CHIB (1993): "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669–679.

ARCIDIACONO, P., G. FOSTER, N. GOODPASTER, AND J. KINSLER (2012): "Estimating spillovers using panel data, with an application to the classroom," *Quantitative Economics*, 3, 421–470.

BALLESTER, C., A. CALVÓ-ARMENGOL, AND Y. ZENOU (2006): "Who's Who in Networks. Wanted: The Key Player," *Econometrica*, 74, 1403–1417.

BLUME, L. E., W. A. BROCK, S. N. DURLAUF, AND Y. M. IOANNIDES (2011): "Identification of social interactions," in *Handbook of social economics*, Elsevier, vol. 1, 853–964.

BOUCHER, V. AND B. FORTIN (2016): "Some Challenges in the Empirics of the Effects of Networks," *Handbook on the Economics of Networks*, 45–48.

BOUCHER, V. AND A. HOUNDETOUNGAN (2022): *Estimating peer effects using partial network data*, Centre de recherche sur les risques les enjeux économiques et les politiques . . . .

BOUCHER, V., S. TUMEN, M. VLASSOPOULOS, J. WAHBA, AND Y. ZENOU (2021): "Ethnic Mixing in Early Childhood: Evidence from a Randomized Field Experiment and a Structural Model," Tech. rep., CEPR Discussion Paper No.15528.

BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2020): "Peer effects in networks: A survey," *Annual Review of Economics*, 12, 603–629.

BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

CALVÓ-ARMENGOL, A., E. PATACCHINI, AND Y. ZENOU (2009): "Peer Effects and Social Networks in Education," *The Review of Economic Studies*, 76, 1239–1267.

CONTI, G., A. GALEOTTI, G. MUELLER, AND S. PUDNEY (2013): "Popularity," *Journal of Human Resources*, 48, 1072–1094.

CORNELISSEN, T., C. DUSTMANN, AND U. SCHÖNBERG (2017): "Peer effects in the workplace," *American Economic Review*, 107, 425–456.

DAVEZIES, L., X. D'HAULTFOEUILLE, AND D. FOUGÈRE (2009): "Identification of peer effects using group size variation," *The Econometrics Journal*, 12, 397–413.

DE GIORGI, G., M. PELLIZZARI, AND S. REDAELLI (2010): "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2, 241–275.

DE PAULA, Á. (2017): "Econometrics of network models," in *Advances in economics and econometrics: Theory and applications, eleventh world congress,* Cambridge University Press, Cambridge, 268–323.

——— (2020): "Econometric models of network formation," *Annual Review of Economics*, 12, 775–799.

DUNCAN, G. J., J. BOISJOLY, AND K. MULLAN HARRIS (2001): "Sibling, peer, neighbor, and school-mate correlations as indicators of the importance of context for adolescent development," *Demography*, 38, 437–447.

DURLAUF, S. N. AND Y. M. IOANNIDES (2010): "Social interactions," *Annu. Rev. Econ.*, 2, 451–478.

DZEMSKI, A. (2019): "An empirical model of dyadic link formation in a network with unobserved heterogeneity," *Review of Economics and Statistics*, 101, 763–776.

EPPLE, D. AND R. E. ROMANO (2011): "Peer effects in education: A survey of the theory and evidence," in *Handbook of Social Economics*, Elsevier, vol. 1, 1053–1163.

FORTIN, B. AND M. YAZBECK (2015): "Peer effects, fast food consumption and adolescent weight gain," *Journal of Health Economics*, 42, 125–138.

FRUEHWIRTH, J. C. (2013): "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap," *Quantitative Economics*, 4, 85–124.

——— (2014): "Can achievement peer effect estimates inform policy? a view from inside the black box," *Review of Economics and Statistics*, 96, 514–523.

GOLDSMITH-PINKHAM, P. AND G. W. IMBENS (2013): "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 31, 253–264.

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

——— (2017): "An econometric model of network formation with degree heterogeneity," *Econometrica*, 85, 1033–1063.

GRIFFITH, A. (2022): "Name your friends, but only five? the importance of censoring in peer effects estimates using social network data," *Journal of Labor Economics*, 40, 779–805.

HASTIE, T. J. (2017): "Generalized additive models," in *Statistical models in S*, Routledge, 249–307.

HAUSMAN, J. A. (1978): "Specification tests in econometrics," *Econometrica: Journal of the econometric society*, 1251–1271.

HONG, S. C. AND J. LEE (2017): "Who is sitting next to you? Peer effects inside the classroom," *Quantitative Economics*, 8, 239–275.

HORN, R. A. AND C. R. JOHNSON (2012): *Matrix analysis*, Cambridge University Press.

HOUNDETOUNGAN, E. A. (2022): "Count data models with social interactions under rational expectations," Tech. rep., SSNR.

HSIEH, C.-S. AND L. F. LEE (2016): "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, 31, 301–319.

HSIEH, C.-S. AND X. LIN (2017): "Gender and racial peer effects with endogenous network formation," *Regional Science and Urban Economics*, 67, 135–147.

JACKSON, M. O. AND Y. ZENOU (2015): "Games on networks," in *Handbook of game theory with economic applications*, Elsevier, vol. 4, 95–163.

JOCHMANS, K. (2023): "Peer effects and endogenous social interactions," *Journal of Econometrics*, 235, 1203–1214.

JOHNSSON, I. AND H. R. MOON (2021): "Estimation of peer effects in endogenous social networks: control function approach," *Review of Economics and Statistics*, 103, 328–345.

KELEJIAN, H. H. AND I. R. PRUCHA (1998): "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *The Journal of Real Estate Finance and Economics*, 17, 99–121.

KLINE, B. AND E. TAMER (2020): "Econometric analysis of models with social interactions," in *The Econometric Analysis of Network Data*, Elsevier, 149–181.

LANCASTER, T. (2000): "The incidental parameter problem since 1948," *Journal of Econometrics*, 95, 391–413.

LEE, L.-F. (2004): "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

——— (2007): "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 140, 333–374.

LEE, L.-F., X. LIU, AND X. LIN (2010): "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, 13, 145–176.

LIN, X. (2010): "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables," *Journal of Labor Economics*, 28, 825–860.

MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, 531–542.

MAS, A. AND E. MORETTI (2009): "Peers at work," *American Economic Review*, 99, 112–145.

NEWEY, W. K. AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

ROSE, C. D. (2017): "Identification of peer effects through social networks using variance restrictions," *The Econometrics Journal*, 20, S47–S60.

SACERDOTE, B. (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*, Elsevier, vol. 3, 249–277.

THEIL, H. (1957): "Specification errors and the estimation of economic relationships," *Revue de l'Institut International de Statistique*, 25, 41–51.

WHITE, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica: journal of the Econometric Society*, 817–838.

YAN, T., B. JIANG, S. E. FIENBERG, AND C. LENG (2019): "Statistical inference in a directed network model with covariates," *Journal of the American Statistical Association*, 114, 857–868.

A Appendix: Proofs

A.1 Uniqueness of the Nash Equilibrium

By replacing the GPA with its expression given by Equation (1), we obtain a new payoff function $\hat{u}_{s,i}(e_{s,i}, \mathbf{e}_{s,-i})$ that does not depend on the GPA. The new payoff function is

$$\hat{u}_{s,i}(e_{s,i}, \mathbf{e}_{s,-i}) = (c_s + \mathbf{x}'_{s,i}\boldsymbol{\beta} + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma} + \varepsilon_{s,i})(\alpha_s + \mathbf{x}'_{s,i}\boldsymbol{\theta} + \delta e_{s,i} + \eta_{s,i}) - \frac{e_{s,i}^2}{2} + \lambda e_{s,i}\mathbf{g}_{s,i}\mathbf{e}_s. \quad \text{(A.1)}$$

The first-order condition of the maximization of $\hat{u}_{s,i}(e_{s,i}, \mathbf{e}_{s,-i})$ with respect to the effort $e_{s,i}$ gives

$$e_{s,i} = \delta c_s + \lambda \mathbf{g}_{s,i}\mathbf{e}_s + \delta \mathbf{x}'_{s,i}\boldsymbol{\beta} + \delta \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma} + \delta \varepsilon_{s,i}. \quad \text{(A.2)}$$

If we write Equation (A.2) at the school level, we get the best response functions of all students:

$$\mathbf{e}_s = \delta c_s \mathbf{1}_{n_s} + \lambda \mathbf{G}_s \mathbf{e}_s + \delta \mathbf{X}_s \boldsymbol{\beta} + \delta \mathbf{G}_s \mathbf{X}_s \boldsymbol{\gamma} + \delta \boldsymbol{\varepsilon}_s, \quad \text{(A.3)}$$

where $\mathbf{1}_{n_s}$ is an n_s-vector of ones and $\boldsymbol{\varepsilon}_s = (\varepsilon_{s,1}, \ldots, \varepsilon_{s,n_s})'$. Equation (A.3) is a system of n_s linear equations in the effort. This system has a unique solution if $|\mathbf{I}_{n_s} - \lambda \mathbf{G}_s| \neq 0$, where $\mathbf{I}_{n_s}$ is the $n_s \times n_s$ identity matrix. The condition $|\mathbf{I}_{n_s} - \lambda \mathbf{G}_s| \neq 0$ is equivalent to saying that 1 is not an eigenvalue for $\lambda \mathbf{G}_s$. As $\mathbf{G}_s$ is a row-normalized matrix, the eigenvalues of $\lambda \mathbf{G}_s$ are in the closed interval $[-|\lambda|, |\lambda|]$.[21] Thus, if $|\lambda| < 1$, then $|\mathbf{I}_{n_s} - \lambda \mathbf{G}_s| \neq 0$ and the solution of Equation (A.3) is

$$\mathbf{e}_s = (\mathbf{I}_{n_s} - \lambda \mathbf{G}_s)^{-1}(\delta c_s \mathbf{1}_{n_s} + \delta \mathbf{X}_s \boldsymbol{\beta} + \delta \mathbf{G}_s \mathbf{X}_s \boldsymbol{\gamma} + \delta \boldsymbol{\varepsilon}_s). \quad \text{(A.4)}$$

As a result, the game described by the payoff function (A.1) has a unique NE given by (A.4).

A.2 Reduced form equation of the GPA

Let $\boldsymbol{\eta}_s = (\eta_{s,1}, \ldots, \eta_{s,n_s})'$ be the vector of the idiosyncratic error terms in Equation (1). Let also $\mathbf{y}_s = (y_{s,1}, \ldots, y_{s,n_s})'$ be the GPAs' vector. From Equation (1), we have $e_{s,i} = (y_{s,i} - \alpha_s - \mathbf{x}'_{s,i}\boldsymbol{\theta} - \eta_{s,i})/\delta$. By replacing this expression in Equation (A.2), we get

$$\frac{y_{s,i} - \alpha_s - \mathbf{x}'_{s,i}\boldsymbol{\theta} - \eta_{s,i}}{\delta} = \frac{\lambda \mathbf{g}_{s,i}(\mathbf{y}_s - \alpha_s \mathbf{1}_{n_s} - \mathbf{X}_s\boldsymbol{\theta} - \boldsymbol{\eta}_s)}{\delta} + \delta(c_s + \mathbf{x}'_{s,i}\boldsymbol{\beta} + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma} + \varepsilon_{s,i}),$$

$$y_{s,i} = \kappa_{s,i} + \lambda \mathbf{g}_{s,i}\mathbf{y}_s + \mathbf{x}'_{s,i}\tilde{\boldsymbol{\beta}} + \mathbf{g}_{s,i}\mathbf{X}_s\tilde{\boldsymbol{\gamma}} + (\boldsymbol{\omega}_{s,i} - \lambda \mathbf{g}_{s,i})\boldsymbol{\eta}_s + \delta^2 \varepsilon_{s,i},$$

---

[21] This is a direct implication of the Gershgorin circle theorem (Horn and Johnson, 2012).

where $\kappa_{s,i} = \delta^2 c_s + (1 - \lambda \mathbf{g}_{s,i} \mathbf{1}_{n_s}) \alpha_s$, $\tilde{\boldsymbol{\beta}} = \delta^2 \boldsymbol{\beta} + \boldsymbol{\theta}$, $\tilde{\boldsymbol{\gamma}} = \delta^2 \boldsymbol{\gamma} - \lambda \boldsymbol{\theta}$, and $\boldsymbol{\omega}_{s,i}$ is a row-vector of dimension n_s in which all the elements are equal to zero except the i-th element, which is one.

A.3 Proof of Proposition 3.1

We show that the reflection problem is addressed under Conditions (i) and (ii) of Assumption 3.2. Assume that $\mathbb{E}(\mathbf{J}_s \mathbf{G}_s \mathbf{y}_s | \mathbf{G}_s, \mathbf{X}_s)$ is perfectly collinear with $\mathbf{J}_s \mathbf{X}_s$ and $\mathbf{J}_s \mathbf{G}_s \mathbf{X}_s$. For any $i \in \mathcal{V}_s^{NI}$, we have $\mathbb{E}(\mathbf{g}_{s,i} \mathbf{y}_s - y_s^{NI} | \mathbf{G}_s, \mathbf{X}_s) = (\mathbf{x}'_{s,i} - \mathbf{x}_s^{NI\prime}) \dot{\boldsymbol{\beta}} + (\mathbf{g}_{s,i} \mathbf{X}'_s - \bar{\mathbf{x}}_s^{NI\prime}) \dot{\boldsymbol{\gamma}}$, where y_s^{NI}, $\mathbf{x}_s^{NI}$, and $\bar{\mathbf{x}}_s^{NI}$ are the respectively averages of $\mathbf{g}_{s,i} \mathbf{y}_s$, $\mathbf{x}_{s,i}$, and $(\mathbf{g}_{s,i} \mathbf{X}_s)'$ within $\mathcal{V}_s^{NI}$, and $\dot{\boldsymbol{\beta}}$, $\dot{\boldsymbol{\gamma}}$ are unknown parameters. The variables $\mathbf{g}_{s,i} \mathbf{y}_s$, $\mathbf{x}_{s,i}$, and $\mathbf{g}_{s,i} \mathbf{X}_s$ are taking in deviation with respect to their average in $\mathcal{V}_s^{NI}$ because of the matrix $\mathbf{J}_s$ that multiplies the terms of Equation (7). Let us take the previous equation in difference between two students i_1, j from $\mathcal{V}_s^{NI}$, where j is i_1's friend. This implies:

$$\bar{y}_{s,i_1}^e - \bar{y}_{s,j}^e = (\mathbf{x}'_{s,i_1} - \mathbf{x}'_{s,j}) \dot{\boldsymbol{\beta}} + (\bar{\mathbf{x}}'_{s,i_1} - \bar{\mathbf{x}}'_{s,j}) \dot{\boldsymbol{\gamma}}, \tag{A.5}$$

where $\bar{y}_{s,i}^e = \mathbb{E}(\mathbf{g}_{s,i} \mathbf{y}_s | \mathbf{G}_s, \mathbf{X}_s)$ and $\bar{\mathbf{x}}'_{s,i} = \mathbf{g}_{s,i} \mathbf{X}_s$ for all i. Assume an increase in $\mathbf{x}_l$ for all l that is separated from i_1 by a link of distance three, *ceteris paribus*. Such an l exists by Condition (ii) of Assumption 3.2. As j is i_1's friend, l cannot be j's friend, otherwise, it would be possible to find a link of distance two from l to i_1. Thus, an increase in any $\mathbf{x}_l$ has no influence on $\mathbf{x}'_{s,i_1}$, $\mathbf{x}'_{s,j}$, $\bar{\mathbf{x}}'_{s,i_1}$, and $\bar{\mathbf{x}}'_{s,j}$. Therefore, the right-hand side (RHS (A.5) would not be influenced and we would have

$$\Delta^l \bar{y}_{s,i_1}^e - \Delta^l \bar{y}_{s,j}^e = 0, \tag{A.6}$$

where the operator Δ^l measures the variation after the increase in $\mathbf{x}_l$.

Using a proof by contradiction, we will now show that the condition $\Delta^l \bar{y}_{s,i_1}^e = \Delta^l \bar{y}_{s,j}^e$ for all j who is i_1's friend is not possible. By applying the operator Δ^l to every term of Equation (6), we have $\Delta^l \mathbf{y}_s = \lambda \mathbf{G}_s (\Delta^l \mathbf{y}_s) + (\Delta^l \mathbf{X}_s) \tilde{\boldsymbol{\beta}} + \mathbf{G}_s (\Delta^l \mathbf{X}_s) \tilde{\boldsymbol{\gamma}}$. This implies that $\Delta^l \mathbf{y}_s = (\mathbf{I} - \lambda \mathbf{G}_s)^{-1} \big((\Delta^l \mathbf{X}_s) \tilde{\boldsymbol{\beta}} + \mathbf{G}_s (\Delta^l \mathbf{X}_s) \tilde{\boldsymbol{\gamma}} \big)$. As $(\mathbf{I} - \lambda \mathbf{G}_s)^{-1} = \sum_{k=0}^{\infty} \lambda^k \mathbf{G}_s^k$, we can also write

$$\Delta^l \mathbf{y}_s = (\Delta^l \mathbf{X}_s) \tilde{\boldsymbol{\beta}} + \sum_{k=0}^{\infty} \lambda^k \mathbf{G}_s^{k+1} (\Delta^l \mathbf{X}_s)(\lambda \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}}). \tag{A.7}$$

Equation (A.7) implies the GPA is influenced by the contextual variables if and only if $\lambda \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}} \neq 0$. By premultiplying (A.7) by $\mathbf{g}_{s,i_1}$ and taking the expectation conditional on $\mathbf{G}_s$ and $\mathbf{X}_s$, we have

$$\Delta^l \bar{y}_{s,i_1}^e = \mathbf{g}_{s,i_1} \sum_{k=1}^{\infty} \lambda^k \mathbf{G}_s^{k+1} (\Delta^l \mathbf{X}_s)(\lambda \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}}). \tag{A.8}$$

Indeed $\mathbf{g}_{s,i_1}(\Delta^l \mathbf{X}_s) = 0$ and $\mathbf{g}_{s,i_1} \mathbf{G}_s (\Delta^l \mathbf{X}_s) = 0$ (since l is separated from i_1 by a link of distance

three, l is not i_1's friend, nor i_1's friend's friend).

By premultiplying each term of Equation (A.7) by $\lambda \mathbf{G}_s$, we obtain $\sum_{k=1}^{\infty} \lambda^k \mathbf{G}_s^{k+1} (\Delta^l \mathbf{X}_s)(\lambda \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}}) = \lambda \mathbf{G}_s \Delta^l \mathbf{y}_s - \lambda \mathbf{G}_s (\Delta^l \mathbf{X}_s) \tilde{\boldsymbol{\beta}}$. By replacing the previous equation in (A.8), we get

$$\Delta^l \bar{y}_{s,i_1}^e = \lambda \mathbf{g}_{s,i_1} \mathbf{G}_s \Delta^l \mathbf{y}_s, \tag{A.9}$$

because $\mathbf{g}_{s,i_1} \mathbf{G}_s (\Delta^l \mathbf{X}_s) = 0$. As $\mathbf{G}_s \Delta^l \mathbf{y}_s = (\Delta^l \bar{y}_{s,1}^e, \ldots, \Delta^l \bar{y}_{s,n_s}^e)'$, the term $\mathbf{g}_{s,i_1} \mathbf{G}_s \Delta^l \mathbf{y}_s$ in the RHS of Equation (A.9) is the average of $\Delta^l \bar{y}_{s,j}^e$ among students j who are i_1's friends. If Equation (A.6) holds true, that is, if $\Delta^l \bar{y}_{s,j}^e = \Delta^l \bar{y}_{s,i_1}^e$ for any j who is i_1's friend, this would mean that $\mathbf{g}_{s,i_1} \mathbf{G}_s \Delta^l \mathbf{y}_s = \Delta^l \bar{y}_{s,i_1}^e$ and Equation (A.9) would imply that $\Delta^l \bar{y}_{s,i_1}^e = \lambda \Delta^l \bar{y}_{s,i_1}^e$. This is where the contradiction would come from. Indeed, the previous equation is not compatible with Equation (A.6) since $\lambda \neq 1$ by Assumption 2.1, and $\Delta^l \bar{y}_{s,i_1}^e \neq 0$ because $\lambda \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}} \neq 0$ (see Equation (A.7)). As a result, the model does not suffer from the reflection problem.

Let $\tilde{\mathbf{X}}_s = \mathbf{J}_s[\mathbf{X}_s, \ \mathbf{G}_s\mathbf{X}_s]$, $\mathbf{R}_s = [\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s, \ \tilde{\mathbf{X}}_s]$, $\mathbf{Z}_s = [\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s, \ \tilde{\mathbf{X}}_s]$, $\mathbf{R}'\mathbf{Z} = \sum_{s=1}^{S} \mathbf{R}_s'\mathbf{Z}_s$, and $\mathbf{Z}'\mathbf{Z} = \sum_{s=1}^{S} \mathbf{Z}_s'\mathbf{Z}_s$. We set the following identification conditions.

Assumption A.1. *The matrices $\mathbf{R}'\mathbf{Z}/n$ and $\mathbf{Z}'\mathbf{Z}/n$ converge to full rank matrices as S grows to infinity. Moreover, $\sum_{s=1}^{S} \mathbf{Z}_s'((\mathbf{I}_{n_s} - \lambda\mathbf{G}_s)\boldsymbol{\eta}_s + +\delta^2 \boldsymbol{\varepsilon}_s)/n = o_p(1)$.*

The first half of Assumption A.1 suggests that the columns of design matrix $\mathbf{R} = [\mathbf{R}_1', \ \ldots, \mathbf{R}_S']'$ and those of the instrument matrix $\mathbf{Z} = [\mathbf{Z}_1', \ \ldots, \mathbf{Z}_S']'$ are linearly independent for large S. The second condition of the assumption comes from the exogeneity of $\mathbf{X}_s$ and $\mathbf{G}_s$ with respect to $\boldsymbol{\eta}_s$ and ε_s. Under Assumptions 2.1, 3.1, 3.2, and A.1, the design matrix of Equation (7) is full rank for large n, and the identification of $\boldsymbol{\psi}$ follows.

A.4   Proof of Proposition 3.2

In this section, we use different notations for the parameters and their true values; that is their values in the data-generating process. We denote by $\boldsymbol{\psi}_0$, $\sigma_{0\eta}$, $\sigma_{0\epsilon}$, and ρ_0 the true values of $\boldsymbol{\psi}$, σ_η, σ_ϵ, and ρ, respectively.

However, the log-likelihood cannot be written directly because the transformation we apply to eliminate the fixed effects κ_s^{NI} and κ_s^I makes the covariance matrix $\mathbb{E}(\boldsymbol{v}_s\boldsymbol{v}_s'|\mathbf{G}_s)$ singular (for example, we have $\mathbf{1}_{n_s}'\boldsymbol{v}_s = 0$). In other words, we cannot invert the covariance matrix $\mathbb{E}(\boldsymbol{v}_s\boldsymbol{v}_s'|\mathbf{G}_s)$, which is a necessary task to write the log-likelihood. To address this issue, we use a similar approach to that of Lee et al. (2010). Let $[\mathbf{F}_s, \ \boldsymbol{\ell}_s^I/\sqrt{n_s^I}, \ \boldsymbol{\ell}_s^{NI}/\sqrt{n_s^{NI}}]$ be the orthonormal matrix of $\mathbf{J}_s$, where the columns in $\mathbf{F}_s$ are eigenvectors of $\mathbf{J}_s$ corresponding to the eigenvalue one.[22] To ease the notational burden, we

---

[22]The eigenvalues of $\mathbf{J}_s$ are zero and one. The multiplicity of the eigenvalue one is $n_r - 2$ if the school s has students in both $\mathcal{V}_s^I$ and $\mathcal{V}_s^{NI}$, and $n_r - 1$ if s has students in either $\mathcal{V}_s^I$ or $\mathcal{V}_s^{NI}$.

assume the school s has students in both $\mathcal{V}_s^I$ and $\mathcal{V}_s^{NI}$. We have $\mathbf{F}_s\mathbf{F}_s' = \mathbf{J}_s$ and $\mathbf{F}_s'\mathbf{F}_s = \mathbf{I}_{n_s-2}$. As $\mathbf{F}_s$ does not depend on unknown parameters, maximizing the log-likelihood of $\hat{\boldsymbol{v}}_s$ is equivalent to maximizing that of $\mathbf{F}_s'\hat{\boldsymbol{v}}_s$.[23] The log-likelihood of $\mathbf{F}_s'\hat{\boldsymbol{v}}_s$ is given by

$$\hat{L}(\sigma_\eta^2, \sigma_\epsilon^2, \rho) = -\sum_{s=1}^S \frac{n_s-2}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\sum_{s=1}^S \log|\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)| - \sum_{s=1}^S \frac{1}{2\sigma_\epsilon^2}\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\hat{\boldsymbol{v}}_s, \quad \text{(A.10)}$$

where $\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho) = \mathbf{I}_{n_s-2} + \tau^2\mathbf{F}_s'\mathbf{W}_s\mathbf{W}_s'\mathbf{F}_s + \rho\tau\mathbf{F}_s'(\mathbf{W}_s + \mathbf{W}_s')\mathbf{F}_s$, $\mathbf{W}_s = \mathbf{I}_{n_s} - \hat{\lambda}\mathbf{G}_s$, and $\tau = \sigma_\eta/\sigma_\epsilon$. The first-order conditions of the maximization of (A.10) imply that σ_ϵ^2 can be substituted with $\tilde{\sigma}_\epsilon^2(\tau, \rho) = \sum_{s=1}^S \frac{\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\hat{\boldsymbol{v}}_s}{n-2S}$. This leads to a simpler concentrated log-likelihood that does not depend on σ_ϵ^2 and is easier to maximize. We also define the following log-likelihood by replacing $\hat{\boldsymbol{v}}_s$ and $\hat{\lambda}$ in $\hat{L}(\sigma_\eta^2, \sigma_\epsilon^2, \rho)$ with their true value:

$$L(\sigma_\eta^2, \sigma_\epsilon^2, \rho) = -\sum_{s=1}^S \frac{n_s-2}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\sum_{s=1}^S \log|\boldsymbol{\Omega}_s(\lambda_0, \tau, \rho)| - \sum_{s=1}^S \frac{1}{2\sigma_\epsilon^2}\boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s,$$

where $\boldsymbol{v}_s = \mathbf{J}_s(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \delta^2\mathbf{J}_s\boldsymbol{\varepsilon}_s$. Let $\pi_{\min}(.)$ be the smallest eigenvalue and $\pi_{\max}(.)$ be the largest eigenvalue. The operator $\|.\|_2$ applied to a matrix is the operator norm induced by the ℓ^2-norm. We also denote by $\boldsymbol{\Theta}$ the space of $(\sigma_\eta^2,\ \sigma_\epsilon^2,\ \rho)$.

The proof is done in several steps.

Step 1

We show that $\frac{1}{n-2S}(\hat{L}(\sigma_\eta^2, \sigma_\epsilon^2, \rho) - L(\sigma_\eta^2, \sigma_\epsilon^2, \rho))$ converges in probability to zero uniformly in $\boldsymbol{\Theta}$. This proof would imply that we can focus on $L(\sigma_\eta^2, \sigma_\epsilon^2, \rho)$ for the identification and consistency instead of $\hat{L}(\sigma_\eta^2, \sigma_\epsilon^2, \rho)$. To do so, we set the following assumptions.

Assumption A.2. *(i) $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^3$ and (ii) $\lim_{S\to\infty}\pi_{\min}(\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)) > 0$ for any s.*

Assumption A.3. *(i) $\mathbb{E}(\eta_{s,i}^4|\mathbf{G}_s, \mathbf{X}_s)$, $\mathbb{E}(\varepsilon_{s,i}^4|\mathbf{G}_s, \mathbf{X}_s)$, and $\mathbb{E}(\eta_{s,i}^2\varepsilon_{s,i}^2|\mathbf{G}_s, \mathbf{X}_s)$ exist; (ii) $\max_{s,i}|\eta_{s,i}| = O_p(1)$, $\max_{s,i}|\varepsilon_{s,i}| = O_p(1)$, and $\max_{s,i}\|\mathbf{x}_{s,i}\|_2 = O_p(1)$.*

Condition (i) of Assumption A.2 is required in many econometric models. It allows for generalizing pointwise convergences to uniform convergences. Condition (ii) of Assumption A.2 generalizes the nonsingularity of the matrix $\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)$ to large samples (when S grows to infinity). Assumption A.3 sets further conditions regarding the distribution of $(\eta_{s,i},\ \varepsilon_{s,i})'$ and ensures that $\mathbf{x}_{s,i}$ and the i-th component of $\boldsymbol{v}_s$ are bounded.

Because $\mathbf{G}_s$ is row-normalized and bounded in column sum (Assumption 3.3), then for all τ and ρ, $\tilde{\boldsymbol{\Omega}}_s(\hat{\lambda}, \tau, \rho) := \mathbf{I}_{n_s} + \tau^2\mathbf{W}_s\mathbf{W}_s' + \rho\tau(\mathbf{W}_s + \mathbf{W}_s')$ is also absolutely bounded in both row and

---

[23]Unlike for $\boldsymbol{v}_s$, the covariance matrix of $\mathbf{F}_s'\boldsymbol{v}_s$ is not singular. Indeed, $\mathbb{E}(\mathbf{F}_s'\boldsymbol{v}_s\boldsymbol{v}_s'\mathbf{F}_s|\mathbf{G}_s) = \mathbf{F}_s'\mathbf{J}_s\mathbb{E}(\tilde{\boldsymbol{v}}_s\tilde{\boldsymbol{v}}_s'|\mathbf{G}_s)\mathbf{J}_s\mathbf{F}_s$, where $\tilde{\boldsymbol{v}}_s = (\mathbf{I}_{n_s} - \lambda\mathbf{G}_s)\boldsymbol{\eta}_s + \delta^2\boldsymbol{\varepsilon}_s$. For any $n_s - 2$ vector $\mathbf{u}_s \neq 0$, $\mathbf{u}_s'\mathbf{F}_s'\mathbf{J}_s\mathbb{E}(\tilde{\boldsymbol{v}}_s\tilde{\boldsymbol{v}}_s'|\mathbf{G}_s)\mathbf{J}_s\mathbf{F}_s\mathbf{u}_s > 0$ because $\mathbf{J}_s\mathbf{F}_s\mathbf{u}_s = \mathbf{F}_s\mathbf{u}_s \neq 0$ and $\mathbb{E}(\tilde{\boldsymbol{v}}_s\tilde{\boldsymbol{v}}_s'|\mathbf{G}_s)$ is positive definite (except for special cases where $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$ are collinear).

column sums, and $\pi_{\max}(\tilde{\boldsymbol{\Omega}}_s(\hat{\lambda}, \tau, \rho)) < \infty$.[24] Moreover, as $\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho) = \mathbf{F}_s'\tilde{\boldsymbol{\Omega}}_s(\hat{\lambda}, \tau, \rho)\mathbf{F}_s$, we have $\pi_{\max}(\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)) \leqslant \pi_{\max}(\tilde{\boldsymbol{\Omega}}_s(\hat{\lambda}, \tau, \rho)) < \infty$. Thus, $\frac{1}{n-2S}\sum_{s=1}^{S}\log|\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)| < \infty$ for all $\sigma_\eta^2, \sigma_\epsilon^2$, and ρ. As a result, $\frac{1}{n-2S}\sum_{s=1}^{S}\log|\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)| - \frac{1}{n-2S}\sum_{s=1}^{S}\log|\boldsymbol{\Omega}_s(\lambda_0, \tau, \rho)| = o_p(1)$ (because the determinant is continuous).

Besides, $\hat{\boldsymbol{v}}_s = \boldsymbol{v}_s + \Delta\hat{\boldsymbol{v}}_s$, where $\Delta\hat{\boldsymbol{v}}_s = \mathbf{R}_s(\boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}})$. As $\max_{s,i}\|\mathbf{x}_{s,i}\|_2 = O_p(1)$ and $(\boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}}) = O_p(n^{-1/2})$, then each component of $\Delta\hat{\boldsymbol{v}}_s$ is $O_p(n^{-1/2})$ and $\|\Delta\hat{\boldsymbol{v}}_s\|_2 = O_p((n_s/n)^{1/2})$. On the other hand, as $\max_{s,i}|\eta_{s,i}| = O_p(1)$ and $\max_{s,i}|\varepsilon_{s,i}| = O_p(1)$, we have $\|\boldsymbol{v}_s\|_2 = O_p(n_s^{1/2})$. We also have

$$\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\hat{\boldsymbol{v}}_s = \boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s + 2\Delta\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s + \Delta\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\Delta\hat{\boldsymbol{v}}_s.$$

The submultiplicativity property of the operator norm implies that $\sum_{s=1}^{S}|\Delta\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s| = O_p(n^{1/2})$ and $\sum_{s=1}^{S}|\Delta\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\Delta\hat{\boldsymbol{v}}_s| = O_p(1)$ because $\|\mathbf{F}_s\|_2 = 1$ and $\|\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\|_2 = O_p(1)$ (Assumption A.2). Thus, $\frac{1}{n-2S}(\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\hat{\boldsymbol{v}}_s - \boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s) = o_p(1)$. We also have $\frac{1}{n-2S}(\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\hat{\boldsymbol{v}}_s - \boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s) = o_p(1)$ because $\boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s$ is a continuous function of $\hat{\lambda}$.

As a result, $\frac{1}{n}(\hat{L}(\sigma_\eta^2, \sigma_\epsilon^2, \rho) - L(\sigma_\eta^2, \sigma_\epsilon^2, \rho)) = o_p(1)$. The convergence is uniform because the log-likelihoods can be expressed as a polynomial function in $(\sigma_\eta^2, \sigma_\epsilon^2, \rho)$. We can now focus on the log-likelihood $L(\sigma_\eta^2, \sigma_\epsilon^2, \rho)$ for the identification and the consistency of $(\sigma_{0\epsilon}^2, \tau_0, \rho_0)$, where $\tau_0 = \sigma_{0\eta}/\sigma_{0\epsilon}$.

Step 2

The first-order conditions (foc) of the maximization of $L(\sigma_\eta^2, \sigma_\epsilon^2, \rho)$ imply that σ_ϵ^2 can be replaced with $\hat{\sigma}_\epsilon^2(\tau, \rho) = \sum_{s=1}^{S}\frac{\boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s}{n-2S}$. This leads to a concentrated log-likelihood given by $L_c(\tau, \rho) = -\frac{n-2S}{2}\hat{\sigma}_\epsilon^2(\tau, \rho) - \frac{1}{2}\sum_{s=1}^{S}\log|\boldsymbol{\Omega}_s(\lambda_0, \tau, \rho)| - \frac{n-2S}{2}$ that does not depend on σ_ϵ^2. Let $L^*(\sigma_\eta^2, \sigma_\epsilon^2, \rho) = \mathbb{E}\left(L(\sigma_\eta^2, \sigma_\epsilon^2, \rho)|\mathbf{G}_1, \ldots, \mathbf{G}_S\right)$. As for $L(\sigma_\eta^2, \sigma_\epsilon^2, \rho)$, we can also replace σ_ϵ^2 with $\mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau, \rho)|\mathbf{G}_1, \ldots, \mathbf{G}_S\right)$.

$$\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau, \rho)|\mathbf{G}_1, \ldots, \mathbf{G}_S) = \frac{\sigma_{0\varepsilon}^2}{n-2S}\sum_{s=1}^{S}\mathbb{E}(\mathrm{Tr}(\boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s)|\mathbf{G}_1, \ldots, \mathbf{G}_S),$$
$$\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau, \rho)|\mathbf{G}_1, \ldots, \mathbf{G}_S) = \frac{\sigma_{0\varepsilon}^2}{n-2S}\sum_{s=1}^{S}\mathbb{E}(\mathrm{Tr}(\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\mathbf{F}_s'\boldsymbol{v}_s\boldsymbol{v}_s'\mathbf{F}_s)|\mathbf{G}_1, \ldots, \mathbf{G}_S),$$
$$\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau, \rho)|\mathbf{G}_1, \ldots, \mathbf{G}_S) = \frac{\sigma_{0\varepsilon}^2}{n-2S}\sum_{s=1}^{S}\mathrm{Tr}(\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\mathbb{E}(\mathbf{F}_s'\boldsymbol{v}_s\boldsymbol{v}_s'\mathbf{F}_s|\mathbf{G}_1, \ldots, \mathbf{G}_S)),$$
$$\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau, \rho)|\mathbf{G}_1, \ldots, \mathbf{G}_S) = \frac{\sigma_{0\varepsilon}^2}{n-2S}\sum_{s=1}^{S}\mathrm{Tr}(\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\boldsymbol{\Omega}_{0,s}), \tag{A.11}$$

where $\boldsymbol{\Omega}_{0,s} = \boldsymbol{\Omega}_s(\lambda_0, \tau_0, \rho_0)$.

We obtain the concentrated log-likelihood $L_c^*(\tau, \rho) = -\frac{n-2S}{2}\tilde{\sigma}_\epsilon^{2*}(\tau, \rho) - \frac{1}{2}\sum_{s=1}^{S}\log|\boldsymbol{\Omega}_s(\lambda_0, \tau, \rho)| - \frac{n-2S}{2}$, where $\tilde{\sigma}_\epsilon^{2*}(\tau, \rho) = \frac{\sigma_{0\varepsilon}^2}{n-2S}\sum_{s=1}^{S}\mathrm{Tr}\left(\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\boldsymbol{\Omega}_{0,s}\right)$. We show that $\frac{1}{n}(L_c(\tau, \rho) - L_c^*(\tau, \rho))$ converges to zero uniformly.

Although $\mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau, \rho)|\mathbf{G}_1, \ldots, \mathbf{G}_S\right) = \tilde{\sigma}_\epsilon^{2*}(\tau, \rho)$, this does not implies that $\mathrm{plim}\,\hat{\sigma}_\epsilon^2(\tau, \rho) = \tilde{\sigma}_\epsilon^{2*}(\tau, \rho)$.

---

[24]We state and show in OA C.1 basic properties used throughout the paper. See properties P.5 and P.6.

We also need to show that the variance of $\hat{\sigma}_\epsilon^2(\tau, \rho)$ converges to zero as S grows to infinity. This is especially important in our framework because the components of $\boldsymbol{v}_s$ are connected through the network and also because n_s is not necessarily bounded. This is why we impose that fourth-order moments of $\eta_{s,i}$ and $\varepsilon_{s,i}$ exist in Assumption A.3. In fact, the variance of $\hat{\sigma}_\epsilon^2(\tau, \rho)$ involves up to the fourth power of the components of $\boldsymbol{v}_s$. We provide the proof in OA C.2. The uniform convergence of $\frac{1}{n}(L_c(\tau, \rho) - L_c^*(\tau, \rho))$ to zero directly follows.

Step 3

To establish the identification of the consistency, we need to show $L_c^*(\tau, \rho)$ is maximized at a single point which is (τ_0, ρ_0) (see Newey and McFadden, 1994). We set the following assumption.

Assumption A.4. *If* $(\tau, \rho) \neq (\tau_0, \rho_0)$, *then* $\dfrac{\lim\limits_{n\to\infty} (\sum_{s=1}^S (\log|\tilde{\sigma}_\epsilon^{2*}(\tau, \rho)\boldsymbol{\Omega}_s(\lambda_0, \tau, \rho)| - \log|\sigma_{0\varepsilon}^2\boldsymbol{\Omega}_{0,s}|)}{n} \neq 0.$

The intuition of Assumption A.4 is as follows. After replacing σ_ϵ^2 with $\tilde{\sigma}_\epsilon^2(\tau, \rho)$ in Equation (A.10), the variable part of the concentrated log-likelihood is proportional to $\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^2(\tau, \rho)\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)|$, which is asymptotically equivalent to $\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau, \rho)\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)|$. Assumption A.4 implies that the value of $\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau, \rho)\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)|$ at (τ_0, ρ_0) cannot be reached at another point as S grows to infinity. This assumption adapts Assumption 9 of Lee (2004) or Assumption 5.1 of Lee et al. (2010) to our framework.[25]

Let $(\boldsymbol{e}_s)_s$ be a process normally distributed of zero mean and covariance matrix $\sigma_{0\epsilon}^2\boldsymbol{\Omega}_s(\tau_0, \rho_0)$. Let $L^0(\sigma_\eta^2, \sigma_\epsilon^2, \rho) = -\sum_{s=1}^S \frac{n_s-2}{2}\sigma_\epsilon^2 - \frac{1}{2}\sum_{s=1}^S \log|\boldsymbol{\Omega}_s(\lambda_0, \tau, \rho)| - \sum_{s=1}^S \frac{1}{2\sigma_\epsilon^2}\boldsymbol{e}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0, \tau, \rho)\mathbf{F}_s'\boldsymbol{e}_s$. By Jensen's inequality, we have $\mathbb{E}(L^0(\sigma_\eta^2, \sigma_\epsilon^2, \rho) - L^0(\sigma_{0\eta}^2, \sigma_{0\epsilon}^2, \rho_0)|\mathbf{G}_1, \ldots, \mathbf{G}_S) \leqslant 0$. This suggests that (τ_0, ρ_0) is a global maximizer of $\text{plim} \frac{1}{n}L_c^*(\tau, \rho)$. The uniqueness of the maximizer is guaranteed by Assumption A.4. If (τ_0, ρ_0) is not the unique maximizer, then there would be another $(\tau_+, \rho_+) \in \boldsymbol{\Theta}$ such that $\text{plim} \frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau_0, \rho_0)\boldsymbol{\Omega}_s(\lambda_0, \tau_0, \rho_0)| = \text{plim} \frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau_+, \rho_+)\boldsymbol{\Omega}_s(\lambda_0, \tau_+, \rho_+)|$. This would violate Assumption A.4. As a result, (τ_0, ρ_0) is globally identified and $(\hat{\tau}, \hat{\rho})$ is a consistent estimator of (τ_0, ρ_0). The consistency of $\hat{\sigma}_\eta^2$ comes from Equation (A.11). We have $\text{plim} \hat{\sigma}_\epsilon^2(\hat{\tau}, \hat{\rho}) = \mathbb{E}(\hat{\sigma}_\epsilon^2(\tau_0, \rho_0)|\mathbf{G}_1, \ldots, \mathbf{G}_S) = \sigma_{0\epsilon}^2$; thus, $\text{plim} \hat{\sigma}_\eta^2 = \tau_0^2\sigma_{0\epsilon}^2 = \sigma_{0\eta}^2$.

B Supplementary Results on the Application

Table B.1 presents detailed estimation results when we exclude isolated students. Table B.2 presents detailed estimation results controlling for network endogeneity.

---

[25]Although we cannot connect Assumption A.4 to the fundamental elements of the model, we can explain why the covariance matrix of $\boldsymbol{v}_s$ conditionally on $\mathbf{G}_s$ captures much nonlinearity to allow identifying $(\sigma_{0\epsilon}^2, \tau_0, \rho_0)$. As shown in online Appendix (OA) C.3, a crucial requirement for identification is that $\mathbf{J}_s, \mathbf{J}_s(\mathbf{G}_s + \mathbf{G}_s')\mathbf{J}_s$, and $\mathbf{J}_s\mathbf{G}_s\mathbf{G}_s'\mathbf{J}_s$ are linearly independent. This holds under the following two conditions: (1) there are four students who have friends in a certain school, those students are not directly linked and only two of them have common friends; (2) there are four students who have friends in a certain school and only two of them are linked. We present an example of a common network structure under which the conditions are verified (see Figure C.1).

Table B.1: Detailed estimation results after excluding isolated students

| | Model 2 | | Model 3 | | Model 4 | | Model 2$'$ | |
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.561	0.030	0.788	0.042	0.878	0.044	0.581	0.034
Has friends			−2.449	0.145				
Own effects								
Female	0.182	0.006	0.165	0.007	0.165	0.007	0.186	0.007
Age	−0.008	0.004	−0.047	0.004	−0.045	0.004	−0.009	0.004
Hispanic	−0.096	0.011	−0.094	0.011	−0.086	0.011	−0.092	0.012
Race								
Black	−0.113	0.013	−0.124	0.013	−0.102	0.015	−0.076	0.016
Asian	0.199	0.014	0.183	0.015	0.173	0.015	0.175	0.015
Other	−0.030	0.011	−0.032	0.011	−0.029	0.012	−0.023	0.012
Lives with both parents	0.098	0.008	0.088	0.008	0.083	0.008	0.098	0.008
Years in school	0.032	0.003	0.025	0.003	0.023	0.003	0.027	0.003
Member of a club	0.169	0.013	0.158	0.013	0.150	0.013	0.182	0.015
Mother's education								
< High	−0.072	0.009	−0.066	0.009	−0.062	0.009	−0.072	0.010
> High	0.146	0.008	0.125	0.008	0.118	0.008	0.132	0.008
Missing	0.017	0.013	0.012	0.013	0.013	0.013	0.007	0.014
Mother's job								
Professional	0.040	0.009	0.037	0.010	0.034	0.010	0.029	0.010
Other	−0.035	0.008	−0.032	0.008	−0.031	0.008	−0.041	0.008
Missing	−0.070	0.012	−0.064	0.012	−0.061	0.012	−0.078	0.013
Contextual effects								
Female	−0.122	0.012	−0.111	0.013	−0.127	0.013	−0.137	0.012
Age	−0.082	0.004	0.029	0.005	0.028	0.006	−0.075	0.005
Hispanic	0.049	0.017	0.094	0.019	0.086	0.021	0.023	0.017
Race								
Black	−0.010	0.017	0.067	0.019	0.055	0.021	−0.040	0.019
Asian	−0.051	0.022	−0.111	0.026	−0.129	0.028	−0.052	0.023
Other	−0.040	0.020	−0.008	0.021	−0.001	0.022	−0.028	0.019
Lives with both parents	−0.047	0.016	−0.011	0.017	−0.021	0.018	−0.023	0.016
Years in school	0.032	0.004	−0.010	0.005	−0.008	0.006	0.039	0.005
Member of a club	−0.160	0.028	−0.065	0.028	−0.091	0.029	−0.180	0.029
Mother's education								
< High	−0.043	0.016	0.020	0.018	0.026	0.019	−0.022	0.016
> High	0.014	0.017	−0.026	0.020	−0.036	0.021	0.022	0.016
Missing	−0.068	0.024	−0.033	0.026	−0.031	0.026	−0.046	0.023
Mother's job								
Professional	−0.062	0.018	−0.026	0.019	−0.036	0.020	−0.037	0.017
Other	−0.103	0.014	−0.021	0.016	−0.021	0.016	−0.070	0.014
Missing	−0.110	0.021	0.004	0.024	0.010	0.024	−0.089	0.020
σ^2_η			0.296		0.292			
σ^2_ϵ	0.493		0.099		0.047		0.480	
ρ			0.168		0.485			
Weak instrument F	160		100		105		112	
Sargan test prob.	0.000		0.095		0.493		0.000	
Number of schools	139		139		139		139	
Number of students	61,183		61,183		61,183		53,529	

Models 2, 3, and 4 are estimated using the subsample excluding fully isolated students (students who nominate no friends and who have not been nominated by others), whereas Model 2$'$ is estimated using the sample excluding any isolated friends (students who nominate no friends). Models 2 and 2$'$ control for unobserved school heterogeneity but do not include dummy variables capturing isolated students (i.e., $\alpha_s = 0$ and c_s varies across schools). Model 3 includes a single fixed effect per school with a dummy variable capturing isolated students (i.e., α_s may not be zero but is constant across schools, whereas c_s varies). Model 4 is our structural model. The columns "Coef" report the coefficient estimates followed by their corresponding standard errors in the columns "Sd Err".

Table B.2: Detailed estimation results controlling for network endogeneity

	Model 2a		Model 2b		Model 3a		Model 3b		Model 4a		Model 4b	
	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err	Coef	Sd Err
Peer Effects	0.672	0.036	0.478	0.029	0.729	0.042	0.717	0.042	0.828	0.044	0.826	0.044
Has friends					−2.164	0.164	−2.250	0.141				
Own effects												
Female	0.173	0.006	0.178	0.006	0.171	0.006	0.167	0.006	0.169	0.006	0.167	0.006
Age	−0.032	0.003	−0.016	0.003	−0.044	0.003	−0.045	0.003	−0.043	0.003	−0.044	0.003
Hispanic	−0.099	0.010	−0.101	0.010	−0.101	0.010	−0.100	0.010	−0.092	0.010	−0.091	0.010
Race												
Black	−0.123	0.012	−0.120	0.012	−0.119	0.012	−0.128	0.012	−0.107	0.014	−0.109	0.013
Asian	0.210	0.013	0.217	0.013	0.201	0.013	0.203	0.013	0.194	0.014	0.194	0.014
Other	−0.029	0.011	−0.033	0.011	−0.034	0.011	−0.035	0.011	−0.031	0.011	−0.033	0.011
Lives with both parents	0.097	0.007	0.105	0.007	0.095	0.007	0.095	0.007	0.090	0.007	0.090	0.007
Years in school	0.029	0.003	0.031	0.003	0.026	0.003	0.027	0.003	0.024	0.003	0.024	0.003
Member of a club	0.151	0.013	0.167	0.012	0.143	0.013	0.160	0.012	0.149	0.014	0.150	0.012
Mother's education												
< High	−0.068	0.009	−0.072	0.009	−0.067	0.009	−0.068	0.009	−0.064	0.009	−0.065	0.009
> High	0.142	0.008	0.156	0.007	0.138	0.008	0.138	0.008	0.130	0.008	0.129	0.008
Missing	0.028	0.012	0.030	0.012	0.026	0.012	0.025	0.012	0.027	0.012	0.025	0.012
Mother's job												
Professional	0.035	0.009	0.036	0.009	0.034	0.009	0.034	0.009	0.030	0.009	0.030	0.009
Other	−0.040	0.008	−0.044	0.007	−0.040	0.008	−0.041	0.008	−0.039	0.008	−0.040	0.008
Missing	−0.075	0.011	−0.081	0.011	−0.075	0.011	−0.076	0.011	−0.071	0.011	−0.073	0.011
Contextual effects												
Female	−0.108	0.012	−0.101	0.012	−0.099	0.013	−0.096	0.013	−0.118	0.013	−0.117	0.013
Age	−0.015	0.004	−0.072	0.004	0.023	0.005	0.023	0.005	0.024	0.006	0.025	0.006
Hispanic	0.078	0.017	0.044	0.017	0.092	0.018	0.094	0.019	0.081	0.020	0.081	0.020
Race												
Black	0.048	0.017	−0.004	0.015	0.077	0.018	0.077	0.017	0.076	0.020	0.069	0.020
Asian	−0.087	0.023	−0.033	0.021	−0.101	0.025	−0.095	0.024	−0.127	0.027	−0.126	0.027
Other	−0.026	0.020	−0.046	0.020	−0.013	0.021	−0.012	0.021	−0.002	0.022	−0.002	0.022
Lives with both parents	−0.027	0.016	−0.034	0.016	−0.004	0.017	−0.002	0.017	−0.015	0.018	−0.014	0.018
Years in school	0.003	0.004	0.029	0.004	−0.010	0.005	−0.010	0.005	−0.008	0.006	−0.008	0.006
Member of a club	−0.110	0.027	−0.140	0.028	−0.053	0.028	−0.053	0.028	−0.084	0.029	−0.084	0.029
Mother's education												
< High	−0.008	0.017	−0.049	0.016	0.017	0.018	0.016	0.018	0.025	0.019	0.024	0.019
> High	−0.012	0.018	0.033	0.017	−0.017	0.020	−0.013	0.020	−0.027	0.021	−0.026	0.021
Missing	−0.049	0.024	−0.064	0.024	−0.029	0.025	−0.029	0.025	−0.027	0.026	−0.027	0.026
Mother's job												
Professional	−0.044	0.018	−0.057	0.018	−0.024	0.019	−0.024	0.019	−0.035	0.019	−0.036	0.019
Other	−0.059	0.014	−0.109	0.014	−0.026	0.016	−0.026	0.016	−0.026	0.016	−0.026	0.016
Missing	−0.047	0.022	−0.117	0.021	0.000	0.023	−0.001	0.023	0.006	0.024	0.005	0.024
σ_{η}^2					0.282		0.283		0.281		0.282	
σ_{ϵ}^2	0.510		0.500		0.116		0.124		0.054		0.056	
ρ					0.205		0.173		0.547		0.525	
Weak instrument F	131		190		113		113		114		114	
Sargan test prob.	0.039		0.000		0.076		0.090		0.447		0.453	
Number of schools	141		141		141		141		141		141	
Number of students	68,430		68,430		68,430		68,430		68,430		68,430	

Models 2a, 3a, and 4a use logit fixed effect estimations for $\mu_{s,i}^{out}$ and $\mu_{s,i}^{in}$ in the first stage, whereas Models 2b, 3b, and 4b consider the Bayesian probit random estimate. Models 2a and 2b control for unobserved school heterogeneity but do not include dummy variables capturing isolated students ($\alpha_s = 0$ and c_s varies across schools). Models 3a and 3b include a single fixed effect per school with a dummy variable capturing isolated students (i.e., α_s may not be zero but is constant across schools, whereas c_s varies). Models 4a and 4b are specified as our structural model. The columns "Coef" report the coefficient estimates followed by their corresponding standard errors in the columns "Sd Err".

C   Online Appendix

Supplementary material related to this paper can be found online at https://ahoundetoungan. com/files/Papers/PEEffort_OA.pdf