Identifying Peer Effects on Student Academic Effort

Aristide Houndetoungan* and Cristelle Kouame†

September 17, 2023

Abstract

Peer influences on students' academic effort are often studied using GPA as a proxy variable for the effort when the latter is unobserved. We present an alternative method that does not require this approximation. Our identification strategy distinguishes unobserved shocks exerted on GPA without influencing the effort from unobserved students' preference shocks. We show that our estimate may be significantly different from the classical estimate (where the effort is approximated) if the network includes isolated students. Applying our approach to Add Health data shows that the peer effect estimate using a proxy for the effort is 1.6 times lower.

Keywords: Social networks, Peer effects, Academic achievement, Nonparametric methods

JEL classification: C14, C31, J24

*Thema, Cy Cergy Paris Université, Corresponding author, Email: aristide.houndetoungan@cyu.fr

†World Bank Group, Email: ckouame1@worldbank.org

We are grateful to Vincent Boucher and Michael Vlassopoulos for their comments on this manuscript. This research uses data from Add Health, a program directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is given to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available on the Add Health website (`www.cpc.unc.edu/addhealth`). No direct support was received from Grant P01-HD31921 for this research. Replication codes can be found at `https://github.com/ahoundetoungan/PeerEffectsEffort`.

1 Introduction

Over the last two decades, there has been a growing interest in the impact of peers on educational outcomes (Sacerdote, 2011). Peer effects on students' academic effort may involve a social multiplier effect; therefore, knowing whether students are influenced by their friends and the size of this influence is crucial for evaluating policies aimed at inducing changes in academic achievement (Manski, 1993). However, because the effort is unobserved in general, most studies use the grade point average (GPA) as a proxy variable. Such an approximation disregards that the GPA is not solely influenced by the effort but also by many factors, including unobserved student and school characteristics. It is essential to obtain a valid estimate of peer effects originating from academic efforts, and not from GPA, as students cannot directly control their GPA, but rather through their effort levels. To illustrate, a majority of microeconomic models exploring the impact of peer interactions on academic achievements are based on effort (e.g., see Calvó-Armengol et al., 2009; Fruehwirth, 2013). GPA is typically incorporated into empirical models only when the effort is unobserved.

In this paper, we develop a structural model with social interactions, where students decide on their academic effort, taking into account that of their peers (friends). We introduce a complementary between students' and peers' efforts. We also present a trade-off between GPA, which increases student payoff, and effort, which results in disutility (Boucher and Fortin, 2016). Without using a proxy for student academic effort, we show that the model allows for identifying peer effects on academic effort and we provide an interpretation of the effects in terms of GPA. We also show that the size of the effect may differ from that estimated using the GPA as a proxy. Our identification strategy consists of controlling for two types of unobserved GPA shocks to disentangle peer effects on academic effort from other common effects captured by the GPA. First, we 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, we account for common shocks affecting students' preferences, such as sensitization to caring more about academic achievement. These shocks influence both academic effort and GPA.

We show that approximating student effort with the GPA may result in a biased estimate of the peer effects when some students in the network do not have friends. We derive from our structural model a reduced-form equation for GPA that differs from the classical linear-in-means peer effect specification. We demonstrate that students who have friends are not affected by both types of GPA shocks in the same way as those who do not have friends. Shocks that directly influence GPA, irrespective of the effort level, do not involve a social multiplier effect, whereas preference shocks that affect effort and GPA may imply a social multiplier effect within students with friends. Standard approaches using GPA as a proxy do not disentangle the two types of shock, and therefore, lead to biased estimates of peer effects when the network includes students who have no friends.

We extend the estimation strategy to the case of endogenous networks and present an empirical analysis using the National Longitudinal Study of Adolescent Health (Add Health) data.[1] We find 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. We also find that the network is endogenous. Controlling for endogeneity reduces the bias of the standard linear-in-means model.

The network endogeneity occurs because we do not observe some student characteristics, such as intellectual quotient (IQ), that can influence students' likelihood to form links with others and their GPA. We control for those unobserved characteristics using a two-stage estimation approach. Our method is nonparametric as in Johnsson and Moon (2021). We impose no parametric restrictions regarding the way the unobserved characteristics affect GPA. Our approach is similar to generalized additive models (GAM), widely employed in the nonparametric regression literature (see Hastie, 2017).

Our model also controls for unobserved school heterogeneity. In particular, the two types of unobserved GPA shocks are introduced at the school level as common shocks. This feature of the model renders the well-known *reflection problem* more complex.[2] In the case of the standard linear-in-means model, Bramoullé et al. (2009) find straightforward conditions under which the reflection problem is fixed. Their conditions involve the network structure and can be readily tested in practice. We extend their identification analysis to our framework under similar conditions. Our main condition is that the network has at least two students separated by a link of distance three, which is slightly stronger than the identification condition of Bramoullé et al. (2009).

Our model can also be used 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., Cornelissen et al., 2017). The observed outcome is generally worker's productivity, whereas peer effects take source in the effort.

Related Literature

This paper contributes to the extensive literature on social interaction models. Manski (1993) proposes a model in which an agent's action is a linear function of the average action of their group. Ballester et al. (2006) analyze a noncooperative game with linear-quadratic utilities, in which each player decides

[1] Add Health data set comprises 22% of students without friends, including 11% who are not fully isolated, in the sense that they are friends of others.

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

how much effort they exert. In a similar game, Calvó-Armengol et al. (2009) use GPA as a proxy for the exerted effort. We contribute to this literature by proposing a structural model in which students choose an effort level taking into account their peers' efforts. We disentangle unobserved common shocks that directly impact GPA without affecting the effort from unobserved common shocks that impact both academic effort and GPA.

Our paper is closest to Fruehwirth (2013) who also assumes that peer effects take source in academic effort. Our study differs from her framework in two ways. First, we allow students to have no friends, and we show that this setting implies a different econometric model. In our application using the Add Health data, there are 22% of students without friends, including 11% who are not fully isolated because they are friends of others. We find that the peer effect estimate obtained using a standard linear-in-means model is 1.6 times lower than that obtained through our approach. Second, as friends are self-nominated in the Add Health data, we control for network endogeneity, and we find a slight decrease in the peer effect estimate.

Our paper makes a valuable contribution to the econometric literature on peer effects (De Paula, 2017; Kline and Tamer, 2020). A key challenge in this field is the reflection problem (Manski, 1993). Many studies propose to fix this problem using group size variation, whereas others impose conditions on the network structure (Davezies et al., 2009; Lee, 2007; Bramoullé et al., 2009; Rose, 2017). We study the reflection problem in a setting where the GPA is influenced by various types of common shocks at the school level and where students are allowed to have no peers. Our analysis is similar to that of Houndetoungan (2022) for nonlinear models, although here we account for unobserved heterogeneity across schools. As in Bramoullé et al. (2009), our main identification condition involves the network structure and can be readily tested in practice.

The paper also contributes to the econometric literature on peer effects by addressing the network endogeneity issue. We allow for unobserved attributes to influence students' likelihood to form links with others and their GPA (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021). Whereas many studies impose a strong parametric restriction between unobserved attributes and the peer effect model, we use a nonparametric approach to connect the 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, even those with nonlinear specifications. Specifically, we employ tensor products of cubic B-splines, which are used in GAM, to establish a smooth function between the unobserved attributes and GPA (Hastie, 2017).

Finally, the paper contributes to the large suite of empirical literature on peer effects on education outcome (e.g., Sacerdote, 2001; Zimmerman, 2003; Hanushek et al., 2003; Zabel, 2008; Calvó-Armengol et al., 2009; Carrell et al., 2009; Lin, 2010; Sacerdote, 2011; Black et al., 2013; Burke and Sass, 2013; Hatami et al., 2015; Marotta, 2017; Dickerson et al., 2018; Fruehwirth and Gagete-Miranda, 2019;

Gagete-Miranda, 2020; Norris, 2020; Boucher et al., 2022). In our empirical application using the Add Health data set, we control for several sources of common school-level shocks by allowing peer influence to stem from academic effort instead of GPA. This setting leads to a different econometric model and allow to identify peer effects on students' academic effort. Our findings suggest that increasing the average GPA of peers by one point leads to a 0.856 point increase in students' GPA, controlling for network endogeneity. In contrast, the peer effect estimate obtained using the standard linear-in-means model by approximating effort by GPA is 1.6 times lower.

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 concludes this paper.

2 The Structural Model

In this section, we present a structural model in which students decide on their effort level and obtain a GPA. Both effort and GPA yield a utility that characterizes students' preferences; the effort results in a disutility, whereas the GPA increases the student's utility (see Fruehwirth, 2013; Boucher and Fortin, 2016).[3] We consider S independent schools and denote by $n_s \geqslant 3$ the number of students in the s-th, $s \in \{1, \dots, S\}$. Let n be the total number of students, that is, $n = \sum_{s=1}^{S} n_s$. Throughout, we use the subscript $s,\ i$ denotes the i-th student in the school s. Let $y_{s,i}$ be the GPA and $\mathbf{x}_{s,i}$ be a K-vector of observable student characteristics. 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 a $i's$ 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 only interact with peers from their school and have no peers in other schools (see Calvó-Armengol et al., 2009). Finally, 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 i.

Student i chooses their effort $e_{s,i}$, which in turn implies their GPA. We assume that the effort is observed by peers (for example, each student knows how much effort their friends put into their academic success). Student GPA is linked to their 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), the relationship is given by the production equation

[3] The model can also be used to study peer effects on other outcomes that are not directly observed. An example is the exerted effort such as developing healthy diet habits or engaging in physical exercise to improve BMI. It is easier to observe the BMI than the effort. Another example is the worker's effort, which leads to observed productivity.

[4] Whether or not the random term $\eta_{s,i}$ is observed by i or by the peers does not matter in the game.

$$y_{s,i} = \alpha_{0,s} + \delta_0 e_{s,i} + \mathbf{x}'_{s,i}\boldsymbol{\theta}_0 + \eta_{s,i}, \tag{1}$$

where $\delta_0 > 0$ and $\alpha_{0,s}$, $\boldsymbol{\theta}_0$ are unknown parameters. The positive sign of δ_0 has support in the education literature and suggests that an increase in effort would increase GPA (see e.g., Plant et al., 2005; Brint and Cantwell, 2010). The parameter $\alpha_{0,s}$ captures unobserved school heterogeneity in GPA as fixed effects. We are aware that the use of a linear production function is restrictive. Although the microfoundation can be generalized to a nonlinear or nonparametric production function (e.g., see Fruehwirth, 2013), the resulting econometric model may be complex, especially when the model controls for unobserved school heterogeneity. The effort exerted and the GPA obtained provide students with a benefit that we describe using a payoff function. As in Boucher and Fortin (2016) and Calvó-Armengol et al. (2009), we assume a linear-quadratic payoff function given by

$$u_{s,i}(e_{s,i}, \mathbf{e}_{s,-i}, y_{s,i}) = \underbrace{(c_{0,s} + \mathbf{x}'_{s,i}\boldsymbol{\beta}_0 + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0 + \varepsilon_{s,i})y_{s,i} - \frac{e^2_{s,i}}{2}}_{\text{private sub-payoff}} + \underbrace{\lambda_0 e_{s,i}\mathbf{g}_{s,i}\mathbf{e}_s}_{\text{social sub-payoff}}, \tag{2}$$

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 peers' effort, and $c_{0,s}$, $\boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_0$ are unknown parameters. The parameter λ_0 captures the endogenous peer effects. The payoff function (2) is separable into two components: a private sub-payoff and a social sub-payoff. The term $(c_{0,s} + \mathbf{x}'_{s,i}\boldsymbol{\beta}_0 + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0 + \varepsilon_{s,i})$ 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 $\mathbf{g}_{s,i}\mathbf{X}_s$ called *contextual variables* (see Manski, 1993). The benefit also accounts for school unobserved heterogeneity through the parameter $c_{0,s}$. The second component of the private sub-payoff, $e^2_{s,i}/2$ is the cost of exerting an effort to achieve a specific GPA. The social sub-payoff $\lambda_0 e_{s,i}\mathbf{g}_{s,i}\mathbf{e}_s$ 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 $\lambda_0 \neq 0$. When $\lambda_0 > 0$, the payoff function (2) implies complementary between students' and peers' efforts, whereas $\lambda_0 < 0$ suggests a substitute in the efforts.

The parameters $\alpha_{0,s}$ and $c_{0,s}$ control for unobserved shocks at the school level. In terms of policy implications, they are conceptually different, and this needs to be discussed. First, $\alpha_{0,s}$ captures unobserved shocks on GPA without affecting student effort. This includes shocks on teaching quality and school management, which would influence GPA regardless of student effort.[5] Besides, the parameter $c_{0,s}$ captures shocks on student preferences, especially on the marginal payoff. For instance, sensitizing

[5]This analysis holds true because the GPA is unbounded. An increase in $\alpha_{0,s}$ necessarily results in a higher GPA, and thus a higher payoff, regardless of the effort level. If we consider a framework in which the GPA is bounded, an increase in $\alpha_{0,s}$ may decrease the effort for the students with a high GPA close to the upper bound (e.g., see Fruehwirth, 2013). In any case, however, an increase in $\alpha_{0,s}$ does not have the same implication as an increase in $c_{0,s}$ and this is what matters in our framework.

students to be more aware of their academic achievement could influence the marginal payoff. Such a shock can be captured by $c_{0,s}$ and would influence effort and GPA. We show that $\alpha_{0,s}$ and $c_{0,s}$ do not impact GPA in the same way (see Section 3.1).

By replacing $y_{s,i}$ of Equation (2) with its expression in Equation (1), we obtain a payoff function that does not depend on GPA (see Appendix A.1). This new payoff function characterizes a static game with complete information, in which students simultaneously choose their effort as to maximize their payoff. The students' best response function can be expressed as

$$e_{s,i} = \delta_0 c_{0,s} + \lambda_0 \mathbf{g}_{s,i}\mathbf{e}_s + \delta_0 \mathbf{x}'_{s,i}\boldsymbol{\beta}_0 + \delta_0 \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0 + \delta_0 \varepsilon_{s,i}. \tag{3}$$

In Equation (3), students' levels of effort are expressed as a function of the average effort of 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_0 represents the impact of peers on a student's effort level. A positive value for λ_0 indicates that a student's effort level increases if their peers put in more effort. This is similar to the effect of coworkers on an individual's productivity in a workplace (Cornelissen et al., 2017). Furthermore, Equation (3) shows that the optimal effort level (and thus the resulting GPA) is influenced by shocks at the school level (on $c_{0,s}$) that affect student preferences. However, shocks that are directly exerted on GPA through the parameter $\alpha_{0,s}$ do not impact the effort.

The best response function (3) in a matrix form can be expressed as $\mathbf{e}_s = \delta_0 c_{0,s}\mathbf{1}_{n_s} + \lambda_0 \mathbf{G}_s\mathbf{e}_s + \delta_0 \mathbf{X}_s\boldsymbol{\beta}_0 + \delta_0 \mathbf{G}_s\mathbf{X}_s\boldsymbol{\gamma}_0 + \delta_0\boldsymbol{\varepsilon}_s$, where $\boldsymbol{\varepsilon}_s = (\varepsilon_{s,1}, \dots, \varepsilon_{s,n_s})'$ and $\mathbf{1}_{n_s}$ is an n_s-vector of ones. A solution of this equation in $\mathbf{e}_s$ is the game's Nash equilibrium (NE). As $\mathbf{G}_s$ is row-normalized, the NE is unique if the following restriction holds (see Appendix A.1).

Assumption 2.1. $|\lambda_0| < 1$.

The condition $|\lambda_0| < 1$ means that students do not increase (in absolute value) their effort greater than the increase in the effort of their peers. Put differently, when the average effort of a student's friends increases by one, the increase/decrease in the student's effort is less than one.

3 The Econometric Model and Identification Strategy

If the effort were observed, we could estimate the peer effect parameter from Equation (3) using the general method of moments (GMM) approach. As we do not observe the effort, the equation cannot be estimated directly. We derive from this equation an estimable equation based on GPA. We show that this equation is econometrically different from the effort equation (3), if we proxy the effort by the GPA. We also present an identification strategy to identify λ_0 in the effort equation.

3.1 Reduced-Form Equation of GPA

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

$$y_{s,i} = \kappa_{s,i} + \lambda_0 \mathbf{g}_{s,i}\mathbf{y}_s + \mathbf{x}'_{s,i}\tilde{\boldsymbol{\beta}}_0 + \mathbf{g}_{s,i}\mathbf{X}_s\tilde{\boldsymbol{\gamma}}_0 + (\boldsymbol{\omega}_{s,i} - \lambda_0\mathbf{g}_{s,i})\boldsymbol{\eta}_s + \delta_0^2\varepsilon_{s,i}, \tag{4}$$

where $\kappa_{s,i} = \delta_0^2 c_{0,s} + (1-\lambda_0\mathbf{g}_{s,i}\mathbf{1}_{n_s})\alpha_{0,s}$, $\tilde{\boldsymbol{\beta}}_0 = \delta_0^2\boldsymbol{\beta}_0 + \boldsymbol{\theta}_0$, $\tilde{\boldsymbol{\gamma}}_0 = \delta_0^2\boldsymbol{\gamma}_0 - \lambda_0\boldsymbol{\theta}_0$, 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. If the effort were approximated by the GPA in Equation (3), the resulting reduced-form equation of GPA would be $y_{s,i} = \delta_0 c_{0,s} + \lambda_0\mathbf{g}_{s,i}\mathbf{y}_s + \delta_0\mathbf{x}'_{s,i}\boldsymbol{\beta}_0 + \delta_0\mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0 + \tilde{\varepsilon}_{s,i}$, where $\tilde{\varepsilon}_{s,i}$ is the approximation error.We will refer to this specification as the *standard* (*classical*) *model*.

Let $\hat{\mathcal{V}}_s$ be the subsample of students of the school s who have friends and $\bar{\mathcal{V}}_s$ be the subsample of students of the school s who have no friends. As $\mathbf{G}_s$ is row-normalized, we have $\mathbf{g}_{s,i}\mathbf{1}_{n_s} = 1$ if $i \in \hat{\mathcal{V}}_s$ and $\mathbf{g}_{s,i}\mathbf{1}_{n_s} = 0$ otherwise. Thus, $\kappa_{s,i} = \bar{\kappa}_{0,s}$ if $i \in \bar{\mathcal{V}}_s$ and $\kappa_{s,i} = \hat{\kappa}_{0,s}$ if $i \in \hat{\mathcal{V}}_s$, where $\bar{\kappa}_{0,s} = \delta_0^2 c_{0,s} + \alpha_{0,s}$ and $\hat{\kappa}_{0,s} = \delta_0^2 c_{0,s} + (1-\lambda_0)\alpha_{0,s}$. If there are no isolated students in the network, $\kappa_{s,i}$ would be a simple school fixed effects and our framework would be equivalent to the classical model.[6] For the rest of this paper, we assume a general context where some students may be isolated. It follows that the difference between the standard model and our framework is that the standard model has a single intercept term per school. In Equation (4), $\kappa_{s,i}$ controls for unobserved school-level heterogeneity depending on whether the student i has friends or not.[7]

To account for the unobserved factor $\kappa_{s,i}$ in Equation (4), the above discussion suggests that we need to control for school fixed effects as well as *student's social status* (a dummy variable indicating whether the student has friends or not) at the school level. The school fixed effects require including S school dummy variables as explanatory variables and controlling for student's social status at the school level requires additional S dummy variables. Each of the latter dummy variables is associated with one school and takes one if the student has friends. This specificity comes from the shocks $\alpha_{0,s}$ and $c_{0,s}$, that is, the fact that we distinguish between shocks exerted on GPA without influencing the effort and preference shocks at the school level. To understand why disentangling between these shocks requires controlling for students' social status, it is important to analyze the implications of both shocks. As per Equations (3) and (4), an increase in $\alpha_{0,s}$ (e.g., by improving teaching quality) suggests an increase in GPA without affecting effort. Importantly, this increase does not depend on

[6] 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.

[7] We would have $\hat{\kappa}_{0,s} = \bar{\kappa}_{0,s}$ if and only if $\alpha_{0,s} = 0$, $\lambda_0 = 0$, or if every student has friends. These conditions may not be verified in many settings.

the student's social status and does not involve a social multiplier effect.[8] In contrast, a preference shock on $c_{0,s}$ could result in a social multiplier effect on the effort (by Equation (3)), and thus on the 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 $\alpha_{0,s}$, the impact of $c_{0,s}$ on GPA is contingent upon the student's social status.

The main difference between the standard model and Equation (4) can be described by a variable omission problem in the standard model. These variables are $-\lambda_0\alpha_{s,0}\mathbf{g}_{s,i}\mathbf{1}_{n_s}$, for $s = 1, \dots, S$, and may imply a significant difference between the peer effect estimates from both models. By using the argument of variable omission 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 $-\lambda_0\alpha_{s,0}\mathbf{g}_{s,i}\mathbf{1}_{n_s}$ and $\mathbf{g}_{s,i}\mathbf{y}_s$. In most cases, it will be the sign of $-\lambda_0\alpha_{s,0}$ if $\mathbf{y}_s$ is nonnegative. This suggests that peer effects estimated using the standard model would be lower than the estimate from our specification if $\lambda_0\alpha_{s,0} > 0$.[9]

3.2 Identification and Estimation

As the effort $e_{s,i}$ is not observed, we cannot identify all the parameters of the structural model. The parameter δ_0 cannot be identified because it only appears in Equation (4) through a product with other parameters. Similarly, $\boldsymbol{\theta}_0$ and $\boldsymbol{\gamma}_0$ cannot be identified separately, as they only appear in the equation through $\tilde{\boldsymbol{\gamma}}_0 = \delta_0^2\boldsymbol{\gamma}_0 - \lambda_0\boldsymbol{\theta}_0$. Consequently, we set the following identification restrictions.

Restriction 3.1 (Identification restrictions)**.** $\delta_0 = 1$ *and* $\boldsymbol{\theta}_0 = 0$.

It is worth noting that this non-identification problem limits the implementation of some counterfactual analyses. Indeed, Equation (4) suggests that increasing $c_{0,s}$ by $\Delta c_{0,s}$ (preference shocks) leads to an increase of $\delta_0^2\Delta c_{0,s}(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)^{-1}\mathbf{1}_{n_s}$ in $\mathbf{y}_s$. The effect of the increase in $c_{0,s}$ on the GPA depends on δ_0, and involves a social multiplier effect only when $\delta_0^2/(1-\lambda_0) > 1$, that is, if λ_0 or δ_0 is sufficiently large. Unfortunately, identifying δ_0 requires additional data and goes beyond the scope of this paper. Under Condition 3.1, Equation (4) can be expressed as

$$y_{s,i} = \kappa_{s,i} + \lambda_0\mathbf{g}_{s,i}\mathbf{y}_s + \mathbf{x}'_{s,i}\boldsymbol{\beta}_0 + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0 + (\boldsymbol{\omega}_{s,i} - \lambda_0\mathbf{g}_{s,i})\boldsymbol{\eta}_s + \varepsilon_{s,i}, \tag{5}$$

where $\kappa_{s,i} = \bar{\kappa}_{0,s}$ if $i \in \bar{\mathcal{V}}_s$ and $\kappa_{s,i} = \hat{\kappa}_{0,s}$ if $i \in \hat{\mathcal{V}}_s$. Let $\hat{n}_s$ be the number of students in the school

[8] Equation (4) implies that the variation in $\mathbf{y}_s$, denoted $\Delta^\alpha\mathbf{y}_s$, following an increase $\Delta\alpha_{0,s}$ in $\alpha_{0,s}$ is such that $\Delta^\alpha\mathbf{y}_s = \Delta\alpha_{0,s}(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\mathbf{1}_{n_s} + \lambda_0\mathbf{G}_s\Delta^\alpha\mathbf{y}_s$. This implies that $(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\Delta^\alpha\mathbf{y}_s = \Delta\alpha_{0,s}(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\mathbf{1}_{n_s}$ and thus, $\Delta^\alpha\mathbf{y}_s = \Delta\alpha_{0,s}\mathbf{1}_{n_s}$. Hence, the increase in $\alpha_{0,s}$ 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 $(\boldsymbol{\omega}_{s,i} - \lambda_0\mathbf{g}_{s,i})\boldsymbol{\eta}_s$. However, we should point out that this second difference does not involve 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 the correlated effects leads to a consistent estimator (Kelejian and Prucha, 1998).

s with peers and $\bar{n}_s$ be the number of students in the school s without peers. Let also $\hat{\boldsymbol{\ell}}_s = \mathbf{G}_s\mathbf{1}_{n_s}$, $\bar{\boldsymbol{\ell}}_s = \mathbf{1}_{n_s} - \hat{\boldsymbol{\ell}}_s$. Equation (5) can be written in a matrix form at the school level as

$$\mathbf{y}_s = \bar{\kappa}_{0,s}\bar{\boldsymbol{\ell}}_s + \hat{\kappa}_{0,s}\hat{\boldsymbol{\ell}}_s + \lambda_0\mathbf{G}_s\mathbf{y}_s + \mathbf{X}_s\boldsymbol{\beta}_0 + \mathbf{G}_s\mathbf{X}_s\boldsymbol{\gamma}_0 + (\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \boldsymbol{\varepsilon}_s. \tag{6}$$

The number of unknown parameters to be estimated in Equation (6) grows to infinity with the number of schools (they 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 $\bar{\kappa}_{0,s}$ and $\hat{\kappa}_{0,s}$. To do so, we define $\mathbf{J}_s := \mathbf{I}_{n_s} - \frac{1}{\bar{n}_s}\bar{\boldsymbol{\ell}}_s\bar{\boldsymbol{\ell}}_s' - \frac{1}{\hat{n}_s}\hat{\boldsymbol{\ell}}_s\hat{\boldsymbol{\ell}}_s'$ and impose by convention that $\frac{1}{\bar{n}_s}\bar{\boldsymbol{\ell}}_s\bar{\boldsymbol{\ell}}_s' = 0$ if $\bar{n}_s = 0$, and that $\frac{1}{\hat{n}_s}\hat{\boldsymbol{\ell}}_s\hat{\boldsymbol{\ell}}_s' = 0$ if $\hat{n}_s = 0$. Note that $\hat{\boldsymbol{\ell}}_s'\hat{\boldsymbol{\ell}}_s = \hat{n}_s$, $\bar{\boldsymbol{\ell}}_s'\bar{\boldsymbol{\ell}}_s = \bar{n}_s$, $\hat{\boldsymbol{\ell}}_s'\bar{\boldsymbol{\ell}}_s = 0$, $\bar{\boldsymbol{\ell}}_s'\hat{\boldsymbol{\ell}}_s = 0$. Thus, $\mathbf{J}_s\bar{\boldsymbol{\ell}}_s = \mathbf{J}_s\hat{\boldsymbol{\ell}}_s = 0$. One can eliminate the term $\bar{\kappa}_{0,s}\bar{\boldsymbol{\ell}}_s + \hat{\kappa}_{0,s}\hat{\boldsymbol{\ell}}_s$ by premultiplying each term of Equation (6) by the matrix $\mathbf{J}_s$.[10] This implies that

$$\mathbf{J}_s\mathbf{y}_s = \lambda_0\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s + \mathbf{J}_s\mathbf{X}_s\boldsymbol{\beta}_0 + \mathbf{J}_s\mathbf{G}_s\mathbf{X}_s\boldsymbol{\gamma}_0 + \mathbf{J}_s(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \mathbf{J}_s\boldsymbol{\varepsilon}_s. \tag{7}$$

The random term $\boldsymbol{v}_s := \mathbf{J}_s(\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \mathbf{J}_s\boldsymbol{\varepsilon}_s$ is characterized by two error vectors $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$. To consistently estimate $\boldsymbol{\psi}_0 := (\lambda_0,\ \boldsymbol{\beta}_0',\ \boldsymbol{\gamma}_0')'$, we impose that the following assumption.

Assumption 3.1. *For any $s = 1, \dots, S$, $\mathbb{E}(\boldsymbol{\eta}_s|\mathbf{G}_s, \mathbf{X}_s) = 0$ 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 3.3).

Identification and Estimation of $\boldsymbol{\psi}_0$

We impose the following identification assumption.

Assumption 3.2. *(i) $\lambda_0\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0 \neq 0$; (ii) There are students in the network separated by a link of distance three.*

Assumption 3.2 deals with the reflection problem. Bramoullé et al. (2009) established similar identification conditions in the standard model. It is important to note, however, that their identification analysis cannot be directly applied to our framework due to the presence of two types of social status: students with friends and students without friends. One crucial assumption in the identification analysis conducted by Bramoullé et al. (2009) is that every agent has at least one friend (see their

[10] By premultiplying each term by $\mathbf{J}_s$, we consider Equation (6) in deviation to the average within the student group, that is, $\hat{\mathcal{V}}_s$ or $\bar{\mathcal{V}}_s$. This eliminates the parameters $\bar{\kappa}_{0,s}$ and $\hat{\kappa}_{0,s}$.

main assumption in Section 2.1). We extend their identification analysis to our framework, under similar (but slightly stronger) conditions, by allowing some students to have no friends. We adapt the identification analysis conducted by Houndetoungan (2022) to our model. Condition (i) is equivalent to saying that GPA is influenced by at least one contextual variable (see Equation A.7). With several characteristics in $\mathbf{X}_s$, this condition can be satisfied. Moreover, Condition (ii) is easy to verify in practice. A weaker version of this condition, as set by Bramoullé et al. (2009), is that $\mathbf{I}_{n_s}$, $\mathbf{G}_s$, $\mathbf{G}_s^2$, and $\mathbf{G}_s^3$ are linearly independent. As we allow students to have no friends, our condition is sufficient and may not be necessary. In the case of the standard model, Bramoullé et al. (2009) demonstrate that condition that $\mathbf{I}_{n_s}$, $\mathbf{G}_s$, $\mathbf{G}_s^2$, and $\mathbf{G}_s^3$ are linearly independent is necessary for identifying .

We estimate $\boldsymbol{\psi}_0$ 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). A natural instrument for addressing this endogeneity is $\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s$ (see Kelejian and Prucha, 1998; Bramoullé et al., 2009). Indeed, by premultiplying each term of Equation (6) by $\mathbf{J}_s\mathbf{G}_s$, one can notice that $\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s$ and $\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s$ are correlated. Because $\mathbf{J}_s\mathbf{G}_s^2\mathbf{X}_s$ does not enter Equation (7) as a regressor and is exogenous with respect to $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$, it can be used as an external instrument for $\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s$. Let $\hat{\boldsymbol{\psi}}$ be the GMM estimator of $\boldsymbol{\psi}_0$.

Proposition 3.1. *Under Condition 3.1 and Assumptions 2.1–3.2 and A.1 (stated in Appendix A.3), $\boldsymbol{\psi}_0$ is globally identified, $\hat{\boldsymbol{\psi}}$ is a consistent estimator of $\boldsymbol{\psi}_0$, and $\sqrt{n}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \xrightarrow{d} \mathcal{N}\big(0, \lim_{n\to\infty} n\,\mathbb{V}(\hat{\boldsymbol{\psi}})\big)$.*

The consistency and asymptotic normality of $\hat{\boldsymbol{\psi}}$ are directly derived from Kelejian and Prucha (1998).[11] We later discuss how the asymptotic variance of $\hat{\boldsymbol{\psi}}$ can be estimated.

The identification of $\boldsymbol{\psi}_0$ is similar to the identification analysis in Houndetoungan (2022). Below, we provide an intuition behind this approach and a formal proof in Appendix A.3. Figure 1 presents an example in which Condition (ii) of Assumption 3.2 is verified (i_1 and i_2 are separated by a link of distance three). 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 $\mathbf{J}_s\mathbf{G}_s\mathbf{X}_s$, that is, for all $i \in \hat{\mathcal{V}}_s$, $\mathbb{E}(\mathbf{g}_{s,i}\mathbf{y}_s - \hat{\bar{y}}_s|\mathbf{G}_s, \mathbf{X}_s) = (\mathbf{x}_{s,i} - \hat{\mathbf{x}}_s)'\dot{\boldsymbol{\beta}} + (\mathbf{g}_{s,i}\mathbf{X}_s - \hat{\bar{\mathbf{x}}}_s)'\dot{\boldsymbol{\gamma}}$, where $\hat{\bar{y}}_s$, $\hat{\mathbf{x}}_s$, and $\hat{\bar{\mathbf{x}}}_s$ 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 $\hat{\mathcal{V}}_s$, 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 $\hat{\mathcal{V}}_s$ because of the matrix $\mathbf{J}_s$ that premultiplies the terms of Equation (7). If we take the previous equation in difference between the 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 $\mathbf{g}_{s,i_3}\mathbf{X}_s = \mathbf{x}_{s,i_4}$. and Equation (8) implies $y^e_{i_3} - y^e_{i_4} = (\mathbf{x}_{s,i_1} - \mathbf{x}_{s,i_3})'\dot{\boldsymbol{\beta}} + (\mathbf{x}_{s,i_3} - \mathbf{x}_{s,i_4})'\dot{\boldsymbol{\gamma}}$, where

[11] See also Hansen (1982) for formal results on the large sample properties of GMM estimators.

$y_i^e = \mathbb{E}(y_{s,i}|\mathbf{G}_s, \mathbf{X}_s)$. 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 $\mathbf{x}_{s,i_2}$. Put differently, an increase in $\mathbf{x}_{s,i_2}$(*ceteris paribus*) 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 a contextual variable 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. 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 $\mathbf{J}_s\mathbf{G}_s\mathbf{X}_s$.

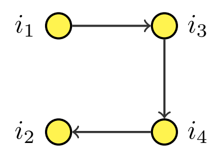

Figure 1: Solving the reflection problem

Note: $\rightarrow$ 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. 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 $\mathbb{V}(\hat{\boldsymbol{\psi}})$ 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 $\hat{\mathbf{D}}_n = (\mathbf{R}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}\operatorname{diag}\{\mathbf{Z}_1'\hat{\boldsymbol{v}}_1\hat{\boldsymbol{v}}_1'\mathbf{Z}_1, \dots, \mathbf{Z}_S'\hat{\boldsymbol{v}}_S\hat{\boldsymbol{v}}_S'\mathbf{Z}_S\}(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{R}'\mathbf{Z})'/n$. However, for this estimator, we need to impose that n_s is bounded, that is, the network is composed of many bounded and independent schools. Therefore this method cannot deal with dense networks when both n_s and S can grow to infinity. This assumption is important for $\mathbf{Z}_s'\hat{\boldsymbol{v}}_s$ be $O_p(1)$ and has a second-order moment.

We also present another estimator based on the covariance structure of the error term $\boldsymbol{v}_s$. This estimator allows dense networks as both n_s and S can grow to infinity. Accommodating dense networks is important to handle, for instance, the issue of endogenous networks (see Section 3.3). For this second approach, we set the following assumptions, which require, among others, homoskedasticity.

Assumption 3.3. *(i)* $\mathbf{G}_s$ *is uniformly bounded in column sum; (ii)* $\lambda_0 \neq 1$*; (iii)* $\mathbb{E}(\eta_{s,i}^2|\mathbf{G}_s, \mathbf{X}_s) = \sigma_{\eta 0}^2 > 0$, $\mathbb{E}(\varepsilon_{s,i}^2|\mathbf{G}_s, \mathbf{X}_s) = \sigma_{\epsilon 0}^2 > 0$, $\mathbb{E}(\varepsilon_{s,i}\eta_{s,i}|\mathbf{G}_s, \mathbf{X}_s) = \rho_0\sigma_{\eta 0}\sigma_{\epsilon 0}$*, 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. For instance, if the indegree of a student j (the number of students who have j as a peer) grows to infinity with n_s, the condition ensures that the total influence of j in the network, which can be measured by the sum of the j-th column of $\mathbf{G}_s$, is "limited." As $\mathbf{G}_s$ is row-normalized, this means that students who have j as a friend also have many other friends; thus, the average influence of j vanishes asymptotically. This condition is also considered by (Lee, 2004) in the case of the standard model. Condition (iii) accounts for the correlation between $\eta_{s,i}$ and $\varepsilon_{s,i}$. This is important as $\eta_{s,i}$ and $\varepsilon_{s,i}$ characterized the same student i. The parameter ρ_0 measures the correlation between $\eta_{s,i}$ and $\varepsilon_{s,i}$. We impose no restrictions on the joint distribution of $(\eta_{s,i},\ \varepsilon_{s,i})$. Imposing that $(\eta_{s,i},\ \varepsilon_{s,i})$'s are independent across i allows for estimating $\sigma^2_{\eta 0}$, $\sigma^2_{\epsilon 0}$, and ρ_0. The restriction $\lambda_0 \neq 0$ of Condition (ii) is necessary for identifying $\sigma_{\eta 0}$, $\sigma_{\epsilon 0}$, and ρ_0. If $\lambda_0 = 0$, the disturbance of Equation (7) would be $\mathbf{J}_s\boldsymbol{\eta}_s + \mathbf{J}_s\boldsymbol{\varepsilon}_s$, and one cannot disentangle $\sigma_{\eta 0}$, $\sigma_{\epsilon 0}$, and ρ_0.

Using the covariance structure of Assumption 3.3, we can estimate $(\sigma^2_{\eta 0},\ \sigma^2_{\epsilon 0},\ \rho_0)$ 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_0 is replaced by its estimator $\hat{\lambda}$. Importantly, we do not require $\boldsymbol{v}_s$ to be actually normally distributed.

However, the log-likelihood cannot be written directly because the transformation we apply to eliminate the fixed effects $\hat{\kappa}_s$ and $\bar{\kappa}_s$ 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 necessary task to write the log-likelihood. To address this issue, we use a similar approach to that of Lee et al. (2010). Let $[\mathbf{F}_s,\ \bar{\boldsymbol{\ell}}_s/\sqrt{\bar{n}_s},\ \hat{\boldsymbol{\ell}}_s/\sqrt{\hat{n}_s}]$ 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.[12] To ease the notational burden, we assume the school s has students in both $\bar{\mathcal{V}}_s$ and $\hat{\mathcal{V}}_s$. 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 $\mathbf{F}_s'\hat{\boldsymbol{v}}_s$.[13] The log-likelihood of $\mathbf{F}_s'\hat{\boldsymbol{v}}_s$ is given by

$$\hat{L}(\sigma^2_\eta, \sigma^2_\epsilon, \rho) = -\sum_{s=1}^S \frac{n_s-2}{2}\log(\sigma^2_\epsilon) - \frac{1}{2}\sum_{s=1}^S \log|\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho)| - \sum_{s=1}^S \frac{1}{2\sigma^2_\epsilon}\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\hat{\boldsymbol{v}}_s, \quad (9)$$

where σ_η, σ_ϵ, and ρ denote arbitrary values of $\sigma_{0\eta}$, $\sigma_{0\epsilon}$, and ρ_0, $\boldsymbol{\Omega}_s(\hat{\lambda}, \tau, \rho) = \mathbf{I}_{n_s-2} + \tau^2\mathbf{F}_s'\mathbf{W}_s\mathbf{W}_s'\mathbf{F}_s + \rho\tau\mathbf{F}_s'(\mathbf{W}_s + \mathbf{W}_s')\mathbf{F}_s$, $\mathbf{W}_s = \mathbf{I}_{n_s} - \hat{\lambda}\mathbf{G}_s$, and $\tau = \sigma_\eta/\sigma_\epsilon$. The first-order conditions of the maximization of (9) imply that σ^2_ϵ can be substituted with $\tilde{\sigma}^2_\epsilon(\tau, \rho) = \sum_{s=1}^S \frac{\hat{\boldsymbol{v}}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\hat{\lambda}, \tau, \rho)\mathbf{F}_s'\hat{\boldsymbol{v}}_s}{n-2S}$. This leads to a simpler concentrated log-likelihood that does not depend on σ^2_ϵ and easier to be maximized. Let $(\hat{\sigma}^2_\eta,\ \hat{\tau},\ \hat{\rho})$ be the QML estimator of $(\sigma^2_{\varepsilon 0},\ \tau_0,\ \rho_0)$. We establish the following result.

Proposition 3.2. *Under Proposition 3.1 and Assumptions A.2–A.4,* $(\sigma^2_{\epsilon 0},\ \tau_0,\ \rho_0)$ *is globally identified,*

[12] 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 $\bar{\mathcal{V}}_s$ and $\hat{\mathcal{V}}_s$, and $n_r - 1$ if s has students in either $\bar{\mathcal{V}}_s$ or $\hat{\mathcal{V}}_s$.

[13] 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 $\tilde{\boldsymbol{v}}_s = (\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \boldsymbol{\varepsilon}_s$. 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).

and $(\hat{\sigma}^2_{\eta},\ \hat{\tau},\ \hat{\rho})$ *is a consistent estimator of* $(\sigma^2_{\epsilon 0},\ \tau_0,\ \rho_0)$.

We can now consistently estimate $\lim_{n\to\infty} n\,\mathbb{V}(\hat{\boldsymbol{\psi}})$ using the estimator $(\hat{\sigma}^2_{\eta},\ \hat{\tau},\ \hat{\rho})$.
Let also $\mathbf{B}_0 = \text{plim}(\mathbf{R}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{R}'\mathbf{Z})'/n$ (this exists by Assumption A.1), where plim stands for the limit in probability as n grows to infinity. We assume 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 $\mathbf{D}_0$. It follows that $\lim_{n\to\infty} n\,\mathbb{V}(\hat{\boldsymbol{\psi}}) = \sigma^2_{\epsilon 0}\mathbf{B}_0^{-1}\mathbf{D}_0\mathbf{B}_0^{-1}$. We can obtain a consistent estimator of the variance by replacing $\sigma^2_{\epsilon 0}$, $\sigma^2_{\eta 0}$, and ρ_0 with their estimator, and $\mathbf{B}_0$ and $\mathbf{D}_0$ with their empirical counterpart.

3.3 Extension to Endogenous Networks

In practice, the assumption of an exogenous network may not hold because of some students' characteristics that we do not observe. For example, the intellectual quotient (IQ) and the sociability level of the student may not be included in $\mathbf{X}_s$ because they are not available in the data set. These variables being likely to explain the GPA would therefore be captured by the error terms $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$. The network is endogenous (that is, it depends on $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$) if it also depends on IQ or sociability level. The problem of network endogeneity can then be regarded as an omission of important factors in the design matrix $\mathbf{X}_s$ that can explain both GPA and network. This also suggests that $\mathbf{X}_s$ may be endogenous because $\mathbf{X}_s$ is likely to be correlated to IQ or sociability level.

To address network endogeneity, we control for the omitted factors in $\mathbf{X}_s$ that explain the network. We use a two-stage estimation method similar to the control function approach proposed by Johnsson and Moon (2021). We first estimate the unobserved factors using a network formation model. In the second stage, we control for the estimated factors in the peer effect model.

We consider a network formation model with degree heterogeneity (see Graham, 2017; Dzemski, 2019; Yan et al., 2019).[14] For two students i and j in the same school s, the conditional probability to observe a link from i to j (that is, i declares that j is a friend) can be defined as

$$P_{s,ij} := \mathbb{P}(a_{s,ij} = 1 | \ddot{\mathbf{x}}_{s,ij}, \ddot{\boldsymbol{\beta}}_0, \mu^{out}_{0,s,i}, \mu^{in}_{0,s,j}) = \Phi(\ddot{\mathbf{x}}'_{s,ij}\ddot{\boldsymbol{\beta}}_0 + \mu^{out}_{0,s,i} + \mu^{in}_{0,s,j}), \tag{10}$$

where $\ddot{\mathbf{x}}_{s,ij}$ is a vector of observed dyad-specific variables (e.g., $\ddot{\mathbf{x}}_{s,ij}$ may contain the distance between students i and j's characteristics), $\ddot{\boldsymbol{\beta}}_0$ is an unknown parameter, $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,j}$ are unobserved attributes, and Φ is either the normal or logistic distribution function depending on whether the network formation model follows a probit or logit model. The term $\ddot{\mathbf{x}}'_{s,ij}\ddot{\boldsymbol{\beta}}_0$ is a measure of social distance between students i and j that drives the homophily of linking decisions. Equation (10) implies that the matrix of link probabilities $\mathbf{P}_s := [P_{s,ij}]_{ij}$ is nonsymmetric. For any i, $\mu^{out}_{0,s,i}$ only influences

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

the probabilities for other students to be i's friends, whereas $\mu^{in}_{0,s,i}$ only influences the probabilities for i to be another students' friend.

Equation (10) includes two parameters $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,j}$ for each pair $(i,\ j)$. This implies more than $2n_s$ parameters to be estimated per school. Yan et al. (2019) show that the standard logit estimators of $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,j}$ are consistent if the network is dense.[15] In this logit model, $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,j}$ 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 $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,j}$ (see Albert and Chib, 1993). However, this approach treats $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,j}$ as random effects.[16]

The problem of network endogeneity occurs because the adjacency matrix is dependent on $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$ through $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$. We impose the following assumption regarding how $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$ are related to $\boldsymbol{\eta}_s$ and $\boldsymbol{\varepsilon}_s$.

Assumption 3.4. *$\eta_{s,i} = h_\eta(\mu^{out}_{0,s,i}, \mu^{in}_{0,s,i}) + \eta^*_{s,i}$ and $\varepsilon_{s,i} = h_\epsilon(\mu^{out}_{0,s,i}, \mu^{in}_{0,s,i}) + \varepsilon^*_{s,i}$, where $\eta^*_{s,i}$ and $\varepsilon^*_{s,i}$ are new error terms independent of $(\mu^{out}_{0,s,1}, \dots, \mu^{out}_{0,s,n_s})'$ and $(\mu^{in}_{0,s,1}, \dots, \mu^{in}_{0,s,n_s})'$, and h_η and h_ϵ are smooth functions. We also adapt Assumption 3.1 to $\eta^*_{s,i}$ and $\varepsilon^*_{s,i}$.*

As in Johnsson and Moon (2021), we control for any smooth function of $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$ to link $\eta_{s,i}$ and $\varepsilon_{s,i}$ to the new error terms $\eta^*_{s,i}$ and $\varepsilon^*_{s,i}$. Assumption 3.4 is weak, as it mainly requires that $\mathbf{E}(\eta_{s,i}|\mu^{out}_{0,s,i}, \mu^{in}_{0,s,i})$ and $\mathbf{E}(\varepsilon_{s,i}|\mu^{out}_{0,s,i}, \mu^{in}_{0,s,i})$ exist and are smooth functions in $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$. By replacing $\eta_{s,i}$ and $\varepsilon_{s,i}$ in Equation (7) with their expression in Assumption 3.4, we have

$$\mathbf{J}_s\mathbf{y}_s = \lambda\mathbf{J}_s\mathbf{G}_s\mathbf{y}_s + \mathbf{J}_s\mathbf{X}_s\boldsymbol{\beta}_0 + \mathbf{J}_s\mathbf{G}_s\mathbf{X}_s\boldsymbol{\gamma}_0 + \mathbf{J}_s(\mathbf{h}^\eta_s + \mathbf{h}^\epsilon_s - \lambda\mathbf{G}_s\mathbf{h}^\eta_s) + \mathbf{J}_s(\mathbf{I}_{n_s} - \lambda\mathbf{G}_s)\boldsymbol{\eta}^*_s + \mathbf{J}_s\boldsymbol{\varepsilon}^*_s, \quad (11)$$

where $\mathbf{h}^\eta_s$ and $\mathbf{h}^\epsilon_s$ are the vectors of $h_\eta(\mu^{out}_{0,s,i}, \mu^{in}_{0,s,i})$'s and $h_\epsilon(\mu^{out}_{0,s,i}, \mu^{in}_{0,s,i})$'s at the school level. The main issue about Equation (11) is that $\mathbf{h}^\eta_s$ and $\mathbf{h}^\epsilon_s$ are unknown. In a similar model, Johnsson and Moon (2021) use a control function approach to eliminate these variables from the equation. Their approach consists of subtracting the influence of $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,i}$ from the other variables of the equation using a nonparametric method. Given the complexity of the disturbance structure and the presence of school heterogeneity, it would be challenging to apply the same technique to our framework.

A similar approach to that of Johnsson and Moon (2021) is to replace $\mathbf{h}^\eta_s$ and $\mathbf{h}^\epsilon_s$ with nonparametric approximations. Specifically, we approximate $\mathbf{h}^\eta_s$ and $\mathbf{h}^\epsilon_s$ using tensor products of cubic B-splines. This approach is widely used in nonparametric regressions and consists of approximating any smooth

[15] 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.

[16] We present a Gibbs sampler to simulate the posterior distributions of $\mu^{out}_{0,s,i}$ and $\mu^{in}_{0,s,j}$ in Online Appendix (OA) S.2.

function by a combination of piecewise cubic polynomials called *basis* (where the "B" of B-spline comes from). The polynomials are defined such that the approximation fits the function well everywhere on its support. The advantage of this method is that it is straightforward and can be easily applied to complex models (see Hastie, 2017). The baseline idea of this approach stems from the Weierstrass theorem, which states that any real-valued continuous function defined on a compact can be *well approximated* using polynomials.[17]

Controlling for endogeneity results in plugging new regressors (that is, the bases of the polynomials) into the initial specification. Therefore, identification and consistency results established in Section 3.2 still hold if the new matrix of explanatory variables verifies the regulatory conditions set in Assumption A.1 and A.3. In contrast, the asymptotic normality becomes more complicated and cannot be generalized without new restrictions. The asymptotic properties of the estimator of $\boldsymbol{\psi}_0$ under endogenous networks depend on the uncertainty related to the estimation of $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ in the first stage. 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 S.1.4). For instance, this restriction is set in the Lipschitz condition of Assumption 8 in Johnsson and Moon (2021) and in Assumption S.1 of Houndetoungan (2022). Alternatively, a bootstrap approach can be used to approximate the asymptotic distribution of the second-stage estimator accounting for the uncertainty of the first stage.

4 Empirical Analysis

In this section, we present an empirical illustration of the model using a unique and now widely used data set provided by the National Longitudinal Study of Adolescent Health (Add Health).[18]

4.1 Data

We use the Wave I in-school Add Health data, collected between September 1994 and April 1995, which gathered national representative information on 7th–12th graders in the US. The surveyed sample comprises 90,118 students from 145 middle, junior high, and high schools. The data provides information on the social and demographic characteristics of students as well as their friendship links (that is, best friends, up to 5 females and up to 5 males), education level, occupation of parents, etc.

After removing observations with missing data, the sample used for our empirical analysis has 68,430 students from 141 schools. The number of students per school ranges from 18 to 2,027, with

[17] Our approach is flexible and can accommodate many smooth functions h_η and h_ϵ. For instance, $\mathbf{h}s^{\eta} + \mathbf{h}s^{\epsilon} - \lambda \mathbf{G}s\mathbf{h}s^{\eta}$ is approximated using a combination of up to 390 piecewise polynomial functions, which also account for the interactions between $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$. We replace $\mu_{0,s,i}^{out}$ and $\mu_{0,s,i}^{in}$ with their estimator from the first stage (see OA S.1.4).

[18] The data are available on the website `https://addhealth.cpc.unc.edu/documentation/study-design/`.

an average of 485. On average, each student had 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 friends of 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 can be ignored, as shown by Boucher and Houndetoungan (2022).

The dependent variable, GPA, is the average grade of four subjects: mathematics, science, English/language arts, and history or social science. It is calculated on the basis of students' grades (A, B, C, and D or lower) obtained in the four subjects. We re-coded these grades as A = 4; B = 3; C = 2; and D = 1. We control for several other potential factors that can influence GPA (Duncan et al., 2001; Lin, 2010). These factors include sex, age, whether the student is Hispanic, race, whether the student lives with both parents, the number of years spent at their current school by the student, whether the student is a school club member, mother's education, and mother's profession. We also control for contextual variables associated with the factors, that is, the average of friends' control variables.

Table 1 defines the variables included in the empirical analysis and presents the data summary. The average GPA is 2.8 and the average friend's GPA is 2.2. The average age of the students is about 15 years, they have attended their current school for 2.5 years on average. The sample includes 48.7% boys, 16.4% Hispanics, 64.7% whites, and 16.8% Black. Asians account for 7%, and 9.5% of the sample are of other races. About 74.1% of the students live with both parents. The highest education level achieved is high school (HS) for about 30.6% of the student's mothers, beyond HS for 16.9% of the mothers, and less than HS for 41.9% of the mothers. As for mothers' occupations, Add Health provides a detailed list with more than 15 categories. We combine these occupations into four broader categories along with a missing indicator. Specifically, 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.

As an important feature of the model is to control for whether students have friends on not, we also present a data summary in the subsample of students who have no friends. The average GPA is slightly lower than that of students who have friends. We also notice that students who have no peers are often from small minorities, such as Hispanics, Blacks, and Asians.

Table 1: Variables and summary statistics

Variable	Definition	Own characteristics: All students Mean	SD	Own characteristics: Without peers Mean	SD	Average friend's characteristics 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	2.246	1.287
Gender							
(Male)	1 if male; 0 otherwise	0.487	0.500	0.582	0.493	0.359	0.339
Female	1 if female; 0 otherwise	0.513	0.500	0.418	0.493	0.423	0.359
Age	Student's age in year	15.074	1.680	15.364	1.656	11.760	6.355
Hispanic	1 if Hispanic; 0 otherwise	0.164	0.370	0.228	0.420	0.117	0.262
Race							
(White)	1 if White; 0 otherwise	0.647	0.478	0.550	0.498	0.525	0.449
Black	1 if Black; 0 otherwise	0.168	0.374	0.205	0.404	0.126	0.302
Asian	1 if Asian; 0 otherwise	0.070	0.254	0.085	0.279	0.052	0.176
Other	1 if the student's race is not in the list above; 0 otherwise	0.095	0.293	0.120	0.325	0.067	0.177
Lives with both parents	1 if the student lives with both of their parents; 0 otherwise	0.741	0.438	0.691	0.462	0.596	0.398
Year in school	Number of years spent in the current school	2.506	1.422	2.361	1.351	2.042	1.511
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	0.743	0.412
Mother's education							
(High)	1 if mother's education is high school; 0 otherwise	0.450	0.306	0.461	0.282	0.241	0.284
$<$ High	1 if mother's education is greater than high school; 0 otherwise	0.169	0.375	0.191	0.393	0.119	0.222
$>$ High	1 if mother's education is less than high school; 0 otherwise	0.484	0.419	0.493	0.376	0.349	0.345
Missing	1 if mother's education is missing; 0 otherwise	0.106	0.308	0.151	0.358	0.073	0.167
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	0.173	0.240
(Stay home)	1 if mother is a homemaker, retired, or does not work; 0 otherwise	0.206	0.405	0.214	0.410	0.153	0.229
Other	1 if mother's job is not in the list above; 0 otherwise	0.435	0.496	0.406	0.491	0.346	0.322
Missing	1 if mother's job is missing; 0 otherwise	0.157	0.364	0.208	0.406	0.110	0.203

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

4.2 Estimation Results

In this section, we present our empirical results on Add Health data.[19] We study the case of exogenous and endogenous networks. We estimate the model with school fixed effects given by Equation (7). We also consider a specification without school heterogeneity. In this case, we include a new explanatory dummy variable that takes one if the student has friends and zero otherwise. Including this variable disentangles the GPA shock parameter $\alpha_{0,s}$ from the preference shock parameter $c_{0,s}$. In addition, we estimate a classical linear-in-means peer effect model, controlling for school heterogeneity as fixed effects and then without controlling for it. As discussed earlier, we expect the peer effect estimate from the classical model to be different the estimate from our model.

4.2.1 Assuming an Exogenous Network

The estimation results under an exogenous networks is presented in Table 2. The statistics of the weak instrument test suggest that the specifications do not suffer from a weak instrument issue. However, Models 1–3 suffer from an overidentification issue, which is addressed when we allow for additional heterogeneity in Model 4 (the main model). Estimation results for Model 4 imply that increasing the average GPA of peers by one point results in a 0.856 point increase in students' GPA. This finding is aligned with the empirical literature (Sacerdote, 2011) highlighting the importance of peers as determinants of student performance. The peer effect estimate in our specification is larger by over 1.6 times the size of the coefficient in the classical models (0.502 in Model 1 without unobserved school heterogeneity and 0.507 in Model 3 when we allow for unobserved school heterogeneity). 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 $\alpha_{0,s}$ and $c_{0,s}$. This distinction is important, as it deeply affects the causal interpretation of peer effects.[20]

School fixed effects do not appear to significantly influence the endogenous peer effect parameter. The estimates are similar for Models 1 and 3 (based on the standard specification) and for Models 2 and 4 (based on our proposed specification). It is worth mentioning that the parameter associated with the dummy variable "Has friends" in Model 3 is not the shift in the GPA for students having friends. In Equation (4), the coefficient associated with this variable is $-\lambda_0\alpha_{s,0}$. On the one hand, its significance suggests the presence of common shocks that directly affect GPA, irrespective of effort. Those shocks are captured by $\alpha_{s,0}$. On the other hand, the negative estimate means that $\lambda_0\alpha_{s,0} > 0$, which is in line with the negative bias of Models 1 and 3. Indeed, as argued in Section 3.1, peer effects estimated using standard models would be biased downward if $\lambda_0\alpha_{s,0} > 0$.

[19] Replication codes can be found at `https://github.com/ahoundetoungan/PeerEffectsEffort`.

[20] As pointed out above, it is challenging to compute the social multiplier effect associated with the estimate 0.856 because we do not identify δ_0^2.

The intuition of this bias is as follows. The standard model only accounts for preference shocks. It wrongly considers direct shocks to GPA without influencing the effort as preference shocks leading to a social multiplier effect. This overestimates the impact of direct shocks. Since co-movements in students' and 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.

We also find that several students' own characteristics and contextual variables influence their GPA. Female students score 0.165 grade points higher than male students. Older students tend to do worse, and students who have been in the current school for longer periods tend to do better. 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. Mother's education is an important determinant of student GPA, and the relationship is positive. Specifically, children of mothers having less than a high school education score 0.068 points lower than those of high-school-graduate mothers, whereas children of mothers who have an education beyond high school achieve 0.124 points higher. Compared with students whose mothers do not work, children of teachers, lawyers, or other professional job holders are more successful, whereas children of other job holders have lower grades.

The coefficients are significant for a number of contextual variables. In particular, students who have relatively more female peers tend to perform worse than students who have male peers. This result suggests that a higher fraction of girls among a student's peers could provide a greater distraction for teenage boys. Coleman (1966) argued that there may be a distraction inherent in mixed-gender educational settings for adolescents. He pointed to the strong emphasis on "rating and dating" in American high school culture, with mixed peer groups having a negative effect on the achievement of both girls and boys. For example, a student's GPA decreases when their peers are Asian, participate in club activities, and when their peers' mother's job is professional. On the contrary, the student's GPA rises with the mean age of their peers, or when their peers are Black and Hispanic.

4.2.2 Controlling for Network Endogeneity

Figure 2 shows the 95% confidence interval of the peer effect estimates after controlling for network endogeneity. We approximate $h_\eta(.)$ and $h_\epsilon(.)$ using cubic B-spline curves allowing interactions between the spline bases (tensor products of cubic B-splines). This leads to additional explanatory variables we then plug into the model. These variables are globally significant, suggesting that the network is endogenous.[21] However, in our model, the endogeneity does not significantly influence the peer effect

[21] Full tables for some specifications including the case of the Bayesian random model are presented in OA S.3 (see Tables S.1–S.3).

Table 2: 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	0.832	0.038	0.507	0.028	0.856	0.044
Has friends			−2.569	0.132				
Own effects								
Female	0.183	0.006	0.168	0.006	0.176	0.006	0.165	0.006
Age	0.007	0.003	−0.042	0.003	−0.015	0.003	−0.043	0.003
Hispanic	−0.110	0.010	−0.111	0.010	−0.101	0.010	−0.091	0.010
Race								
Black	−0.144	0.012	−0.159	0.012	−0.131	0.012	−0.121	0.013
Asian	0.206	0.013	0.189	0.014	0.218	0.013	0.194	0.014
Other	−0.024	0.011	−0.026	0.011	−0.026	0.011	−0.026	0.011
Lives with both parents	0.116	0.008	0.099	0.007	0.107	0.007	0.091	0.008
Years in school	0.023	0.003	0.027	0.003	0.033	0.003	0.027	0.003
Member of a club	0.166	0.013	0.148	0.012	0.157	0.012	0.138	0.012
Mother’s education								
$<$ High	−0.076	0.009	−0.071	0.009	−0.076	0.009	−0.068	0.009
$>$ High	0.163	0.007	0.133	0.008	0.151	0.007	0.124	0.008
Missing	0.032	0.013	0.025	0.012	0.031	0.012	0.026	0.012
Mother’s job								
Professional	0.044	0.009	0.036	0.009	0.039	0.009	0.032	0.009
Other	−0.041	0.007	−0.037	0.008	−0.040	0.007	−0.037	0.008
Missing	−0.083	0.011	−0.075	0.011	−0.078	0.011	−0.070	0.011
Contextual effects								
Female	−0.138	0.012	−0.130	0.012	−0.108	0.012	−0.123	0.013
Age	−0.067	0.003	0.030	0.004	−0.073	0.004	0.024	0.006
Hispanic	0.033	0.016	0.097	0.018	0.050	0.017	0.087	0.020
Race								
Black	−0.001	0.015	0.104	0.017	−0.007	0.015	0.070	0.020
Asian	−0.051	0.019	−0.136	0.022	−0.043	0.021	−0.135	0.027
Other	−0.048	0.020	−0.003	0.021	−0.046	0.020	−0.001	0.022
Lives with both parents	−0.040	0.017	−0.016	0.018	−0.040	0.016	−0.019	0.018
Years in school	0.003	0.004	−0.020	0.004	0.028	0.004	−0.009	0.006
Member of a club	−0.175	0.029	−0.081	0.028	−0.142	0.028	−0.084	0.029
Mother’s education								
$<$ High	−0.044	0.017	0.031	0.018	−0.050	0.016	0.025	0.019
$>$ High	0.062	0.016	−0.024	0.020	0.027	0.017	−0.032	0.021
Missing	−0.057	0.025	−0.026	0.026	−0.070	0.024	−0.031	0.026
Mother’s job								
Professional	−0.058	0.018	−0.029	0.019	−0.056	0.018	−0.034	0.020
Other	−0.131	0.014	−0.026	0.016	−0.105	0.014	−0.022	0.016
Missing	−0.144	0.022	0.005	0.024	−0.116	0.021	0.008	0.024
σ_η^2			0.295				0.286	
σ_ϵ^2	0.515		0.086		0.503		0.046	
ρ			0.305				0.605	
Weak instrument F	141		80		209		122	
Sargan test prob.	0.000		0.016		0.000		0.223	

Models 1 and 2 do not control for school heterogeneity, but Model 2 controls for heterogeneity associated with having peers. Model 3 includes a single fixed effect per school (classical model), and Model 4 is our structural model. The columns "Coef" report the coefficient estimates followed by their corresponding standard errors in the columns "Sd Err". Models 1 and 3 have a single disturbance. The statistic of the weak instrument test and the p-value of Sargan’s overidentifying test are reported in the last two rows.

estimate. This result is robust to the number of knots used for the cubic B-spline approximation. It also supports 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; Houndetoungan, 2022). As in those studies, the peer effect estimate decreases slightly (from 0.856 to 0.820). This decline occurs because of unobserved attributes such as IQ that are positively linked to GPA. For instance, an exogenous shock to IQ would simultaneously influence students' and peers' GPA and could be mistakenly considered as peer effects. The specification with endogenous network separates peer effects from these co-movements in GPA.

In the case of the standard model, controlling for network endogeneity significantly increases the peer effect estimate from 0.507 to 0.692. Importantly, this also resolves the problem of overidentification when the unobserved attributes are estimated using a fixed-effect logit model, and the functions $h_\eta(.)$ and $h_\epsilon(.)$ are approximated by tensor products of cubic B-splines (see Table S.2). The overidentification issue arises because of the endogeneity of the network, which appears more concerning when student efforts are approximated by GPA. Nevertheless, the increase in the peer effect estimate is at odds with popular evidence that peer effect estimates tend to decrease after controlling for network endogeneity. The intuition of the increase is as follows. The standard approach biases the peer effect estimate because it does not distinguish between the common shock parameters $\alpha_{0,s}$ and $c_{0,s}$. We show that one way to distinguish between the shocks is to control for whether the students have friends or not. Addressing network endogeneity can also solve this issue (or at least reduces the bias) because it is equivalent to controlling for student attributes that influence their likelihood of forming links. These attributes are likely to be correlated to whether the student has friends or not. The increased peer effect estimate can then be explained by the standard approach biasing the peer effects downward, and controlling for network endogeneity is a means of reducing this bias.

4.2.3 Excluding "Fully Isolated" Students

The Add Health data set comprises 22% students without friends, including 11% who are not fully isolated (in the sense that they are friends of others because the adjacency matrix is directed), and 11% who have no friends and are not others' friends ("fully isolated" students). We construct a new subsample by excluding the "fully isolated" students from the sample. The new subsample includes 61,185 students from 139 schools. Using this sample, we estimate the proposed structural model and the standard peer effect model, which approximates student effort by GPA. We examine the two situations where the network is exogenous and endogenous.

Excluding the 22% isolated students from the sample will generate missing values in the network because the students who are not fully isolated are friends of others. As shown by Boucher and Houndetoungan (2022), this can result in a biased estimate of peer effects. Moreover, our structural

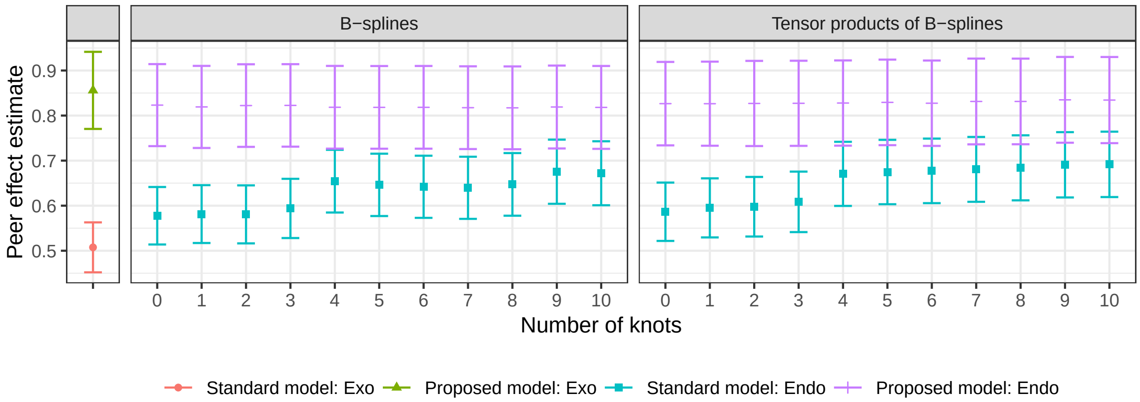

Figure 2: Peer effect estimates after controlling for network endogeneity

Note: This figure shows the 95% confidence interval of the peer effect estimates. The first column presents the estimates when we assume that the network is exogenous (Exo). The second and third columns present the estimates after controlling for the network endogeneity (Endo). We approximate $h_\eta(.)$ and $h_\epsilon(.)$ using cubic B-splines (second column) and the tensor products of cubic B-splines (third column). The unobserved attributes are estimated using a logit model with individual fixed effects. We define the nodes to evenly space the range of the estimators of $\mu^{in}_{0,s,i}$ and $\mu^{out}_{0,s,i}$.

model is equivalent to the standard linear-in-means specification when the network does not include isolated students. In contrast, the exclusion of the "fully isolated" students allows conducting a robustness analysis, as it does not involve a missing network data issue.

Figure 2 shows the 95% confidence interval of the peer effect estimates.[22] The estimate of the peer effect parameter using our structural model without controlling for network endogeneity is 0.878, which is close to the previous estimate (0.856) using the full sample. Moreover, the estimate using the standard linear-in-means model increases from 0.507 for the full sample to 0.561. This increase can be explained by the reduced proportion of isolated students. Indeed, the larger the proportion of isolated students, the greater the bias of the standard approach.

The peer effect estimate using the structural model decreases slightly to 0.854 after controlling for network endogeneity, whereas the estimate using the standard linear-in-means model increases to 0.728. The increase in the estimate using the standard approach has the same interpretation as that of the full sample. The bias of the standard approach is smaller than that obtained for the full sample. However, the difference is not substantial, suggesting that the bias is more likely because of isolated students who are friends of others than "fully isolated" students. Overall, the results are similar to those presented in Figure 2.

[22] Full tables for some specifications are presented in OA S.3 (see Tables S.4–S.5).

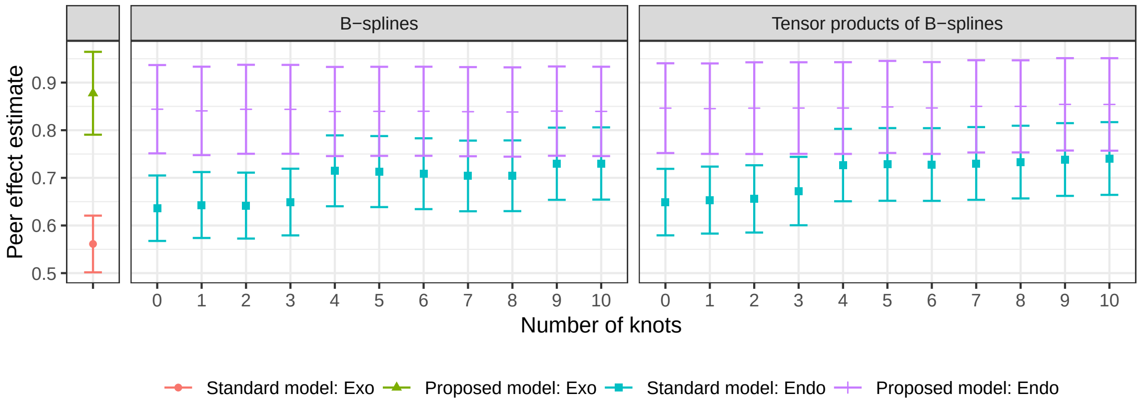

Figure 3: Peer effect estimates excluding "fully isolated" students

Note: This figure shows the 95% confidence interval of the peer effect estimates for the data excluding "fully isolated" students. The first column presents the estimates when we assume that the network is exogenous (Exo). The second and third columns present the estimates after controlling for network endogeneity (Endo). We approximate $h_\eta(.)$ and $h_\epsilon(.)$ using cubic B-splines (second column) and tensor products of cubic B-splines (third column). The unobserved attributes are estimated using a logit model with individual fixed effects. We define the nodes to evenly space the range of the estimators of $\mu^{in}_{0,s,i}$ and $\mu^{out}_{0,s,i}$.

5 Conclusion

This paper proposes a peer effect model in which students choose their level of effort and receive a corresponding GPA. Unlike standard models used in the literature to estimate peer effects on GPA, our model accounts for two types of common shocks at the school level and allows for identifying the effects on the effort, even though it is unobserved. We introduce common shocks that directly influence the GPA, irrespective of the effort level, and common shocks affecting students' preferences.

We show that both shocks do not have the same impact on GPA. Shocks exerted directly on the GPA without influencing academic effort do not involve a social multiplier, whereas preference shocks may involve a social multiplier effect. We provide an interpretation of the estimate of peer effects on academic efforts in terms of GPA. We also show that failure to disentangle both shocks results in a biased estimate of peer influence on GPA. Accounting for the difference between the shocks is equivalent to controlling for student heterogeneity on the basis of whether they have friends or not.

Our model leads to an econometric specification that raises identification issues. This occurs in particular because of the presence of unobserved school heterogeneity and students with no peers in the network. We establish the identification under straightforward conditions that can be verified in practice. We 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 analysis 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 1.6 times lower than that obtained from our proposed approach. Controlling for network endogeneity in the standard models reduces the bias.

Our framework is general and can be used to study peer effects on outcomes that cannot be directly controlled. 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' efforts. The observed outcome is generally the worker's productivity, whereas peer effects stem from the effort.

References

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

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's Who in Networks. Wanted: The Key Player," *Econometrica*, 74, 1403–1417.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2013): "Under pressure? The effect of peers on outcomes of young adults," *Journal of Labor Economics*, 31, 119–153.

Boucher, V., F. A. Dedewanou, and A. Dufays (2022): "Peer-induced beliefs regarding college participation," *Economics of Education Review*, 90, 102307.

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 . . . .

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

Brint, S. and A. M. Cantwell (2010): "Undergraduate Time Use and Academic Outcomes: Results from the University of California Undergraduate Experience Survey 2006," *Teachers College Record*, 112, 2441–2470.

Burke, M. A. and T. R. Sass (2013): "Classroom peer effects and student achievement," *Journal of Labor Economics*, 31, 51–82.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer Effects and Social Networks in Education," *The Review of Economic Studies*, 76, 1239–1267.

Carrell, S., R. L. Fullerton, and J. West (2009): "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27, 439–464.

Coleman, J. S. (1966): "Equality of educational opportunity," *Washington: U.S. Govt. Printer's Office*, [summary report].

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 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.

Dickerson, A., K. Maragkou, and S. McIntosh (2018): "The causal effect of secondary school peers on educational aspirations," *Centre for Vocational Education Research Discussion Paper*, 17.

Duncan, G. J., J. Boisjoly, and K. Mullan Harris (2001): "Sibling, peer, neighbor, and schoolmate correlations as indicators of the importance of context for adolescent development," *Demography*, 38, 437–447.

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

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.

Fruehwirth, J. C. and J. Gagete-Miranda (2019): "Your peers' parents: Spillovers from parental education," *Economics of Education Review*, 73, 101910.

Gagete-Miranda, J. (2020): "An aspiring friend is a friend indeed: school peers and college aspirations in Brazil," *Manuscript*.

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

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the Econometric Society*, 1029–1054.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003): "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18, 527–544.

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

Hatami, M., A. Kazemi, and T. Mehrabi (2015): "Effect of peer education in school on sexual health knowledge and attitude in girl adolescents," *Journal of Education and Health Promotion*, 4, 78.

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.

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.

Marotta, L. (2017): "Peer effects in early schooling: Evidence from Brazilian primary schools," *International Journal of Educational Research*, 82, 110–123.

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

Norris, J. (2020): "Peers, parents, and attitudes about school," *Journal of Human Capital*, 14, 290–342.

Plant, E. A., K. A. Ericsson, L. Hill, and K. Asberg (2005): "Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance," *Contemporary Educational Psychology*, 30, 96–116.

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

Sacerdote, B. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates*," *The Quarterly Journal of Economics*, 116, 681–704.

——— (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.

Zabel, J. E. (2008): "The Impact of Peer Effects on Student Outcomes in New York City Public Schools," *Education Finance and Policy*, 3, 197–249.

Zimmerman, D. J. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *The Review of Economics and Statistics*, 85, 9–23.

A Appendix: Proofs

A.1 Uniqueness of the Nash Equilibrium

By replacing the GPA with its expression given by Equation (1), we get 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

$$\hat{u}_{s,i}(e_{s,i}, \mathbf{e}_{s,-i}) = (c_{0,s} + \mathbf{x}'_{s,i}\boldsymbol{\beta}_0 + \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0 + \varepsilon_{s,i})(\alpha_{0,s} + \mathbf{x}'_{s,i}\boldsymbol{\theta}_0 + \delta_0 e_{s,i} + \eta_{s,i}) - \frac{e^2_{s,i}}{2} + \lambda_0 e_{s,i}\mathbf{g}_{s,i}\mathbf{e}_s. \quad \text{(A.1)}$$

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

$$e_{s,i} = \delta_0 c_{0,s} + \lambda_0 \mathbf{g}_{s,i}\mathbf{e}_s + \delta_0 \mathbf{x}'_{s,i}\boldsymbol{\beta}_0 + \delta_0 \mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0 + \delta_0\varepsilon_{s,i}. \quad \text{(A.2)}$$

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

$$\mathbf{e}_s = \delta_0 c_{0,s}\mathbf{1}_{n_s} + \lambda_0\mathbf{G}_s\mathbf{e}_s + \delta_0\mathbf{X}_s\boldsymbol{\beta}_0 + \delta_0\mathbf{G}_s\mathbf{X}_s\boldsymbol{\gamma}_0 + \delta_0\boldsymbol{\varepsilon}_s, \quad \text{(A.3)}$$

where $\mathbf{1}_{n_s}$ is an n_s-vector of ones and $\boldsymbol{\varepsilon}_s = (\varepsilon_{s,1}, \dots, \varepsilon_{s,n_s})'$. Equation (A.3) is a system of n_s linear equations in the effort. This system has a unique solution if $|\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s| \neq 0$, where $\mathbf{I}_{n_s}$ is the $n_s \times n_s$ identity matrix. The condition $|\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s| \neq 0$ is equivalent to saying that 1 is not an eigenvalue for $\lambda_0\mathbf{G}_s$. As $\mathbf{G}_s$ is a row-normalized matrix, the eigenvalues of $\lambda_0\mathbf{G}_s$ are in the closed interval $[-|\lambda_0|,\ |\lambda_0|]$.[23] Thus, if $|\lambda_0| < 1$, then $|\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s| \neq 0$ and the solution of Equation (A.3) is

$$\mathbf{e}_s = (\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)^{-1}(\delta_0 c_{0,s}\mathbf{1}_{n_s} + \delta_0\mathbf{X}_s\boldsymbol{\beta}_0 + \delta_0\mathbf{G}_s\mathbf{X}_s\boldsymbol{\gamma}_0 + \delta_0\boldsymbol{\varepsilon}_s). \quad \text{(A.4)}$$

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 $\boldsymbol{\eta}_s = (\eta_{s,1}, \dots, \eta_{s,n_s})'$ be the vector of the idiosyncratic error terms in Equation (1). Let also $\mathbf{y}_s = (y_{s,1}, \dots, y_{s,n_s})'$ be the GPAs' vector. From Equation (1), we have $e_{s,i} = (y_{s,i} - \alpha_{0,s} - \mathbf{x}'_{s,i}\boldsymbol{\theta}_0 - \eta_{s,i})/\delta_0$.

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

By replacing this expression in Equation (A.2), we get

$$\frac{y_{s,i}-\alpha_{0,s}-\mathbf{x}'_{s,i}\boldsymbol{\theta}_0-\eta_{s,i}}{\delta_0}=\frac{\lambda_0\mathbf{g}_{s,i}(\mathbf{y}_s-\alpha_{0,s}\mathbf{1}_{n_s}-\mathbf{X}_s\boldsymbol{\theta}_0-\boldsymbol{\eta}_s)}{\delta_0}+\delta_0(c_{0,s}+\mathbf{x}'_{s,i}\boldsymbol{\beta}_0+\mathbf{g}_{s,i}\mathbf{X}_s\boldsymbol{\gamma}_0+\varepsilon_{s,i}),$$
$$y_{s,i}=\kappa_{s,i}+\lambda_0\mathbf{g}_{s,i}\mathbf{y}_s+\mathbf{x}'_{s,i}\tilde{\boldsymbol{\beta}}_0+\mathbf{g}_{s,i}\mathbf{X}_s\tilde{\boldsymbol{\gamma}}_0+(\boldsymbol{\omega}_{s,i}-\lambda_0\mathbf{g}_{s,i})\boldsymbol{\eta}_s+\delta_0^2\varepsilon_{s,i},$$

where $\kappa_{s,i}=\delta_0^2c_{0,s}+(1-\lambda_0\mathbf{g}_{s,i}\mathbf{1}_{n_s})\alpha_{0,s}$, $\tilde{\boldsymbol{\beta}}_0=\delta_0^2\boldsymbol{\beta}_0+\boldsymbol{\theta}_0$, $\tilde{\boldsymbol{\gamma}}_0=\delta_0^2\boldsymbol{\gamma}_0-\lambda_0\boldsymbol{\theta}_0$, 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 $i\in\hat{\mathcal{V}}_s$, we have $\mathbb{E}(\mathbf{g}_{s,i}\mathbf{y}_s-\hat{\bar{y}}_s|\mathbf{G}_s,\mathbf{X}_s)=(\mathbf{x}'_{s,i}-\hat{\mathbf{x}}'_s)\dot{\boldsymbol{\beta}}+(\mathbf{g}_{s,i}\mathbf{X}'_s-\hat{\bar{\mathbf{x}}}'_s)\dot{\boldsymbol{\gamma}}$, where $\hat{\bar{y}}_s$, $\hat{\mathbf{x}}_s$, and $\hat{\bar{\mathbf{x}}}_s$ 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 $\hat{\mathcal{V}}_s$, 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 $\hat{\mathcal{V}}_s$ because of the matrix $\mathbf{J}_s$ that premultiplies the terms of Equation (7). let us take the previous equation in difference between two students i_1, j from $\hat{\mathcal{V}}_s$, where j is i_1's friend. This implies

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

where $\bar{y}^e_{s,i}=\mathbb{E}(\mathbf{g}_{s,i}\mathbf{y}_s|\mathbf{G}_s,\mathbf{X}_s)$ 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 path 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) of Equation (A.5) would not be influenced and we would have

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

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 $\Delta^l\bar{y}^e_{s,i_1}=\Delta^l\bar{y}^e_{s,j}$ 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 $\Delta^l\mathbf{y}_s=\lambda_0\mathbf{G}_s(\Delta^l\mathbf{y}_s)+(\Delta^l\mathbf{X}_s)\boldsymbol{\beta}_0+\mathbf{G}_s(\Delta^l\mathbf{X}_s)\boldsymbol{\gamma}_0$. This implies that $\Delta^l\mathbf{y}_s=(\mathbf{I}-\lambda_0\mathbf{G}_s)^{-1}\big((\Delta^l\mathbf{X}_s)\boldsymbol{\beta}_0+\mathbf{G}_s(\Delta^l\mathbf{X}_s)\boldsymbol{\gamma}_0\big)$. As $(\mathbf{I}-\lambda_0\mathbf{G}_s)^{-1}=\sum_{k=0}^{\infty}\lambda_0^k\mathbf{G}_s^k$, we can also write

$$\Delta^l\mathbf{y}_s=(\Delta^l\mathbf{X}_s)\boldsymbol{\beta}_0+\textstyle\sum_{k=0}^{\infty}\lambda_0^k\mathbf{G}_s^{k+1}(\Delta^l\mathbf{X}_s)(\lambda_0\boldsymbol{\beta}_0+\boldsymbol{\gamma}_0).\tag{A.7}$$

Equation (A.7) implies the GPA is influenced by the contextual variables if and only if $\lambda_0\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0 \neq 0$. 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

$$\Delta^l \bar{y}^e_{s,i_1} = \mathbf{g}_{s,i_1} \textstyle\sum_{k=1}^{\infty} \lambda_0^k \mathbf{G}_s^{k+1}(\Delta^l \mathbf{X}_s)(\lambda_0\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0). \tag{A.8}$$

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 $\lambda_0\mathbf{G}_s$, we obtain $\sum_{k=1}^{\infty} \lambda_0^k \mathbf{G}_s^{k+1}(\Delta^l \mathbf{X}_s)(\lambda_0\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0) = \lambda_0\mathbf{G}_s\Delta^l\mathbf{y}_s - \lambda_0\mathbf{G}_s(\Delta^l\mathbf{X}_s)\boldsymbol{\beta}_0$. By replacing the previous equation in (A.8), we get

$$\Delta^l \bar{y}^e_{s,i_1} = \lambda_0 \mathbf{g}_{s,i_1}\mathbf{G}_s\Delta^l\mathbf{y}_s, \tag{A.9}$$

because $\mathbf{g}_{s,i_1}\mathbf{G}_s(\Delta^l\mathbf{X}_s) = 0$. As $\mathbf{G}_s\Delta^l\mathbf{y}_s = (\Delta^l\bar{y}^e_{s,1}, \dots, \Delta^l\bar{y}^e_{s,n_s})'$, 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 $\Delta^l\bar{y}^e_{s,j}$ among students j who are i_1's friends. If Equation (A.6) holds true, that is, if $\Delta^l\bar{y}^e_{s,j} = \Delta^l\bar{y}^e_{s,i_1}$ for any j who is i_1's friend, this would mean that $\mathbf{g}_{s,i_1}\mathbf{G}_s\Delta^l\mathbf{y}_s = \Delta^l\bar{y}^e_{s,i_1}$ and Equation (A.9) would imply that $\Delta^l\bar{y}^e_{s,i_1} = \lambda_0\Delta^l\bar{y}^e_{s,i_1}$. This is where the contradiction would come from. Indeed, the previous equation is not compatible with Equation (A.6) since $\lambda_0 \neq 1$ by Assumption 2.1, and $\Delta^l\bar{y}^e_{s,i_1} \neq 0$ because $\lambda_0\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0 \neq 0$ (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 condition.

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, $\sum_{s=1}^{S}\mathbf{Z}_s'((\mathbf{I}_{n_s} - \lambda_0\mathbf{G}_s)\boldsymbol{\eta}_s + \boldsymbol{\varepsilon}_s)/n = o_p(1)$.*

The first half of Assumption A.1 suggests that the columns of design matrix $\mathbf{R} = [\mathbf{R}_1',\ \dots, \mathbf{R}_S']'$ and those of the instrument matrix $\mathbf{Z} = [\mathbf{Z}_1',\ \dots, \mathbf{Z}_S']'$ 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 Condition 3.1 and 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 $\boldsymbol{\psi}_0$ follows.

A.4 Proof of Proposition 3.2

The log-likelihood can be written as

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

where σ_η, σ_ϵ, and ρ denote arbitrary values of $\sigma_{0\eta}$, $\sigma_{0\epsilon}$, and ρ_0, $\mathbf{\Omega}_s(\hat\lambda,\tau,\rho) = \mathbf{I}_{n_s-2} + \tau^2\mathbf{F}_s'\mathbf{W}_s\mathbf{W}_s'\mathbf{F}_s + \rho\tau\mathbf{F}_s'(\mathbf{W}_s+\mathbf{W}_s')\mathbf{F}_s$, $\mathbf{W}_s = \mathbf{I}_{n_s} - \hat\lambda\mathbf{G}_s$, and $\tau = \sigma_\eta/\sigma_\epsilon$.
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:

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

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 $\mathbf{\Theta}$ 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 $\mathbf{\Theta}$. 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) $\mathbf{\Theta}$ is a compact subset of $\mathbb{R}^3$ and (ii) $\lim_{S\to\infty}\pi_{\min}(\mathbf{\Omega}_s(\hat\lambda,\tau,\rho)) > 0$ 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 $\mathbf{\Omega}_s(\hat\lambda,\tau,\rho)$ 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, $\tilde{\mathbf{\Omega}}_s(\hat\lambda,\tau,\rho) := \mathbf{I}_{n_s} + \tau^2\mathbf{W}_s\mathbf{W}_s' + \rho\tau(\mathbf{W}_s+\mathbf{W}_s')$ is also absolutely bounded in both row and column sums, and $\pi_{\max}(\tilde{\mathbf{\Omega}}_s(\hat\lambda,\tau,\rho)) < \infty$.[24] Moreover, as $\mathbf{\Omega}_s(\hat\lambda,\tau,\rho) = \mathbf{F}_s'\tilde{\mathbf{\Omega}}_s(\hat\lambda,\tau,\rho)\mathbf{F}_s$, we have $\pi_{\max}(\mathbf{\Omega}_s(\hat\lambda,\tau,\rho)) \leqslant \pi_{\max}(\tilde{\mathbf{\Omega}}_s(\hat\lambda,\tau,\rho)) < \infty$. Thus, $\frac{1}{n-2S}\sum_{s=1}^S\log|\mathbf{\Omega}_s(\hat\lambda,\tau,\rho)| < \infty$ for all σ_η^2, σ_ϵ^2, and ρ. As a result, $\frac{1}{n-2S}\sum_{s=1}^S\log|\mathbf{\Omega}_s(\hat\lambda,\tau,\rho)| - \frac{1}{n-2S}\sum_{s=1}^S\log|\mathbf{\Omega}_s(\lambda_0,\tau,\rho)| = o_p(1)$ (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

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

[24] We state and show in OA S.1.1 basic properties used throughout the paper. See properties P.5 and P.6.

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 $(\sigma_{\epsilon 0}^2,\ \tau_0,\ \rho_0)$.

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 $\mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S\right)$.

$$
\begin{aligned}
\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S) &= \tfrac{\sigma_{\varepsilon 0}^2}{n-2S}\textstyle\sum_{s=1}^{S}\mathbb{E}(\mathrm{Tr}(\boldsymbol{v}_s'\mathbf{F}_s\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\mathbf{F}_s'\boldsymbol{v}_s)|\mathbf{G}_1,\ldots,\mathbf{G}_S),\\
\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S) &= \tfrac{\sigma_{\varepsilon 0}^2}{n-2S}\textstyle\sum_{s=1}^{S}\mathbb{E}(\mathrm{Tr}(\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\mathbf{F}_s'\boldsymbol{v}_s\boldsymbol{v}_s'\mathbf{F}_s)|\mathbf{G}_1,\ldots,\mathbf{G}_S),\\
\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S) &= \tfrac{\sigma_{\varepsilon 0}^2}{n-2S}\textstyle\sum_{s=1}^{S}\mathrm{Tr}(\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\mathbb{E}(\mathbf{F}_s'\boldsymbol{v}_s\boldsymbol{v}_s'\mathbf{F}_s|\mathbf{G}_1,\ldots,\mathbf{G}_S)),\\
\mathbb{E}(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S) &= \tfrac{\sigma_{\varepsilon 0}^2}{n-2S}\textstyle\sum_{s=1}^{S}\mathrm{Tr}(\boldsymbol{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\boldsymbol{\Omega}_{0,s}),
\end{aligned}
\tag{A.10}
$$

where $\boldsymbol{\Omega}_s(\lambda_0,\tau_0,\rho_0)$.

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

Although $\mathbb{E}\left(\hat{\sigma}_\epsilon^2(\tau,\rho)|\mathbf{G}_1,\ldots,\mathbf{G}_S\right) = \tilde{\sigma}_\epsilon^{2*}(\tau,\rho)$, we cannot directly assert that $\operatorname{plim}\hat{\sigma}_\epsilon^2(\tau,\rho) = \tilde{\sigma}_\epsilon^{2*}(\tau,\rho)$. 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 S.1.2. 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 $(\tau_0,\ \rho_0)$ (see Newey and McFadden, 1994).

We set the following assumption.

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

The intuition of Assumption A.4 is as follows. After replacing σ_ϵ^2 with $\tilde{\sigma}_\epsilon^2(\tau,\rho)$ in Equation (9), the variable part of the concentrated log-likelihood is proportional to $\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^2(\tau,\rho)\mathbf{\Omega}_s(\hat{\lambda},\tau,\rho)|$, which is asymptotically equivalent to $\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau,\rho)\mathbf{\Omega}_s(\hat{\lambda},\tau,\rho)|$. Assumption A.4 implies that the value of $\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau,\rho)\mathbf{\Omega}_s(\hat{\lambda},\tau,\rho)|$ at $(\tau_0,\ \rho_0)$ 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 $\sigma_{\epsilon 0}^2\mathbf{\Omega}_s(\tau_0,\ \rho_0)$. Let $L^0(\sigma_\eta^2,\sigma_\epsilon^2,\rho) = -\sum_{s=1}^S \frac{n_s-2}{2}\sigma_\epsilon^2 - \frac{1}{2}\sum_{s=1}^S \log|\mathbf{\Omega}_s(\lambda_0,\tau,\rho)| - \sum_{s=1}^S \frac{1}{2\sigma_\epsilon^2}\boldsymbol{e}_s'\mathbf{F}_s\mathbf{\Omega}_s^{-1}(\lambda_0,\tau,\rho)\mathbf{F}_s'\boldsymbol{e}_s$. By Jensen's inequality, we have $\mathbb{E}(L^0(\sigma_\eta^2,\sigma_\epsilon^2,\rho) - L^0(\sigma_{\eta 0}^2,\sigma_{\epsilon 0}^2,\rho_0)|\mathbf{G}_1,\dots,\mathbf{G}_S) \leqslant 0$. This suggests that $(\tau_0,\ \rho_0)$ is a local maximizer of $\operatorname{plim}\frac{1}{n}L_c^*(\tau,\rho)$. The identification is guaranteed by Assumption A.4. If $(\tau_0,\ \rho_0)$ is not the global maximizer, then there would be another $(\tau_+,\ \rho_+) \in \mathbf{\Theta}$, such that $\operatorname{plim}\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau_0,\rho_0)\mathbf{\Omega}_s(\lambda_0,\tau_0,\rho_0)| = \operatorname{plim}\frac{1}{n}\sum_{s=1}^S \log|\tilde{\sigma}_\epsilon^{2*}(\tau_+,\rho_+)\mathbf{\Omega}_s(\lambda_0,\tau_+,\rho_+)|$, and Assumption A.4 would be violated. As a result, $(\tau_0,\ \rho_0)$ is globally identified and $(\hat{\tau},\ \hat{\rho})$ is a consistent estimator of $(\tau_0,\ \rho_0)$. The consistency of $\hat{\sigma}_\eta^2$ comes from Equation (A.10). We have $\operatorname{plim}\hat{\sigma}_\epsilon^2(\hat{\tau},\hat{\rho}) = \mathbb{E}(\hat{\sigma}_\epsilon^2(\tau_0,\rho_0)|\mathbf{G}_1,\dots,\mathbf{G}_S) = \sigma_{\epsilon 0}^2$; thus, $\operatorname{plim}\hat{\sigma}_\eta^2 = \tau_0^2\sigma_{\epsilon 0}^2 = \sigma_{\eta 0}^2$.

B Online Appendix

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

[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 $(\sigma_{\epsilon 0}^2,\ \tau_0,\ \rho_0)$. As shown in online Appendix (OA) S.1.3, 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 S.1).