Friendship Networks and Academic Success: The Impact of Class Size and Gender Composition*

Winfried Pohlmeier[a]
University of Konstanz, Zeppelin University

Livia Shkoza[b]
Kreditanstalt für Wiederaufbau

Derya Uysal[c]
University of Munich, CESifo

December 15, 2025

Abstract

This study examines how heterogeneity in classroom characteristics affects the influence of peer interactions on educational outcomes. Using unique friendship network data from German upper secondary schools, we investigate the roles of class size and gender composition in shaping academic achievement. To capture these dynamics, we propose a Multivariate Instrumental Variable (MVIV) approach that extends conventional homogeneous peer effects models and accounts for error heterogeneity across networks.

Our findings demonstrate that neglecting classroom-specific heterogeneity leads to biased conclusions. By integrating these variations, we uncover new structural insights into how classroom composition influences peer effects and, in turn, student performance.

Keywords: Network Heterogeneity, Peer Effects, Instrumental Variables, Educational Achievement, Class Size, Gender Composition

JEL classification: D85, L14, C3, I21

---

[*] Financial support by the Graduate School of Decision Sciences (GSDS), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984 and the IAAE is gratefully acknowledged. Derya Uysal gratefully acknowledges financial support by DFG through CRC TRR 190. The usual disclaimer applies. The authors report there are no competing interests to declare.

[c] Department of Economics, Ludwigstr. 33, D-80539 Munich, Germany. email: Derya.Uysal@econ.lmu.de.

[b] Kreditanstalt für Wiederaufbau (KfW), Palmengartenstraße 5-9, D-60325 Frankfurt am Main, Germany. email: livia.shkoza@kfw.de.

[a] Department of Economics, Universitätsstraße 1, D-78457 Konstanz, Germany. email: Winfried.Pohlmeier@uni-konstanz.de.

1 Introduction

Understanding how peers' outcomes influence an individual's educational achievements is a fundamental and widely studied question. Peer effects can operate across multiple dimensions, yet the extent to which this heterogeneity impacts student performance has received limited attention. For instance, peer effects may vary depending on the size of the peer group or the average performance of its members. Moreover, individuals within the same peer group may be affected differently due to differences in their socio-economic background. This paper examines the heterogeneity of peer effects at the network level, focusing specifically on the role of class size and gender composition. To investigate this, we introduce a novel Multivariate Instrumental Variable (MVIV) approach, which facilitates the estimation of heterogeneous peer effects while maintaining standard identification assumptions in network data.

Peer effects in education are often analyzed within a reduced-form framework, despite their inherently network-based nature. The limited number of studies adopting a network-theoretic approach is largely attributable to the scarcity of suitable network data. To the best of our knowledge, most existing research on network peer effects in education relies on the National Longitudinal Study of Adolescent to Adult Health (Add Health) (see Bramoullé et al., 2009; Boucher et al., 2014). In this paper, we examine heterogeneous peer effects on school grades using unique network data from 99 secondary school classes in Germany. By leveraging previously unused network data, our study provides new insights into how specific network characteristics influence individual educational performance. Specifically, we analyze the impact of gender composition and class size on student grades through peer interactions.

Peer effects in educational contexts are often analysed using Manski's linear-in-means framework (Manski, 1993). In a linear-in-means model, the individual outcome is modelled as a linear function of own characteristics and the average outcome of the peer group. The coefficient on the average peer outcome is referred to as the endogenous peer effect. It measures how an individual's outcome changes when the average achievement of her friends increases. Because peers influence each other simultaneously, this coefficient is not identified by OLS in general.

Observed friendship networks provide additional structure that can be exploited to identify and estimate endogenous peer effects. Bramoullé et al. (2009) show that, in linear-in-means models with observed networks, endogenous peer effects can be identified by using averages of the exogenous characteristics of second-order (and higher-order) peers as instruments. The key property is that the covariates of "peers of peers" shift friends' outcomes but, conditional on the network, do not directly affect the individual's own outcome. This generates exogenous variation that identifies the endogenous peer effect, in close analogy to IV–GMM strategies used in spatial lag models (e.g., Kelejian and Prucha, 1998, 1999); see also Bramoullé et al. (2020) for a recent survey.

The standard linear-in-means model assumes that peer effects operate exclusively through the average outcome of one's peers. While this specification is widely used, it imposes a strong and somewhat ad hoc restriction: it presumes that only the peer-group mean matters for individual behaviour. At the same time, a different strand of the literature models peer influence as operating through the aggregate (or total) outcome of the peer group, reflecting the idea that peers work as complements and that the productivity of a student's own effort increases with the overall quality of her friends. Relying on only one of these mechanisms is also a restrictive modelling choice.

The composite peer-effects model of Liu et al. (2014) provides a unifying framework that nests both channels. In this model, peers can affect an individual through the average outcome of her friends and through the aggregate (total) outcome of the peer group. Accordingly, the composite specification contains two endogenous peer effects: (i) a *local-*

average peer effect, capturing the influence of the average achievement of one's friends, and (ii) a *local-aggregate peer effec*t, capturing the influence of the total achievement of the peer group. These components arise naturally from utility-maximising behaviour in network games (e.g. Calvó-Armengol et al., 2009; Blume et al., 2015). Within this framework, the local-average peer effect has a behavioural interpretation as a desire to conform to the group norm, whereas the local-aggregate peer effect reflects the strength or overall resources of the peer group.

In our application, we extend this composite model by allowing both peer-effect components to vary with class characteristics, specifically class size and gender composition, which enables us to examine how the strength and direction of peer effects change across different classroom environments.

In most peer effect studies, the implicit assumption is that endogenous peer effects are homogeneous across networks. Consequently, empirical evidence on how network attributes affect peer effects in education, particularly those exploiting network information, is rather limited. Notable exceptions are Calvó-Armengol et al. (2009) and Lin (2014). Calvó-Armengol et al. (2009) investigate the relationship between peer effects and network topology, providing graphical evidence that the strength of the network effect varies with certain structural network measures, such as density, asymmetry, and redundancy. Similarly to us, Lin (2014) examines how network heterogeneity affects peer effects in education. She divides the sample into subsamples based on class size and gender to estimate peer effects for each group separately. She finds significant differences between peer effects in large and small classes but does not find a significant difference between the peer effects of the two subsamples by gender proportion. In contrast to these studies, we examine the link between network characteristics and endogenous peer effects in a structural framework using a regression approach. Masten (2018) also explores peer effect heterogeneity in education, but in his study, the endogenous peer effect is pair-specific and purely random, not driven by observable network attributes, which distinguishes our approach.

A more frequently studied type of heterogeneity in peer effect models is heterogeneity by gender of the peers. The idea is that female and male peers may have different peer effects, and their influence might also vary, effectively leading to a four-dimensional peer effect. For example, using a laboratory experiment on individual performance at work, Beugnot et al. (2019) estimate gender-specific peer effects based on the model by Arduini et al. (2014). From a technical perspective, the approach of Arduini et al. (2014) is similar to the composite model, as both allow for different peer effects arising from multiple networks. However, Arduini et al. (2014) propose a model with two endogenous variables, two adjacency matrices, and both within- and between-group interactions.

One strand of the literature on gender effects in school performance focuses on differences in outcomes for girls in single-sex and coeducational classes (see, for a review Mael et al., 2005; Morse, 1999). Observational studies yield somewhat mixed results: some provide evidence of positive effects of single-sex schools, while others suggest no significant difference. Another line of research examines the gender peer effect using exogenous variation in gender due to experimental or quasi-experimental research designs. For instance, Hoxby (2002) and Lavy and Schlosser (2011) find that the proportion of female students positively impacts students' cognitive achievement. However, none of these studies explicitly considers the network structure within classrooms. The common thread across these studies is that gender (or the gender ratio) is included as a regressor in reduced-form equations. In contrast, our structural approach considers the gender ratio's impact on academic success through the endogenous peer effect. This indirect effect has a clear structural interpretation: observed differences in academic success between classes with different gender compositions stem from distinct collaborative patterns captured by peer effects.

Finally, our study contributes to the ongoing debate on the impact of class size on academic achievement, a topic for which empirical evidence remains inconclusive. For example, Hanushek (1996) and Hoxby (2000) find no significant effect of class size reduction on student performance. Similarly, Dobbelsteen et al. (2002) report that students in smaller classes do not necessarily achieve better academic outcomes and, in some cases, even perform worse than their peers in larger classes. In contrast, studies by Angrist and Lavy (1999) and Krueger (1999), along with more recent research by Heinesen (2010) and Fredriksson et al. (2012), find a substantial positive effect of smaller class sizes on academic achievement. Much like research on gender effects, most empirical studies on class size focus on its direct impact on academic success within reduced-form frameworks and therefore do not identify the channels through which class size operates in determining student achievement. By explicitly modelling heterogeneous peer effects, our approach sheds light on how class size shapes peer behaviour and, in turn, academic achievement.

The outline of this paper is as follows. In Section 2, we introduce the composite network model and elaborate on its identification conditions. Moreover, we present the new MVIV approach for modelling and estimating heterogeneous peer effects. In Section 3, we describe our network data and discuss further implementation issues. Finally, Section 5 concludes and provides an outlook for future research.

2 The Network Model and Estimation

As a baseline model, we rely on the composite peer effects model by Liu et al. (2014). We consider a set of L independent networks, each consisting of n_l agents (the network size) and an overall number of observations $N = \sum_{l=1}^{L} n_l$. The social connections for network l are indicated by the adjacency matrix $A_l = [a_{ij,l}]$, where $a_{ij,l} = 1$ if agent i in network l is connected to agent j, and $a_{ij,l} = 0$ otherwise. The diagonal elements $a_{ii,l}$ are set to zero. The reference group of agent i in network l is the set of their peers, and the size of the reference group is the outdegree $a_{i,l} = \sum_{j=1}^{n_l} a_{ij,l}$. Let $G_l = [g_{ij,l}]$ be the row-normalized adjacency matrix of network l, with elements $g_{ij,l} = a_{ij,l}/a_{i,l}$, where, by construction, $0 \leq g_{ij,l} \leq 1$ and $\sum_{j=1}^{n_l} g_{ij,l} = 1$. For agent i, the component $(G_l y_l)_i = \sum_j g_{ij,l} y_{j,l}$ is the average outcome of her peers, whereas $(A_l y_l)_i = \sum_j a_{ij,l} y_{j,l}$ is the sum (aggregate) of peers' outcomes.

The econometric specification for the composite peer effects model is of the form

$$y_{i,l} = \eta_l + x'_{i,l}\gamma + \sum_{j=1}^{n_l} g_{ij,l} x'_{j,l}\gamma_g + \beta_{a,l} \sum_{j=1}^{n_l} a_{ij,l} y_{j,l} + \beta_{g,l} \sum_{j=1}^{n_l} g_{ij,l} y_{j,l} + \varepsilon_{i,l}, \tag{1}$$

for $i = 1, \ldots, n_l$ and $l = 1, \ldots, L$ and $\mathrm{E}[\varepsilon_l | \, x_l, A_l, \eta_l] = 0$[1]. In this specification, γ captures the direct effect of an individual's own characteristics $x_{i,l}$, while γ_g measures *contextual* effects of the average peer characteristics $\sum_j g_{ij,l} x_{j,l}$. The coefficients $\beta_{g,l}$ and $\beta_{a,l}$ are the network-specific endogenous peer-effect parameters. The coefficient $\beta_{g,l}$ captures the local-average peer term $\sum_j g_{ij,l} y_{j,l}$ and therefore measures how $y_{i,l}$ responds to a change in the average outcome of i's peers, holding the size of the peer group fixed. The coefficient $\beta_{a,l}$ captures the local-aggregate peer term $\sum_j a_{ij,l} y_{j,l}$ and therefore measures how $y_{i,l}$ responds to a change in the total outcome of the peer group, which depends both on peer achievement and on the number of peers. Finally, the network-specific intercept η_l represents correlated effects at the classroom level.

Bramoullé et al. (2009) assume that error terms in (1) are independent across observations but heteroskedastic. Here we also impose within-network independence of the error terms but allow for clusters at the network level.

Under homogeneity of the peer effects, i.e., $\beta_{a,l} = \beta_a$ and $\beta_{g,l} = \beta_g$, our model reduces to the (homogeneous) composite model of Liu et al. (2014). The composite model nests two specifications, the local-aggregate ($\beta_{g,l} = 0$) and the local-average model ($\beta_{a,l} = 0$).

---

[1] This exogeneity assumption is strong but rather common. It requires that there are no unobserved factors that affect both the link formation and the outcome variable. Among others, Bramoullé et al. (2009) and Liu et al. (2014) rely on the same assumption in their applications to the schooling context. Alternatively, one could rely on a model with endogenous link formation. See for example Auerbach (2022) and Johnsson and Moon (2021) for endogenous link formation approach. However, incorporating heterogeneous peer effects into such models is not straightforward. We leave this for future research and focus on the heterogeneity aspect of peer effects.

In matrix notation, the general specification (1) for network l takes the form

$$\mathbf{Y}_l = \eta_l \iota_{n_l} + X_l \gamma + X_l^g \gamma_g + \beta_{a,l} y_l^a + \beta_{g,l} y_l^g + \varepsilon_l, \tag{2}$$

where $\mathbf{Y}_l = (y_{1,l}, \ldots, y_{n_l,l})'$, $y_l^a = A_l \mathbf{Y}_l$, $y_l^g = G_l \mathbf{Y}_l$, $X_l = \left[x_{1,l}\ x_{2,l} \cdots\ x_{n_l,l}\right]'$ and $X_l^g = G_l X_l$, while $\varepsilon_l = (\varepsilon_{1,l}, \ldots, \varepsilon_{n_l,l})'$ and ι_{n_l} is an $n_l \times 1$ vector of ones.

We partial out the network-specific level effects via transformation by multiplying (2) by $J_l = I_{n_l} - \frac{1}{n_l} \iota_{n_l} \iota_{n_l}'$ from left. Because $J_l \iota_{n_l} = 0$, the transformed model is

$$\tilde{\mathbf{Y}}_l = \tilde{X}_l \gamma + \tilde{X}_l^g \gamma_g + \beta_{a,l} \tilde{y}_l^a + \beta_{g,l} \tilde{y}_l^g + \tilde{\varepsilon}_l\,, \tag{3}$$

where the $\tilde{}$ symbol denotes left multiplication by J_l, i.e., $\tilde{\mathbf{Y}} = J_l \mathbf{Y}_l$. The regressor matrix $\mathbf{W}_l = \left[\tilde{X}_l\ \tilde{X}_l^g\ \tilde{y}_l^a\ \tilde{y}_l^g\right]$ with dimension $n_l \times k_w$, where $k_w = 2(1 + k_x)$, has column full rank.

Following Bramoullé et al. (2009) and Liu et al. (2014), the endogenous peer terms in (3) are instrumented using (i) the *aggregate exogenous characteristics of peers*, $A_l X_l$, for the local–aggregate component, and (ii) the *average characteristics of second–order peers*, $G_l^2 X_l$, for the local–average component. The key restriction is that, conditional on an individual's own covariates and on the observed composition of her direct peer group (X_l and $G_l X_l$), neither $A_l X_l$ nor $G_l^2 X_l$ affects $y_{i,l}$ directly. Instead, these variables influence outcomes only through peers' outcomes: $A_l X_l$ shifts the aggregate quality of the peer group, whereas $G_l^2 X_l$ shifts peers' peers and therefore changes the mean behaviour of the direct peers. This provides excluded variation in $A_l Y_l$ and $G_l Y_l$, the two endogenous peer channels, while satisfying the exclusion restriction. The formal identification conditions for the local–average and local–aggregate components are given in Bramoullé et al. (2009) and Liu et al. (2014).

Intuitively, the following happens: each student's outcome depends on (i) the average outcome of her friends, (ii) the aggregate outcome of her friends, (iii) her own characteristics, and (iv) the average characteristics of her friends. If there are students who are friends of her friends but not directly connected to her, then variation in the exogenous characteristics of these "friends of friends" affects her friends' outcomes (because each friend's outcome depends on the characteristics of *their* peer group). This in turn shifts the peer–average $G_l Y_l$ but does not affect her outcome directly, making $G_l^2 X_l$ a valid instrument for the local–average peer effect. We demonstrate this formally in Appendix A.1.

A parallel argument applies to the aggregate channel. The term $A_l X_l$ captures the total exogenous characteristics of a student's peer group, who the peers are and how many there are. Holding fixed the observed composition of the peer group, changes in $A_l X_l$ shift the aggregate peer outcome $A_l Y_l$ but not the individual's outcome directly, once the average peer composition is controlled for.

Finally, successful use of both instruments requires sufficient heterogeneity in network structure. Identification hinges on network sparsity: friends of friends must not be the same individuals as one's own friends, and students must differ in their outdegree (the size of their peer group) and in the composition of their second–order peers. When the network is sufficiently sparse and these dimensions vary across students, $G_l^2 X_l$ and $A_l X_l$ generate exogenous shifts in peers' outcomes along distinct dimensions, allowing separate identification of the local–average and local–aggregate peer-effect parameters.

In order to introduce network heterogeneity we assume next that the parameters $\beta_{a,l}$ and $\beta_{g,l}$ for network-specific peer effects can be explained by a set of network-specific observable factors

$$\beta_{j,l} = \beta_{0,j} + v_l' \beta_j \text{ for } \qquad j = a, g \,, \tag{4}$$

where v_l is a $k_v \times 1$ vector of network-specific characteristics, and $\beta_{0,a}$ and $\beta_{0,g}$ denote the intercept terms for $j = a, g$, respectively. Inserting (4) into (3) gives in matrix notation

$$\tilde{\mathbf{Y}}_l = \mathbf{X}_l \, \theta + \tilde{\varepsilon}_l \, , \tag{5}$$

with:

$$\begin{aligned} \mathbf{X}_l &= \big[\tilde{X}_l \quad \tilde{X}_l^g \quad \tilde{y}_l^a \quad \tilde{y}_l^g \quad \tilde{y}_l^a v_l' \quad \tilde{y}_l^g v_l' \big] \\ \theta &= \big(\gamma', \gamma_g', \beta_{0,a}, \beta_{0,g}, \beta_a', \beta_g' \big)' \end{aligned}$$

It is important to note that the $n_l \times k$ regressor matrix $\mathbf{X}_l$ with $k = 2(1 + k_x + k_v)$ has no full column rank because the interaction terms with v_l imply perfect multicollinearity. Consequently, GMM-estimation of θ for a single network becomes infeasible (see Appendix A.2 for a more detailed exposition).

Multivariate Instrumental Variable Estimation (MVIV)

To estimate the network system in (1), we propose a GMM approach—Multivariate Instrumental Variable Estimation (MVIV)—which recovers the parameter vector θ under heterogeneous peer effects for both types of peer behaviour.

By stacking the single network representations (5) into a hyper-system, we obtain:

$$\mathbf{Y} = \mathbf{X} \, \theta + \boldsymbol{\varepsilon} \tag{6}$$

where:

$$\tilde{\mathbf{Y}} = \begin{pmatrix} \tilde{\mathbf{Y}}_1 \\ \tilde{\mathbf{Y}}_2 \\ \vdots \\ \tilde{\mathbf{Y}}_L \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_L \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_L \end{pmatrix}.$$

Moreover, define $\mathbf{Z} = \text{diag}\big[\,\mathbf{Z}_l\,\big]$ as the $\big((\sum_l n_l) \times qL\big)$-dimensional block-diagonal instrument matrix with $\mathbf{Z}_l$ as the $n_l \times q$-dimensional network-specific instrument matrix, where $q \geq k$. Then, using the efficient GMM estimation approach for the system of networks of the parameters in (6) under network-specific heteroskedasticity leads to the following MVIV estimator:

$$\hat{\theta}_{MVIV} = \left(\sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l\right)^{-1} \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \tilde{\mathbf{Y}}_l \,, \tag{7}$$

where $\mathbf{P}_l = \mathbf{Z}_l \left[\, \mathbf{Z}_l' \, \mathbf{D}_l \, \mathbf{Z}_l \,\right]^{-1} \mathbf{Z}_l'$ and $\mathbf{D}_l = \text{diag}\big[e_l^2\big]$. The specific model assumption and the proof are given in Appendix A.3.

Note that, allowing the peer–effect coefficients to vary with class–level characteristics v_l does not introduce new sources of endogeneity. The only endogenous objects remain the two peer aggregates $A_l Y_l$ and $G_l Y_l$. The heterogeneous specification introduces additional endogenous regressors of the form $\tilde{y}_l^a v_l$ and $\tilde{y}_l^g v_l$, but these are simply the original endogenous peer variables multiplied by the class–level characteristics v_l, which are constant within each classroom and exogenous. Because v_l enters only as a known scalar multiplier, introducing it does not create any additional endogenous component beyond $A_l Y_l$ and $G_l Y_l$. The original instrument set $\{A_l X_l,\ G_l^2 X_l\}$ therefore remains valid for the heterogeneous model and, as long as $q \geq k$ and there is sufficient variation in class characteristics, provides enough exogenous variation to shift both the level and the interacted endogenous peer terms through peers' and second–order peers' characteristics. Interacting the instruments with v_l (e.g. $v_l A_l X_l$) would not add identifying variation and would create perfect multicollinearity, because v_l is constant within each classroom and factors out of the corresponding first–stage moment conditions. Thus, the heterogeneous specification inherits the identification conditions of the homogeneous composite model and additionally requires sufficient cross-class variation in the network-specific factors for system-level identification.

In the Web-appendix to this paper we demonstrate that our model can also be estimated by an instrumental variable minimum distance (MDIV) strategy. This alternative approach is asymptotically equivalent to MVIV and simple to implement.

3 Data

Our empirical study uses data from the *Gymnasiasten-Studie* (CAESR, 2007), a longitudinal survey of 3,385 10th-grade students enrolled in upper secondary schools (*Gymnasium*) in North Rhine-Westphalia, Germany, during the 1969/1970 academic year.[2] The initial dataset includes 121 classes from 68 randomly selected upper secondary schools, providing information on students' prior school grades and individual characteristics such as gender and age. Additionally, during the initial data collection period, a standard psychometric Intelligence Structure Test (IST) was administered in the classroom. Approximately ten years after the original survey, students' grades were retrieved from school archives. Central to our study is the network data collected through the Sociometric Test in the *Gymnasiasten-Studie* during the initial survey. To construct the adjacency matrices A_l and G_l for each class, we use data on each student's assessment of whom they liked in the class based on the following question:[3]

> "*In every class, there are fellow students who one likes more than others in the class. Some others one finds quite unpleasant, and that is quite normal. Kindly first list the students who you personally like a lot.*"

As previously noted, most empirical studies on network peer effects in education rely on

---

[2] Although the survey is longitudinal, classroom network data were collected only at the initial survey. Three follow-up surveys were conducted, each conducted approximately ten years apart, but these later surveys experienced significant panel attrition.

[3] The original question in German is "*In jeder Klasse gibt es Mitschüler, die man sympathisch findet und die man mehr als andere in der Klasse gut leiden kann. Einige findet man sicher recht unsympathisch, und das ist auch ganz normal so. Würden sie nun zunächst einmal die Schüler nennen, die Sie persönlich gut leiden können.*"

data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Our unique network dataset differs from Add Health in several key aspects. One major distinction is that in our dataset, students nominate their friends within the classroom, whereas in Add Health, nominations are made at the school level. In the year the survey was conducted, 10th-grade students in German upper secondary schools were taught in stable class units with relatively fixed compositions over multiple years. Consequently, collecting friendship nominations at the class level is likely more appropriate.

More importantly, the Add Health survey restricts respondents to naming up to ten best friends (five male and five female), which may introduce a truncation issue that our dataset avoids. In fact, Griffith (2021) demonstrates that peer effects estimated from censored data tend to be underestimated.

We construct our sample by merging information from three different sources: student surveys (Hummell et al., 1970), administrative data from school archives (n et al., 1986), and the sociometric test (Hummell et al., 2018). Among the 68 schools that initially participated in the survey, 6 did not agree to the collection of administrative data due to privacy concerns. The remaining 62 schools accounted for 91% of the initial sample of students (3,010 students). Due to incomplete information in the school archives, it was only possible to collect information on grades for approximately 2,700 students. Furthermore, around 150 students neither answered the student survey nor participated in the sociometric test. Lastly, we delete an observation when any of the variables used in the empirical model was missing. This leaves a sample of 2,409 students and 101 classes. In network data, removing individuals due to missing information might have different consequences than in cross-sectional data. Specifically, it might lead to mismeasurement in the network if a friendship link is affected by this. Following the identification approach by Bramoullé et al. (2009), an implicit assumption in the literature, which we also adopt here, is that the links among individuals are perfectly observed. The econometric implications of measurement error in networks remain largely unexplored. However, Lewbel et al. (2023) demonstrate in a recent study that relatively small measurement errors in network data can be safely ignored in estimation. In particular, they show that the instrumental variable estimators like Bramoullé et al. (2009) and their standard errors, remain consistent and valid, as long as the number and size of measurement errors in an observed adjacency matrix is relatively small. Although they do not consider a hybrid model with local-aggregate and local-average effects together, they show for both cases separately that the usual asymptotics provide a good approximation for inference. We believe that in our case the measurement errors in the adjacency matrix caused by deleting individuals due to missing observations are small enough to be ignored safely as they suggested. To keep the network structure in classes as close as possible to the original one we further excluded two extreme classes which lost a significant portion of the network. Thus the final estimation sample comprises 2,398 students across 99 classes.[4] Table A1 provides summary statistics for the variables used in our empirical analysis, along with some network statistics. The left panel of the table shows the variables before applying the class size restriction, demonstrating that the sample means are not significantly affected by this restriction.

Academic performance is measured by the average of the final grades (GPA) for all compulsory and elective courses at the end of the school year 1969/70. As mentioned before, we use the administrative data collected from the school archives to construct the GPA. At the time of the survey, the choice of different courses within a class was very limited, i.e., all students of the same class had to take the same courses. Selection of certain specializations (e.g., languages, mathematics and sciences, humanistic secondary school) occurred at the time of choosing the specific secondary school. Therefore, the GPA within a given class is based on mostly the same subjects. The grades are measured in terms of the German grading system: with 1 ('very good') being the best grade and 6

---

[4] We estimated the model excluding smaller classes using different thresholds for robustness checks, but the results remained qualitatively consistent. The results are available upon request.

('insufficient') as the worst grade. In addition to the overall GPA, we also closely examine scores in Mathematics and German to detect potential differences in peer behaviour across subjects.

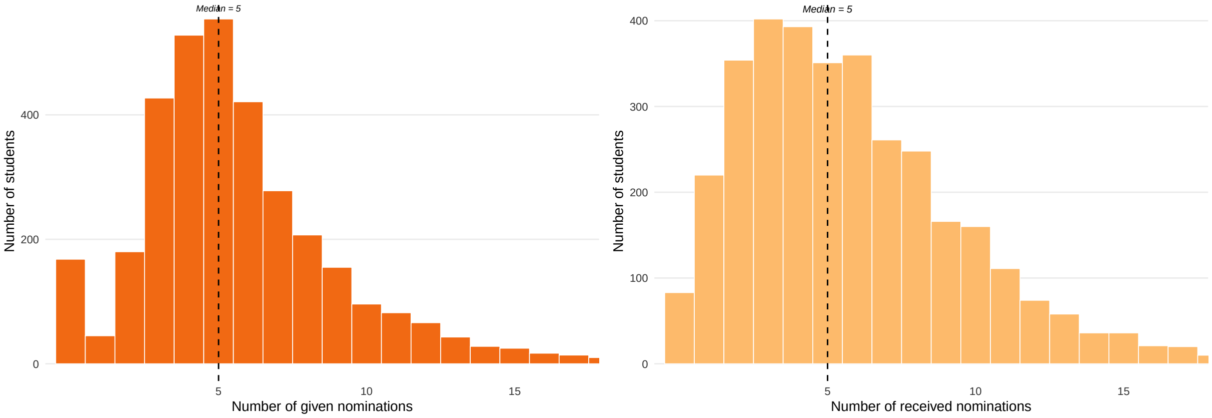
Individual heterogeneity in our model is captured by the student's IQ score, the GPA from the previous school year, and the student's age. The IQ is constructed as a standardized, normed composite score combining four IQ subtests, scaled using intercorrelations and age-specific norms, where higher values indicate higher intelligence (Hummell et al., 1979). We control for the GPA from previous year as a proxy for the overall school performance at the beginning of the survey year. In order to account for network heterogeneity in the peer effects of the local-average and local-aggregate models, we allow the two peer effect parameters $\beta_{a,l}$ and $\beta_{g,l}$ in Equation (2) to depend on class-specific factors. As such factors we use the relative class size, i.e., the size of the network relative to the largest class size, and the fraction of girls in the class.

As pointed out above, the literature on the effects of the class size and gender (ratio) on school outcomes is very rich. In general, the main consideration is the direct causal link from class size to the outcome. Experimental evidence about gender diversity and performance shows that team collaboration is greatly improved by the presence of women in the group (see, for example, Bear and Woolley, 2011, and references therein). Therefore, it seems reasonable to look for a potential indirect link through heterogeneous peer effects.

The network density is defined as the ratio of all connections in a network to the number of potential connections. Thus, the denser a network is, the closer its density is to unity. In our sample, network density varies between 0.05 and 0.35, with 58 classes having a lower density than the mean, indicating rather sparse networks.

To obtain a better understanding of the network structure and its potential role for peer effects, we take a look at the summary statistics of the friends' nominations. The average number of friends a student names (outdegree) is 5.48, indicating that students give considerable thought to nominating their friends. Figure 1 depicts the distribution of friendship links. The distributions of outdegrees and indegrees—where the indegree is the number of times a student is named by others—are consistent with the sparsity of the networks. Most students name around five classmates they like, and very few name more than ten peers.

Figure 1: Distribution of Friendship Nominations (Outdegrees and Indegrees)



Histograms of the number of friends each student names (outdegrees) and the number of times each student is named as a friend (indegrees). Both distributions have a median of 5 and a mean of 5.8.

4 Empirical Results

The primary specification in our study is the heterogeneous composite model given by (1), which incorporates the local-aggregate and the local-average peer effects model as nested specifications. Although most of the empirical studies focus on peer effects resulting from norm behaviour, and therefore favoring the local-average model, ex-ante, both hypotheses on how peers affect individual educational achievement are reasonable. In fact, the two effects may complement or even counteract each other.

In our empirical study we mainly focus on the GPA as the students' outcome variable. However, since peer effects may operate differently depending on the subject, we also study the peer effects for Math and German (see Tables A2 and A3 in the Appendix). As predetermined or exogenous explanatory variables we use the GPA of the previous year, IQ, and age, as well as their counterparts for the student's peer group. As network-specific characteristics we consider the relative class size and the fraction of girls.

The exogeneity assumption on the friendship links conditional on observed characteristics and network fixed effects might fail if there are unobserved factors affecting both link formation and the outcome variable. We believe that by including previous GPA, IQ, and age we control for the most important factors (or proxies for such factors) that might affect current grades as well as the formation of friendship links. Specifically, we argue that previous GPA reflects overall school performance at the beginning of 10th grade, IQ captures cognitive ability, and age serves as a proxy for age-specific skill differences[5].

Table 1 summarizes the estimation results for the composite, local-aggregate, and local-average models with heterogeneous peer effects estimated using the MVIV approach with globally differenced variables. As explained in Section 2, the identification of the peer effects results from instruments that shift peers' outcomes but do not directly affect the individual: (i) the average characteristics of second-order peers for the local-average component and (ii) the aggregate characteristics of peers for the local-aggregate component. These variables affect a student's friends through their own peer environments, thereby generating exogenous variation in peer achievement and in the interaction terms with class characteristics. Because second-order peers and aggregate peer characteristics affect the individual only through their influence on direct peers, they plausibly satisfy the exclusion restriction. This provides the necessary variation to identify the endogenous peer effect and its heterogeneity with respect to class size and gender composition. To facilitate the interpretation of the estimation results, we centered the network-specific characteristics around their means. Consequently, the two intercept terms in (4) represent the aggregate

---

[5] Since we are studying 10th graders, age variation largely reflects grade retention.

and average peer effects for a class with average characteristics.

First and foremost, our results show that accounting for heterogeneity in peer effects at the network level is crucial. We find strong evidence that peer effects vary with class size and gender composition across subjects. Several studies that assume homogeneous peer effects fail to find sufficient statistical evidence supporting their existence (e.g., Boucher et al., 2014; Liu et al., 2014). Similarly, our estimates remain mostly insignificant when network heterogeneity is ignored (see Table A4 in the Appendix).

Second, the mechanism through which peers influence student performance is also important, as both the local-aggregate and local-average mechanisms positively influence student achievement in a representative classroom with average relative size and gender composition. The Wald statistics in the last row of Table 1 indicate that the heterogeneous local models are rejected in favor of the heterogeneous composite model. In what follows, we therefore discuss only the coefficients estimated for the composite model.

In the composite model, the local-average intercept is strongly significant, whereas the local-aggregate intercept is only significant at the 10% level.

This implies that an individual's performance improves when their peers perform better, either individually or on average in a typical class, i.e. in a class with average female share and average class size. The local-aggregate peer effect is proportional to the number of peers (the student's outdegree), meaning that a larger peer group amplifies the impact on an individual's performance. To quantify this effect, we calculate the impact of a one-unit change in the GPA of the peers on the individual's GPA for both models. Nevertheless, the local-average intercept is so much larger than the local-aggregate one that no reasonable number of peers will make the local-aggregate effect exceed the local-average effect. More specifically, in a class with average characteristics, the local-aggregate effect would exceed the local-average effect only if the student had more than 360 peers. Given that the median outdegree in our sample is 5 (see Figure 1), it is safe to say that the local-average peer effect dominates the local-aggregate effect.

Table 1 shows for GPA that class-level characteristics shape the local-aggregate and local-average peer effects in different ways. The effect of the female share is insignificant in the local-average part but significantly negative in the local-aggregate part. Interpreting the local average effect as the one reflecting norm behavior in peer groups, we can conclude that the norm behaviour in classrooms does not change due to gender composition. By contrast, the local-aggregate effect reflects performance spillovers from the overall achievement level of classmates. The negative coefficient on female share in this component indicates that these achievement-based peer spillovers become weaker as the proportion of female students in the class increases.

Class size has a significantly negative effect on the local-average peer effect, but a positive and significant effect on the local-aggregate peer effect. A positive coefficient on a class characteristic in a given peer-effect component means that this peer effect becomes stronger as the characteristic increases, whereas a negative coefficient implies that the corresponding peer effect weakens when the characteristic is larger. Taken together, these estimates indicate that the local-aggregate effect tends to become stronger as class size increases, while the local-average effect tends to become weaker, and that the opposite pattern holds for the female share. Behaviourally, this suggests that larger classes strengthen achievement-based spillovers but dampen norm-related peer influences, whereas variation in gender composition mainly operates by weakening achievement-based spillovers.

To give a sense of magnitudes of the effects, with the female share held at its average, the local-aggregate peer effect increases from a value not statistically different from zero for the smallest class size to about 0.0019 for the largest.[6] The local-average peer effect is estimated to be 0.2384 (standard error 0.0342) for the smallest class and 0.1260 (standard

---

[6] For the smallest class size, the estimated local-aggregate peer effect is -0.0009 with a standard error of 0.0009, so it is not statistically different from zero. For the largest class size, the estimate increases to 0.0019 with a standard error of 0.0006, holding the female share at its average.

Table 1: MVIV Estimation Results for GPA

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0005*	0.0005	
	(0.0003)	(0.0004)	
Relative Class Size	0.0045**	0.0044*	
	(0.0022)	(0.0025)	
Female Share	-0.0029***	-0.0022***	
	(0.0008)	(0.0009)	
Local-average peer effect			
Intercept	0.1810***		0.2144***
	(0.0293)		(0.0363)
Relative Class Size	-0.1789***		-0.2243***
	(0.0594)		(0.0701)
Female Share	0.0206		0.0082
	(0.0198)		(0.0206)
Own characteristics			
IQ	-0.0025***	-0.0026***	-0.0026***
	(0.0003)	(0.0004)	(0.0004)
Previous GPA	0.7426***	0.7492***	0.7405***
	(0.0072)	(0.0085)	(0.0085)
Age	-0.0189***	-0.0204***	-0.0148***
	(0.0042)	(0.0050)	(0.0049)
Peers' characteristics			
IQ	0.0011**	0.0012**	0.0008
	(0.0005)	(0.0005)	(0.0006)
Previous GPA	-0.0615**	0.0653***	-0.0788**
	(0.0279)	(0.0163)	(0.0343)
Age	-0.0227***	-0.0137**	-0.0222***
	(0.0049)	(0.0054)	(0.0058)
Wald statistics (d.f.)		49.25 (3)	18.71(3)
p-value		0.0000	0.0003

Estimates of the three model variants obtained by MVIV (multi-variable IV) estimation. The first column reports the composite model, the second the local-aggregate model, and the third the local-average model. Following Liu et al. (2014), we use the average characteristics of second-order peers as instruments for the local-average component and the aggregate characteristics of peers for the local-aggregate component. The Wald statistics test the composite model against each local model. Robust standard errors (heteroskedasticity-consistent) are reported in parentheses. ${}^{*}p < 0.1$; ${}^{**}p < 0.05$; ${}^{***}p < 0.01$. $N = 2{,}398$, $L = 99$.

error 0.0353) for the largest class. Turning to the female share, the estimates for the local-aggregate component suggest some variation across the female-share distribution, whereas the local-average component is essentially unchanged. At the average class size, the local-aggregate peer effect is about 0.002 at the minimum observed female share and about -0.0010 at the maximum, but is not statistically different from zero.[7] By contrast, the local-average peer effect remains around 0.16–0.17 and does not vary significantly with the female share.

We view this pattern as descriptive evidence that the strength of different peer-effect components depends on the broader classroom context in which friendships are embedded. One potential interpretation is that, in larger classes, teachers have less scope for individualized support, so students rely more on friends for academic help, which makes the total performance of friends (local aggregate) more influential. At the same time, norms and social comparisons may be less tightly tied to a single small group in larger classes, which could weaken the role of the average performance of friends (local average). However, the data do not allow us to clearly distinguish this channel from other potential mechanisms, so we view this interpretation as suggestive rather than conclusive.

It is important to emphasize that this class size effect is novel in the literature, as it operates through peer behaviour. This effect complements the direct impact of class size on student performance, which is traditionally examined in studies on educational achievement. To our knowledge, class size effects on educational performance have only been studied as level effects. In our model we account for the presence of such effects by partialling out specific network level effects represented by η_l (see Eq. (3)). Therefore, the effect of class size on peer behaviour represents an additional channel for the impact of class size on educational achievement. Unlike the conventional direct effect of class size in reduced-form specifications, its influence through peer behavior highlights the role of

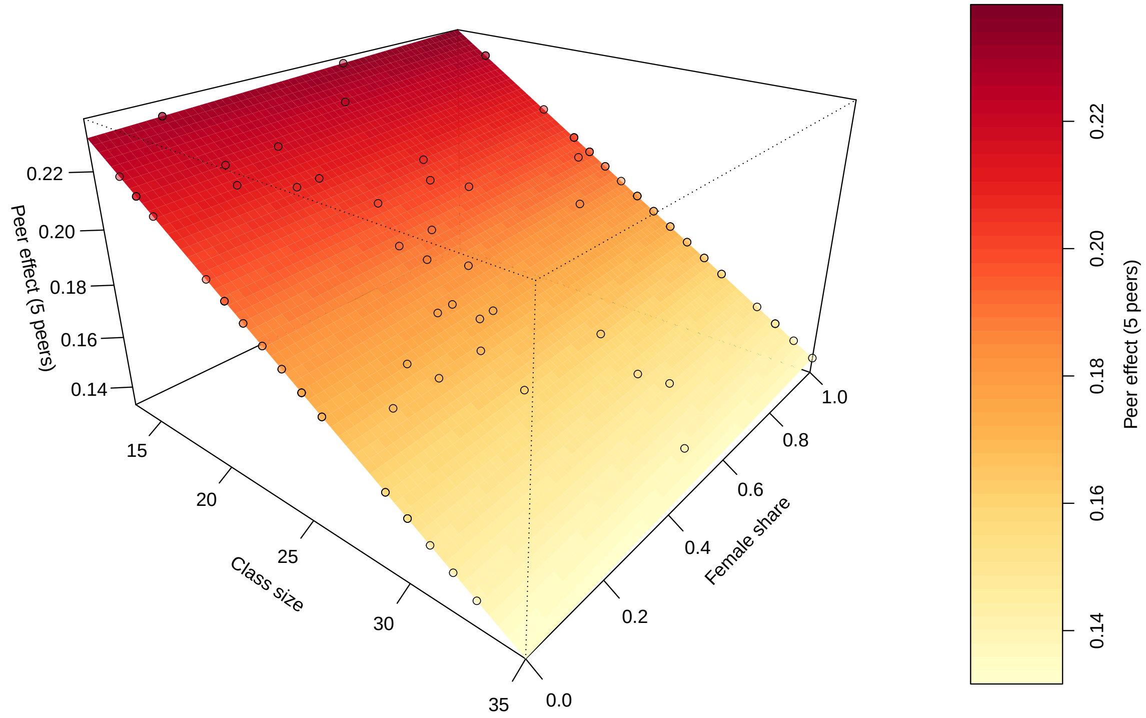
---

[7] For the local-aggregate peer effect, the estimate at the maximum female share is -0.0010, with a 95% confidence interval of $[-0.0021, 0.0001]$.

social interactions, which in turn partially shape individual performance. In this sense, our approach also offers a specific structural explanation for variations in class performance.

Figure 2 illustrates the variation in the combined peer effect across different class sizes and gender compositions. In larger classes with a high proportion of female students, the peer effect is at the lower end of this range. Overall, the combined peer effect ranges from about 0.13 to 0.24.

Figure 2: Peer Effects by Class Size and Gender Composition (GPA)



Surface of the estimated combined peer effect as a function of class size and gender composition for a student with an outdegree of 5, based on the parameter estimates of the composite model in Table 1. The peer effect measures the change in a student's GPA resulting from a one-unit increase in the GPA of all five nominated peers. Circles indicate the observed class combinations of class size and female share, plotted at their corresponding estimated peer effects.

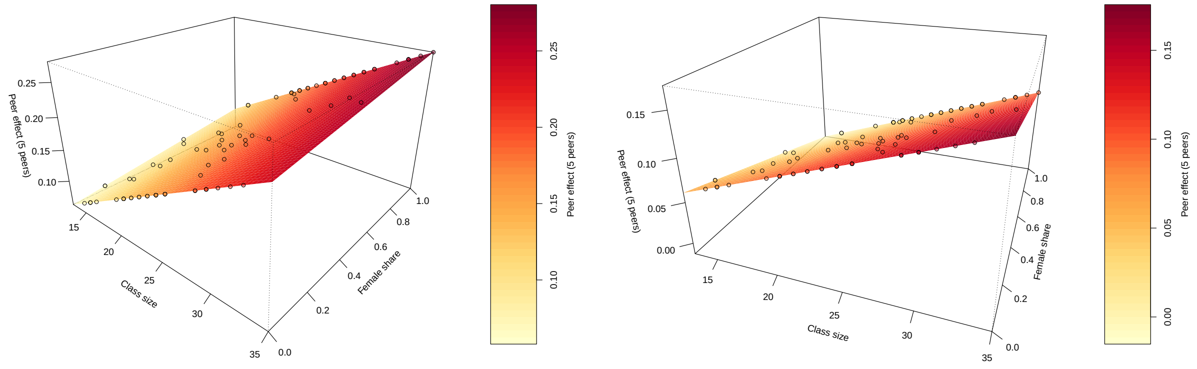
The coefficients on own IQ and own previous GPA have the expected signs. As anticipated, last year's GPA is a strong predictor of current performance: students with worse previous grades tend to perform worse again. Students with a higher IQ also perform better, and we find evidence that older students achieve slightly better outcomes. Recall that in the German grading system lower numerical values of GPA indicate better performance.

Our results point out to rather complex and puzzling interactions between the students and their peers when considering jointly the (endogenous) peer effects as discussed above with the exogenous peer effects (γ_g in (1)). The peers' IQ and peers' GPA significantly affect individual performance, and their signs are opposite to the corresponding own effects: higher peer IQ is associated with worse performance, whereas higher peer GPA (i.e. worse grades) is associated with better performance, conditional on own characteristics. In contrast, having older peers is associated with better individual performance, in line with the positive effect of own age.

The MVIV estimates of the heterogeneous composite and local models with grades in Math and German as dependent variables are reported in Tables A2 and A3 in the Appendix. For both subjects, we again find heterogeneity in peer effects across class characteristics and transmission mechanisms, although the patterns are less pronounced than for GPA. In the local-aggregate component, the intercept is 0.0019 (standard error 0.0006) in German and 0.0012 (standard error 0.0008) in Math; the former is precisely estimated, whereas the latter is not statistically different from zero, and in both subjects the coefficients on class size and female share are estimated imprecisely. By contrast, the local-average component displays clearer subject-specific differences: in German, the local-average effect is large and precisely estimated and increases with class size, whereas in Math the local-average effect is smaller and accompanied by a significantly negative coefficient on female share, with the class-size coefficient remaining imprecisely estimated. Coefficients on exogenous peer characteristics are generally comparable across subjects and broadly consistent with the GPA results. Figures 3 illustrate how the combined peer effect varies with class size and gender composition for Math and German.

Across the three outcomes, the role of gender composition appears to be subject specific.

Figure 3: Peer Effects by Class Size and Gender Composition (German and Math)



Surfaces of the estimated combined peer effect as functions of class size and gender composition for an outdegree of 5, based on the parameter estimates of the composite model in Tables A2 (German, left) and A3 (Math, right). The peer effect denotes the change in a student's German (Math) score resulting from a one-unit increase in the German (Math) score of all five nominated peers. Circles indicate the observed class combinations of class size and female share, plotted at their corresponding estimated peer effects.

For GPA, the total peer effect is positive across classroom profiles and changes little with the female share, suggesting that gender mix is not a key determinant of overall achievement spillovers. In German, the only systematic heterogeneity arises through class size: peer effects are substantially stronger in larger classes, while the female-share coefficients are not statistically meaningful. In Math, by contrast, peer effects vary with gender composition: the local-average component declines significantly with the female share, whereas the class-size coefficient is estimated imprecisely. Although we cannot identify the underlying mechanisms, these patterns indicate that the strength of peer effects depends on classroom gender composition and class size in ways that differ across subjects.

For completeness, Table A4 reports the MVIV estimates for the homogeneous peer-effects specifications for GPA, German, and Math. The Wald tests at the bottom of the table compare each homogeneous model with its heterogeneous counterpart. For GPA and the German grade, the composite specification rejects homogeneity at conventional significance levels, confirming that allowing peer effects to vary with class characteristics provides a significantly better fit than a model with constant peer effects. For the Math grade, the statistical evidence against the homogeneous model is weaker. Overall, however, the composite results point to empirically relevant heterogeneity in peer effects across all three subjects. It is also important to note that the homogeneous peer-effect estimates obtained with MVIV for GPA and Math are broadly similar in magnitude and sign to those reported by Liu et al. (2014) for "study effort." As in our case, the peer-effect coefficients in Liu et al. (2014) are estimated imprecisely.

For comparison, Table A5 reports the OLS estimates of the composite peer-effects model. As expected, these coefficients on the endogenous peer terms differ from the MVIV results. Because peers' outcomes are jointly determined, OLS cannot recover the causal peer parameter and therefore delivers biased and often attenuated estimates. In contrast, the MVIV approach corrects this endogeneity and yields, in most cases, larger and statistically significant peer effects, while the coefficients on own characteristics remain broadly similar across the two methods.

5 Conclusions

This paper contributes to the expanding body of empirical literature on social networks. Specifically, we examine the impact of heterogeneity on network peer effects by accounting for both network-specific factors and the various mechanisms driving peer behaviour. We consider a network model that allows for both network-specific heterogeneity and multiple peer-effect mechanisms and propose a Multivariate Instrumental Variable (MVIV) approach to estimate this parametrically rich model. We demonstrate the versatility of our approach by studying the role of class size and gender composition for peer effects in the classroom, using a unique dataset comprising friendship networks of 99 secondary school classes in Germany.

We find that network-specific factors, such as the size of the network (e.g., a school class) and its gender composition, significantly influence the effects that peers exert on students' educational outcomes. By contrast, models that assume homogeneous peer effects yield mostly insignificant results. In addition to network characteristics, behavioural heterogeneity in the educational context plays a crucial role: a student's attainment is shaped by both the size of their peer group (the local-aggregate mechanism) and normative behaviour within that group (the local-average mechanism).

Our empirical analysis yields several broader insights for the study of peer effects and for education economics. First, this type of network-specific heterogeneity within a structural framework has, to our knowledge, not been examined previously. Second, most existing empirical evidence is based on Add Health data, whereas we contribute new findings using previously unexploited European classroom network data. Third, our study adds to the extensive literature on the determinants of educational achievement by demonstrating that class size affects peer behaviour in systematic ways. In particular, increasing class size can diminish positive peer behaviour in some subjects, while in others it strengthens the influence of peers. Unlike many reduced-form approaches, our method provides a structural interpretation of why class size and gender composition matter and how these channels differ across subjects. In this sense, we identify peer behaviour as a specific mechanism through which class size influences educational attainment.

We view our study as a promising foundation for more realistic modelling of heterogeneous network behaviour and for a deeper understanding of network dynamics. Future research should explore more advanced approaches to modelling network heterogeneity, such as nonlinear or nonparametric specifications of heterogeneous peer effects. For example, the two-step nonparametric GMM procedure of Hong et al. (2024), developed for varying coefficients in spatial dynamic panel models, may offer a flexible way to extend our framework. Additionally, further work should examine how properties of network structure (e.g., features of adjacency matrices) influence the identification of peer effects.

A further step would be to allow for heterogeneous peer effects in a correlated random-coefficients framework. Recent work by Masten (2018) shows that random peer-effect parameters can be identified in simple linear-in-means models, but it is still unclear how such results carry over to network settings. Adapting network-based identification strategies to this type of random-coefficient structure is therefore an attractive direction for future research.

Declarations

The authors have no relevant financial or non-financial interests to disclose.

References

Angrist, J., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. Quarterly Journal of Economics 114, 533–575. doi:`10.1162/003355399556061`.

Arduini, T., Patacchini, E., Rainone, E., 2014. Identification and Estimation of Outcome Response with Heterogeneous Treatment Externalities. EIEF Working Papers Series 1407. Einaudi Institute for Economics and Finance (EIEF).

Auerbach, E., 2022. Identification and estimation of a partially linear regression model using network data. Econometrica 90, 347–365. doi:`10.3982/ECTA19794`.

Bear, J.B., Woolley, A.W., 2011. The role of gender in team collaboration and performance. Interdisciplinary Science Reviews 36, 146–153. doi:`10.1179/030801811X13013181961473`.

Beugnot, J., Fortin, B., Lacroix, G., Villeval, M.C., 2019. Gender and peer effects on performance in social networks. European Economic Review 113, 207–224. doi:`10.1016/j.euroecorev.2019.01.002`.

Blume, L.E., Brock, W.A., Durlauf, S.N., Jayaraman, R., 2015. Linear social interactions models. Journal of Political Economy 123, 444–496. doi:`10.1086/679496`.

Boucher, V., Bramoullé, Y., Djebbari, H., Fortin, B., 2014. Do peers affect student achievement? Evidence from Canada using group size variation. Journal of Applied Econometrics 29, 91–109. doi:`10.1002/jae.2299`.

Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. Journal of Econometrics 150, 41–55. doi:`10.1016/j.jeconom.2008.12.021`.

Bramoullé, Y., Djebbari, H., Fortin, B., 2020. Peer effects in networks: A survey. Annual Review of Economics 12, 603–629. doi:`10.1146/annurev-economics-020320-033926`.

CAESR, 2007. Dataset Gymnasiastenstudie. Central Archive for Empirical Social Research, Cologne, Germany.

Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. The Review of Economic Studies 76, 1239–1267. doi:`10.1111/j.1467-937X.2009.00550.x`.

Dobbelsteen, S., Levin, J., Oosterbeek, H., 2002. The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition. Oxford Bulletin of Economics and Statistics 64, 17–38. doi:`10.1111/1468-0084.00003`.

Fredriksson, P., Öckert, B., Oosterbeek, H., 2012. Long-term effects of class size. The Quarterly Journal of Economics 128, 249–285. doi:`10.1093/qje/qjs048`.

Griffith, A., 2021. Name your friends, but only five? the importance of censoring in peer effects estimates using social network data. Journal of Labor Economics 40, 779–805. doi:`10.1086/717935`.

Hanushek, E.A., 1996. Measuring investment in education. Journal of Economic Perspectives 10, 9–30. doi:`10.1257/jep.10.4.9`.

Heinesen, E., 2010. Estimating class-size effects using within-school variation in subject-specific classes. The Economic Journal 120, 737–760. doi:`10.1111/j.1468-0297.2009.02301.x`.

Hong, H., Ju, G., Li, Q., Yan, K.X., 2024. Varying-coefficient spatial dynamic panel data models with fixed effects: Theory and application. Journal of Econometrics 245, 105883. doi:`https://doi.org/10.1016/j.jeconom.2024.105883`.

Hoxby, C.M., 2000. The effects of class size on student achievement: New evidence from population variation. The Quarterly Journal of Economics 115, 1239–1285. doi:`10.1162/003355300555060`.

Hoxby, C.M., 2002. The power of peers: How does the makeup of a classroom influence achievement. Education Next Summer 2, 57–63.

Hummell, H.J., Klein, M., Wieken-Mayser, M., Ziegler, R., 1970. Structure analysis of the school (schoolchildren survey). GESIS Data Archive, Cologne. ZA0600 Data file Version 1.0.0. doi:`10.4232/1.0600`.

Hummell, H.J., Klein, M., Wieken-Mayser, M., Ziegler, R., 1979. Workbook: Social origins and school career (high school). (ZA1032; Version 1.0.0) [Data set]. GESIS, Cologne. doi:`10.4232/1.1032`.

Hummell, H.J., Klein, M., Wieken-Mayser, M., Ziegler, R., 2018. Structure analysis of the school (schoolchildren survey - sociometric test). GESIS Data Archive, Cologne. ZA0942 Data file Version 2.0.0. doi:`10.4232/1.13036`.

Johnsson, I., Moon, H.R., 2021. Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach. The Review of Economics and Statistics 103, 328–345. doi:`10.1162/rest_a_00870`.

Kelejian, H.H., Prucha, I.R., 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. doi:`10.1023/A:1007707430416`.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533. doi:`10.1111/1468-2354.00027`.

Krueger, A.B., 1999. Experimental estimates of education production functions. The Quarterly Journal of Economics 114, 497–532. doi:`10.1162/003355399556052`.

Lavy, V., Schlosser, A., 2011. Mechanisms and impacts of gender peer effects at school. American Economic Journal: Applied Economics 3, 1–33. doi:`10.1257/app.3.2.1`.

Lewbel, A., Qu, X., Tang, X., 2023. Ignoring measurement errors in social networks. The Econometrics Journal 27, 171–187. doi:`10.1093/ectj/utad028`.

Lin, X., 2014. Network attributes and peer effects. Economics Bulletin 34, 2060–2079.

Liu, X., Patacchini, E., Zenou, Y., 2014. Endogenous peer effects: Local aggregate or local average? Journal of Economic Behavior and Organization 103, 39–59. doi:`10.1016/j.jebo.2014.03.025`.

Mael, F., Alonso, A., Gibson, D., Rogers, K., Smith, M., 2005. Single-sex versus coeducational schooling: A systematic review. Doc# 2005-01. Technical Report.

Manski, C.F., 1993. Identification of endogenous social effects: The reflection problem. The Review of Economic Studies 60, 531–542. doi:`10.2307/2298123`.

Masten, M.A., 2018. Random coefficients on endogenous variables in simultaneous equations models. The Review of Economic Studies 85, 1193–1250. doi:`10.1093/restud/rdx047`.

Morse, S., 1999. Separated by sex: A critical look at single-sex education for girls. Journal of Chemical Education 76, 615. doi:`https://pubs.acs.org/doi/abs/10.1021/ed076p615`.

n, H., Wieken-Mayser, M., Wiese, W., 1986. Social origins and school career of high school students. GESIS Data Archive, Cologne. ZA1440 Data file Version 1.0.0. doi:`10.4232/1.1440`.

A Appendix

A.1 Identification of homogeneous local average model

For ease of exposition the results presented in this appendix are based on the local-average model with heterogeneous peer effects ($\beta_{a,l} = 0, \beta_{g,l} = \beta_l$). To simplify the notation, we assume that the correlated effect η_l is zero or has been partialled out. Moreover, we concentrate on one exogenous factor ($k_x = 1$) and one factor driving the peer effects parameter $k_v = 1$. Thus, the local-average model in vector notation is given by:

$$\mathbf{Y}_l = \gamma x_l + \gamma_g x_l^g + \beta_l \, y_l^g + \varepsilon_l \,, \qquad (l = 1, \ldots, L) \tag{8}$$

where $\mathbf{Y}_l$ is the $n_l \times 1$ vector of outcome variables, G_l the row normalized adjacency matrix, x_l the exogenous factor, $x_l^g = G_l x_l$ the average exogenous factor of the peers and $y_l^g = G_l \, \mathbf{Y}_l$ as the the $n_l \times 1$ vector of average outcome of the peers. Solving (8) for $\mathbf{Y}_l$ yields the reduced form expression:

$$\mathbf{Y}_l = \big(I_{n_l} - \beta_l G_l\big)^{-1} \big(\gamma I_{n_l} + \gamma_g G_l\big) x_l + \big(I_{n_l} - \beta_l G_l\big)^{-1} \varepsilon_l \,. \tag{9}$$

In the terminology of instrumental variables estimation the corresponding first stage regression for the local average specification can be obtained by premultiplying (9) by the adjacency matrix G_l. This gives for the expected value:

$$\mathrm{E}[y_l^g \,|\, x_l, G_l] = G_l \big(I_{n_l} - \beta_l G_l\big)^{-1} \big(\gamma I_{n_l} + \gamma_g G_l\big) x_l \,. \tag{10}$$

In order to see that the peers outcome y_l^g is a function of the exogenous characteristics of the peers of peers (higher order peers) note that for $|\beta_l| < 1$ the matrix inverse can be expressed as an infinite series, $\big(I_{n_l} - \beta_l G_l\big)^{-1} = I_{n_l} + \beta_l G_l + \beta_l^2 G_l^2 + \ldots$

Replacing in (10) the inverse by the infinite series gives the desired result:

$$\mathrm{E}[y_l^g|\, x_l, G_l] = \gamma x_l^g + (\gamma_g + \gamma\beta_l) \sum_{k=0}^{\infty} \beta^k G_l^{k+2} x_l \tag{11}$$

If $\gamma_g + \gamma\beta_l \neq 0$ and I_{n_l}, G_l, and G_l^2 are linearly independent, the parameters are identified (see Bramoullé et al., 2009, Proposition 1). Linear independence between matrix G_l and G_l^2 is obtained if the network is sufficiently sparse.

For the identification conditions for the local aggregate and the composite model we refer to Liu et al. (2014, Proposition 2). These conditions are generally weaker than the ones for the local average model given above.

A.2 (Non-) Identification of a single network l

Based on (11), it is apparent that $(G_l^2 x_l, G_l^3 x_l, \ldots)$ can serve as instruments, thus the IV estimator of parameters in (8) is consistent and asymptotically normal for $n_l \to \infty$.

Assume now that the heterogeneous peer effects parameter β_l depends linearly on the observable factor v_l: $\beta_l = \beta_0 + \beta_1 v_l$.

Even without the presence of endogeneity the $k \times 1$-dimensional overall parameter vector $\theta = (\gamma, \gamma_g, \beta_0, \beta_1)'$ of the structural equation

$$\mathbf{Y}_l = \gamma x_l + \gamma_g x_l^g + \beta_0 y_l^g + \beta_1 v_l y_l^g + \varepsilon_l = \mathbf{X}_l\, \theta + \varepsilon_l \tag{12}$$

are not identified on the network level since the y_l^g and $v_l y_l^g$ are proportional such that the $n_l \times k$ dimensional regressor matrix $\mathbf{X}_l = \begin{bmatrix} x_l & x_l^g & y_l^g & v_l y_l^g \end{bmatrix}$ has no full column rank.

A.3 MVIV Estimation

Assumption A.1 (Set of Peer Effects Networks)

Consider the model for the l-th network as defined in (12) and assume that the following conditions hold:

(i) $\mathrm{E}[\varepsilon_{i,l}|\, \mathbf{z}_{i,l}] = 0$

(ii) $\mathrm{V}[\varepsilon_{i,l}|\, \mathbf{z}_{i,l}] = \sigma_l^2(\mathbf{z}_{i,l})$

(iii) $\mathrm{E}\big[\mathbf{z}_{i,l}\, \mathbf{x}'_{i,l}\big] = Q_{zx,l}$ is a finite $q \times k$ - matrix

(iv) $\mathrm{E}\big[\mathbf{z}_{i,l}\, \mathbf{z}'_{i,l}\big] = Q_{zz,l}$ is a finite, non-singular $q \times q$ - matrix

(v) $\mathrm{E}\big[\mathbf{x}_{i,l}\, \mathbf{x}'_{i,l}\big] = Q_{xx,l}$ is a finite $k \times k$ - matrix

(vi) Let $N = \sum_{l=1}^{L} n_l$ and $\omega_l = \frac{n_l}{N} \to \omega_l^0$ with $0 < \omega_l^0 < 1$ for $n_l \to \infty$.

Definition A.1 (Moment function of a single network)

Let $\mathcal{S}_l$ denote the set of all observations i belonging to network l and let $d_{il} = \mathbb{1}\,(i \in \mathcal{S}_l)$ be a independent binary variable indicating whether observation i belongs to network l with $\mathrm{E}[d_{il}] = \omega_l$. Then

(i) $\psi_{il} = \mathbf{z}_{il}\, \varepsilon_{il} d_{il}$ with $d_{il} = \mathbb{1}\,(i \in \mathcal{S}_l)$ is well-defined moment function for network l with moment restriction $\mathrm{E}[\psi_{il}] = \mathrm{E}[d_{il}]\, \mathrm{E}[\mathbf{z}_{il}\ \varepsilon_{il}] = 0$.

(ii) $\mathrm{V}[\psi_{il}] = \mathrm{E}[\psi_{il}\psi'_{il}] = \omega_l\, \mathrm{E}[\varepsilon_{il}^2\, \mathbf{z}_{il}\, \mathbf{z}'_{il}]$

The moment functions as defined in Definition A.1 can be stacked into an $(q \cdot L)$-dimensional vector of moment function ψ_i as defined below.

Definition A.2 (Moment Function of the Network)

Let $\psi_i = \big(\psi'_{i1}, \psi'_{i2}, \ldots, \psi'_{iL}\big)'$ be the $q \cdot L$-dimensional vector of moment functions for observation i with variance-covariance $\Phi_0 \equiv \mathrm{V}[\psi_i] = \mathrm{diag}\big[\omega_l\, \mathbf{V}_l\,\big]$, with $\mathbf{V}_l = \mathrm{E}[\varepsilon_{il}^2\, \mathbf{z}_{il}\, \mathbf{z}'_{il}]$

Definitions A.1 and A.2 allow us to formulate the empirical moment function of the network in terms of the overall sample size.

$$\bar{\psi} = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N} \psi_{i1} \\ \sum_{i=1}^{N} \psi_{i2} \\ \vdots \\ \sum_{i=1}^{N} \psi_{iL} \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_{i1}\,\varepsilon_{i1} d_{i1} \\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_{i2}\,\varepsilon_{i2} d_{i2} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_{iL}\,\varepsilon_{iL} d_{iL} \end{pmatrix} = \frac{1}{N}\begin{pmatrix} \mathbf{Z}_1'\,\boldsymbol{\varepsilon}_1 \\ \mathbf{Z}_2'\,\boldsymbol{\varepsilon}_2 \\ \vdots \\ \mathbf{Z}_L'\,\boldsymbol{\varepsilon}_L \end{pmatrix} = \frac{1}{N}\,\mathbf{Z}'\,\boldsymbol{\varepsilon} \tag{13}$$

Our proposed MVIV estimator is the asymptotically efficient two-step GMM estimator based on the vector of empirical moment functions (13):

$$\hat{\theta}_{MVIV} = \arg\min_{\theta} \bar{\psi}'\,\hat{\Phi}^{-1}\bar{\psi}\,, \tag{14}$$

where $\hat{\Phi}$ is a consistent estimate of the optimal weighting matrix Φ_0 given by:

$$\hat{\Phi} = \text{diag}\big[\hat{\omega}_l^{-1}\hat{V}_l^{-1}\big]$$

with $\hat{\omega}_l = n_l/N$ and $\hat{V}_l^{-1} = \frac{1}{n_l}\,\mathbf{Z}_l'\,\mathbf{D}_l\,\mathbf{Z}_l$ and $\mathbf{D}_l = \text{diag}\big[e_l^2\big]$, where e_l are the residuals from first step. Due to the linearity of the moment functions the GMM estimator takes the well-known form:

$$\begin{aligned} \hat{\theta}_{MVIV} =& \big(\mathbf{X}'\,\mathbf{Z}\,\hat{\Phi}^{-1}\,\mathbf{Z}'\,\mathbf{X}\big)^{-1}\,\mathbf{X}\,\mathbf{Z}\,\hat{\Phi}^{-1}\,\mathbf{Z}_1'\,\mathbf{Y} \\ =& \big(\sum_{l=1}^{L}\mathbf{X}_l'\,\mathbf{P}_l\,\mathbf{X}_l\big)^{-1}\sum_{l=1}^{L}\mathbf{X}_l'\,\mathbf{P}_l\,\mathbf{Y}_l \end{aligned} \tag{15}$$

with $\mathbf{P}_l = \mathbf{Z}_l\,\big[\,\mathbf{Z}_l'\,\mathbf{D}_l\,\mathbf{Z}_l\,\big]^{-1}\,\mathbf{Z}_l'$.

For a single network ($L = 1$), the matrix $\mathbf{X}_1'\,\mathbf{P}_1\,\mathbf{X}_1$ is singular and the estimator would not be identified. However, the sum

$$\sum_{l=1}^{L} \mathbf{X}_l' \mathbf{P}_l \mathbf{X}_l$$

is non-singular if the network-specific variables differ across networks. Thus, identification of the network-specific peer-effect parameters crucially depends on variation in v_l across networks. See the Web Appendix for a numerical illustration for $L = 2$ and for a comparison of MVIV and MDIV, which rely on the same identification condition.

B Tables

Table A1: Summary Statistics

	Entire Sample		Estimation Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Outcome Variables				
GPA^a	3.20	0.48	3.20	0.48
German	3.46	0.76	3.46	0.76
Math	3.50	0.96	3.50	0.96
Individual characteristics				
IQ	110.51	11.42	110.49	11.42
Previous GPA^a	3.19	0.49	3.19	0.49
Age	15.39	0.87	15.39	0.87
Network measures				
Class size	28.92	5.72	28.90	5.72
Relative Class size^b	0.74	0.15	0.74	0.15
Sample Class size^c	25.31	5.29	25.41	5.13
Female share	0.53	0.43	0.53	0.43
Sample female share^c	0.53	0.43	0.53	0.43
Density	0.17	0.06	0.17	0.06
N	2,409		2,398	
L	101		99	

Note: Own calculations. We exclude classes with fewer than 13 students in the estimation sample. *a*: Better grades are represented by lower values. *b*: relative class size represents the class size divided by the size of the largest class. *c*: Sample class size refers to the number of remaining students in the classroom after deleting the individuals with missing information either in the survey data, administrative data, or in the sociometric test. The same principle applies to the calculation of the female share measure. The network density is defined as the ratio of all connections in a network to the number of potential connections.

Table A2: MVIV Estimation Results: German

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0019***	0.0020***	
	(0.0006)	(0.0007)	
Relative Class Size	0.0018	0.0047	
	(0.0034)	(0.0045)	
Female Share	-0.0019	-0.0016	
	(0.0015)	(0.0016)	
Local-average peer effect			
Intercept	0.1598***		0.2215***
	(0.0258)		(0.0323)
Relative Class Size	0.2763**		0.3392**
	(0.1199)		(0.1463)
Female Share	0.0479		0.0690
	(0.0371)		(0.0458)
Own characteristics			
IQ	-0.0058***	-0.0050***	-0.0053***
	(0.0006)	(0.0008)	(0.0008)
Previous GPA	0.7970***	0.8058***	0.8063***
	(0.0139)	(0.0175)	(0.0177)
Age	-0.0720***	-0.0719***	-0.0760***
	(0.0079)	(0.0096)	(0.0102)
Peers' characteristics			
IQ	0.0025***	0.0027**	0.0019*
	(0.0010)	(0.0011)	(0.0012)
Previous GPA	-0.0201	0.1309***	-0.0739*
	(0.0332)	(0.0318)	(0.0420)
Age	-0.0564***	-0.0511***	-0.0496***
	(0.0095)	(0.0101)	(0.0120)
Wald statistics (d.f.)		40.86 (3)	16.07 (3)
p-value		0.0000	0.0011

Estimates of the three model variants obtained by MVIV (multi-variable IV) estimation. The first column reports the composite model, the second the local-aggregate model, and the third the local-average model. Following Liu et al. (2014), we use the average characteristics of second-order peers as instruments for the local-average component and the aggregate characteristics of peers for the local-aggregate component. The Wald statistics test the composite model against each local model. Robust standard errors (heteroskedasticity-consistent) are reported in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. $N =$ 2,398, $L = 99$.

Table A3: MVIV Estimation Results: Math

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0012	0.0005	
	(0.0008)	(0.0010)	
Relative Class Size	0.0017	0.0078	
	(0.0057)	(0.0067)	
Female Share	-0.0010	-0.0014	
	(0.0019)	(0.0020)	
Local-average peer effect			
Intercept	0.0783***		0.1096***
	(0.0286)		(0.0357)
Relative Class Size	0.1705		0.3387***
	(0.1063)		(0.1126)
Female Share	-0.0696**		-0.0704*
	(0.0348)		(0.0384)
Own characteristics			
IQ	-0.0169***	-0.0170***	-0.0164***
	(0.0009)	(0.0011)	(0.0011)
Previous GPA	0.8482***	0.8573***	0.8655***
	(0.0178)	(0.0228)	(0.0220)
Age	-0.0343***	-0.0398***	-0.0258*
	(0.0110)	(0.0135)	(0.0135)
Peers' characteristics			
IQ	0.0036***	0.0030**	0.0038***
	(0.0012)	(0.0015)	(0.0015)
Previous GPA	-0.0686	-0.0237	-0.0980*
	(0.0453)	(0.0435)	(0.0546)
Age	-0.0014	0.0157	-0.0050
	(0.0126)	(0.0150)	(0.0155)
Wald statistics (d.f.)		11.49(3)	3.48(3)
p-value		0.0095	0.322

Estimates of the three model variants obtained by MVIV (multi-variable IV) estimation. The first column reports the composite model, the second the local-aggregate model, and the third the local-average model. Following Liu et al. (2014), we use the average characteristics of second-order peers as instruments for the local-average component and the aggregate characteristics of peers for the local-aggregate component. The Wald statistics test the composite model against each local model. Robust standard errors (heteroskedasticity-consistent) are reported in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. $N = 2{,}398$, $L = 99$.

Table A4: Homogeneous Model Estimates

	GPA			**German**			**Math**		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Aggregate Peer Effect	0.0007	0.0009		0.0030***	0.0029***		0.0016	0.0016	
	(0.0006)	(0.0006)		(0.0011)	(0.0011)		(0.0013)	(0.0013)	
Average Peer Effect	0.4376		0.2970	-0.3629		-0.3180	0.1353		0.1078
	(0.3517)		(0.3660)	(0.4292)		(0.4433)	(0.3778)		(0.3879)
Own characteristics									
IQ	-0.0028***	-0.0028***	-0.0028***	-0.0053***	-0.0053***	-0.0053***	-0.0177***	-0.0176***	-0.0177***
	(0.0006)	(0.0006)	(0.0006)	(0.0013)	(0.0013)	(0.0013)	(0.0017)	(0.0017)	(0.0017)
Previous GPA	0.7309***	0.7339***	0.7324***	0.7847***	0.7809***	0.7856***	0.8380***	0.8331***	0.8369***
	(0.0147)	(0.0143)	(0.0146)	(0.0298)	(0.0292)	(0.0299)	(0.0363)	(0.0353)	(0.0363)
Age	-0.0229***	-0.0272***	-0.0250***	-0.0830***	-0.0783***	-0.0842***	-0.0286	-0.0258	-0.0298
	(0.0089)	(0.0079)	(0.0089)	(0.0180)	(0.0166)	(0.0181)	(0.0215)	(0.0212)	(0.0215)
Peers' characteristics									
IQ	0.0007	0.0014	0.0009	0.0044*	0.0035*	0.0042*	0.0032	0.0021	0.0028
	(0.0011)	(0.0010)	(0.0011)	(0.0024)	(0.0020)	(0.0024)	(0.0037)	(0.0024)	(0.0038)
Previous GPA	-0.2712	0.0759***	-0.1580	0.4221	0.1022*	0.3992	-0.1902	-0.0704	-0.1534
	(0.2827)	(0.0275)	(0.2940)	(0.3855)	(0.0558)	(0.3991)	(0.3541)	(0.0687)	(0.3621)
Age	-0.0312**	-0.0204**	-0.0269**	-0.0406**	-0.0471**	-0.0404**	0.0155	0.0287	0.0187
	(0.0127)	(0.0096)	(0.0130)	(0.0204)	(0.0197)	(0.0204)	(0.0423)	(0.0241)	(0.0431)
Wald statistics (d.f.)	23.52 (4)	11.47 (2)	7.39 (2)	9.88 (4)	8.52 (2)	1.66 (2)	6.35(4)	9.41 (2)	1.49 (2)
p-value	0.0001	0.0032	0.0247	0.0423	0.0141	0.4344	0.1743	0.0090	0.4738

Estimates of the three homogeneous-effects models obtained by MVIV (multi-variable IV) estimation. For each outcome variable, the first column reports the composite model, the second the local-aggregate model, and the third the local-average model. Following Liu et al. (2014), we use the average characteristics of second-order peers as instruments for the local-average component and the aggregate characteristics of peers for the local-aggregate component. The Wald statistics test the heterogeneous specifications against their homogeneous counterparts. Robust standard errors (heteroskedasticity-consistent) are reported in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. $N = 2{,}398$, $L = 99$.

Table A5: OLS Estimates: Composite Peer-Effects Model

OLS Estimates: Composite Peer-Effects Model			
Parameter	GPA	German	Math
Local-aggregate peer effect			
Intercept	0.0006	0.0027**	0.0010
	(0.0007)	(0.0014)	(0.0016)
Relative Class Size	0.0027	0.0029	0.0054
	(0.0043)	(0.0083)	(0.0112)
Female Share	-0.0014	-0.0032	-0.0013
	(0.0014)	(0.0027)	(0.0032)
Local-average peer effect			
Intercept	0.0948*	-0.0084	0.0028
	(0.0516)	(0.0430)	(0.0460)
Relative Class Size	-0.2000	0.0924	-0.0209
	(0.1242)	(0.2304)	(0.2279)
Female Share	0.0042	-0.0187	-0.0846
	(0.0394)	(0.0673)	(0.0716)
Own characteristics			
IQ	-0.0028***	-0.0054***	-0.0175***
	(0.0006)	(0.0013)	(0.0017)
Previous GPA	0.7336***	0.7819***	0.8324***
	(0.0143)	(0.0292)	(0.0354)
Age	-0.0265***	-0.0786***	-0.0261
	(0.0079)	(0.0166)	(0.0212)
Peers' characteristics			
IQ	0.0013	0.0034*	0.0019
	(0.0010)	(0.0020)	(0.0024)
Previous GPA	0.0044	0.1026	-0.0662
	(0.0520)	(0.0637)	(0.0799)
Age	-0.0218**	-0.0464**	0.0262
	(0.0096)	(0.0199)	(0.0245)

Ordinary least squares estimates of the composite peer-effects model for GPA, German, and Math. Robust standard errors (heteroskedasticity-consistent) are reported in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. $N = 2{,}398$, $L = 99$.

Web-based supporting materials

“Friendship Networks and Academic Success: The Impact of Class Size and Gender Composition”

by

Winfried Pohlmeier, Livia Shkoza, Derya Uysal

December 15, 2025

Invertibility of the Pooled Moment Matrices

In general, there is no simple closed-form expression for

$$(A_1 + A_2)^{-1}$$

in terms of A_1^{-1} and A_2^{-1}, even if A_1 and A_2 are symmetric.

(i) Under Invertibility:

If A_1 (or alternatively A_2) is invertible, the matrix inversion lemma (also known as the Woodbury formula) applies. The general form is:

$$(A_1 + A_2)^{-1} = A_1^{-1} - A_1^{-1}(I + A_2A_1^{-1})^{-1}A_2A_1^{-1}. \tag{WA.1}$$

Alternatively, this can also be written symmetrically as:

$$(A_1 + A_2)^{-1} = A_2^{-1} - A_2^{-1}(I + A_1A_2^{-1})^{-1}A_1A_2^{-1}. \tag{WA.2}$$

Both equations are equivalent when A_1 and A_2 are invertible. However in our case A_1 and A_2 are both singular. The Woodbury formulas (WA.1) and (WA.2) do **not** apply for our case.

(ii.) Under Non-Invertibility:

In our case we have to take a closer look at the structure of $X_1'X_1 + X_2'X_2$. Let us take a look at the moment matrix of network l:

$$\frac{1}{n_l}X_l'X_l = \frac{1}{n_l}\begin{pmatrix} \sum_i x_{i,l}^2 & v_l \sum_i x_{i,l}^2 \\ v_l \sum_i x_{i,l}^2 & v_l^2 \sum_i x_{i,l}^2 \end{pmatrix} = \begin{pmatrix} a_l & v_l a_l \\ v_l a_l & v_l^2 a_l \end{pmatrix}$$

where $a_l = \frac{1}{n_l}\sum_i x_{i,l}^2 > 0$. For the sum of the moment matrices we get:

$$M = \frac{1}{n_1}X_1'X_1 + \frac{1}{n_2}X_2'X_2 = \begin{pmatrix} a_1 + a_2 & v_1 a_1 + v_2 a_2 \\ v_1 a_1 + v_2 a_2 & v_1^2 a_1 + v_2^2 a_2 \end{pmatrix} \tag{WA.3}$$

The determinant of the moment matrix M is given by:

$$\begin{aligned} \det(M) =& (a_1 + a_2)(v_1^2 a_1 + v_2^2 a_2) - (v_1 a_1 + v_2 a_2)^2 \\ =& a_1 a_2 (v_1 - v_2)^2, \end{aligned}$$

which is positive for $v_1 \neq v_2$, i.e. if the network specific factors differ.

Minimum Distance IV Estimator

In the following we show that our proposed MVIV estimator has a Minimum Distance (MD) representation where the single network parameters are estimated by IV-GMM in the first stage stage.

The first estimation stage of the Minimum Distance Instrumental Variable (MDIV) estimator is based on the IV regression:

$$\mathbf{Y}_l \quad = \quad \mathbf{W}_l \, \pi_l + \varepsilon_l \, , \qquad (l = 1, \ldots, L) \tag{WA.4}$$

where $\pi_l = \big(\gamma, \gamma_g, \beta_{g,l}\big)'$ is the first-stage ("reduced form") parameter vector. This equation is in fact same as (8). The $n_l \times k_w$ dimensional regressor matrix $\mathbf{W}_l = \big[x_l \; x_l^g \; y_l^g\big]$ has full column rank.

Assuming heteroskedasticity within-network the IV-GMM estimator for network l is:

$$\hat{\pi}_l = \big(\, \mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l \,\big)^{-1} \mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{Y}_l \tag{WA.5}$$

with $\hat{\mathrm{V}}\,[\hat{\pi}_l] = \big(\, \mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l \,\big)^{-1}$.

The restriction between the first-stage reduced form parameter vector π_l and the structural form parameter vector θ is given by

$$\pi_l(\theta) = M_l \, \theta \, , \tag{WA.6}$$

with

$$M_l = \begin{bmatrix} I_{k_x} & 0 & 0 & 0 \\ 0 & I_{k_x} & 0 & 0 \\ 0 & 0 & 1 & v_l \end{bmatrix} .$$

where $k_x = 1$ for the simplified model with only one regressor. Stacking the restrictions between the L reduced form parameter vectors π_l and the structural form parameter θ

given by (WA.6) into a $(k_w L \times 1)$-vector containing the complete set of restrictions gives:

$$\pi(\theta) = M\theta\,, \tag{WA.7}$$

with $\pi(\theta) = (\pi_1(\theta)', \pi_2(\theta)', \ldots, \pi_L(\theta)')'$ and $M = (M_1', M_2', \ldots, M_L')'$.

In the second step we estimate θ by minimizing the weighted quadratic distance between the estimated reduced form parameter vector $\hat{\pi}(\theta) = \big(\hat{\pi}_1', \hat{\pi}_2', \ldots, \hat{\pi}_L'\big)'$ and $M\theta$ with respect to the structural parameter vector θ based on the estimated optimal weight matrix $\hat{\Omega}^{-1} = \hat{\mathrm{V}}\,[\hat{\pi}]^{-1}$:

$$\hat{\theta}_{MDIV} \equiv \arg\min_{\theta}\,\big[\hat{\pi} - M\theta\big]'\hat{\Omega}^{-1}\big[\hat{\pi} - M\theta\big], \tag{WA.8}$$

Because of the independence of the networks, the estimated weighting matrix is block-diagonal of the form: $\hat{\Omega}^{-1} = \mathrm{diag}\big[\hat{\mathrm{V}}\,[\hat{\pi}_l]^{-1}\big] = \mathrm{diag}\big[\,\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\,\big]$. Moreover, due to the linearity of the restriction between π and θ, the MDIV can be represented as a generalized least square estimator of a regression of $\hat{\pi}$ on M:

$$\begin{aligned}\hat{\theta}_{MDIV} =& \big(M'\hat{\Omega}^{-1}M\big)^{-1}M'\hat{\Omega}^{-1}\hat{\pi}\\ =& \Big(\sum_{l=1}^{L} M_l'\hat{\mathrm{V}}\,[\hat{\pi}_l]^{-1}\,M_l\Big)^{-1}\Big(\sum_{l=1}^{L} M_l'\hat{\mathrm{V}}\,[\hat{\pi}_l]^{-1}\,\hat{\pi}_l\Big)\,.\end{aligned} \tag{WA.9}$$

Equivalence of MVIV and MDIV

Proposition 1 (Equivalence of Estimators)

For the a heteroskedastic network design given by Assumption A.1 the Minimum Distance estimator $\hat{\theta}_{MDIV}$ given by (WA.9) and the the MVIV estimator $\hat{\theta}_{MVIV}$ given by (15) are numerically equivalent.

Proof 1 (Equivalence of MVIV and MDIV)

Decomposing the first stage estimate of the reduced form parameter vector and using the restriction $\pi_l = M_l\theta$:

$$\hat{\pi}_l = \pi_l + \left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l \tag{WA.10}$$

$$= M_l\theta + \left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l \tag{WA.11}$$

yields the first stage estimator in terms of θ. Inserting (WA.11) into (WA.9) gives:

$$\begin{aligned}
\hat{\theta}_{MDIV} =& \Big(\sum_{l=1}^{L} M_l'\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1} M_l\Big)^{-1} \cdot \Big(\sum_{l=1}^{L} M_l'\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1}\left[M_l\theta + \left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l\right]\Big) \\
=& \,\theta + \Big(\sum_{l=1}^{L} M_l'\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1} M_l\Big)^{-1} \cdot \Big(\sum_{l=1}^{L} M_l'\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1}\left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l\Big) \\
=& \,\theta + \Big(\sum_{l=1}^{L} \mathbf{X}'\,\mathbf{P}_l\,\mathbf{X}_l\Big)^{-1} \cdot \Big(\sum_{l=1}^{L} \mathbf{X}'\,\mathbf{P}_l\,\varepsilon_l\Big)
\end{aligned}$$

where we used to obtain the final equation the following relationships:

- $\mathbf{W}_l\,M_l = \mathbf{X}_l$
- $M_l'\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1} M_l = M_l[\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l]M_l = \mathbf{X}_l'\,\mathbf{P}_l\,\mathbf{X}_l$
- $M_l'\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1}\left(\mathbf{W}_l'\,\mathbf{P}_l\,\mathbf{W}_l\right)^{-1}\mathbf{W}_l'\,\mathbf{P}_l\,\varepsilon_l = \mathbf{X}_l'\,\mathbf{P}_l\,\varepsilon_l$

Inserting (12) into (15) yields the decomposed $\hat{\theta}_{GMM}$ and the desired result:

$$\hat{\theta}_{MVIV} = \theta + \big(\sum_{l=1}^{L}\mathbf{X}_l'\,\mathbf{P}_l\,\mathbf{X}_l\big)^{-1}\sum_{l=1}^{L}\mathbf{X}_l'\,\mathbf{P}_l\,\varepsilon_l = \hat{\theta}_{MDIV}$$

□

Practical Note on the Equivalence of MVIV and MDIV: One can verify the equivalence numerically by constructing P_l from a common set of classwise reduced-form residuals $Y_l = W_l\pi_l + \varepsilon_l$ and plugging the same P_l into both expressions. However, the natural implementation of the two estimators leads to different first-stage choices, and therefore to different weighting matrices. In the MVIV formulation, the most natural first stage is a system-level regression that stacks all classes and estimates

$$\hat{\theta}_1 = \left(\sum_{l=1}^{L} X_l' P_{1,l} X_l \right)^{-1} \left(\sum_{l=1}^{L} X_l' P_{1,l} Y_l \right),$$

after which the class-specific residuals $e_l = Y_l - X_l \hat{\theta}_1$ are used to form P_l. In contrast, the MDIV approach naturally relies on classwise reduced-form first stages, estimating each $\hat{\pi}_l$ separately from $Y_l = W_l \pi_l + \varepsilon_l$ and constructing P_l from the corresponding class-level residuals.

Because these two first-stage strategies differ, the resulting weighting matrices P_l also differ in practice, and the two-step GMM estimators no longer coincide in finite samples even though they remain asymptotically equivalent. Thus, the equality $\hat{\theta}^{MVIV} = \hat{\theta}^{MDIV}$ in Proposition 1 applies only to a specific implementation that forces both procedures to use exactly the same weighting matrices; it is not a statement that the two estimators coincide for all reasonable two-step implementations.