

Tatsi, Eirini

**Conference Paper**

## Endogenous Social Interactions: Which Peers Matter?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Understanding the nature of peer effects, No. E02-V1

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Tatsi, Eirini (2015) : Endogenous Social Interactions: Which Peers Matter?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Understanding the nature of peer effects, No. E02-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/113168>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Endogenous Social Interactions: Which Peers Matter?

Eirini Tatsi\*

SAFE and Goethe University

Job Market Paper

January 21, 2015

## Abstract

This paper compares endogenous social interactions models to determine which one fits the classroom reality best. The analysis uses data from German 9th-graders and considers the effect of the best and worst peers' scores, the peers' sum and peers' average scores on own achievement. Although each model seems plausible when estimated separately, comparison and a selection test point to the classmates' average model, meaning that group-based policies are effective. The worst peers' model comes second, followed by the best peers' and the sum of peers' models. Examination of different-ability students' responses to increases of average peer achievement reveals either competition for the first place or last-place aversion. Conditional on own course preferences, own and peer characteristics, spillovers transmit only through cognitive ability. Therefore, regrouping on the basis of characteristics such as immigration background is obsolete. Policies should aim at low achievers in small enough classrooms because only then single-student influences can change the social norm. By improving the average through the worst, the best become even brighter.

**Keywords:** education peer effects, star networks, social multiplier, spatial econometrics

**JEL Classification:** C31, I20, Z13

---

\*Center of Excellence SAFE, Goethe University Frankfurt am Main, House of Finance, Room 4.13, Tel. +49 1767 2448968, Grüneburgplatz 1, 60323 Frankfurt am Main, Germany, Email: tatsi@safe.uni-frankfurt.de, Website: sites.google.com/site/etatsi.

Previous version: 10.12.2014 and 13.11.2014. I thank Horst Entorf, Michael Kosfeld, Ester Faia, Sascha Baghestanian, Nathanael Vellekoop, Marie Lalanne, Katerina Gradeva, participants of the Brown Bag Seminar in Goethe University, of the Chair of Econometrics Seminar in Goethe University and of the SOFI AME Seminar in Stockholm University for useful comments. All errors are my own.

# 1 Introduction

This paper examines different models of endogenous social interactions to understand how achievement spillovers occur in 9th-grade German classrooms. Social interactions refer to interactions among individuals that are not regulated by a price mechanism (Glaeser and Scheinkman, 2000). In the economics of education, the outcome of interest is unobserved ability - proxied by observed achievement in courses or test scores. Effects emanating from peers' ability, i.e. endogenous social effects, are important because they mask dependencies responsive to current behavior. Each plausible model of social interactions has different educational policy implications, pointing mainly to single or group-based rules. In turn, educational policy is interesting because of its close relation to human capital formation and long-term economic growth (Hanushek and Woessmann, 2010).

The relevant literature extensively studies the *linear-in-means* or *peers' average* model due to its technical and interpretational simplicity: what matters for own outcomes is the mean of peers' outcome or background, meaning that individuals tend to conform to the social norm. Nevertheless, there is a wide range of alternative social interactions channels. Hoxby and Weingarth (2005) and Sacerdote (2011) provide a taxonomy of 8 competing models. This paper estimates models that go beyond the average and sheds light on the role of aggregate influences and star networks, namely the peers' sum and the best (stars) or worst (bad apples) students in the classroom. The aim is threefold: first, discover which peers are more important for the regular classmate; second, discover how peer effects transmit; third, exploit the prevailing student influences to inform policy. To exclude spurious endogenous social effects due to nonrandom assignment of students into classes or common environmental factors, I drop schools that practice ability tracking and include school fixed effects in the econometric models. The dataset includes test scores in mathematics, reading comprehension and natural sciences with a negligible number of missing values.

The literature criticizes conformity, i.e. the peers' average, as the most important model (see Hoxby and Weingarth, 2005, and review in Sacerdote, 2011). Lazear (2001) introduces a model that underpins the importance of bad apples due to congestion. Hoxby and Weingarth (2005) reject the linear-in-means model in favor of tracking. Lavy, Silva, and Weinhardt (2012) explore the role of the mean as well as the bottom and top 5% performers in English schools to find negative predetermined-ability effects from low achievers. In a similar fashion, Lavy, Paserman, and Schlosser (2012) use the proportion of classroom repeaters in Israel to uncover negative effects on regular students. But more recently, Liu, Patacchini, and Zenou (2014) with friendship networks in the U.S. and Tatsi (2014) with classroom interactions for Germany provide evidence that favors the average against single individuals (peers' sum) in

simultaneous academic achievement spillovers. Still, little is known about how other forms of *simultaneous* peer influence such as star networks compare with peers' average *in the classroom*.

To this purpose, I apply for the first time the [Tao and Lee \(2014\)](#) extreme order statistic (best and worst performers) in an educational context.<sup>1</sup> At a first step, I show that - when estimated separately - each of the four social interactions models seems plausible. The *invidious model* is implicitly rejected because the dependence among students' scores is positive. Then, I estimate models including more than one effect: combinations of extreme order statistics (stars or bad apples), the peers' sum and the traditional peers' average. Ultimately, I pick the one that fits data best through a model selection test ([Kelejian, 2008](#); [Kelejian and Piras, 2011](#)). Furthermore, I explore possible heterogeneity in simultaneous spillover effects because students at the extremes or the middle of the achievement distribution may not be affected the same way by their peers. Finally, I specify a simultaneous equations model that allows for estimating the effect of each competency on the other two. To the best of my knowledge, this is the first paper to exploit system estimation in a purely educational context.

The paper presents a variety of results: both stars and bad apples models seem to provide a better fit to the data than the peers' aggregate effect, while bad apples are more important than classroom stars. Nevertheless, the linear-in-means model outperforms specifications that consider specific individuals to matter more than the average. Thus, 9th-grade German students tend to conform to the classroom norm, meaning that educational policy design is challenging. The most effective policies are those targeting at changing the peer group's norm (group-based policies). Quantile regression reveals heterogeneity in response to increases in average peer achievement: students at the top of the ability distribution respond the most to increases in average peer performance. Finally, I estimate a system of simultaneous equations so as to see exactly how effort in one subject affects effort in other subjects. Unlike recent system estimation ([Cohen-Cole, Liu, and Zenou, 2012](#); [Liu, 2014](#)), this paper discerns two sources of decaying feedback loops when calculating impacts: one from the system itself and the other from the social multiplier. A change in any explanatory variable, say in the mathematics equation, has an effect not only on the mathematics score but also on the reading comprehension and natural sciences scores. Also, a change in any explanatory variable for a single student will affect not only his/her own scores but also his/her classmates' scores. Based on impacts incorporating both feedback loops, I find that conditional on a student's

---

<sup>1</sup>[Tao and Lee \(2014\)](#) identify endogenous social interactions parameters stemming from the best and the average of players' performance in college basketball. The peers' average performance measures are not significant but the best performers' endogenous effect is significant and negative, signalling a highly competitive environment.

gender, age and family background characteristics, only the classmates' cognitive ability affects own academic achievement and only test scores in mathematics have an effect on another subject (reading comprehension).

The findings suggest that regrouping or reassigning students (see [Fruehwirth, 2014](#)) according to observable predetermined traits, e.g. immigration background, are not valid policies in the case of the German educational system. Educational policies should target at groups formed on the basis of ability alone. Since any effective policy for improving performance entails changing the classroom norm, I propose improving the achievement of the bad apples, for instance, through remedial classes or mentoring in mathematics, for small enough classroom sizes. The reason is that policy effectiveness declines in classroom size, and bad apples matter more than stars.

The rest of the paper continues with Section 2 describing the data and basic summary statistics. Section 3 provides sufficient information to understand identification and estimation of the econometric models, while Section 4 discusses the results. Section 5 concludes.

## 2 Data

The empirical analysis uses the 2006 extended German Programme for International Student Assessment (PISA). The German PISA Consortium enlarges the international PISA survey by testing two 9th-grade classrooms from the same schools chosen for the international sample. The data are available at the Research Data Centre (FDZ) of the Institute for Educational Progress (IQB).<sup>2</sup> The students complete a questionnaire concerning their background and a test based on several items. The scores reported on reading comprehension, mathematics and natural sciences are plausible values given the students' performance. Furthermore, the school principals fill out a questionnaire regarding school characteristics.

The plausible values in the three subjects serve as dependent variables and capture different aspects of verbal and quantitative skills or literacy. I use individual and family background characteristics as relevant explanatory variables for academic achievement. As in [Tatsi \(2014\)](#) and [Entorf and Tatsi \(2014\)](#), regressors include the student's gender, age, immigration background and the PISA highest international socio-economic index of occupational status in the family (HISEI). Additionally, I exploit new information from the 2006 PISA wave, namely the score from a cognitive ability test (CAT). Gender and immigration background are indicator variables. The former takes value 1 if the student is female and the latter if at least one of the parents/student was born abroad or the student speaks a foreign language at home. In Germany students attending the 9th grade are around 15 years old.

---

<sup>2</sup>Details in English can be found under <http://www.iqb.hu-berlin.de/fdz>.

Thus, observed variations in age reflect grade skippers or repeaters. HISEI is a continuous variable and roughly translates parents' occupation into income.<sup>3</sup> The cognitive ability test is also a continuous variable and captures a significant part of a student's unobserved ability. The latter is important for identification as the PISA data are cross-sectional, meaning student fixed effects are excluded.

The 2006 PISA wave includes the marks students received in German language, mathematics, natural sciences, physics, biology and chemistry at school. Unfortunately, the missing information on physics, biology and chemistry marks amounts to more than 1,000 observations and for natural sciences to more the 4,000. Therefore, I use only the German language and mathematics marks to construct excluded instruments. Nevertheless, the advantage of using PISA scores for the identification and estimation of endogenous social effects is that scores are comparable not only across schools but also across countries. The PISA emerged from the need to perform international comparisons in 15-year-old students' basic literacies and competencies. Thus, estimates are not subject to teachers' preferences or grade inflation as the school marks would be.

The German educational system thrives on early tracking at the age of 10. Hence, students are nonrandomly assigned to schools at this stage. A student's transition to schools offering vocational training or university education is based on the student's academic ability (marks), teachers' recommendations and the family's wishes (Entorf and Tatsi, 2014). In order to tackle further nonrandom assignment of students into classrooms, I exclude schools whose principal reports ability tracking. First, the school principals provide information whether they organize teaching by performance either to different classes or within the same class. Possibilities include "in all subjects", "in some subjects", "for no subject" or "missing". Second, these two questions are combined to form information about "ability grouping". I keep a school in the sample only if the principal reports ability grouping "for no subject".

After deleting missing values, the working sample corresponds to approximately 95.45% of the original observations. Also, around 90% of the classrooms contain more than 90% of original observations. Most importantly, none of the (second) best/worst students' observations are missing. Therefore, I expect contamination in the estimates due to missing values to be negligible. Approximately 52% of the subsample belongs to the "Gymnasium" schools, which lead to university education, 41% to the "Realschule" type - the second best academic quality schools in Germany - and only the remaining 7% comes from the "Hauptschule" type offering vocational training paths.

Table 1 provides summary statistics covering the mean and standard deviation for all

---

<sup>3</sup>The OECD glossary provides an exact definition: <http://stats.oecd.org/glossary/detail.asp?ID=5405>.

variables and different quantiles of the distribution for the continuous variables.<sup>4</sup> For the latter, distributions are almost symmetric as the mean and median assume close values. The three PISA scores have similar means and standard deviations. Among them, natural sciences has the highest mean and standard deviation. In the sample, around 52% of the students is female. The age of the average student is 15 years and 7 months old; 10% of the sample is younger than 15 years old, and another 10% older than 16 years and 5 months. Around 1 out of 5 students reports having some immigration background. 8 of them were born abroad while both parents were born in Germany. The mean of the highest index of occupational status in the family is around 53 points with variation of 15.7 points. Students at the lower percentile have an occupational-status index below 33 points and those at the upper percentile above 73. The observed mean for the cognitive ability test is around 0.4 - a bit lower than the median. A student belongs to the 90th percentile if the CAT score is greater than 2.17 and to the 10th percentile if the score is smaller than -1.16 points.

### 3 Econometric Models

For identification and estimation of social interaction parameters - especially endogenous - I employ a spatial autoregressive type model as in [Lee \(2007a\)](#) and [Lee, Liu, and Lin \(2010\)](#):

$$Y_s = \lambda W_s Y_s + X_s \beta_1 + W_s X_s \beta_2 + \iota_{m_s} \alpha_s + \varepsilon_s, \quad s = 1, \dots, S \quad (3.1)$$

in which  $s$  denotes the school and  $m_s$  the number of observations in each school;  $Y_s$  is the  $m_s \times 1$  vector of observations on the dependent variable, i.e. PISA scores on mathematics, reading comprehension or natural sciences;  $W_s$  denotes the  $m_s \times m_s$  socio-matrix mapping influences among students so that  $W_s Y_s$  is the  $m_s \times 1$  vector of peers' scores. The scalar parameter  $\lambda$  - coined by [Manski \(1993\)](#) as the *endogenous* social interactions coefficient - captures the effect of peers' academic achievement on own achievement.  $X_s$  is the  $m_s \times k$  matrix of observations on the  $k$  exogenous explanatory variables, namely gender, age, immigration background, highest index of occupational status in the family and scores of the cognitive ability test; thus, the  $k \times 1$  vector coefficient  $\beta_1$  measures the effect of own characteristics on academic achievement.  $W_s X_s$  represents peers' characteristics, so that the  $k \times 1$  vector coefficient  $\beta_2$  captures *exogenous* social interactions or *contextual* effects. The latter makes sense if  $W_s$  is nonstochastic and  $X_s$  is predetermined. The  $m_s \times 1$  vector of error terms,  $\varepsilon_s$ , is independent and identically distributed with mean zero and variance  $\sigma_\varepsilon^2$ . The model also allows for school-specific effects through the  $m_s \times 1$  vector  $\iota_{m_s} \alpha_s$ . The school

---

<sup>4</sup>The data provider (FDZ) prohibits presenting aggregate results for less than 5 students; hence, I can not mention minimum and maximum values as these represent single observations.

effects can be correlated with any of the elements in  $X_s$  and, hence, are treated as fixed.

For a generic socio-matrix, the equation in (3.1) represents a stable Nash Equilibrium provided  $|\lambda| < 1$ . Following the spatial econometrics literature I discern three cases for the parameter space of  $\lambda$ :

1.  $\lambda \in (-1, 0)$ . Negative social influences describe individuals whose utility decreases if the peer group performs well. Hence, an increase in peers' scores reduces own scores because students get discouraged, unmotivated or even intimidated by higher-performing classmates. This channel of social interactions resembles the invidious model (Sacerdote, 2011). Values very close to  $-1$  signify intense behavioral dissimilarity.
2.  $\lambda = 0$ , meaning endogenous social interactions are absent as no specific pattern of similarity or dissimilarity emerges.
3.  $\lambda \in (0, 1)$ . Positive values of  $\lambda$  are very common in the educational context, because they signify similarity in classmates' achievement. The closer the value is to 1, the higher the resemblance among peers. What is more, this case triggers a virtuous circle of interdependencies that propagate already existing influences.<sup>5</sup> As a result, aggregate coefficients do not correspond to the sum of individual coefficients due to the *social multiplier*. Consecutively, the social multiplier serves as a *policy tool*: target a policy only at a subset of the classroom and expect it to affect every classmate.

The socio-matrix for each school is block-diagonal:

$$W_s = \begin{pmatrix} W_{c_1} & \mathbf{0} \\ \mathbf{0} & W_{c_2} \end{pmatrix}, \quad s = 1, \dots, S \quad (3.2)$$

because the sample includes two classrooms per school,  $c_1$  and  $c_2$ , and interactions occur at the classroom level. Below I provide the socio-matrix structure in the classroom for each of the social interactions models under consideration.

## Stars and Bad Apples

Estimation of the stars' and bad apples' behavioral effect follows the extreme order statistic model developed by Tao and Lee (2014). I order the PISA score for each subject and classroom and identify the first and second best(worst) performing students. The best (worst)

---

<sup>5</sup>It can also trigger a vicious circle through which educational inequality multiplies and enhances low social mobility (see Entorf and Tatsi, 2014).



student affects all classmates and is affected by the second best (worst). In a sense, the extreme order statistic is a special case of a star network. To illustrate, the stars socio-matrix for the classroom is:

$$W_{c,star} = \begin{pmatrix} \overset{worst}{0} & 0 & \dots & \overset{2nd\ best}{0} & \overset{best}{1} \\ 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & \dots & 0 & 1 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}. \quad (3.3)$$

For a generic observation in classroom  $c = 1, 2, \dots, C$  equation (3.1) is described by the following two equations:

$$y_{i_{-best,c}c} = \lambda y_{best,c} + x_{i_{-best,c}c} \beta_1 + \varepsilon_{i_{-best,c}c} \quad (3.4)$$

$$y_{best,c} = \lambda y_{second\ best,c} + x_{best,c} \beta_1 + \varepsilon_{best,c} \quad (3.5)$$

An analogous expression holds for the bad apples socio-matrix and corresponding single-observation equations. The socio-matrix in this case is not exogenous because the order of the students' PISA scores decides the structure of student influences. Furthermore, contextual effects are not meaningful for an extreme order statistic model, meaning  $\beta_2 = 0$  in equation (3.1) because they model the characteristics of the best/worst student alone. The specific form of the adjacency matrix in equation (3.3) does not guarantee that condition  $|\lambda| < 1$  is satisfied. Therefore, the proposed estimation method is instrumental variables (IV), two-stage least squares (2SLS) or generalized method of moments (GMM) - instead of quasi-maximum likelihood (QML). Moments estimation does not require the existence of a unique stable solution as ML. The former is a valid estimation method as long as appropriate instrumental variables exist. The instruments proposed by [Kelejian and Prucha \(1998\)](#) as well as [Lee \(2003\)](#) serve as relevant and valid excluded instruments. [Kelejian and Prucha \(1998\)](#) suggest using higher order spatial lags of the exogenous variables,  $(W^2X, W^3X, \dots)$ . In social interactions,  $W^2$  denotes the second order effect or the effect through the peers' peers. Thus, it describes a channel of influence that reaches the student only indirectly. [Lee \(2003\)](#) refines the [Kelejian and Prucha \(1998\)](#) instruments with a two-step procedure to obtain the Best IV.<sup>6</sup> Including

---

<sup>6</sup>Calculation of the Best IV entails two steps. First, estimate equation (3.1) with 2SLS and  $IV = (X \ WX \ W^2X)$  to obtain  $(\hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2)$ . Second, the Best IV is simply  $BIV = W(I - \hat{\lambda}W)^{-1}(X\hat{\beta}_1 + WX\hat{\beta}_2)$ . The same procedure applies on within-transformed data.

the second and third powers of the Best IV improves efficiency, especially when comparing alternative socio-matrices specifications. Additionally, an exclusion restriction comes from the second or third order social effects of German language and mathematics school marks. Note that instruments are constructed using the average socio-matrix shown below.

## Peers' Sum and Average

Aggregate and average social interactions in the classroom form a complete network with uniform weights. The socio-matrix for a generic classroom is:

$$W_{c,aggregate} = \iota_{m_c} \iota_{m_c}' - I_{m_c} = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}, \quad c = 1, \dots, C \quad (3.6)$$

whereas the model for a generic student  $i$  in a generic classroom  $c$  becomes:

$$y_{ci} = \lambda \sum_{j=1, j \neq i}^{m_c} y_{cj} + x_{ci} \beta_1 + \sum_{j=1, j \neq i}^{m_c} x_{cj} \beta_2 + \alpha_s + \varepsilon_{ci} \quad (3.7)$$

in which  $m_c$  denotes classroom size.

The row-normalized version of socio-matrix (3.6) is:

$$W_{c,average} = \frac{1}{m_c - 1} (\iota_{m_c} \iota_{m_c}' - I_{m_c}) = \begin{pmatrix} 0 & \frac{1}{m_c - 1} & \cdots & \frac{1}{m_c - 1} \\ \frac{1}{m_c - 1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{m_c - 1} \\ \frac{1}{m_c - 1} & \cdots & \frac{1}{m_c - 1} & 0 \end{pmatrix}, \quad c = 1, \dots, C \quad (3.8)$$

implying the following generic-student model:

$$y_{ci} = \lambda \left( \frac{1}{m_c - 1} \sum_{j=1, j \neq i}^{m_c} y_{cj} \right) + x_{ci} \beta_1 + \left( \frac{1}{m_c - 1} \sum_{j=1, j \neq i}^{m_c} x_{cj} \right) \beta_2 + \alpha_s + \varepsilon_{ci}. \quad (3.9)$$

Thus, each student is equally affected by every classmate. In the first case, what matters is the *number* of peers, while in the second the *average influence* from peers. Obviously, an increasingly larger classroom size reduces individual influence for the latter but not for the former. Contextual effects or peers' characteristics also generate social effects but only

endogenous social interactions affect current behavior. The peers' average and aggregate models can be estimated by both method of moments (Kelejian and Prucha, 1998; Lee, 2003) and (quasi-)maximum likelihood (Lee, 2007a; Lee, Liu, and Lin, 2010). The row-normalization of the socio-matrix in (3.8) ensures  $|\lambda| < 1$ .<sup>7</sup> To avoid an arbitrary parameter space related to socio-matrix (3.6), I apply the inconsequential normalization suggested by Kelejian and Prucha (2010). The single normalizing factor is simply the maximum number of links. The structure of the aggregate and average socio-matrices does not depend on student achievement, and since the sample excludes classes formed on the basis of ability, it is plausible to assume that both  $W_{c,aggregate}$  and  $W_{c,average}$  are exogenous.

## Quantile Estimation

In order to shed light on the relationship between scores and explanatory variables other than the mean of the distribution, I resort to quantile regression. Let  $\lambda(\tau)$ ,  $\beta_1(\tau)$  and  $\beta_2(\tau)$  denote dependence of the coefficients on quantile  $\tau$ ,  $0 < \tau < 1$ . The quantile estimates reflect effects from endogenous social or exogenous explanatory variables at different points in the conditional distribution of the test scores. I mainly explore how peers' achievement affects low-achieving students (bad apples), those at the 25th, the median, the 75th quantiles and high-achieving students (stars). Kim and Muller (2004), Chernozhukov and Hansen (2006) and Lee (2007b) discuss estimation of a quantile model with an endogenous variable. The first stage estimates the mean regression with Ordinary Least Squares. Lee's (2003) Best instruments serve as excluded instruments. The second stage estimates the quantile regression at  $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$  using the residuals from the first stage as an additional explanatory variable. In a sense, the estimation procedure is a parametric version of Lee's (2007b) control function approach with bootstrapped standard errors.

## System of Simultaneous Equations

To allow for simultaneity effects in the PISA competencies, I estimate a system of equations in the three subject scores. Kelejian and Prucha (2004) discuss system estimation is a spatial setting, while Cohen-Cole, Liu, and Zenou (2012) and Liu (2014) lay down a taxonomy of models and identification conditions. The current application estimates the "square model" with simultaneity effects:

---

<sup>7</sup>Lee (2007a) points out that  $\lambda > 1 - \inf(m_c)$  and Tao and Lee (2014) that  $\lambda > 1$  but  $\lambda \neq 1$ . The condition  $|\lambda| < 1$  ensures that  $(I_n - \lambda W_n)^{-1}$  exists, allowing for the calculation of the Best IV as well as the possibility of a social multiplier ( $n$  denotes the full sample size).

$$Y_{1s} = \lambda_1 W_s Y_{1s} + \rho_2 Y_{2s} + \rho_3 Y_{3s} + X_s \beta_{11} + W_s X_s \beta_{12} + \iota_{m_s} \alpha_s + \varepsilon_{1s} \quad (3.10)$$

$$Y_{2s} = \lambda_2 W_s Y_{2s} + \rho_1 Y_{1s} + \rho_3 Y_{3s} + X_s \beta_{21} + W_s X_s \beta_{22} + \iota_{m_s} \alpha_s + \varepsilon_{2s} \quad (3.11)$$

$$Y_{3s} = \lambda_3 W_s Y_{3s} + \rho_1 Y_{1s} + \rho_2 Y_{2s} + X_s \beta_{31} + W_s X_s \beta_{32} + \iota_{m_s} \alpha_s + \varepsilon_{3s} \quad (3.12)$$

for each school  $s = 1, \dots, S$ . The system is estimated with ML without imposing normality of the error terms. After estimation, I calculate impacts incorporating effects across the system's equations as well as the social multiplier matrix  $(I_n - \lambda W_n)^{-1}$  for the full sample  $n$ .<sup>8</sup> I derive standard errors with the delta method.

## Fixed Effects

The school fixed effects in equation (3.1),  $\iota_{m_s} \alpha_s$ , reflect common environmental factors as well as sorting of students across schools. The general teaching quality and resources in a school affect students' academic performance. Also, parents tend to select neighborhoods on the basis of their income and surrounding school quality. All of the above might induce biases in the peer effects estimates. Thus, a within transformation of the model is necessary. For method-of-moments estimation I demean each observation from its school mean. The demeaning matrix is:

$$J_s = I_{m_s} - \frac{1}{m_s} \iota_{m_s} \iota_{m_s}' = \begin{pmatrix} 1 - \frac{1}{m_s} & -\frac{1}{m_s} & \dots & -\frac{1}{m_s} \\ -\frac{1}{m_s} & 1 - \frac{1}{m_s} & \ddots & -\frac{1}{m_s} \\ \vdots & \ddots & \ddots & \vdots \\ -\frac{1}{m_s} & \dots & -\frac{1}{m_s} & 1 - \frac{1}{m_s} \end{pmatrix}, \quad s = 1, \dots, S \quad (3.13)$$

and, thus, equation (3.1) becomes:

$$J_s Y_s = \lambda J_s W_s Y_s + J_s X_s \beta_1 + J_s W_s X_s \beta_2 + J_s \varepsilon_s \quad s = 1, \dots, S. \quad (3.14)$$

For quasi-maximum likelihood estimation, I within-transform the model using matrix  $F_s$ , the columns of which are the eigenvectors of matrix  $J_s$  corresponding to eigenvalue one. Equation (3.1) becomes:

---

<sup>8</sup>See [Bentler and Freeman \(1983\)](#) and [Tatsi \(2014\)](#).

$$F'_s Y_s = \lambda(F'_s W_s F_s) F'_s Y_s + F'_s X_s \beta_1 + (F'_s W_s F_s) F'_s X_s \beta_2 + F'_s \varepsilon_s \quad s = 1, \dots, S. \quad (3.15)$$

The within-transformed error term,  $F'_s \varepsilon_s$ , contains  $m_s - 1$  linearly independent vectors and its variance matrix is  $\sigma^2 I_{m_s-1}$ .

The main problem from demeaning with respect to the school instead of the classroom, is that I do not eliminate the classroom-teacher specific effect but the general school teaching quality. First, since ability tracking is excluded, there is no reason to believe that more able or experienced teachers get assigned *systematically* to certain classrooms. In general, in the absence of tracking, classrooms should in principle receive similar schooling resources. Second, “teacher shopping” is highly unlikely for students at the age of 15. Third, [Sacerdote \(2011\)](#) considers the teacher impact as a peer effect, referring to externalities in a broader sense. For instance, some teachers might assign group exercises, which leads to more intense interactions among students, or spend time answering questions posed by the bad apples even if the answer is obvious for the majority of the classmates. Finally, for the 9th grade in Germany it is usually the case the same teacher delivers lectures on a specific subject to many classrooms.

## Identification

Identification of endogenous and exogenous social interaction parameters suffers from the so-called reflection problem. Technically, the reflection problem ([Manski, 1993](#)) is a perfect collinearity problem: the expected peer outcome is collinear with the rest of the right-hand-side variables. Thus, the problem describes the inability to disentangle the endogenous from the exogenous social parameter even in the absence of correlated effects. Only functions of the two social effects can be uncovered. Intuitively, it describes the inability to explain dependencies among peers’ outcomes: is it because of peer interaction or because of peer characteristics? Furthermore, peer group members may experience exposure to some common external stimulus or share similar preferences, which the empiricist might erroneously perceive as the result of social interactions.

Regarding the reflection problem, [Lee \(2007a\)](#) offers a solution in the simplest of settings: a cross-section of data with peers who interact in a group and are affected by common group factors. Identification of the single-equation models stems from *sufficient* group size variation ([Boucher, Bramoullé, Djebbari, and Fortin, 2014](#)), meaning that one classroom has 11 students, another 13, another 15, etc. Insufficient classroom size variation, namely few distinct classroom sizes, manifests in multicollinearity. To increase variation and achieve identifica-

tion, I subtract observations from their respective school mean - instead of classroom. Notice that [Tatsi \(2014\)](#) demonstrates the non-identification of the aggregate effects model in the case of classroom interactions with classroom fixed effects.

For system identification, a sufficient condition with a row-normalized socio-matrix requires that matrices  $J_n$ ,  $J_n W_n$ ,  $J_n W_n^2$ ,  $J_n W_n^3$  are linearly independent. (see [Cohen-Cole, Liu, and Zenou, 2012](#)). I verify the condition by checking the rank of matrix  $Q = (J_n J_n W_n J_n W_n^2 J_n W_n^3)$ . Furthermore, since the system in equations (3.10)-(3.12) is nonrecursive, I verify its stability through the eigenvalues of matrix  $B$ : let  $B$  be the square coefficient matrix containing the effects of the endogenous variables. The system is stable if the absolute value of the largest eigenvalue of  $B$  is less than 1.

Apart from technical considerations, there are many reasons why the empirical evidence presented in this paper is reliable. First, I do not have to assume the possibility of student interactions because they emerge naturally from the setting. As [Tatsi \(2014\)](#) argues, students spend many hours a day with their classmates in the same space. The learning procedure is affected, for instance, by a noisy classmate who does not let others follow the teacher or a classmate who poses smart questions. What is more, students can directly observe the outcome of their peers, herein achievement. This argument relates to peer group misspecification as is the case when considering endogenous social interactions at the school level: it implies that hundreds of students know each others' academic performance, which is highly unlikely. Second, estimated endogenous social effects can not reflect exposure to common shocks because the econometric models include school fixed effects. Third, in this application selection does not occur at the individual level because the 9th grade is compulsory and, moreover, the student does not choose which school or classroom to attend. Selection is more problematic for students attending non-compulsory education, e.g. university. Fourth, any estimated dependence among students' scores can not reflect endogenous group formation or common preferences because I exclude schools forming classrooms on the basis of ability. To illustrate the nature of the bias imagine a classroom comprised only by high-achievers in mathematics. Then, the estimated endogenous effects parameter might simply capture a pre-existing propensity to excel in mathematics and not students imitating or learning from each other. The former is an example of endogenous group formation while the latter of social interactions. In this sample, students are not systematically assigned to classrooms according to ability; therefore, the endogenous effect indeed represents interactions. Fifth, estimating a system of simultaneous equations in the three scores allows for the first time to control for own preferences regarding courses, which might also be responsible for observed dependence among scores. Sixth, endogenous effects can not reflect measurement error in the outcome because the PISA scores are absolutely comparable across schools.

## 4 Empirical Results

### Which Students Matter?

The analysis starts with estimating social interaction models including only one socio-matrix at a time to match the usual practice in the literature so far. Then I compare four competing social interactions models: the two star networks, the aggregate effects and the traditional average. Comparison pertains to the endogenous effects parameter  $\lambda$  in equation (3.1). Panel A of Table 2 shows that endogenous effects parameters are statistically significant if the econometric equation includes a single socio-matrix (with the sole exception of aggregate effects in reading comprehension). Thus, all four alternatives would separately appear as a plausible social interactions model. This result implies first, that the peers' average effect is free from the "mechanical" bias in the Angrist (2014) sense because the estimated  $\lambda$  is positive and significant for models that go beyond the average; second, the invidious model of social interactions is not valid in this case because of the positive sign of the estimated  $\lambda$ . Regarding policy, the choice of model is clear: pick any of the individual influences - aggregate effects, stars and bad apples - because they imply individual-based as opposed to group-based rules (peers' average channel). Of course, policies aiming at single students are much easier and less costly to implement; moreover, they reach every classmate through the social multiplier. With more than one endogenous social effect, though, technical problems arise in estimation, especially with respect to efficiency and instruments' relevance. A good solution is to include only two socio-matrices at a time and obtain a preliminary picture of which one fits the data best. An even better solution is to resort to model selection by means of a J-Test (Kelejian, 2008; Kelejian and Piras, 2011, and Liu, Patacchini, and Zenou, 2014).

Each specification in Panel B of Table 2 compares two social interaction channels so that estimation entails two endogenous variables. The peers' aggregate estimated parameter is insignificant in all combinations. Students are not so responsive to changes in the sum of generic classmates' scores as they are to specific classmates at the extremes of the distribution, i.e. the stars or the bad apples. Notwithstanding the lack of statistical significance in the subject of mathematics, the stars' and bad apples' endogenous effects have positive signs. Liu, Patacchini, and Zenou (2014) compare local average with local aggregate social effects and find evidence to support the former in the educational context by means of a spatial J-test. Also, according to Tatsi (2014), the linear-in-means model slightly outperforms the linear-in-sums model based on the log-likelihood values. The peers' average endogenous parameter estimate is always positive and statistically significant, apart from the specification including the stars' reading comprehension scores. These simple models show that endogenous social effects from stars, bad apples or the peers' sum do not matter once I control for the peers'

average.

Perhaps the most interesting result of Table 2 is the specification with the two star networks: bad apples are more important than stars for all three subjects. Improving the performance of the 2 worst students has a positive effect on own scores. One plausible explanation is that bad apples tend to be disruptive and noisy (see Lazear, 2001); therefore they impede the learning process for the rest. A second plausible explanation is that they slow down the learning process, either because they need more time to finish an assigned task or because they receive most of the teacher’s attention. A third explanation is simply “last-place aversion” (see Kuziemko, Buell, Reich, and Norton, 2013): when the performance of the worst students improves, the rest get motivated; otherwise they risk becoming the classroom’s new bad apples.

Although Table 2 conveys a clear message on the appropriateness of group-based policies, ideally each specification should include three or all four competing endogenous social effects. As the latter would be a formidable task, I turn to a standard way of testing nonnested models, namely the spatial J-test. Under the null hypothesis the peers’ average model is the true model. I test the null hypothesis against multiple alternative models (stars, bad apples or peers’ sum). At a first step I estimate the stars, bad apples and peers’ sum specifications with 2SLS including school fixed effects and obtain the predicted values. At a second step, I estimate the peers’ average model with 2SLS including school fixed effects and the predicted values from the stars, bad apples and the peers’ sum as additional regressors. Statistical significance of the estimated parameter on the predicted value(s) provides evidence against the null. Table 3 presents the J-test results. None of the predicted values bears statistical significance, which verifies the main conclusions of Table 2.

Hence, the socio-matrix that fits the data best for classroom interactions among German 9th-graders is the average effects matrix. This is not surprising, given that students in Germany have already experienced tracking at the age of 10, dividing them into higher and lower ability institutions. The early tracking system does not leave much room for policy, since specific individuals do not matter significantly more than the average. Large variations in ability have already vanished.

## Inside Peers’ Average

Table 4 gathers the coefficient estimates for the social interactions model according to a student’s placement in the achievement distribution, namely the 10th, 25th, 50th, 75th and 90th quantiles. For mathematics, the effect from the peers’ average increases as we move across the quantiles of the distribution. Thus, everybody benefits from an increase in the average performance but students at the 90th and especially the 75th quantile the most. For



reading comprehension, students at the bottom of the distribution do not react to changes in the average scores but those at the median and especially the upper quantiles respond positively. In natural sciences, only those at the extremes of the distribution are affected by changes in mean classroom performance. Overall, the magnitude of the endogenous parameter estimates reveal that average peer achievement spillovers are more intense for students at the 75th and 90th quantiles in the three subjects. The results in Table 4 are in line with much of the literature that favors heterogeneous social interactions (see [Sacerdote, 2011](#), for a comprehensive review and results in [Fruehwirth, 2013](#)).

In reality, a student’s abilities and competences are inherently intertwined. Preferences for a career in engineering motivate some students towards the study of mathematics and natural sciences, in the same way the desire to become a literature teacher motivates others to focus on the German language course. The time a student spends studying for the German language course takes away effort and time from other subjects such as mathematics or science. On the other hand, when a student develops verbal abilities through the study of German language and quantitative abilities through mathematics, he/she simultaneously improves understanding of natural sciences by easily comprehending the textbook. Thus, in principle performance in any course is interdependent with performance in the rest of the courses. [Fruehwirth \(2014\)](#) emphasizes that peer achievement masks unobserved characteristics. To partially control for individual preferences and efforts in the three subjects, I estimate the simultaneous effect of own scores in the other two subjects by means of a nonrecursive system of equations.

Table 5 presents marginal effects after system estimation. After including feedback within the system’s equations, the endogenous social effects parameter estimates are 0.250 for mathematics, 0.373 for reading comprehension and 0.265 for natural science - all significant at 1%.<sup>9</sup> The average own impacts are calculated as the partial derivative of student  $i$ ’s score with respect to a change in student  $i$ ’s explanatory variable, and peer impacts as the partial derivative of student  $i$ ’s score with respect to a change in student  $j$ ’s explanatory variable. Both contain feedback not only from student  $j$  to student  $i$  but also from student  $i$  back to student  $j$  through other classmates. Since  $|\lambda| < 1$ , the feedback through the social multiplier matrix  $(I_n - \lambda W_n)^{-1} = I_n + \lambda W_n + \lambda^2 W_n^2 + \dots$  eventually vanishes. The sum of the two impacts gives an average total measure of changes in an explanatory variable from both  $i$  and  $j$ , and makes sense when both own and peer average impacts are economically and statistically significant.

The PISA score in mathematics has a positive own and peer impact on reading comprehension scores. Thus, other things equal, putting more effort in mathematics or being among

---

<sup>9</sup>Results of single equation estimation as well as the system’s direct and indirect effects are available upon request.

classmates who are better in mathematics positively affects own reading literacy. The impacts for natural sciences are also positive but lack statistical significance. Interestingly, the impact of reading comprehension scores is zero for mathematics and very small for natural sciences, regardless of the fact that estimates are statistically insignificant. The natural sciences impacts on the other two scores have a positive sign but again they are not statistically different from zero. Notice that, although it would be interesting to see the cross-peer effect, meaning impacts from the peers' other two scores on the third, identification of the non-recursive system is not feasible unless some of the own characteristics differ across equations (exclusion restriction). With the simultaneous effects, the conclusion drawn is that students should put more effort in the study of mathematics because it improves their performance in another subject.

Coming to predetermined characteristics, there is a significant gender differential for own impacts. In mathematics and natural sciences the sign is negative, while for reading comprehension is positive. Girls outperform boys only in verbal skills. For the latter there is a significant positive peer impact: being among females increases reading comprehension scores more than putting own effort. Although insignificant, the peer impact is positive for natural sciences as well. An explanation could be that girls are more disciplined and well-behaved. Less disruption helps the classmates' learning process. A student's age has adverse effects only for himself/herself. The result makes sense because much older students are grade repeaters and much younger grade skippers. Similarly, immigration background students score significantly less than native classmates. The effect is less prominent in mathematics, a subject in which solid knowledge of the German language is not so important. It is striking to see that a student's socio-economic status does not play a role for mathematics scores and does not generate any peer spillovers. From all predetermined characteristics, only cognitive ability test scores impact directly and indirectly PISA scores. Actually, for mathematics and natural sciences the peer impact is greater in magnitude than the own impact. Hence, own scores improve substantially in the presence of classmates with higher cognitive ability, reaching a total impact of around 38 points.

Understanding the lack of peer spillovers from individual traits - other than cognitive ability - lies in the core of the German educational system and the specific working subsample. As mentioned in Section 2, only 373 observations out of 5144 come from the low achievement school type, the so-called "Hauptschule". [Entorf and Tatsi \(2014\)](#), with a similar dataset for year 2003, report that older students as well as students with immigration and low socio-economic status are more likely to attend the "Hauptschule" type than the "Realschule" or the "Gymnasium". This fact is due to the early tracking at the age of 10. The low achievers, whose parents are most probably immigrants or low-income earners, end up more frequently in a

path that leads to vocational training rather than a university degree. Therefore, immigration background students or students of lower socio-economic status who reach the “Gymnasium” are very similar to the rest with respect to academic achievement. Early tracking has already filtered out large variations in ability.

## Policy Implications

The working subsample consists of classes formed - more or less - randomly. At a first glance, regrouping policies would make sense because gains are free from the effect of matching higher ability teachers to higher ability students. Regrouping students usually relies on some observed predetermined characteristic, for instance immigration background. Then, given students are not systematically assigned to classrooms, a suggestion is to compare among classes with higher and lower immigration status variations and decide whether immigration background homogeneity or heterogeneity is better for the regular student. But the main results of this paper show that none of the observed predetermined peer characteristics matter for own achievement. Therefore, regrouping policies are obsolete; for instance, a student’s performance will not increase by placing him/her in a classroom with a low proportion of immigrants, exactly because there are no spillovers from immigration background classmates. A similar argument holds for the other variables. The quantile estimation reveals that behavioral effects are heterogeneous, with the classroom stars benefiting the most from increases in average peer performance. This nonlinearity further complicates calculating expected gains from regrouping. The message of Table 5 regarding peer effects is clear: conditional on own performance in the three subjects and other relevant individual and family background characteristics, classmates affect each other only through their cognitive abilities. Even if reassignment or regrouping takes place according to ability, under the linear-in-means model the resulting welfare is zero, provided classroom sizes are equal (see [Hoxby and Weingarth, 2005](#)). The results by [Carrell, Sacerdote, and West \(2013\)](#) provide yet another argument against regrouping on the basis of average peer ability: after regrouping, endogenous group formation might take place in unpredictable ways and to the detriment of the group whose improvement was the initial goal.

The empirical findings uncover the superiority of the linear-in-means model against a number of competitors: the star, the bad apple and the aggregate effects models. The result implies, though, that policy implementation is challenging because it has to change the classroom norm. Group-based policies are appropriate as students perceive individual peer influence only as a fraction of the total. Nevertheless, for small enough classroom sizes indi-

vidual influence can be used to alter the norm.<sup>10</sup> In conjunction with bad apples being more important than stars, in small classrooms average performance can change through improving the achievement for the bad apples, for instance, through remedial classes.<sup>11</sup> Remedial classes cost in terms of paying teachers for additional working hours. Therefore, another proposal is mentoring, during which the stars mentor the bad apples. The latter is a policy with zero cost for the school and can be implemented easily through the teacher: assign stars and bad apples to work together. The proposed policies fit well with the results by Lazear (2001): bad apples are either disruptive or take time from the teacher by asking questions from which the rest do not learn. Raising the performance of the bad apples raises the average performance in small enough classrooms, which in turn raises the scores mostly for the stars. Quantile estimation shows that the upper quantiles' scores respond more intensely to increases in peers' average score. Thus, the explanation is that improving the bad apples frees time and efforts for the teacher to speed up the syllabus. Stars benefit themselves from mentoring because explaining or teaching others deepens own understanding.

## 5 Conclusion

Peer effects in the economics of education is an immensely studied field. The linear-in-means model is the channel that has been studied most extensively. Nevertheless, it has received considerable criticism because either it imposes homogeneous effects or ignores effects from specific individuals (Sacerdote, 2011). Thus, I explore several social interactions channels in the classroom, and especially star networks and the peers' sum. The current paper is the first to apply the extreme order statistic estimator by Tao and Lee (2014). The data come from the OECD's extended PISA (FDZ-IQB), and allow for estimation of ability social interaction parameters at the classroom level, where the core of the learning process takes place. Estimates do not reflect selection or endogenous group formation because I drop schools whose principals admit forming classes on the basis of ability. Also, missing values are not an issue of distress: the sample comprises more than 95% of the original dataset and none of the extreme values is a missing observation. Another identification concern comes from common external factors and selection of students into schools on the basis of parental income or preferences. I address biases of this sort by including school fixed effects in the econometric models. Model comparison and selection, either through estimating models with two socio-matrices or by testing directly their predictive power, point to the appropriateness of the traditional linear-in-means model. Furthermore, the bad apples model is second to the

---

<sup>10</sup>I will not argue in favor of smaller classrooms because it entails school spending.

<sup>11</sup>For recently documented benefits of remedial classes see Stillings-Candal (2015).

peers' average as it outperforms the stars and the peers' sum. It seems that the disruption and delay of the lecture's delivery has an adverse effect for the whole classroom that surpasses effects from good peers.

Then, I examine deeper the implications of conformity. In line with much of the previous literature (see [Hoxby and Weingarth, 2005](#); [Sacerdote, 2011](#), and [Fruehwirth, 2013](#)), I verify that endogenous social effects are indeed heterogeneous. Students placed at the upper quantiles of the ability distribution experience interdependencies more intensely than those at the median or the lower quantile. The classroom's good and top students benefit most from increases in their peers' average, possibly because they compete to become the best student in the classroom. Those at the bottom of the distribution also react positively to increases in the peers' average but to a lesser extent and with the exception of reading comprehension. A possible explanation is "last-place aversion": they increase own efforts to avoid the stigma of the lowest achiever in the class.

The literature considers separate equations in each subject score when estimating social interaction parameters. Contrary to this practice, I estimate a system of simultaneous equations in the three core competencies, namely mathematics, reading comprehension and natural sciences. The system not only includes endogenous and exogenous social interaction parameters but also simultaneity effects. Thus, we can see how a student's performance in each subject affects other subjects. Despite the fact that the system is nonrecursive and includes both endogenous and contextual social effects, I ensure that identification conditions are met. Unlike similar studies, e.g. [Cohen-Cole, Liu, and Zenou \(2012\)](#) or [Liu \(2014\)](#), the current paper considers feedback loops from within the system's endogeneity and decaying social multiplier effects when interpreting impacts. Conditional on student characteristics, only peers' cognitive ability test scores affect own achievement for all subjects and, of course, in a positive manner.

Finally, combining all of the empirical findings I suggest the following in order to improve achievement for the regular student: aim at improving the performance of the bad apples in small enough classrooms because only then the influence of single students can effectively change the social norm. Interestingly, by improving the average through the worst, the stars become even brighter.

## 6 Appendix

Table 1: Summary Statistics

Variable	Mean	Quantiles		
		10th	Median	90th
PISA Score in:				
Mathematics	535.26 (78.42)	434.09	535.74	636.62
Reading Comprehension	537.56 (76.47)	435.67	545.53	627.90
Natural Sciences	548.39 (80.03)	444.08	552.11	650.34
Female	0.52 (0.50)	-	-	-
Age	15.58 (0.57)	15	15.5	16.42
Immigration Background	0.19 (0.39)	-	-	-
Highest Index of Occupational Status	53.16 (15.70)	33	53	73
Cognitive Ability Test Scores	0.38 (1.35)	-1.16	0.57	2.17

Notes: Author's own calculations using the FDZ-IQB PISA 2006, Germany.  $n = 5144$ .

Table 2: Endogenous Social Interactions Parameters - Model Comparison

	Mathematics	Reading Comprehension	Natural Sciences
<i>Panel A</i>			
Stars' PISA Score	0.213*** (0.060)	0.321*** (0.092)	0.237** (0.096)
Bad Apples' PISA Score	0.436*** (0.132)	0.355*** (0.111)	0.511*** (0.128)
Peers' Aggregate PISA Score	0.127*** (0.039)	0.030 (0.074)	0.133*** (0.051)
Peers' Average PISA Score	0.383*** (0.086)	0.449*** (0.119)	0.437*** (0.102)
<i>Panel B</i>			
Stars' PISA Score	0.258 (0.185)	0.419*** (0.110)	0.583*** (0.187)
Peers' Aggregate PISA Score	-0.032 (0.121)	-0.099 (0.064)	-0.120 (0.110)
Bad Apples' PISA Score	0.267 (0.210)	0.416*** (0.117)	0.502*** (0.132)
Peers' Aggregate PISA Score	0.066 (0.063)	0.098* (0.057)	0.030 (0.047)
Stars' PISA Score	0.054 (0.092)	0.112 (0.147)	0.186* (0.103)
Bad Apples' PISA Score	0.670*** (0.136)	0.380*** (0.133)	0.461*** (0.142)
Stars' PISA Score	-0.062 (0.116)	0.110 (0.171)	-0.163 (0.136)
Peers' Average PISA Score	0.459*** (0.168)	0.337 (0.222)	0.601*** (0.164)
Bad Apples' PISA Score	0.025 (0.190)	0.072 (0.169)	0.230 (0.191)
Peers' Average PISA Score	0.370*** (0.127)	0.397** (0.186)	0.310** (0.153)
Peers' Aggregate PISA Score	-0.031 (0.066)	-0.013 (0.055)	-0.087 (0.058)
Peers' Average PISA Score	0.438*** (0.145)	0.464*** (0.122)	0.574*** (0.131)

Notes: Author's own calculations using the FDZ-IQB PISA 2006, Germany.  $n = 5144$ . Estimation of single socio-matrix specification with 2SLS and of two socio-matrices with Two-Step GMM including school fixed effects, gender, age, immigration background, highest index of occupational status in the family and cognitive ability test scores. Excluded instruments include the Best IV in mathematics/reading comprehension/natural sciences, their second/third powers and second order social effect of school marks in mathematics and German. All specifications pass tests for underidentification, weak instruments and overidentification. Robust to heteroscedasticity standard errors in parentheses. \*, \*\*, and \*\*\* denote significance at 10%, 5% and 1% level respectively.

Table 3: Endogenous Social Interactions Parameters - J-test for Model Selection

	Mathematics	Reading Comprehension	Natural Sciences
Average Peers' PISA Score	0.483*** (0.154)	0.531** (0.232)	0.645*** (0.176)
Predicted Value from Stars	-0.402 (0.268)	-0.104 (0.261)	-0.181 (0.133)
Predicted Value from Bad Apples	-0.179* (0.105)	-0.204 (0.181)	-0.157* (0.094)
Predicted Value from Peers' Sum	0.046 (0.212)	1.090 (0.934)	-0.222 (0.233)

Notes: Author's own calculations using the FDZ-IQB PISA 2006, Germany.  $n = 5144$ . Estimation with 2SLS including school fixed effects, gender, age, immigration background, highest index of occupational status in the family and cognitive ability test scores. Excluded instruments are the BEST IV in mathematics/reading comprehension/natural sciences as well as second and third powers. The specification passes tests for underidentification, weak instruments and overidentification. Robust to heteroscedasticity standard errors in parentheses. \*, \*\*, and \*\*\* denote significance at 10%, 5% and 1% level respectively.



Table 4: Endogenous Social Interactions Parameters - Quantile Estimation

	10th	25th	Median	75th	90th
Peers' Average PISA Score in Mathematics	0.298** (0.136)	0.226** (0.114)	0.346*** (0.102)	0.582*** (0.115)	0.470*** (0.147)
Peers' Average PISA Score in Reading Comprehension	0.340 (0.265)	0.301 (0.209)	0.325*** (0.121)	0.621*** (0.136)	0.618*** (0.187)
Peers' Average PISA Score in Natural Sciences	0.380** (0.183)	0.379** (0.172)	0.236* (0.121)	0.494*** (0.140)	0.663*** (0.216)

Notes: Author's own calculations using the FDZ-IQB PISA 2006, Germany.  $n = 5144$ . Estimation with 2SLS including school fixed effects, gender, age, immigration background, highest index of occupational status in the family and cognitive ability test scores. Excluded instruments are the BEST IV in mathematics/reading comprehension/natural sciences as well as second and third powers. Bootstrapped standard errors in parentheses. \*, \*\*, and \*\*\* denote significance at 10%, 5% and 1% level respectively.

Table 5: Average Impacts - System Estimation

Average Impacts	Mathematics			Reading Comprehension			Natural Sciences		
	Own	Peer	Total	Own	Peer	Total	Own	Peer	Total
PISA Score in Mathematics	-	-	-	0.676** (0.284)	0.303** (0.144)	0.979** (0.422)	0.642 (0.450)	0.185 (0.165)	0.827 (0.609)
PISA Score in Reading Comprehension	0.000 (0.167)	0.000 (0.045)	0.001 (0.212)	-	-	-	0.185 (0.170)	0.053 (0.053)	0.238 (0.221)
PISA Score in Natural Sciences	0.453 (0.463)	0.122 (0.158)	0.575 (0.618)	0.613* (0.319)	0.275* (0.149)	0.888* (0.462)	-	-	-
Female	-27.094*** (1.226)	-11.934 (7.988)	-39.029*** (7.051)	25.193*** (1.728)	37.979*** (9.073)	63.173*** (9.704)	-15.315*** (1.777)	3.523 (8.802)	-11.791 (10.581)
Age	-7.046*** (1.346)	-2.826 (7.419)	-9.872 (7.326)	-4.564*** (1.467)	-4.775 (8.843)	-9.339 (9.188)	-5.100*** (1.472)	1.438 (9.090)	-3.659 (9.530)
Immigration Background	-14.352*** (2.135)	-7.769 (8.056)	-22.121* (12.137)	-18.955*** (2.226)	1.5193 (15.906)	-17.436 (16.667)	-22.265*** (2.148)	-19.041 (14.573)	-41.306*** (15.412)
Highest Index of Occupational Status	0.078 (0.052)	-0.373 (0.371)	-0.295 (0.363)	0.207*** (0.057)	0.561 (0.471)	0.768 (0.490)	0.262*** (0.056)	-0.140 (0.422)	0.122 (0.429)
Cognitive Ability Test Scores	17.894*** (0.664)	19.916*** (2.805)	37.810*** (9.092)	12.408*** (0.775)	10.563** (5.006)	22.971*** (5.876)	17.570*** (0.858)	20.319*** (4.524)	37.889*** (9.359)

Note: Author's own calculations using the FDZ-IQB PISA 2006, Germany.  $n = 5144$  and  $n^* = 5035$ . System estimation with QML including school fixed effects and peer characteristics. Standard errors in parentheses. \*, \*\*, and \*\*\* denote significance at 10%, 5% and 1% level respectively.

## References

- ANGRIST, J. D. (2014): “The perils of peer effects,” *Labour Economics*, 30(0), 98–108, Special Section articles on "What determined the dynamics of labour economics research in the past 25 years? edited by Joop Hartog and and European Association of Labour Economists 25th Annual Conference, Turin, Italy, 19-21 September 2013 Edited by Michele Pellizzari.
- BENTLER, P., AND E. H. FREEMAN (1983): “Tests for Stability in Linear Structural Equation Systems,” *Psychometrika*, 48(1), 143–145.
- BOUCHER, V., Y. BRAMOULLÉ, H. DJEBBARI, AND B. FORTIN (2014): “Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation,” *Journal of Applied Econometrics*, 29(1), 91–109.
- CARRELL, S. E., B. I. SACERDOTE, AND J. E. WEST (2013): “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation,” *Econometrica*, 81(3), 855–882.
- CHERNOZHUKOV, V., AND C. HANSEN (2006): “Instrumental Quantile Regression Inference for Structural and Treatment Effect Models,” *Journal of Econometrics*, 132(2), 491–525.
- COHEN-COLE, E., X. LIU, AND Y. ZENOU (2012): “Multivariate Choice and Identification of Social Interactions,” Working Paper.
- ENTORF, H., AND E. TATSI (2014): “Peer Effects and Educational Inequality,” Working Paper, Goethe University Frankfurt.
- FRUEHWIRTH, J. C. (2013): “Identifying Peer Achievement Spillovers: Implications for Desegregation and the Achievement Gap,” *Quantitative Economics*, 4(1), 85–124.
- (2014): “Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box,” *The Review of Economics and Statistics*, forthcoming.
- GLAESER, E. L., AND J. SCHEINKMAN (2000): “Non-Market Interactions,” Working Paper 8053, National Bureau of Economic Research.
- HANUSHEK, E., AND L. WOESSMANN (2010): “The High Cost of Low Educational Performance. The Long-Run Economic Impact of Improving PISA Outcomes,” Written in consultation with members of the PISA Governing Board and Schleicher, A., Duval, R., Jakubowski, M. from the OECD Secretariat, Programme for International Student Assessment, Indicators and Analysis Division of the OECD Directorate for Education, OECD.

- HOXBY, C. M., AND G. WEINGARTH (2005): “Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects,” Mimeo, Harvard University.
- KELEJIAN, H. H. (2008): “A Spatial J-Test for Model Specification Against a Single or a Set of Nonnested Alternatives,” Working Paper, University of Maryland, College Park.
- KELEJIAN, H. H., AND G. PIRAS (2011): “An Extension of Kelejian’s J-test for Non-Nested Spatial Models,” *Regional Science and Urban Economics*, 41(3), 281–292.
- 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(1), 99–121.
- (2004): “Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations,” *Journal of Econometrics*, 118(1-2), 27–50.
- (2010): “Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances,” *Journal of Econometrics*, 157(1), 53–67.
- KIM, T.-H., AND C. MULLER (2004): “Two-stage Quantile Regression when the First Stage is Based on Quantile Regression,” *Econometrics Journal*, 7(1), 218–231.
- KUZIEMKO, I., R. BUELL, T. REICH, AND M. I. NORTON (2013): “Last-Place Aversion: Evidence and Redistributive Implications,” *The Quarterly Journal of Economics*.
- LAVY, V., M. D. PASERMAN, AND A. SCHLOSSER (2012): “Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom,” *The Economic Journal*, 122(559), 208–237.
- LAVY, V., O. SILVA, AND F. WEINHARDT (2012): “The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools,” *Journal of Labor Economics*, 30(2), 367–414.
- LAZEAR, E. P. (2001): “Educational Production,” *The Quarterly Journal of Economics*, 116(3), 777–803.
- LEE, L.-F. (2003): “Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances,” *Econometric Reviews*, 22(4), 307–335.
- (2007a): “Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects,” *Journal of Econometrics*, 140(2), 333 – 374.

- LEE, L.-F., X. LIU, AND X. LIN (2010): “Specification and Estimation of Social Interaction Models with Network Structures,” *Econometrics Journal*, 13(2), 145–176.
- LEE, S. (2007b): “Endogeneity in Quantile Regression Models: A Control Function Approach,” *Journal of Econometrics*, 141(2), 1131–1158.
- LIU, X. (2014): “Identification and Efficient Estimation of Simultaneous Equations Network Models,” *Journal of Business and Economic Statistics*, 32(4), 516–536.
- LIU, X., E. PATACCHINI, AND Y. ZENOU (2014): “Endogenous Peer Effects: Local Aggregate or Local Average?,” *Journal of Economic Behavior and Organization*, forthcoming.
- MANSKI, C. F. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *The Review of Economic Studies*, 60(3), 531–542.
- SACERDOTE, B. (2011): “Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” vol. 3 of *Handbook of the Economics of Education*, pp. 249–277. Elsevier.
- STILLINGS-CANDAL, C. (2015): “Match Corps Goes National,” Education Next, <http://educationnext.org/match-corps-goes-national>.
- TAO, J., AND L.-F. LEE (2014): “A Social Interaction Model with an Extreme Order Statistic,” *The Econometrics Journal*, 17(3), 197–240.
- TATSI, E. (2014): “Estimating Classroom Peer Effects with a Spatial Autoregressive Type Model,” Working Paper, Goethe University Frankfurt.