Received: 4 July 2022 | Revised: 18 February 2023 | Accepted: 25 February 2023

DOI: 10.1111/ssqu.13257

ORIGINAL ARTICLE

"Like with like" or "do like?" Modeling peer effects in the classroom

Jonathan Pratschke[1] | **Giovanni Abbiati**[2]

[1]Department of Social Sciences, University of Naples Federico II, Naples, Italy

[2]Department of Economics and Management, University of Brescia, Brescia, Italy

Correspondence
Jonathan Pratschke, Department of Social Sciences, University of Naples Federico II, Vico Monte della Pietà 1, Naples 80138, Italy.
Email: jonathan.pratschke@unina.it

Funding information:
The Italian Ministry for Education, Universities and Research under the PRIN funding programme (2010-11), Grant Number CUPE61J12000220001

Abstract

Objective: The authors discuss the role of peer networks in shaping the decision to enroll at university. Using panel data from Italy, they apply innovative statistical methods to study a sample of students as they complete high school and decide whether or not to attend university.

Methods: The authors use simultaneous autoregressive (SAR) models to analyze a four-wave panel database of Italian students. They explore the role of endogenous, exogenous, and correlated peer effects in relation to the decision about whether or not to enroll at university.

Results: The findings suggest that endogenous peer effects exert a significant influence on the probability of enrolling after controlling for homophilous preferences and a range of variables. Exogenous peer effects do not appear to influence this outcome. Sensitivity tests suggest that the results of the estimation are robust to selection.

Conclusions: This article contributes to an emerging body of literature on the use of SAR models to study peer effects, illustrating its considerable potential in the study of educational outcomes.

KEYWORDS

education, Italy, peer effects, simultaneous autoregressive models, social inequalities, university enrollment

The role and impacts of peer effects have been debated by sociologists for many years (see Lazarsfeld and Merton 1954). Some researchers argue that peers can have a considerable influence on social outcomes, particularly during adolescence (Haynie 2002), while others suggest that peer effects are little more than statistical artifacts (Angrist 2014). Despite this impasse, recent conceptual and methodological advances have led to the development of more powerful research designs, which have the potential to shed new light on this debate.

Our aim in this article is to review the challenges posed by the measurement of peer effects and to assess the potential of simultaneous autoregressive (SAR) models to address these. These models provide a framework for studying peer effects in experimental and non-experimental settings and have the potential to make a valuable contribution to education research. Their main advantage over other approaches is their ability to provide a comprehensive picture of peer effects, distinguishing between the role of social context, homophily, the characteristics of peers, and social interactions. This is particularly useful when studying educational outcomes, as it is important to control for shared learning environments, choice of friends, family characteristics, and reciprocal influence. Using a case study that analyses the propensity of Italian secondary school students to enroll at university, we show how these forms of influence can have different theoretical and policy implications and how the failure to distinguish between them can lead to misleading conclusions.

The existing literature on peer effects is rather fragmented and heterogeneous, both in substantive and technical terms. The resulting diversity makes it difficult to compare different research traditions and to assess the available evidence across disciplines. Many researchers continue to equate peer effects with the influence of fixed groups such as school classes (Bernburg, Thorlindsson, and Sigfusdottir 2009; Crosnoe 2009) or include the characteristics of peers as covariates in standard statistical models (e.g., Kim and Chang 2018). These designs are at risk of confounding, as they conflate distinct types of peer influence (Lauen and Gaddis 2013) and ignore the network structure of micro-level interactions (Steglich, Snijders, and Pearson 2010). Several scholars have drawn attention to the problems that arise when standard statistical methods are used to study the influence of social interactions (Brock and Durlauf 2000; Manski 1993; Moffitt 2001; Shalizi and Thomas 2011).

We start by providing a conceptual overview of peer effects and reviewing a range of approaches. We then show how different designs have been used in empirical research and argue that one of the difficulties researchers face when studying peer effects derives from the fact that it is frequently impossible to quantify one type of effect without simultaneously measuring others. We describe SAR models, including their assumptions and limitations, and present a case study that applies these techniques to a new set of survey data relating to Italian secondary school students. The case study enables us not only to showcase a set of methods that have not previously been used to study peer effects in relation to university enrollment but also to discuss the relationship between the statistical model and the substantive and policy-related concerns of educational researchers.

CONCEPTUAL OVERVIEW

In his influential contribution to the debate on peer effects in economics, Manski (1993, p. 532) distinguishes between *endogenous peer effects* (where "the propensity of an individual to behave in some way varies with the behavior of the group"), *exogenous peer effects* (where "the propensity of an individual to behave in some way varies with the exogenous characteristics of the group"), and *correlated peer effects* (where "individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments").

This three-way distinction implies that grouping individuals into schools, neighborhoods, and friendship networks can generate contextual effects (correlated peer effects), which must be controlled for in order to estimate the impact of social interactions (endogenous peer effects). Given that friendships do not arise at random, it is typically necessary to control for the characteristics of peers (exogenous peer effects) and for the ways in which people choose their friends (selection effects).

Endogenous peer effects are variously referred to as social interactions, contagion, induction, or spillover effects in what is now a vast cross-disciplinary literature on peer influence. They capture the impact of emulation, assimilation, information exchange, social learning, and other mechanisms rooted in micro-level social processes. Exogenous peer effects, by contrast, depend on characteristics such as the socioeconomic position of peers. People can be influenced not just by the behavior of their friends but also by their characteristics. The concept of exogenous peer effects includes the idea that peers can channel influences, opportunities, and resources that originate outside the peer group (Ragan, Osgood, and Feinberg 2014). Abstracting from social interactions and exogenous peer effects, correlated peer effects capture the impact of shared environments, as well as the impact of less intense interactions with a plurality of "familiar others" (Suh, Shi, and Brashears 2017).

Homophily and shared environments are the backcloth against which social interactions occur. Homophily has been defined as "the principle that a contact between similar people occurs at a higher rate than among dissimilar people" (McPherson, Smith-Lovin, and Cook 2001, p. 416). By grouping like with like, the tendency toward social homophily means that network flows tend to be relatively localized and the experiences of social actors tend to be reinforced. Homophily is relevant to the study of peer relations, as it reflects the existence of selection effects that can give rise to confounding (Shalizi and Thomas 2011; VanderWeele 2011).

Another important attribute of peer effects is that they form part of a continuous process of reciprocal influence over time (Steglich et al. 2010), leading to the well-known reflection problem (Manski 1993).[1] As in other areas of research regarding development processes across the life cycle, study designs seek to measure the influence of peers by taking slices through this process. When seeking to understand peer effects in non-experimental settings—as in other situations involving confounding—longitudinal designs are useful. However, latent forms of homophily and the incomplete observation of change in friendships can still lead to bias (Hsieh and Van Kippersluis 2015; Steglich et al. 2010). We will return to these issues below and describe techniques that address this source of confounding.

RESEARCH DESIGNS AND METHODS FOR ASSESSING PEER INFLUENCE

The conceptual and methodological complexities involved in the quantification of peer influence pose great challenges, and a number of innovative research designs have been proposed. In what follows, we briefly review the main approaches that have been applied to observational data: (1) use of group means/proportions, (2) social network analysis (SNA) and actor-based models, (3) multilevel models, (4) instrumental variable (IV) models, and (5) SAR models.

Standard regression models have been used widely in the study of peer effects by including group means or perceptions of peers (e.g., Ennett and Bauman 1993). In this kind of model, network position and the attributes of peers are treated as if they were exogenously determined. This approach is problematic for several reasons. First, it combines within a single parameter several different forms of peer influence. Second, it disregards social relations by implying that all members exert the same influence, an assumption difficult to reconcile with what we know about homophily and networks. Third, the standard "linear-in-means" model suffers from the reflection problem and yields biased estimates in the presence of reciprocal influence (Kenny and Judd 1986).

The second approach relies on SNA, which has transformed the study of peer relations in recent decades. It has become standard practice to discuss peer effects using concepts and terms drawn from SNA, such as ego (the respondent), alters (his or her peers), and egonets (personal networks). SNA focuses on the structure of friendships and other ties, and under its influence, researchers have moved away from using fixed peer groups like school classes and toward the use of overlapping networks that are centered on the individual actor (Wasserman and Faust 1994). Social network analysts are typically more interested in exploring the nature and structure of networks than in statistical estimation, although peer effects have been studied using stochastic actor-based models, which analyze the co-evolution (dynamic interplay) of network ties and actor attributes (Mundt 2013). There are limitations on the number of observations that can be included and longitudinal network data are required.

Another strand of research tackles the measurement of peer effects through the prism of hierarchical data structures (Tranmer, Steel, and Browne 2014). In the study of peer effects, multilevel models

[1] While peers can influence ego, ego can simultaneously influence his or her peers, making it difficult to separate endogenous and contextual effects.

emphasise the way in which individuals are nested within friendship networks and other social and institutional settings. Researchers have integrated multilevel modeling with SNA (Tranmer et al. 2014), treating friendship networks as a distinct level of variation, perhaps cross-classifying by classroom or school. Unfortunately, these techniques are not able to deal with large, sparse networks.

In economics, peer effects are often studied using IV models (An 2015; Bramoullé, Djebbari, and Fortin 2009). These models seek to control for endogeneity by using instruments that exert only an indirect influence on the outcome. IV models rely on assumptions that are difficult to satisfy in observational designs and typically cannot capture the reciprocal nature of peer influence (Staiger and Stock 1997). There are also constraints on the use of standard IV models with multiple peers and complex egonets.

The main methodological challenge posed by the study of peer effects arises from the need to bring together different aspects of the aforementioned techniques within a single model. We need the explanatory power of the multivariate regression model in order to control for family background and individual attributes. The model must be extended in order to control for shared environments and be able to account for reciprocal social influence. It is necessary to specify the network structure of peer relations and to control for the ways in which people choose their friends in the first place. The estimation techniques used must be able to deal with these characteristics while simultaneously making realistic assumptions about underlying mechanisms (see Steglich et al. 2010).

The final strand of research on peer effects that we discuss here seeks to achieve this goal by using SAR models. These models—which originated in geography—have been adapted to the study of peer relations by treating individuals (who have friends) as analogous to geographical areas (which have contiguous regions; Lee, Liu, and Lin 2010). Instead of using a spatial contiguity matrix, a sociomatrix is used to encode information on peer relations. Just as the general SAR model (Cliff and Ord 1973) relies on an autoregressive term to capture interactions between contiguous geographical areas, the SAR model for peer effects captures interactions between individuals.

THE SAR MODELING FRAMEWORK

We believe that the most promising recent development in research on peer effects is the integration of spatial analysis and SNA within the SAR model. Several articles published in recent years include variations on this approach, testifying to the interest that it has attracted across the social sciences (e.g., Ajilore 2015; Bramoullé et al. 2009; Calvó-Armengol, Patacchini, and Zenou 2009; Hsieh and Lee 2016; Macdonald-Wallis et al. 2011).

The SAR model can include (a) an appropriate sociomatrix to capture the reciprocal influence that friends exert on each other, (b) controls for baseline homophily and contextual factors, (c) exogenous peer effects, and (d) longitudinal components to control for inbreeding homophily (see Ajilore 2015; Lee et al. 2010; Lin 2015). It exploits the variability of personal networks across individuals to identify endogenous, exogenous, and correlated peer effects, subject to a reasonable set of assumptions.

By contrast with fixed peer groups (such as classes or schools), personal networks can overlap and vary across subjects. These variations provide sufficient information to statistically identify endogenous peer effects in most situations. For identification, it is sufficient to have individuals in the sample who are friends of an individual's friends but are not indicated as friends by the focal individual. The only way these individuals can influence ego is through his or her alters, and this feature is exploited by the estimator. This condition is satisfied automatically in most of the networks that are commonly studied in social science research (Bramoullé et al. 2009).

In the SAR model, friendships are represented using a sociomatrix that specifies the directed dyadic relationships that are observed in a given sample (e.g., the individuals nominated as friends by each respondent). The purpose of this matrix—denominated **W**—is to provide a concise way of specifying the direct and indirect influence that alters exert on ego. The diagonal elements are equal to zero,[2] and if other

[2] The effect of ego on himself or herself is, by definition, measured by adding individual-level explanatory variables to the model.

elements of the matrix are also equal to 0, then the corresponding spillovers are assumed to be zero. If two different elements on a row are equal to 1—where an individual indicates two friends, for example—then the spillovers are assumed to be equal (Leenders 2002). The weighting matrix **W** is taken to be known and non-stochastic and is typically row-normalized before use. The SAR model can be written in matrix notation as follows:

$$\mathrm{y} = \rho \mathrm{Wy} + \alpha \mathrm{WX}_1 + \beta \mathrm{X}_2 + \gamma \mathrm{X}_3 + \varepsilon \tag{1}$$

$$\varepsilon \sim N\left(0, \delta^2 \mathrm{I}_n\right)$$

where **y** is an $n \times 1$ vector of observations on the dependent variable; **W** is an $n \times n$ spatial-weighting matrix with 0 diagonal elements; **Wy** is an $n \times 1$ vector generally referred to as the "spatial lag," which captures the mean values of peers on the dependent variable; ρ is the corresponding scalar parameter generally referred to as the SAR parameter[3]; $\mathbf{X}_1$ is an $n \times k$ matrix of observations on k right-hand-side exogenous variables, α is the $k \times 1$ parameter vector relating to their spatial lags; $\mathbf{X}_2$ is an $n \times l$ matrix of observations on l right-hand-side exogenous variables, and β is the $l \times 1$ parameter vector relating to their direct effects; $\mathbf{X}_3$ is an $n \times m$ matrix of observations on m groups, and γ is the corresponding $m \times 1$ parameter vector of group-level fixed effects; ε is an $n \times 1$ vector of errors; σ^2 is the scalar noise variance parameter, and $\mathbf{I}_n$ is an $n \times n$ identity matrix.

The SAR model generates a system of equations to reflect the ways in which changes due to exogenous influence percolate through a set of overlapping peer networks.[4] As people influence each other reciprocally, their characteristics are refracted through these networks. The endogenous peer effects are captured by ρ, the exogenous peer effects are captured by α, and the correlated peer effects are captured by γ. The $\mathbf{X}_1$ matrix can also be used to control for the status of alters at a previous point in time, if required.

The SAR model is estimated here using the generalized spatial two-stage least squares estimator (GS2SLS), which uses the spatial-weighting matrix **W** in combination with the attributes recorded in the other matrices. GS2SLS is a method-of-moments estimator that allows for higher-order dependent variable lags. It was derived by Kelejian and Prucha (1998, 1999, 2010) and extended by Arraiz et al. (2008) and Drukker, Egger, and Prucha (2013).

Assumptions of the SAR model of peer effects

The identification of peer effects in the context of the SAR modeling framework rests on a series of assumptions.

SAR-specific assumptions are related to the nature, stability, and relevance of the peer relationships that are investigated. In the case of longitudinal SAR models, it is necessary to assume that the peer network is stable over the period of observation. In our case study, we assume that peer effects are homogeneous across the range of the outcome variable and that the influence of multiple peers can be summarized by the mean. Furthermore, any relevant exogenous or correlated effects must be included. Appropriate groups must be used to define fixed effects to control for the social or institutional context defining baseline homophily (e.g., membership of a classroom). It is important, finally, that these groups be sufficiently small and proximal to the subjects to capture the dynamics responsible for correlated effects.

In common with standard regression models, the exogeneity assumption is the most consequential for the reliable estimation of peer effects. SAR models assume that the influence of peers can be considered

[3] With an appropriately scaled spatial weighting matrix, this parameter will typically be bounded by −1 and 1.

[4] It is sometimes said that in a SAR model, there are as many effects of a covariate as there are units. LeSage and Pace (2009) define the average of these unit-level effects as the covariate effect, and we adopt this approach in this article.

exogenous to selection into the peer network. In other words, selection effects within peer networks are either absent or are controlled for within the model. However, self-selection into friendship networks on unmeasured grounds could be confounded with endogenous peer effects, as the latent factor underlying the choice of peers may simultaneously influence the outcome, thus biasing estimates of peer influence. A lively debate has originated within this research strand about the conditions under which self-selection into friendship networks can be brought under control (Christakis and Fowler 2007; Hsieh and Van Kipperluis 2015; Shalizi and Thomas 2011). The correct specification of correlated peer effects and the inclusion of temporal lags of the outcome variables for ego and alter are indeed important modeling features, but they are considered insufficient grounds for identification (Hsieh and Van Kipperluis 2015; Shalizi and Thomas 2011). Latent factors underlying inbreeding homophily could in theory continue to have an influence over time, generating bias in the estimation of endogenous peer effects. Given that peer networks are endogenous in nature, the most promising way of assessing the robustness of modeling results to selection bias is the use of a sensitivity testing approach as argued by Shalizi and Thomas (2011) and VanderWeele (2011). SAR models allow a wide range of sensitivity tests and alternative specifications to be assessed, providing information on the robustness of the assumptions. For example, the random allocation of friends within clusters and the reversal of the direction of friendship ties can indicate whether selection processes could influence estimates of endogenous peer influence.

In the case study presented below, we control for relevant covariates by including fixed effects for school class and by conditioning on prior status. Because socioeconomic position and academic ability are potential confounders, we include these in the model, together with a number of other covariates. Finally, we test whether the results are robust to the effects of latent inbreeding homophily by comparing the results when the direction of friendships is reversed and when friends are casually assigned within the classroom.

CASE STUDY

We apply the SAR model described above to the study of peer effects in the choice of whether or not to enroll at university after completing secondary school, using panel data from a sample of Italian school students. Improving equity of access to higher education is a high-level policy objective in most European countries, based on a combination of social justice and human capital considerations (OECD 2008). Stratification research in sociology has reached a broad consensus regarding the existence of a slow but detectable decline in educational inequalities in most developed countries in relation to both second-level and third-level educational attainments (Breen et al. 2009). Despite these advances, large disparities remain between social classes in terms of university enrollment. A range of mechanisms relating social origins to educational trajectories has been hypothesized, including risk-aversion and rational calculations (e.g., Breen and Goldthorpe 1997) or some combination of cultural capital and role models (Lareau 2003). Despite decades of research, including numerous attempts to measure peer effects, our understanding of the dynamics underlying these theories remains incomplete.

The Italian education system is characterized by a high degree of differentiation within the upper secondary cycle, as children choose between vocational, technical, and generalist schools at about age 14. The most prestigious schools are the *licei*, which provide a generalist education with the aim of preparing students for university, although there are relatively more prestigious (*liceo classico* and *liceo scientifico*) and less prestigious tracks (e.g., *liceo linguistico*, *liceo delle scienze umane*). The former is often chosen by families from the middle and upper classes, and there is a stark contrast between these elite institutions and the more vocationally oriented *istituti professionali* and *istituti tecnici*, which prepare young people for jobs in industry and services. Choice of upper secondary school stream is left to the discretion of families and is strongly influenced by socioeconomic position, making schools and classes rather homogeneous in terms of their social composition, at least at the extremes (e.g., Guetto and Vergolini 2017). Within the more vocationally oriented schools, pupils are much less likely to acquire the academic skills and study habits required for obtaining high grades and satisfying the entrance requirements of more selective third-level courses.

Our theoretical model posits that the choices of school students are influenced by a range of factors operating at different levels of the social structure. Family background is the most important of these factors, with socioeconomic position playing a central role in creating the conditions for correlated, endogenous, and exogenous peer effects. Schools and teachers also make an important contribution, together with the resources and opportunities they make available to students. Such resources are dependent upon the wider social context, which implies that correlated peer effects should be measured at the level of the school or even class. Third, we hypothesise that students influence each other, including both the reciprocal effect that close friends exert upon each other and the more diffuse impact of wider peer groups. Students are sorted into schools and classes but sort themselves into friendships, which have a complex network topography. They interact intensely with a small number of close friends, giving rise to endogenous peer effects. These networks also have the potential to channel resources from outside the school environment, such as where a best friend's father or mother helps with homework, creating the conditions for exogenous peer effects.

Very few studies have applied SAR models to European data to test these kinds of theoretical hypotheses. The data used in this case study come from a project that provides thick descriptive data on the school-to-university transitions of a cohort of upper secondary school students in Italy. The data collection plan comprised four waves in order to track the attitudes and behavior of students as they completed school and made the transition to work, higher education, other forms of study, or other roles. The schools that participated in the project were selected using a two-stage stratified sampling design. Overall, 62 schools in the provinces of Bologna, Milan, Salerno, and Vicenza were sampled and invited to participate. A total of 9058 students enrolled in the fifth and final year of secondary school completed the first questionnaire, and the response rate was very high at 99 percent (including a small number of students who were absent on the day of the survey but completed the questionnaire upon their return).

The first wave of data collection was carried out in October 2013, at the beginning of the school year, and final-year students in each sampled school filled out a paper-and-pencil questionnaire in class under the supervision of a trained supervisor. The collection of friendship data represents a novelty in relation to standard practices in large-scale surveys in Italy. Each student who participated in the survey was asked to name their best friends in the classroom. The second wave was fielded at the end of the school year (May 2014) when students were surveyed by telephone in relation to their study plans and beliefs about higher education. The third wave took place 6 months after the end of the school year, in November 2014, and the fourth and final wave was conducted in November 2015, when it was possible to record progress at the university, including a retrospective section on university enrollment for those who had not responded to the third wave. In this article, we use data from Waves 1 and 3, with a small amount of additional information coming from the retrospective part of Wave 4.

Individual friendship networks were constructed for each student drawing on responses to the Wave 1 questionnaire where each student could nominate up to three friends. Previous research suggests that most close friends of secondary school students frequent the same class and that most students have three such friends or fewer, suggesting that a threshold of three close friends is reasonable (Mercken et al. 2012; Mundt 2013).

The resulting egonets are specific to each individual student and can overlap. While almost one-fifth of the final sample did not indicate any class friends (17.8 percent), just over one-third (34.1 percent) provided three names. The mean number was 1.80, and we used row normalization to standardize the $\mathbf{W}$ matrix. This implies that peer influence is divided among nominated friends: Students who nominated a larger number of friends are not subject to a stronger endogenous peer effect but one that is spread across a larger number of individuals. Students who do not nominate any friends are included in the model but do not contribute to the estimation of the endogenous and exogenous peer effects.

The dependent variable is *university enrollment* (1 = yes; 0 = no), based on reports collected within a year of leaving school. As we have a dichotomous outcome variable, we specify a linear probability model; the statistical theory necessary to specify a logit link function in SAR models has not yet been developed. The linear probability model provides a good approximation to the true curve of probabilities and is widely used in applied research, particularly where the outcome is relatively balanced as is the case here (Hellevik 2007).

The individual-level covariates used in the model include the following measures: *sex* (1 = male, 0 = female), *dialect spoken at home* (1 = yes, 0 = no), *born in Italy* (1 = yes, 0 = no), *family type* (1 = two parents, 0 = one or no parents present), *family educational background* (1 = at least one parent has a university degree, 0 = neither parent has a university degree), a standardized index of *family economic difficulties*,[5] and final school *diploma examination result*.[6] In addition, there are two measures of cultural engagement, namely, *reading habits* (0 = reads for pleasure less than once a week, 1 = at least once a week) and *participation in cultural activities* (1 = sometimes goes to museums, theatres, or concerts, 0 = never goes). These variables control for baseline homophily, family background, and individual attributes.

Endogenous peer effects are measured by including a first-order autoregressive component, which is defined by the directed friendship ties described above. We assess exogenous peer effects in relation to two variables: (1) *family educational background* and (2) *family economic difficulties*. These variables cover two aspects of young people's socioeconomic background, which have been identified in the literature as important covariates when studying educational achievement and transitions.

As far as *correlated effects* are concerned, we include fixed effects to control for selection to a specific academic track and for the specific social, cultural, and institutional characteristics of the school or class. Using alternative specifications, we assess whether different kinds of fixed effects (type of school, school, class) yield different results when controlling for social context.[7]

To control for the main forms of *inbreeding homophily*, we include a measure that reflects the intention to enroll at university roughly 1 year before the outcome variable was measured (with the following response categories: definitely will attend university, probably will intend, probably will not attend, definitely will not attend, do not know; dummy variables were used to compare all other categories with "definitely will not attend").[8]

Descriptive statistics for the variables included in the model are provided in Table 1. The initial sample of 9058 pupils was reduced to 7212 after excluding individuals who did not participate in any of the later waves of data collection and for whom we have no information on university enrollment.[9] We also dropped 49 cases with large amounts of missing values, leaving a sample of 7163.[10] Item-level missing data were estimated using single imputation via the EM algorithm, involving a very small number of cases.

Table 2 reports the descriptive association between each categorical variable of Table 1 and university enrollment. In line with existing educational research, enrollment is stratified by sociodemographic variables, academic proficiency, and frequency of reading and participation in cultural events. The strongest predictor of university enrollment is parental education: having a highly educated parent increases the rate of enrollment by 30 percentage points. Children of university-educated parents are much less likely to speak dialect at home, more likely to read for pleasure, and to engage in cultural activities, and these characteristics are also associated with enrollment. Moreover, children of highly educated parents typically

[5] This is a scale based on six items derived from a longer list used in the European Union Statistics on Income and Living Conditions (EU-SILC) survey to measure economic strain. The items assess whether a family has encountered economic difficulties over the course of the past year in relation to taking holidays, buying clothes or food, paying bills, eating out once a month, or meeting transport costs.

[6] This examination is sat by all students at the end of the fifth year of upper secondary school as long as they obtain a score of at least 6 out of 10 in all subjects during the final year. The score is computed from grades assigned by teachers and from written as well as oral assessments, and the exam commission includes teachers with direct experience of students. The vast majority of students (97.6 percent in our sample) are admitted to sit the exam and very few fail (99.4 percent pass rate). This is not a standardized assessment at national level but nevertheless provides a measure of intra-class relativities in academic performance.

[7] The model which uses fixed effects for school classes (Model 5c) controls most effectively for the idiosyncratic components of the secondary school system in Italy, including differences in curricula, criteria for forming classes and teacher assignment.

[8] We also checked whether it was possible to control for the prior status of alters, as well as egos, but the results showed that coefficients became unstable due to collinearity problems, so this was not pursued further.

[9] We assume here that any selectivity in attrition rates can be controlled for by the characteristics included in the model. Regression-based analyses suggest that this is an acceptable assumption.

[10] Dropping these cases is a sub-optimal solution to the problem of missing data, but given the very small percentage of missing values, there is little risk of bias.

TABLE 1 Descriptive statistics ($N = 7163$)

Variable name		Percent
Categorical variables		
Male		46.9
Dialect at home		11.5
Born in Italy		94.7
Two-parent family		86.6
At least one parent with a university degree		25.6
Reads for pleasure at least once a week		25.8
Participates in cultural activities		89.9
University enrollment		62.2
Intentions 1 year before: definitely enroll		43.2
Intentions 1 year before: probably enroll		28.0
Intentions 1 year before: probably not enroll		11.2
Intentions 1 year before: definitely not enroll		8.7
Intentions 1 year before: do not know		8.8
Exam result 5 years before: pass		11.9
Exam result 5 years before: good		28.0
Exam result 5 years before: distinction		42.1
Exam result 5 years before: excellent		18.0
Continuous variables		
Family economic difficulties	*Mean*:	24.0
	SD:	22.2
Diploma exam grade	*Mean*:	76.3
	SD:	11.4
Exam score at end of Year 4	*Mean*:	6.9
	SD:	0.9

display better school performance, which is associated with the choice of school. The intentions declared by students at the beginning of the previous year are also highly predictive of university enrollment.

RESULTS

We begin by fitting a simple model with only the peer-lagged values of the dependent variable (Model 1) before adding the individual-level explanatory variables (Model 2), exogenous peer effects (Model 3), a longitudinal component (Model 4), and fixed effects for either type of school (Model 5a), school (Model 5b), or class (Model 5c). Table 3 contains the results.[11]

The coefficient associated with the endogenous peer effect is initially high in Model 1 but drops following the inclusion of individual covariates and fixed effects. This confirms the importance of family resources in the determination of educational choices and underlines the crucial role of sorting within the school system. The initial variations in the estimated effects highlight the scope for bias when models are

[11] It is not possible, with available software, to control for the complex sampling design used in this study, which means that standard errors may be slightly underestimated due to the way in which young students are nested within classes and schools in this sample.

TABLE 2 Enrollment by categories of categorical independent variables ($N = 7163$)

Variable	Category	Enrollment (%)
Sex	Male	55.7
	Female	67.8
Dialect at home	Yes	40.4
	No	64.9
Born in Italy	Yes	63.3
	No	42.6
Two-parent family	Yes	62.8
	No	57.7
At least one parent with a university degree	Yes	84.5
	No	54.5
Reads for pleasure at least once a week	Yes	74.9
	No	57.7
Participates in cultural activities	Yes	65.0
	No	36.7
Intentions 1 year before	Definitely enroll	92.8
	Probably enroll	63.3
	Probably not enroll	13.6
	Definitely not enroll	3.4
	Do not know	24.2
Exam result 5 years before	Pass	25.3
	Good	46.0
	Distinction	72.4
	Excellent	88.5

incorrectly specified. The inclusion of a longitudinal component to control for selection effects has a considerable impact on the coefficients associated with the individual-level explanatory variables. The results suggest that the decision to enroll at university has already been determined, to quite a high degree, before the beginning of the final year of secondary school. As we noted earlier, the choice of track for upper secondary school signals the orientation of students and their families in relation to university enrollment.

As Table 3 shows, the coefficient for the autoregressive term is 0.06 in the final model. If all nominated friends enroll at university, then the probability that a student enrolls increases by 0.06 over the course of their last year at school, controlling for a wide range of individual and family characteristics and for other kinds of peer effect, including various forms of selection. This coefficient captures the contribution of peers to the probability of enrolling at university and implies a significant spillover effect due to interactions between friends. It is interesting that the endogenous peer effect remains stable regardless of whether we specify fixed effects for school type, school, or class. By contrast, the exogenous peer effects for parental education and economic difficulties are not found to have a significant effect. These terms express the influence exerted by the parents of peers, who do not appear to influence this outcome for school students.

The model based on school type (Model 5a) shows that when compared to students attending *licei classici* and *licei scientifici* (who have a similar probability of enrolling), those attending technical schools have a probability of enrolling at university that is lower by 0.14, and those attending professional schools have a probability that is lower by 0.32. In fact, once we account for attendance at a prestigious *liceo*, other aspects of the social and institutional context of schools are relatively unimportant. Only small

TABLE 3 Generalized spatial two-stage least squares estimator estimates for linear probability models of peer effects on university enrollment (standard errors in parentheses)

	Model 1	Model 2	Model 3	Model 4	Model 5a	Model 5b	Model 5c
Endogenous effects							
Enrollment (lag)	0.22 (0.02)***	0.27 (0.02)***	0.22 (0.02)***	0.15 (0.01)***	0.06 (0.02)***	0.06 (0.02)***	0.06 (0.02)***
Exogenous effects							
Parental education (lag)			0.11 (0.02)***	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.02)
Family econ. diff. (lag)			−0.002 (0.0003)***	−0.001 (0.0002)***	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Correlated effects					School type***	School***	Class***
Individual attributes	N	Y	Y	Y	Y	Y	Y
Intentions 1 year before	N	N	N	Y	Y	Y	Y
Constant	0.50 (0.01)***	−0.47 (0.04)***	−0.43 (0.04)***	−0.08 (0.03)**	0.19 (0.04)***	−0.19 (0.05)***	−0.14 (0.09)
Wald $\chi 2$ (model)	109.81***	2536.50***	2648.37***	7262.30***	80,490.62***	8580.41***	10,125.03***
Degrees of freedom	1	10	12	16	22	112	489
Wald $\chi 2$ (spatial terms)	109.81***	269.22***	375.84***	145.79***	18.76***	18.79***	15.78**
Degrees of freedom	1	1	3	3	3	3	3
Pseudo R^2	-	0.24	0.25	0.50	0.53	0.54	0.59

Note: $N = 7163$.
$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

differences are observed whether we use five fixed effects for school type or 473 effects for school class. The direct and indirect effects of the explanatory variables are shown in Table 4, based on average unit-level effects (LeSage and Pace 2009). The "direct effects" indicate the influence of the individual-level explanatory variables net of any peer effects, while the "indirect effects" show how peer relations amplify the influence of these variables on the probability of university enrollment.

The SAR model assumes that there is no confounding due to inbreeding homophily on unmeasured characteristics that influence the outcome. In order to assess whether this assumption is warranted, we modified the **W** matrix to reverse the direction of friendships. This enables us to check whether latent inbreeding homophily (self-selection into friendship networks) could be confounded with endogenous peer effects (see An 2016). As selection effects are symmetrical, if there is confounding then the estimated size of the endogenous peer effect should remain stable regardless of the direction of ties (Christakis and Fowler 2007). Conversely, if the direction of ties is important to peer influence, as our theoretical framework suggests, then the endogenous peer effect should decrease when using the modified matrix. This is indeed the case, as the endogenous peer effect drops from 0.06 to 0.03, a substantively and statistically significant change. All other coefficients in the model remain stable. This reduces the likelihood of significant confounding due to latent inbreeding homophily, which can occur where alters are chosen due to their similarities to ego on grounds that directly influence the outcome variable but are not controlled for in the model.

An (2016) argues that even after rejecting the null hypothesis of no change in this effect using the directionality test, we cannot definitively exclude confounding, implying that we may have to make some additional assumptions before making causal inferences. To address this objection, we simulated peer networks by randomly generating friends for each student in the sample (respecting the original number of nominations). In this case, the endogenous peer effect became indistinguishable from 0, suggesting that the observed effect is indeed attributable to the specific configuration of personal friendship networks.

CONCLUSION

The SAR models presented in this article shed considerable light on peer influence in relation to university enrollment. They testify to the complexity of peer processes, providing evidence of different kinds of peer effects even when a longitudinal component and individual-level covariates are included. The model relies on purpose-designed estimators and can be applied to large data sets. Using a data set on Italian secondary school students, we quantify peer effects in relation to university enrollment during the last year of secondary education. The models include individual-level covariates and components that control for endogenous, exogenous, and correlated peer effects as well as baseline and manifest inbreeding homophily. The evidence suggests that endogenous peer effects have a statistically and substantively significant influence on the probability of enrolling at university. The results of our sensitivity tests suggest that this finding is robust to the effects of selection.

SAR models for peer effects have a number of advantages over alternative approaches and can be applied to a very wide range of research questions as long as data on social relationships are available. They do not rely on artificially created peer groups and do not require the researcher to specify IVs. They provide a wealth of policy-relevant information on different forms of social influence because of the equivalence between the terms specified in the regression equation and the substantive concepts at the center of scientific and policy-related debates.

Further research is required on the dynamics of friendships among school students, including tendencies toward homophily and change over time, the composition of peer networks within and outside the classroom, the role of individual attributes in friendship networks, asymmetries in friendship, and perceptions of wider peer group structures. It would also be helpful to explore how respondents define their peer groups, including the number and types of friends that they consider important. The results presented here suggest that these perceptions are relevant and that peer groups should not be inferred from group membership or through random assignment.

TABLE 4 Direct, indirect, and total effects of individual characteristics on university enrollment

Direct effects	**Coefficient**	**Std. error**
Male	0.025**	0.009
Dialect at home	−0.040**	0.014
Born in Italy	0.050**	0.018
Two-parent family	0.012	0.012
Parental education	0.038***	0.010
Family economic difficulties	−0.001***	0.0002
Diploma exam grade	0.005***	0.0004
Reads for pleasure	0.021**	0.009
Cultural participation	0.031*	0.013
Intentions: probably no	0.040*	0.018
Intentions: probably yes	0.427***	0.016
Intentions: definitely yes	0.557***	0.018
Intentions: do not know	0.185***	0.019
Indirect effects		
Male	0.0012*	0.0006
Dialect at home	−0.002*	0.0009
Born in Italy	0.002*	0.001
Two-parent family	0.0006	0.0006
Parental education	0.003	0.013
Family economic difficulties	−0.0002	0.0002
Diploma exam grade	0.0002**	0.0001
Reads for pleasure	0.001	0.001
Cultural participation	0.0015*	0.0008
Intentions: probably no	0.002	0.001
Intentions: probably yes	0.021***	0.006
Intentions: definitely yes	0.027***	0.008
Intentions: do not know	0.009**	0.003
Total effects		
Male	0.026**	0.010
Dialect at home	−0.041**	0.015
Born in Italy	0.052**	0.019
Two-parent family	0.013	0.012
Parental education	0.040*	0.017
Family economic difficulties	−0.001***	0.0003
Diploma exam grade	0.005***	0.0004
Reads for pleasure	0.022*	0.010
Cultural participation	0.032*	0.014
Intentions: probably no	0.042*	0.019
Intentions: probably yes	0.448***	0.018
Intentions: definitely yes	0.584***	0.020
Intentions: do not know	0.194***	0.020

Note: The estimates in this table relate to Model 5c in Table 3.
*$p < 0.05$;
**$p < 0.01$;
***$p < 0.001$. $N = 7163$.

The models presented in this article have some limitations that should be borne in mind. First, they are based on a representative sample of upper secondary school students in four Italian provinces, which means that the results may not generalize to the country as a whole or to different educational systems. The analysis uses observational data, and all inferences are conditional upon the assumptions incorporated within the model regarding the nature of peer effects and relevant control variables. As we described earlier, we use a linear probability model for our dichotomous dependent variable and assume that peer effects are homogeneous, averaging across peers. Finally, friendships are measured (with possible errors) at the start of the study, so we cannot assess the impact of changes in egonets over time. We believe that these limitations do not detract from the results and that the rigorous controls implemented render our estimates of endogenous peer effects slightly conservative.

The models suggest that peers generate spillover effects in relation to participation in higher education. These effects tend to reinforce existing socioeconomic inequalities but could potentially act as multipliers in the context of external interventions. This provides evidence of the ways in which macro-level and micro-level social processes are linked, suggesting that both can contribute to the reproduction of social inequalities over time. Key features of the school system associated with these inequalities include the socioeconomic differentiation of upper secondary schools, the benefits provided by attending more prestigious schools, and the spillover effects generated by interactions between friends.

Research on peer effects can contribute to improvements in educational outcomes and improved access to higher education if it enables policymakers and administrators to achieve a better understanding of their determinants. Policies cannot determine how young people choose their friends, but they can have an impact on the composition of the tie pool from which classroom friends are drawn, and they can take into account how multiplier effects influence outcomes. Further research is needed in order to ascertain whether and how peer effects vary and potentially combine across different social and educational contexts. This holds the promise of achieving a better understanding of peer effects in relation to a range of outcomes and potentially enhancing the educational trajectories of students from disadvantaged backgrounds.

ACKNOWLEDGMENTS

Open Access Funding provided by Universita degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

Ajilore, O. 2015. “Identifying Peer Effects Using Spatial Analysis: The Role of Peers on Risky Sexual Behavior.” *Review of Economics of the Household* 13(3):635–52.

An, W. 2015. "Instrumental Variables Estimates of Peer Effects in Social Networks." *Social Science Research* 50:382–94.

An, W. 2016. "On the Directionality Test of Peer Effects in Social Networks." *Sociological Methods & Research* 45(4):635–50.

Angrist, J. D. 2014. "The Perils of Peer Effects." *Labour Economics* 30:98–108.

Arraiz, I., D. M. Drukker, H. H. Kelejian, and I. R. Prucha. 2008. "A Spatial Cliff-Ord-Type Model with Heteroskedastic Innovations: Small and Large Sample Results." *CESIFO Working Paper* 2485.

Bernburg, J. G., T. Thorlindsson, and I. D. Sigfusdottir. 2009. "Relative Deprivation and Adolescent Outcomes in Iceland: A Multilevel Test." *Social Forces* 87(3):1223–50.

Bramoullé, Y., H. Djebbari, and B. Fortin. 2009. "Identification of Peer Effects through Social Networks." *Journal of Econometrics* 150(1):41–55.

Breen, R., and J. Goldthorpe. 1997 "Explaining Educational Differentials: Towards a Formal Rational Action Theory," *Rationality and Society* 9(3):275–305.

Breen, R., R. Luijkx, W. Müller, and R. Pollak. 2009 "Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries." *American Journal of Sociology* 114(5):1475–521.

Brock, W. A., and S. N. Durlauf. 2000. "Interactions-Based Models." In *Handbook of Econometrics* (Vol. 5), edited by J. Heckman and E. Leamer, 3297–380. Amsterdam: North-Holland.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou. 2009. "Peer Effects and Social Networks in Education." *The Review of Economic Studies* 76(4):1239–67.

Christakis, N. A., and J. H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *New England Journal of Medicine* 2007(357):370–79.

Cliff, A. D., and J. K. Ord. 1973. *Spatial Autocorrelation*. London: Pion.

Crosnoe, R. 2009. "Low-Income Students and the Socioeconomic Composition of Public High Schools." *American Sociological Review* 74(5):709–30.

Drukker, D. M., P. Egger, and I. R. Prucha. 2013. "On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors." *Econometric Reviews* 32(5-6):686–733.

Ennett, S. T., and K. E. Bauman. 1993. "Peer Group Structure and Adolescent Cigarette Smoking: A Social Network Analysis." *Journal of Health and Social Behavior* 34(3):226–36.

Guetto, R., and L. Vergolini. 2017. "Educational Expansion without Equalization: A Reappraisal of the 'Effectively Maintained Inequality' Hypothesis in Children's Choice of the Upper Secondary Track." *European Societies* 19(1):1–27.

Haynie, D. L. 2002. "Friendship Networks and Delinquency: The Relative Nature of Peer Delinquency." *Journal of Quantitative Criminology* 18(2):99–134.

Hellevik, O. 2007. "Linear versus Logistic Regression When the Dependent Variable Is a Dichotomy." *Quality & Quantity*, 43(1):59–74.

Hsieh, C. -S., and L. F. Lee. 2016. "A Social Interactions Model with Endogenous Friendship Formation and Selectivity." *Journal of Applied Econometrics* 31(2):301–19.

Hsieh, C. -S., and H. Van Kippersluis. 2015. "Smoking Initiation: Peers and Personality." *Tinbergen Institute Discussion Paper TI 15–093/V*.

Kim, H. H, and P. Y. Chang. 2018. "The Impact of Delinquent Friendship Networks and Neighborhood Context on Suicidal Ideation among South Korean Youths." *Social Forces* 97(1):347–76.

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.

Kelejian, H. H., and I. R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40(2):509–33.

Kelejian, H. H., and I. R. Prucha. 2010. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances." *Journal of Econometrics* 157(1):553–67.

Kenny, D. A., and C. M. Judd. 1986. "Consequences of Violating the Independence Assumption in Analysis of Variance." *Psychological Bulletin* 99:422–31.

Lareau, A. 2003. *Unequal Childhoods: Class, Race, and Family Life*. Berkeley: University of California Press.

Lauen, D. L., and S. M. Gaddis. 2013. "Exposure to Classroom Poverty and Test Score Achievement: Contextual Effects or Selection?" *American Journal of Sociology* 118(4): 943–979.

Lazarsfeld, P. F., and R. K. Merton. 1954. "Friendship as Social Process." In Freedom and Control in Modern Society*, edited by* M. Berger, T. Abel, and C. Page, 18–66. New York: Octagon.

Lee, L., X. Liu, and X. Lin. 2010. "Specification and Estimation of Social Interaction Models with Network Structures: Estimation of Social Network Models." *Econometrics Journal* 13(2):145–76.

Leenders, R. 2002. "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix." *Social Networks* 24(1):21–47.

LeSage, J., and R. K. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: Chapman & Hall/CRC.

Lin, X. 2015. "Utilizing Spatial Autoregressive Models to Identify Peer Effects among Adolescents." *Empirical Economics* 49(3):929–60.

Macdonald-Wallis, K., R. Jago, A. S. Page, R. Brockman & J. L. Thompson 2011. "School-Based Friendship Networks and Children's Physical Activity: A Spatial Analytical Approach." *Social Science & Medicine* 73(1):6–12.

Manski, C. F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60(3):531–42.

McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27(1):415–44.

Mercken, L., C. Steglich, R. Knibbe, and H. de Vries. 2012. "Dynamics of Friendship Networks and Alcohol Use in Early and Mid-Adolescence." *Journal of Studies on Alcohol and Drugs* 73:99–110.

Moffitt, R. A. 2001. "Policy Interventions, Low-Level Equilibria, and Social Interactions." In Social Dynamics*, edited by* Brookings Institute, 45–82. Cambridge, MA: MIT Press.

Mundt, M. P. 2013. "Social Network Analysis of Peer Effects on Binge Drinking among U.S. Adolescents." In *Social Computing, Behavioral-Cultural Modeling and Prediction*, edited by A. Greenberg, W. G. Kennedy, and N. D. Bos, 123–34. Berlin: Springer-Verlag.

OECD. 2008. *Education at a Glance*. Paris: OECD.

Ragan, D. T., D. W. Osgood, and M. E. Feinberg (2014. "Friends as a Bridge to Parental Influence: Implications for Adolescent Alcohol Use." *Social Forces* 92(3):1061–85.

Shalizi, C. R., and A. C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods & Research* 40(2):211–39.

Staiger, D., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65(3):557–86.

Steglich, C., T. A. B. Snijders, and M. Pearson. 2010. "Dynamic Networks and Behavior: Separating Selection from Influence." *Sociological Methodology* 40(1):329–93.

Suh, C. S., Y. Shi, and M. E. Brashears. 2017. "Negligible Connections? The Role of Familiar Others in the Diffusion of Smoking among Adolescents." *Social Forces* 96(1):423–48.

Tranmer, M., D. Steel, and W. J. Browne. 2014. "Multiple-Membership Multiple-Classification Models for Social Network and Group Dependences." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 177(2):439–55.

VanderWeele, T. J. 2011. "Sensitivity analysis for contagion effects in social networks." *Sociological Methods & Research* 40(2):240–55.

Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

How to cite this article: Pratschke, Jonathan, and Giovanni Abbiati. 2023. "“Like with like” or “do like?” Modeling peer effects in the classroom." *Social Science Quarterly* 00–00. https://doi.org/10.1111/ssqu.13257