CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 603

'Like with Like' or 'Do Like'? Modelling Peer Effects in The Classroom

Giovanni Abbiati and Jonathan Pratschke

February 2021

University of Naples Federico II | University of Salerno | Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
Department of Economics – University of Naples
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

WORKING PAPER NO. 603

'Like with Like' or 'Do Like'? Modelling Peer Effects in The Classroom

Giovanni Abbiati[*] and Jonathan Pratschke[†]

Abstract
This paper reviews the recent peer effects literature and showcases the simultaneous autoregressive model, which integrates aspects of multiple regression modelling, instrumental variables, social network analysis and longitudinal analysis. It describes state of the art techniques for making inferences using survey data, clarifies the assumptions made by statistical models and provides further evidence on the impact of peers in education. The paper includes a case study using data from an Italian survey to study peer effects in relation to university enrollment. The model includes components that control for endogenous, exogenous and correlated peer effects as well as different forms of selection. The evidence presented in the paper suggests that endogenous peer effects have a statistically and substantively significant influence on the probability of enrolling at university, measured over one year. Sensitivity tests suggest that the results of the estimation are robust to confounding due to latent homophily and other potential sources of bias.

Keywords: Peer effects; Simultaneous auto-regressive models; Education; Social inequalities; University enrollment; Italy.

[*] Università di Milano. Email: giovanni.abbiati@unimi.it.

[†] Università di Napoli Federico II and CSEF. Email: jonathan.pratschke@unina.it. (*Corresponding author*)

Table of contents

Introduction

1. *Conceptual Overview*

2. *Research Designs and Methods for Assessing Peer Influence*

3. *The SAR Modeling Framework*

4. *Assumptions and Limitations of SAR Models of Peer Effects*

5. *Case Study*

Conclusions

References

1. Introduction

The role and impacts of peer effects have been debated by sociologists for many years (see, for example, Coleman, Katz, and Menzel 1966; Lazarsfeld and Merton, 1954; Merton 1968; Newcomb 1962). Some researchers argue that peers can have a considerable influence on social outcomes, particularly during adolescence (Haynie 2002), whilst others suggest that peer effects are little more than statistical artefacts (Angrist 2014; Cohen-Cole and Fletcher 2008; Eisenberg 2004). 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 the role of peer effects.

Our aim in this paper is to review the challenges posed by the measurement of peer effects and to assess the potential of simultaneous auto-regressive models (SAR) to overcome these. These models provide a framework for studying peer effects in both 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 such factors as shared learning environments, choice of friends, family characteristics and reciprocal influence. Using a case study which analyses the propensity of Italian secondary school students to enrol at university, we show in this paper 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, heterogeneous and technical in nature, and Graham and Hahn (2005:1) observe that "[…] empirical work continues to be characterized by a plethora of seemingly idiosyncratic approaches to identification and estimation. This diversity of approaches makes it difficult to compare the results of different researchers and hence a clear assessment of the empirical evidence is problematic." Greater multidisciplinary effort is arguably needed in order to build on recent methodological advances and to enable applied researchers across the social sciences to use appropriate research designs and methods when studying peer effects.

Many researchers continue to equate peer effects with the influence of fixed groups such as school classes (Crosnoe 2009; Harding 2003, Bernburg, Thorlindsson, and Sigfusdottir 2009), or include the characteristics of peers as covariates in standard statistical models (e.g. Hyun-soo Kim and Chang, 2018). These designs are at risk of confounding, as they conflate conceptually distinct types of peer influence (Lauen and

Gaddis 2013) and ignore the network structure of micro-level interactions (see, for example, Steglich, Snijders, and Pearson 2010; Carrington, Scott, and Wasserman 2005). 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). These concerns must be addressed comprehensively in order to make further progress in the empirical study of peer influence in the field of education.

We start by providing a conceptual overview of peer effects and by reviewing influential approaches. In the following section we show how different designs have actually been used in empirical research. We 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 identifying and measuring all others. In Section 3 we describe SAR models, which provide a comprehensive framework for achieving this. In Section 4 we set out the assumptions and limitations of these models and in Section 5 we 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 enrolment, but also to discuss the relationship between the statistical model and the substantive and policy-related concerns of educational researchers.

1 Conceptual overview

In his influential contribution to the debate on peer effects within the field of Economics, Manski (1993:532) 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 the grouping together of individuals into schools, neighbourhoods 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 may also be necessary to control for the characteristics of peers (exogenous peer effects) and for the way in which people choose their peers in the first place (selection effects).

Endogenous peer effects are variously referred to as social interactions, contagion, induction, spillover or multiplier 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. The theoretical importance of these mechanisms has led researchers in Sociology and Economics, in particular, to develop methods to measure the effect of social interactions while controlling for confounding factors (e.g. Van den Bulte and

Lilien, 2001; Christakis and Fowler, 2007, 2013; Cohen-Cole and Fletcher, 2008; Fowler and Christakis, 2008; An, 2011; VanderWeele and An, 2013).

Exogenous peer effects 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). These effects are related to broader social inequalities, implying that achieving a better understanding of their influence can facilitate the development of more effective policies for social inclusion.

Abstracting from social interaction and exogenous peer effects, correlated peer effects capture the impact of shared environments, as mentioned above. They also capture the impact of less intense interactions with a plurality of “familiar others” (Suh, Shi, and Brashears 2017). The key insight of scholars such as Manski is that the identification and measurement of peer effects depends on the adoption of appropriate conceptual distinctions and the use of methods which reflect these distinctions. Peer effects form a complex, integrated whole and statistical methods must mirror this if we wish to understand how peers influence social outcomes. In the absence of effective controls, other forms of influence may masquerade as endogenous peer effects.

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 et al. 2001:416). By grouping ‘like with like’, the tendency towards social homophily means that network flows tend to be relatively localized, with the result that the experiences of social actors tend to be reinforced. *Baseline homophily* is created by the demography of the potential tie pool (McPherson et al. 2001), due to the way in which individuals are sorted into certain social or institutional settings (like a classroom), whilst *inbreeding homophily* involves the active expression of preference as relationships are formed (such as classroom friends). Homophily is relevant to the study of peer relations to the extent that it reflects the existence of selection and self-selection effects, which 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, Snijders, and Pearson 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 changes in friendships can still lead to bias (Hsieh and Van

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

Kippersluis 2015; Steglich, Snijders, and Pearson 2010). We will return to these issues below, as well as describing techniques that can addressing the source of confounding.

2 Research designs and methods for assessing peer influence

The sheer conceptual and methodological complexity involved in the quantification of peer influence poses 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[2]: (1) use of group means/proportions, (2) social network analysis and actor-based models, (3) multilevel models, (4) instrumental variable models and (5) simultaneous auto-regressive models. We focus on their applications, strengths and shortcomings as well as highlighting their assumptions.

Standard regression models have been used widely in the study of peer effects by including group means or perceptions of peers, as we mentioned earlier (e.g. Ammermueller and Pischke 2009; Betts and Morell 1999; Buckner et al., 2006; Ennett and Bauman 1993; Huba and Bentler 1980; Lavy, Silva, and Weinhardt 2009). 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, not least

[2] We do not include experimental approaches in this review because they belong to a different methodological tradition and do not rely on survey data. The aim of these designs is to neutralise sorting mechanisms by *artificially* creating peer groups. The external validity of such studies has been questioned (Carrell, Sacerdote and West 2013; Lyle 2007) and the literature on homophily suggests that peers may be most important when they are actually chosen (McPherson et al., 2001). Moreover, randomisation in itself does not solve the reflection problem (Hsieh and Van Kippersluis 2015).

because it combines within a single parameter several different forms of peer influence, which can lead to confounding and misleading inferences. Secondly, the use of group means implies that all members of a group influence the outcomes of its members in the same way, regardless of their dyadic social relations, an assumption that is difficult to reconcile with what we know about homophily and networks (McPherson et al. 2001). Thirdly, 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 to studying peer influence relies on social network analysis (SNA), which has transformed the study of peer relations in recent decades. Indeed, 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 (Wasserman and Faust, 1994), and under its influence researchers have moved away from using fixed, non-overlapping peer groups like school classes and towards the use of networks which are centered on the individual actor.

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; Snijders, Van de Bunt, and Steglich 2010). These models aim to eliminate bias due to the incomplete observation of network topography and use data on changing networks and their influence on one or more outcomes: “the process of co-evolution of network and behaviors from one wave of data to another is simulated as a result of a potentially large number of individually unobserved micro-step changes, and the network and behavior preferences parameters can be estimated” (Mundt 2013:124). The primary assumption of actor-based models is that individuals choose their friendship ties and their behaviors in micro-steps, which are simulated, controlling for individual attributes and aspects of network structure. They condition on the structure of friendships and behaviors at one point in time in order to identify factors that could influence subsequent change, using a longitudinal design to control for selection effects. Available software is currently limited to networks with roughly 1,000 nodes and longitudinal network data are also required.

Another strand of research tackles the measurement of peer effects through the prism of hierarchical data structures (Ryan 2001; Tranmer 2010; Tranmer, Steel, and Browne 2014). Multilevel models were originally developed to identify school and neighbourhood effects by partitioning variance and linking the resulting components with different levels. In the study of peer effects, multilevel models emphasise the way in which individuals are nested within friendship networks and other social and institutional settings. Researchers have also built bridges between different traditions by integrating multilevel modeling with SNA (Tranmer, Steel, and Browne 2014), treating friendship networks as a distinct level of variation whilst cross-classifying by classroom and/or school. These techniques are not currently able to deal with large, sparse networks.

In Economics, peer effects are often studied using instrumental variable (IV) models (An 2015; Bramoullé, Djebbari, and Fortin, 2009; De Giorgi, Pellizzari, and Redaelli 2010; Duncan, Haller, and Portes, 1968; O'Malley et al., 2014). These models seek to control for endogeneity by using instruments which influence the outcome only indirectly (Aral, Muchnik, and Sundararajan 2009). Instrumental variables must satisfy a 'relevance' condition (having a non-zero covariance with the endogenous variable) and an 'exclusion' condition (having a zero covariance with the error term) (An 2015). Scholars have shown that these criteria are difficult to satisfy, and Bound, Jaeger, and Baker (1995) and Staiger and Stock (1997) show that if the correlation between the IV and the endogenous variable is low, bias can result. IV models of peer effects do not provide a satisfactory solution to the reflection problem because they do not capture the reciprocal nature of peer influence. There are also constraints on the use of IV models in the presence of multiple peers, and in the presence of complex egonets.

The main methodological challenge posed by the study of peer effects arises from the need to bring different aspects of the aforementioned techniques together 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 it must 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. Finally, the estimation techniques used must be able to deal with these characteristics of the model whilst making realistic assumptions about the underlying mechanisms (see Steglich, Snijders, and Pearson, 2010).

The final strand of research on peer effects that we will discuss achieves this goal by using simultaneous auto-regressive models. These models—which originated in Geography—have been adapted to the study of peer relations by treating individuals (who have close 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 auto-regressive term to capture interactions between contiguous geographical areas, the simultaneous auto-regressive model for peer effects captures interactions between individuals. We will provide a more detailed description of this modeling approach in the next section, followed by a critical assessment and then our case study.

3 The SAR Modeling framework

We believe that the most promising recent development in research on peer effects is the integration of spatial and social network analysis within the SAR model. Several papers published in 2009-10 include variations on this approach (Bramoullé, Djebbari, and Fortin 2009; Calvó-Armengol, Patacchini, and Zenou 2009; De Giorgi, Pellizzari, and Redaelli 2010; Laschever 2009; Lee, Liu, and Lin 2010). The proliferation of applied papers over the last few years testifies to the interest that this approach has attracted across the social sciences (e.g. Ajilore 2015; De Melo 2014; Hsieh and Lee 2016; Hsieh and Lin 2017; Liu 2014; 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) a way of controlling for baseline homophily and contextual factors, (c) exogenous peer effects and (d) a longitudinal component to control for inbreeding homophily (see Lee, Liu, and Lin, 2010; Lin 2010, 2015; Ajilore 2015). It exploits the variability of personal networks across individuals to identify endogenous, exogenous and correlated peer effects, subject to certain assumptions and conditions, which will be discussed in greater detail below.

By contrast with fixed peer groups (such as classes or schools), personal networks overlap and vary across subjects. These variations provide sufficient information to statistically identify endogenous peer effects in most situations. This is the principal methodological insight associated with the use of SAR models to study peer effects. 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, Djebbari, and Fortin 2009).

In the SAR model, friendships are represented using a sociomatrix which specifies the directed dyadic relationships that are observed in the sample (e.g. the individuals nominated as friends by each respondent). The purpose of this matrix—denominated **W** —is to provide a flexible and concise way to specify the direct and indirect influence that alters can exert on ego. The diagonal elements are equal to zero[3] and if other elements of the matrix are equal to 0, then the corresponding spillover is 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-normalised before use.

The SAR model can be written in matrix notation as follows:

$$y = \lambda Wy + \gamma W X_1 + \beta \mathrm{X}_2 + \alpha X_3 + \epsilon \tag{1}$$

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', and captures the mean values of peers on the dependent variable; λ is the corresponding scalar parameter generally referred to as the 'SAR parameter'[4]; $\mathbf{X}_1$ is an $n \times k$ matrix of observations on k right-hand-side exogenous

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

4 With an appropriately-scaled spatial weighting matrix, this parameter will typically be bounded by -1 and 1.

variables and $\boldsymbol{\gamma}$ is the corresponding $k \times 1$ parameter vector relating to their spatial lags; $\mathbf{X}_2$ is an $n \times l$ matrix of observations on l right-hand-side individual-level covariates and $\boldsymbol{\beta}$ is the corresponding $l \times 1$ parameter vector; $\mathbf{X}_3$ is an $n \times m$ matrix of observations on m groups and $\boldsymbol{\alpha}$ is the corresponding $m \times 1$ parameter vector of group-level fixed effects; ε is an $n \times 1$ vector of errors.

Rather than an individual regression equation, the SAR model implies a system of equations to reflect the way in which changes due to exogenous influence percolate through a set of overlapping peer networks[5]. As people influence each other reciprocally, their characteristics are refracted through these networks. The endogenous peer effects are captured by $\boldsymbol{\lambda}$, the exogenous peer effects are captured by $\boldsymbol{\gamma}$ and the correlated peer effects are captured by $\boldsymbol{\alpha}$. If necessary, the $\mathbf{X}_1$ matrix can be used to control for the status of alters at a previous point in time. The SAR model can also be used with panel data, including fixed or random effects. In this way, it can model time and peer-related lags of the outcome variable as well as peer-related lags of relevant explanatory variables.

The SAR model can be estimated using the generalized spatial two-stage least squares estimator (GS2SLS), which uses the spatial-weighting matrix $\mathbf{W}$ in combination with the individual attributes recorded in the $\mathbf{X}_1$ matrix (Bramoullé, Djebbari, and Fortin 2009; Kelejian & Prucha, 1999). GS2SLS is a method-of-moments estimator which

[5] It is sometimes said that in a simultaneous auto-regressive 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 paper.

allows for higher-order dependent variable lags. It was derived by Kelejian and Prucha (Kelejian and Prucha 1998, 1999, 2010) and extended by Arraiz et al. (2008) and Drukker, Egger, and Prucha (2013). This form of estimation is provided in commercially-available software with a straightforward implementation, which means that the barriers to using these techniques in applied social research are relatively low[6].

4 Assumptions and limitations of SAR models of peer effects

The SAR model has a number of limitations and assumptions which we will discuss in this section. It shares many of the assumptions of regression models, of course, such as linearity in the parameters, specification of the correct functional form for regressors, no confounding and absence of multicollinearity. In addition to these, further assumptions must be made about the peer network, which must not only capture the relevant relationships but must also include “friends of friends who are not friends” in order to ensure the identification of endogenous peer effects, as mentioned earlier. 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 also 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.

In order to obtain unbiased estimates, a further condition of the model is that any relevant exogenous or correlated effects must be included. For example, appropriate

6 The models presented in Section 5 were estimated using the spregress command in Stata 15; for syntax and guidelines see StataCorp (2017).

groups must be used to define fixed effects to control for the social or institutional context and for the effects of baseline homophily (e.g. membership of a classroom). It is important that these groups be sufficiently small and proximal to the subjects to capture the peer dynamics responsible for correlated effects.

The SAR model assumes that selection effects within the peer network are either absent or are controlled for within the model. Selection effects can be theorised using concepts drawn from the study of homophily, as we suggested earlier. Baseline homophily gives rise to correlated peer effects, which can be controlled for by including fixed effects in the model, even if the factors that generated homophily are unmeasured. For example, if students are sorted into schools on the basis of their socioeconomic background, we can control for this by including fixed effects for schools or classes, even if we do not measure socioeconomic background.

In a paper on peer effects in relation to obesity, Christakis and Fowler (2007) argue that including a temporal lag of the outcome variable is sufficient in order to control for selection. By controlling for baseline scores, these researchers argue that it is possible to isolate endogenous peer effects and to measure their impact. This is equivalent to the assumptions made by the stochastic actor-based SNA model. It has been shown, however, that inbreeding homophily is “generically confounded” with endogenous peer effects (Shalizi and Thomas 2011). Self-selection into friendship networks on unmeasured grounds cannot always be distinguished from endogenous peer effects, as the latent factor underlying the choice of peers may simultaneously influence the outcome, what Hsieh and Van Kippersluis (2015) refer to as "individual correlated effects". How to control for this source of confounding is a topic of ongoing debate, and various strategies have been proposed.

If there is latent homophily, and the latent factor influences the outcome variable, this effect may be mediated by past status. In this case, it is sufficient to control for a temporal lag of the outcome variable. Christakis and Fowler (2007) specify cross-lagged effects between ego and alter by including a 'spatial' lag of the outcome variable for alter and a 'spatial' lag of the outcome measured at a previous point in time for both ego and alter[7]. A similar specification can be implemented in SAR models.

As Shalizi and Thomas (2011) observe, the latent factor underlying inbreeding homophily could nevertheless continue to have an influence over time, generating bias in the estimation of endogenous peer effects. If study aptitude influences academic performance, for example, and students with higher aptitude are more likely to be friends, this factor could lead to confounding if it is not included in the model as an individual-level covariate. Although academic performance depends on the incremental acquisition of abilities, this process could itself be influenced by aptitude and it may not be sufficient to control for prior status.

In order to deal with this situation, as Shalizi and Thomas (2011) and VanderWeele (2011) show, it is possible to use a sensitivity testing approach to assess the robustness

7 The term 'spatial' is applied to all variables which have been weighted by **W**, regardless of whether this is a contiguity matrix or sociomatrix.

of results to selection. In certain cases, it may also be possible to argue on theoretical grounds that all relevant confounders have been included in the model. An alternative approach involves extending the SAR model to incorporate an endogenous **W** matrix, for example, or to include latent variables (Hsieh and Van Kippersluis 2015). In the absence of panel data or fixed groups like classes, this may be necessary. However, the theoretical possibility of confounding does not automatically disqualify research on peer effects using observational data, as long as this risk is assessed and controlled for in an appropriate way. One of the great strengths of SAR models is that they are flexible enough to facilitate a range of sensitivity tests and alternative specifications which can provide information on the robustness of the assumptions.

In the case study presented in the next section, we seek to reduce or eliminate the risk of confounding by using a combination of techniques. Firstly, we control for relevant covariates by including fixed effects for school class and 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. We also test whether the results are robust to the effects of latent inbreeding homophily by comparing the results when the direction of friendships is reversed.

5. Case study

We apply the SAR model described above to the study of peer effects in the choice of whether or not to enrol 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 (Clancy 2015; 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, as Breen et al. note, large disparities remain between social classes in relation to university enrolment.

A range of mechanisms relating social origins to educational trajectories have been hypothesised, including risk-aversion and rational calculations (e.g. Breen and Goldthorpe, 1997; Stocké, 2007), or some combination of cultural capital and role models (Bourdieu and Passeron, 1977; Lareau, 2003). Despite decades of research, including numerous attempts to measure peer effects, our understanding of the dynamics underlying these theories remains rather incomplete. Empirical studies using different methods and covering different age groups, education systems and outcomes, can contribute to addressing this gap.

The Italian education system is characterised 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 are typically 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 (Guetto and Vergolini 2017; Panichella and Triventi 2014). Within the more vocationally-oriented schools, pupils are much less likely to acquire the academic skills and study habits that are required for obtaining high grades and satisfying the entrance requirements for 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 particularly central role in creating the conditions for correlated, endogenous and exogenous peer effects to be expressed. Selection into different tracks in secondary schools in Italy is highly patterned along socioeconomic differences and creates the basis for the formation of the tie pool from which friendships are drawn.

Schools and teachers also make an important contribution, together with the resources and opportunities they make available to students (e.g. Hanushek, 1997; Scheerens, 2000). Importantly, such resources are often dependent on the wider social context, which implies that correlated peer effects should be measured at the level of the school or even class. Thirdly, 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 or creates opportunities, thus creating the conditions for exogenous peer effects.

Very few studies have applied simultaneous auto-regressive models to European data to test these kinds of theoretical hypothesis, and researchers have focused on the U.S. National Longitudinal Survey of Adolescent Health (but see De Giorgi et al. 2010). 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 behaviour of students as they completed school and made the transition to work, higher education, other forms of study or other roles. The schools which 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 in the project. A total of 9,058 students enrolled in the fifth and final year of school completed the first questionnaire, and the response rate was very high at 99%, 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, during school-hours, 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 six 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 university, including a retrospective section to collect information on university enrolment for those who had not responded to the third wave. In this paper 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 based on directed ties, drawing on responses to the wave 1 questionnaire, where each student could nominate up to three friends[8]. 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. For example, in the Add Health survey, adolescents could nominate up to ten friends at school, but the mean was just 2.04 in the first wave (Mundt 2013). Other studies show that, when asked, secondary school students report an average of four friends, suggesting that a threshold of three 'close' friends is reasonable (Mercken et al. 2012).

The resulting egonets are specific to each individual student, and can overlap. Whilst almost one-fifth of the final sample did not indicate any friends in the class (17.8%), just over one third (34.1%) provided three names. The mean number was 1.80, and we used row normalisation to standardize the **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 rather 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 endogenous and exogenous peer effects.

[8] The validation and matching of these 'close friends' was carried out using a pattern-matching routine to control for minor spelling mistakes. Roughly 95% of names were matched in this way, with subsequent manual coding to control for difficulties such as serious spelling mistakes (which were relatively frequent, particularly for foreign names), omission of a forename or surname, use of nicknames, inversion of forename and surname and duplicate names within a class. Out of a total of 1,465 individuals who could not be matched automatically, manual coding permitted 1,088 unequivocal matches.

The **dependent variable** is *university enrolment* (1 = yes; 0 = no), based on reported behavior (actual enrolment) 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 is considered to provide 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 (cf. De Giorgi et al. 2010; Hellevik 2007).

The **individual-level covariates** used in the model include the following measures: *gender* (1 = male, 0 = female), *dialect is 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 standardised index of *family economic difficulties*[9] and final school *diploma examination result*[10]. 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

9 This is a scale based on six items derived from a longer list used in the 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.

10 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 commission includes teachers with direct experience of students. The vast majority of students (97.6% in our sample) are admitted to sit the exam and very few fail (99.4% pass rate). This is not a standardised assessment at national level but nevertheless provides a measure of intra-class relativities in academic performance, which is why it is included in our model.

museums, theatres or concerts, 0 = never goes). All of these variables control for baseline homophily, family background and individual attributes.

Endogenous peer effects are measured by including a first-order auto-regressive component which is defined by the directed friendship ties discussed 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. In other words, as well as including these variables as individual attributes, we also assess whether the socioeconomic position of their peers influences young people's examination results or their probability to enrol at university.

As far as **correlated effects** are concerned, we include dummy variables in order to control for selection to a specific academic track and for the specific social, cultural and institutional characteristics of the school or class, including its reputation, 'quality' and catchment area. We assess whether different kinds of fixed effects (type of school, school, class) yield different results when controlling for social context[11].

To control for the main forms of **inbreeding homophily**, we include a measure which reflects the intention to enrol at university roughly one year before the outcome variable

[11] 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.

was measured (with the following response categories: definitely will attend university, probably will intend, probably will not attend, definitely will not attend, don't know; dummy variables were used to compare all other categories with "definitely will not attend")[12].

Descriptive statistics for the variables included in the model are provided in Table 1. Just under half of students are male, roughly one in ten speak a regional dialect at home and only one in 20 was born in another country. More than four fifths (86.6%) live with two parents and just over one quarter read for pleasure at least once a week. Almost three quarters of young people in the sample were 18 years of age at the time of the first questionnaire, although 7.3 per cent were 20 or older. As far as parental educational attainments are concerned, 29.8 per cent of students have a father who completed the lower cycle of secondary school, almost one third have a father who obtained a secondary school diploma and roughly one in seven have a father who graduated from university (14.2%). The figures are similar for mother's education, with 26.6 per cent having a lower secondary education, 35 per cent having a secondary school diploma and 13.9 per cent a university degree. In our sample, almost two thirds (62.2%) of students went on to enrol at university and the mean score at final diploma examinations was 76.3 out of 100 (pass = 60).

[12] 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.

The initial sample of 9,058 pupils was reduced to 7,212 after excluding individuals who did not participate in any of the later waves of data collection and for whom we have no information on examination results or university enrolment[13]. We also dropped 49 cases with large amounts of missing values, leaving a sample of 7,163[14]. Item-level missing data was estimated using single imputation via the EM algorithm, involving very small numbers of cases.

Table 1 Descriptive statistics for all variables (N = 7,163)

Categorical variables

Variable name	%
Male	46.9%
Dialect at home	11.5%
Born in Italy	94.7%
Two parent family	86.6%
At least one parent has 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 one year before: definitely enroll	43.2%
Intentions one year before: probably enroll	28.0%
Intentions one year before: probably not enroll	11.2%
Intentions one year before: definitely not enroll	8.7%
Intentions one year before: don't 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%

[13] 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.

[14] 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.

Continuous variables

Variable name	*Statistic*	
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

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 2 contains the results, while Table 3 shows the average direct, indirect and total effects, taking into account the auto-regressive structure of the model[15].

Model 1: endogenous peer effect only.

Model 2: + individual-level explanatory variables

Model 3: + exogenous peer effects

Model 4: + longitudinal component

Model 5a: + fixed effects for type of school

Model 5b: + fixed effects for school

Model 5c: + fixed effects for class

[15] 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 GS2SLS 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							
Male gender		-0.03 (0.01)***	-0.04 (0.01)***	0.02 (0.01)*	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)**
Dialect at home		-0.13 (0.02)***	-0.12 (0.02)***	-0.05 (0.01)***	-0.06 (0.01)***	-0.04 (0.01)**	-0.04 (0.01)**
Born in Italy		0.11 (0.02)***	0.11 (0.02)***	0.08 (0.02)***	0.04 (0.02)*	0.05 (0.02)**	0.05 (0.02)**
Two parent family		0.02 (0.01)	0.02 (0.01)	0.01 (0.01)	0.003 (0.01)	0.01 (0.01)	0.01 (0.01)
Parental education		0.15 (0.01)***	0.14 (0.01)***	0.06 (0.01)***	0.04 (0.01)***	0.04 (0.01)***	0.04 (0.01)***
Family econ. difficulties		-0.002 (0.0004)***	-0.002 (0.0002)***	-0.001 (0.0002)***	-0.001 (0.00)***	-0.001 (0.0002)***	-0.001 (0.0002)***
Exam grade		0.009 (0.01)***	0.009 (0.0004)***	0.005 (0.0003)***	0.005 (0.0004)***	0.005 (0.0004)***	0.005 (0.0004)***
Reading for pleasure		0.07 (0.01)***	0.07 (0.01)***	0.02 (0.01)*	0.02 (0.01)*	0.02 (0.01)*	0.02 (0.01)*
Cultural participation		0.14 (0.02)***	0.14 (0.02)***	0.05 (0.01)***	0.04 (0.01)**	0.04 (0.01)**	0.03 (0.01)*
Longitudinal							
Intentions = prob. no				0.07 (0.02)***	0.03 (0.02)	0.04 (0.02)*	0.04 (0.02)*
Intentions = prob. yes				0.50 (0.02)***	0.44 (0.02)***	0.44 (0.02)***	0.43 (0.02)***
Intentions = yes				0.70 (0.02)***	0.58 (0.02)***	0.58 (0.02)***	0.56 (0.02)***
Intentions = don't know				0.20 (0.02)***	0.17 (0.02)***	0.18 (0.02)***	0.19 (0.02)***
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***	2,536.50***	2,648.37***	7,262.30***	8,0490.62***	8,580.41***	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

$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. $N = 7{,}163$

Model 1: endogenous peer effect only. **Model 2**: + individual-level explanatory variables

Model 3: + exogenous peer effects **Model 4**: + longitudinal component

Model 5a: + fixed effects for type of school **Model 5b**: + fixed effects for school

Model 5c: + fixed effects for class

Table 3: Direct, indirect and total effects of individual characteristics

Direct effects	*Coefficient*	*Standard Error*
Male gender	0.025**	0.009
Dialect spoken at home	-0.040**	0.014
Born in Italy	0.050**	0.018
Two parent family	0.012	0.012
Parental education (university)	0.038***	0.010
Family economic difficulties	-0.001***	0.0002
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: yes	0.557***	0.018
Intentions: don't know	0.185***	0.019

Indirect effects	*Coefficient*	*Standard Error*
Male gender	0.0012*	0.0006
Dialect spoken at home	-0.002*	0.0009
Born in Italy	0.002*	0.001
Two parent family	0.0006	0.0006
Parental education (university)	0.003	0.013
Family economic difficulties	-0.0002	0.0002
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: probably yes	0.027***	0.008
Intentions: don't know	0.009**	0.003

Total effects	*Coefficient*	*Standard Error*
Male gender	0.026**	0.010
Dialect spoken at home	-0.041**	0.015
Born in Italy	0.052**	0.019
Two parent family	0.013	0.012
Parental education (university)	0.040*	0.017
Family economic difficulties	-0.001***	0.0003
Exam grade		
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: probably yes	0.584***	0.020
Intentions: don't know	0.194***	0.020

Notes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. $N = 7{,}163$.

The estimates in this table relate to Model 5c.

The coefficient associated with the endogenous peer effect is very 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 also highlight the scope for bias that exists when models are 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, choice of track for upper secondary school already signals, at least in part, the orientation of students and their families in relation to university enrollment.

As Table 2 shows, the coefficient for the auto-regressive term is 0.06 ($p < 0.001$) in the final model. If all of his or her 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. As the decision to enroll is most likely taken over a longer time period than this, the overall endogenous peer effect on enrolment is likely to be considerably greater and may include an indirect component. It is interesting to observe that the endogenous peer effect remains stable at 0.06 regardless of whether we specify fixed effects for school type, school or class.

The exogenous peer effects that we specified for parental education level and economic difficulties were not found to have a significant effect. These terms express the influence exerted by the parents of peers, who do not appear to be an important determinant of outcomes for school students. Several individual-level covariates do, however, have significant effects on the probability of enrolling at university, after controlling for the status of the student one year earlier. The total effect of the explanatory variables is shown in Table 3.

As far as correlated peer effects are concerned, the models provide evidence that these are important, once again reflecting the role of academic tracking in Italy. 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 the school are relatively unimportant. Only small differences are observed whether we use 5 dummy variables for school type or 473 dummies for school class.

As mentioned above, 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 assess whether latent inbreeding homophily (self-selection into friendship networks) could be confounded with endogenous peer effects (see An 2016). As selection effects are symmetrical, in the presence of confounding the estimated size of the endogenous peer effect would be expected to 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 W matrix. This is indeed the case, and the endogenous peer effect drops from 0.06 to 0.03, which is a substantively and statistically significant change. All other coefficients in the model remain stable. In other words, when we change the direction of friendship ties, whilst keeping the basic structure of the networks intact, the size of the endogenous peer effect decreases. This rules out the possibility of 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 if we wish to make causal inferences using this model. To address this objection, we also 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 disappeared (becoming indistinguishable from 0), further demonstrating that it is attributable to the specific configuration of personal friendship networks within the sample.

Conclusions

The simultaneous auto-regressive models presented in this paper shed considerable light on peer influence in relation to university enrolment. They testify to the complexity of peer processes, providing evidence for the existence 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 datasets using available software tools. Using a dataset on Italian secondary school students, we quantified peer effects in relation to university enrollment. The models include individual-level covariates and components that control for endogenous, exogenous and correlated peer effects as well as baseline, manifest and latent inbreeding homophily. The evidence suggests that endogenous peer effects have a statistically and substantively significant influence on the probability of enrolling at university, with a coefficient of 0.06 ($p < 0.001$). The results of our sensitivity tests suggest that this finding is robust to the effects of selection.

Simultaneous auto-regressive 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 instrumental variables. They provide a convincing representation of the complexities of peer influence, including indirect effects and reflection. 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 centre of scientific and policy-related debates.

Further research is required on the dynamics of friendships among school students, including tendencies towards homophily and change over time, the composition of peer networks within and outside the classroom, the role of age and gender 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 of great relevance and that peer groups should not be inferred from membership of a social group or random assignment.

The model presented in this paper has some limitations that should be borne in mind. Firstly, it is based on a representative sample of upper secondary school students in four Italian Provinces, which means that the results may not generalise to the country as a whole. The analysis uses observational data and all inferences are conditional upon the assumptions incorporated within the model regarding the nature of peer effects and the required control variables. As we described earlier, we use a linear probability model for our dichotomous dependent variable, and it is not possible to test all theoretically possible forms of confounding. We also assume that peer effects are homogeneous, averaging across heterogeneous peers. Finally, friendships are measured with error and only at the start of the study, so we cannot assess the impact of any changes that might have occurred in egonets over time. We believe that these limitations are acceptable and the rigorous controls implemented within our models are likely to 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 further 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 policy-makers 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.

References

Ajilore, Olugbenga. 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.

Ammermueller, Andreas and Jörn-Steffen Pischke. 2009. “Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study.” *Journal of Labor Economics* 27(3):315–348.

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

An, Weihua. 2016. “On the Directionality Test of Peer Effects in Social Networks.” *Sociological Methods & Research*, 45(4):635-650.

Angrist, Joshua D. 2014. “The Perils of Peer Effects.” *Labour Economics* 30:98–108.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan. 2009. “Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks.” *Proceedings of the National Academy of Sciences*, 106(51):21544-21549.

Arraiz, Irani, David M. Drukker, Harry H. Kelejian, and Ingmar R. Prucha. 2008. “A Spatial Cliff-Ord-Type Model with Heteroskedastic Innovations: Small and Large Sample Results.” *CESIFO Working Paper 2485*.

Bernburg, Jón Gunnar, Thorolfur Thorlindsson, and Inga Dora Sigfusdottir. 2009. “Relative Deprivation and Adolescent Outcomes in Iceland: A Multilevel Test.” *Social Forces* 87(3):1223–50.

Betts, Julian R. and Darlene Morell. 1999. “The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects.” *The Journal of Human Resources* 34(2):268.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. “Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak.” *Journal of the American Statistical Association* 90(430):443–50.

Bourdieu, Pierre and Jean-Claude Passeron. 1977. *Reproduction in Education, Society and Culture*. London, Sage.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. “Identification of Peer Effects through Social Networks.” *Journal of Econometrics* 150(1):41–55.

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

Breen, Richard, Ruud Luijkx, Walter Müller, and Reinhard Pollak (2009) "Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries", *American Journal of Sociology*, 114(5):1475-1521.

Brock, William A. and Steven N. Durlauf. 2000. “Interactions-Based Models.” Pp. 3297–3380 in *Handbook of Econometrics*. Vol. 5, edited by J. Heckman and E. Leamer. Amsterdam: North-Holland.

Buckner, Julia D., Michael A. Mallott, Norman B. Schmidt, and Jeanette Taylor. 2006. “Peer Influence and Gender Differences in Problematic Cannabis Use among Individuals with Social Anxiety.” *Journal of Anxiety Disorders* 20(8):1087–1102.

Van den Bulte, Christophe and Gary L. Lilien. 2001. “Medical Innovation Revisited: Social Contagion versus Marketing Effort.” *American Journal of Sociology* 106(5):1409–35.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2009. “Peer Effects and Social Networks in Education.” *The Review of Economic Studies* 76(4):1239–1267.

Carrell, Scott, Bruce Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81(3):855–82.

Carrington, Peter J., John Scott, and Stanley Wasserman. 2005. *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press.

Christakis, Nicholas A. and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *New England Journal of Medicine* 2007(357):370–379.

Clancy, Peter 2015. *Irish Higher Education: A Comparative Perspective*. Institute of Public Administration.

Cliff, Andrew D. and John K. Ord. 1973. *Spatial Autocorrelation*. London: Pion.

Cohen-Cole, Ethan and Jason M. Fletcher. 2008. "Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic", *Journal of Health Economics* 27(5):1382-1387.

Coleman, James S., Elihu Katz, and Herbert Menzel. 1966. *Medical Innovation: A Diffusion Study*. Indianapolis: The Bobbs-Merrill Company.

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

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of Social Interactions through Partially Overlapping Peer Groups." *American Economic Journal: Applied Economics* 2(2):241–75.

De Melo, Gioia. 2014. "Peer Effects Identified through Social Networks: Evidence from Uruguayan Schools." *Working Papers, Banco de México*.

Drukker, David M., Peter Egger, and Ingmar 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.

Duncan, Otis Dudley, Archibald O. Haller, and Alejandro Portes. 1968. “Peer Influences on Aspirations: A Reinterpretation.” *American Journal of Sociology* 74(2):119–37.

Eisenberg, Daniel. 2004. “Peer Effects for Adolescent Substance Use: Do They Really Exist?” *UC-Berkeley School of Public Health Working Paper*.

Ennett, Susan T. and Karl E. Bauman. 1993. “Peer Group Structure and Adolescent Cigarette Smoking: A Social Network Analysis.” *Journal of Health and Social Behavior* 34(3):226.

Fowler, James H. and Nicholas A. Christakis. 2008. “Estimating Peer Effects on Health in Social Networks” *Journal of Health Economics* 27(5):1400–1405.

Graham, Brian S. and Jinyong Hahn. 2005. “Identification and Estimation of the Linear-in-Means Model of Social Interactions.” *Economics Letters* 88(1):1–6.

Guetto, Raffaele and Loris 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.

Hanushek, E. 1997. "Assessing the effects of school resources on student performance: an update". *Educational Evaluation and Policy Analysis*, 19:141-164.

Harding, David J. 2003. “Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping out and Teenage Pregnancy.” *American Journal of Sociology* 676–719.

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

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

Hsieh, Chih-Sheng and Lung Fei Lee. 2016. “A Social Interactions Model with Endogenous Friendship Formation and Selectivity.” *Journal of Applied Econometrics* 31(2):301–19.

Hsieh, Chih-Sheng and Xu Lin. 2017. “Gender and Racial Peer Effects with Endogenous Network Formation.” *Regional Science and Urban Economics* 67:135–47.

Hsieh, Chih-Sheng and Hans Van Kippersluis. 2015. “Smoking Initiation: Peers and Personality.” *Tinbergen Institute Discussion Paper TI 15-093/V*.

Huba, George J. and Peter M. Bentler. 1980. “The Role of Peer and Adult Models for Drug Taking at Different Stages in Adolescence.” *Journal of Youth and Adolescence* 9(5):449–65.

Hyun-soo Kim, Harris and Paul Y. Chang. 2018. "The Impact of Delinquent Friendship Networks and Neighborhood Context on Suicidal Ideation among South Korean Youths." *Social Forces* 97(1):347–376.

Kelejian, Harry H. and Ingmar 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, Harry H. and Ingmar R. Prucha. 1999. “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model.” *International Economic Review* 40(2):509–533.

Kelejian, Harry H. and Ingmar R. Prucha. 2010. “Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances.” *Journal of Econometrics* 157(1):53–67.

Kenny, David A., and Charles M. Judd. 1986. “Consequences of Violating the Independence Assumption in Analysis of Variance.” *Psychological Bulletin* 99:422–31.

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

Laschever, Ron. 2009. “The Doughboys Network: Social Interactions and the Employment of World War I Veterans.” *Available at SSRN 1205543*.

Lauen, Douglas Lee and S. Michael Gaddis. 2013. “Exposure to Classroom Poverty and Test Score Achievement: Contextual Effects or Selection?” *American Journal of Sociology* 118(4):943–79.

Lavy, Victor, Olmo Silva, and Felix Weinhardt. 2009. "The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools". Working Paper. 15600. Cambridge Mass.: National Bureau of Economic Research.

Lazarsfeld, Paul F., and Robert K. Merton. 1954. “Friendship as Social Process.” Pp. 18–66 in *Freedom and Control in Modern Society*, edited by Monroe Berger, Theodore Abel, and Charles Page. New York: Octagon.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin. 2010. “Specification and Estimation of Social Interaction Models with Network Structures: Estimation of Social Network Models.” *Econometrics Journal* 13(2):145–76.

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

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

Lin, Xu. 2010. “Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables.” *Journal of Labor Economics* 28(4):825–60.

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

Liu, Xiaodong. 2014. “Identification and Efficient Estimation of Simultaneous Equations Network Models.” *Journal of Business & Economic Statistics* 32(4):516–36.

Lyle, David S. 2007. “Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point.” *The Review of Economics and Statistics* 89(2):289–299.

Macdonald-Wallis, Kyle, Russell Jago, Angie S. Page, Rowan Brockman, and Janice L. Thompson. 2011. “School-Based Friendship Networks and Children’s Physical Activity: A Spatial Analytical Approach.” *Social Science & Medicine* 73(1):6–12.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem.” *The Review of Economic Studies* 60(3):531–542.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27(1):415–44.

Mercken, Liesbeth, Christian Steglich, Ronald Knibbe, and Hein 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.

Merton, Robert K. 1968. *Social Theory and Social Structure*. Glencoe: Free Press.

Moffitt, Robert A. 2001. “Policy Interventions, Low-Level Equilibria, and Social Interactions.” *Social Dynamics* 4(45–82):6–17.

Mundt, Marlon P. 2013. "Social Network Analysis of Peer Effects on Binge Drinking among U.S. Adolescents". Pp. 123-134 in *Social Computing,Behavioral-Cultural Modeling and Prediction*, edited by Ariel Greenberg, William G. Kennedy, and Nathan D. Bos. Berling: Springer-Verlag.

Newcomb, Theodore M. 1962. "Student Peer-Group Influence." Pp. 469–88 in *The American College: A Psychological and Social Interpretation of the Higher Learning*, edited by Nevitt Sanford. New York: Wiley.

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

O'Malley, A. James, Felix Elwert, J. Niels Rosenquist, Alan M. Zaslavsky, and Nicholas A. Christakis. 2014. "Estimating Peer Effects in Longitudinal Dyadic Data Using Instrumental Variables: Causal Inference for Peer Effects." *Biometrics* 70(3):506–15.

Panichella, Nazareno and Moris Triventi. 2014. "Social Inequalities in the Choice of Secondary School: Long-Term Trends during Educational Expansion and Reforms in Italy." *European Societies* 16(5):666–93.

Ragan,, Daniel T., D. Wayne Osgood, and Mark E. Feinberg. 2014. "Friends as a Bridge to Parental Influence: Implications for Adolescent Alcohol Use." *Social Forces* 92(3):1061–1085.

Ryan, Allison M. 2001. "The Peer Group as a Context for the Development of Young Adolescent Motivation and Achievement." *Child Development* 72(4):1135–1150.

Scheerens, Jaap 2000. *Improving School Effectiveness. Fundamentals of Educational Planning.* Paris: Unesco.

Shalizi, Cosma Rohilla and Andrew C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods & Research* 40(2):211–239.

Snijders, Tom A. B., Gerhard G. Van de Bunt, and Christian E. G. Steglich. 2010. "Introduction to stochastic actor-based models for network dynamics". *Social Networks* 32: 44-60.

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

StataCorp. 2017. *Stata Spatial Autoregressive Models Reference Manual*. Statistical Software. College Station, TX: StataCorp LLC.

Steglich, Christian, Tom A. B. Snijders, and Michael Pearson. 2010. "Dynamic Networks and Behavior: Separating Selection from Influence." *Sociological Methodology* 40(1):329–393.

Stocké, Volker 2007 "Explaining educational decision and effects of families' social class position: An empirical test of the Breen-Goldthorpe model of educational attainment", *European Sociological Review*, 23(4):505-519.

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

Tranmer, Mark. 2010. "Social Networks in Multilevel Structures." *N-CESS e-Stat Node, Bristol*.

Tranmer, Mark, David Steel, and William 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):439–455.

VanderWeele, Tyler J. 2011. "Sensitivity analysis for contagion effects in social networks." *Sociological Methods & Research* 40(2): 240–255.

VanderWeele, Tyler J. and Weihua An. 2013. "Social Networks and Causal Inference." Pp. 353-374 in *Handbook of Causal Analysis for Social Research*, edited by Stephen Morgan. New York: Springer.

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