Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables

Xu Lin, *Wayne State University*

Disentangling peer effects from other confounding effects is difficult, and separately identifying endogenous and contextual effects is impossible for the linear-in-means model. This study confronts these problems by using spatial autoregressive models with group fixed effects. The nonlinearity introduced by the variations in the peer measurements provides information to identify both endogenous and contextual effects, thus resolving the "reflection problem." The group fixed effects term captures the confounding effects of the common variables. Applying the model to data sets from the National Longitudinal Study of Adolescent Health, I find strong evidence for both endogenous and contextual effects in student academic achievement.

I. Introduction

Peer effects in a school context have received considerable attention, especially since the publication of the Coleman report (Coleman et al. 1966).

This is a substantially revised version of chapter 2 of my doctoral dissertation. I am grateful to the editor John Kennan for very insightful comments. I also thank Lung-fei Lee for his supervision and encouragement. Helpful comments from Patricia Reagan, Bruce Weinberg, and Steven Lehrer as well as from seminar participants at Ohio State University, University of Alberta, Peking University, the Midwest Econometric Group annual meeting, and the Canadian Economics Association annual meeting are also gratefully acknowledged. This research has been supported by National Natural Science Foundation of China under grant

[*Journal of Labor Economics,* 2010, vol. 28, no. 4]

0022-1376/2010/2804-000410.00

The interest in peer effects is largely policy driven. Unlike many market-channeled effects, peer effects represent how an individual's decision or outcome is directly influenced by her peers' outcomes or characteristics. As pointed out by Hoxby (2000), the inherent externality of peer effects provides opportunities for policies aimed at improving social welfare. For instance, in school financing studies (e.g., Epple and Romano 1998; Nechyba 2000), under the assumption that peers' outcomes or characteristics influence educational outcome, researchers demonstrate how an effective policy intervention that internalizes peer effects, like education vouchers, could lead to a more efficient human capital investment profile. In fact, as an important policy variable, peer effects have been at the core of a variety of ongoing debates, including ability tracking versus de-tracking, school segregation versus school busing or magnet schools, racial segregation versus affirmative action, and the like. Understanding the magnitude and nature of peer effects in student outcomes would help inform these debates. For example, if low ability students benefit from the presence of superior peers, while high ability students are not harmed (or not harmed as much) by the presence of disadvantaged peers, then mixing students of different ability levels will generate social gains. In addition, peer effects are a key component in many theoretical economic growth models that study the relationship among macroeconomic growth, heterogeneous human capital, and economic stratification (e.g., Kremer 1993; Benabou 1996).

However, as discussed by Manski (1993, 2000), Moffitt (2001), and Brock and Durlauf (2001), among others, the estimation of social interaction effects is a difficult task. The first issue is the "reflection problem" described by Manski (1993), which refers to the impossibility of separately identifying two types of social effects—endogenous or behavioral effects and contextual or exogenous effects. Endogenous social effects measure how an individual's outcome is affected by her peers' outcomes, and contextual social effects capture the influences of peers' exogenous characteristics. These two types of social interaction effects are distinct in nature and have different policy implications. In particular, endogenous effects generate a social multiplier while contextual effects do not. For

number 70701020. Contact the author at xulin@wayne.edu.

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health Web site (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

example, if endogenous effects are present, a special tutoring program targeting a few students will have multiplier effects: the students directly benefit from the program and will influence the performance of their peers, which in turn will affect the performance of the former, and so on. On the other hand, if only contextual effects are present, then there will be no social multiplier effects in school, and more costly mobility programs that compensate for the composition of neighborhoods will be important for students' academic performance. Therefore, the identification of these two social effects is of great importance to policy implementation and evaluation. In a study on social interactions in labor supply, Grodner and Kniesner (2006) demonstrate the important difference between the impacts of endogenous and contextual effects on the estimations of labor supply and wage elasticities. Their study highlights the necessity of identifying different types of social effects. The clear identification of endogenous effects is also essential for evaluating the impact of the deworming program in Kenya as studied in Miguel and Kremer (2004), where treatment externalities across schools are present. Several recent studies, including Bramoullé, Djebbari, and Fortin (2009) and Laschever (2009), also emphasize the importance of the identification between contextual and endogenous effects. Unfortunately, in the standard linear-in-means social interaction model in Manski (1993), separating endogenous effects from contextual effects is impossible without making strong exclusion assumptions. Most existing studies simply circumvent the "reflection problem" by considering only endogenous or contextual effects.

Another difficulty in social interaction estimation is the separation of social effects from other confounding effects. As noted by many authors, the formation of a peer group is not random. Thus, similarity in the outcomes of group members could be due to the effects of unobservables or common variables faced by the same group members, which are termed correlated effects in Manski (1993), instead of social interaction effects. Many strategies have been proposed to address this omitted variable bias problem, including the instrumental variable method (e.g., Evans, Oates, and Schwab 1992; Rivkin 2001), family fixed effect strategy (e.g., Aaronson 1998; Plotnick and Hoffman 1999), and experiment type strategy (e.g., Sacerdote 2001; Zimmerman 2003). However, as will be discussed in Section II, the validity of these strategies is open to question; therefore, the empirical studies relying on these methods are subject to a wide range of interpretations.

Besides the conceptual problems, the estimation of social interaction effects is also plagued by data limitations. Most data sets do not provide adequate information about an individual's reference group. Consequently, the specification of a reference group in many studies is determined by data availability rather than any substantive criteria, as discussed in Manski (1993) and Durlauf (2003), among others. For example, in the study of peer effects many researchers specify peer groups at rather broad levels, such as all the other students in the same school or grade.[1] This might prevent the detection of social interaction effects. A student is not equally affected by all the others in the same school or grade; instead, she is more likely to be significantly influenced by some of them, such as her friends. Thus, the ideal model should contain the weighted average of the peer variables, with weight determined by the importance of a friend, as opposed to a mean peer variable. This type of data is rarely available. However, richer data sets will undoubtedly contribute much to the social interaction analysis.

This study confronts these conceptual problems by using a spatial autoregressive (SAR) model with both endogenous and contextual effects as well as group fixed effects, a generalized version of the model in Lee (2007b). We estimate peer effects in students' grade point average (GPA) by using data sets from the National Longitudinal Study of Adolescent Health (Add Health). As an important indicator of academic achievement, GPA has been shown by many researchers to be a key determinant of labor market outcomes in adulthood. Many studies, including Loury and Garman (1995) and Jones and Jackson (1990), report a positive and significant link between college GPA and earnings. The importance of high school GPA has also been investigated by many researchers. For instance, Newcomb and Bentler (1986) and Brook and Newcomb (1995) show that high school academic attainment predicts workforce participation and college involvement. Morell (1993) and Betts and Morell (1999) find a strong relationship between high school GPA and college GPA for undergraduate students at the University of California, San Diego.

However, the lack of empirical consensus about the determinants of GPA, especially the effects of peer groups on GPA, is striking. The conceptual and data problems inherent in social interaction estimations are a key contributor to this uncertainty. As shall be shown, the nonlinearity introduced by variations in the measurements of the peer variables in the SAR model provides valuable information to identify both endogenous and contextual effects. Therefore, unlike the linear-in-means model, the SAR model is free of the "reflection problem." The group fixed effects term in the model captures the effects of the common variables faced by the same group members—both observed and more importantly, unob-

[1] For example, Gaviria and Raphael (2001) specify peer groups by school, and Hanushek et al. (2003) by grade. Some exceptions are Henderson, Mieszkowski, and Sauvageau (1978), which specifies peer groups at classroom level, and Zimmerman (1999) and Sacerdote (2001), which uses roommates as peers. In a study of social interaction effects in female labor supply, Woittiez and Kapteyn (1998) rely on survey information to construct reference groups. Several recent studies, such as Calvó-Armengol, Patacchini, and Zenou (2005) and Bramoullév et al. (2009), specify peers based on friendship networks.

served—that tend to confound peer effects. With its special design for analyzing social interaction effects in adolescents' development, Add Health provides information on characteristics and outcomes of not only the respondents but also their friends, making it possible to specify peer groups at the more specific and more relevant level of friendship networks.

The article is organized as follows. Section II reviews the identification problems and existing solutions in the estimation of social interactions. Section III presents the SAR model with individual specific social interactions and the associated maximum likelihood estimation. Section IV demonstrates the empirical framework and results. Concluding remarks are provided in Section V.

II. The Linear-in-Means Model

A. Identification Problems

A standard social interaction model relates an individual's outcome to her own characteristics, contextual influences from her peers' characteristics, and endogenous effects from her peers' outcomes. The model most widely used in existing social interaction studies is Manski's (1993) linear-in-means model, which is given by:

$$y_{ir} = \lambda_0 E(y_r|r) + \beta_{10} x_{ir} + \beta_{20} E(x_r|r) + \epsilon_{ir}, \tag{1}$$

where y_{ir} denotes outcome of individual i from group r with characteristics x_{ir}. The ϵ_{ir}'s are innovations. The terms $E(y_r|r)$ and $E(x_r|r)$ are the group mean outcome and characteristics. In the student academic achievement setting, the endogenous effects coefficient λ_0 measures how a student's test score is affected by her peers' test scores, and the contextual effects coefficient β_{20} captures the impacts of the characteristics of a student's peers, such as family background, race, gender, and the like, on her test score.

The reduced form of the model is given by:

$$y_{ir} = \beta_{10} x_{ir} + \left(\frac{\lambda_0 \beta_{10} + \beta_{20}}{1 - \lambda_0}\right) E(x_r|r) + \epsilon_{ir}. \tag{2}$$

Equation (2) is a representation of the "reflection problem": only β_{10} and $(\lambda_0 \beta_{10} + \beta_{20})/1 - \lambda_0$ can be identified; one cannot separate the endogenous effects coefficient λ_0 from the contextual effects coefficient

β_{20}.[2] In fact, since the group mean outcome $E(y_r|r)$ is a linear function of the group mean characteristics $E(x_r|r)$, any correlations between an individual's outcome and her reference group's mean outcome may simply reflect the effects of the group mean characteristics. This is pessimistic because, as discussed earlier, these two types of social interaction effects are distinct in nature and have different policy inferences. The inherent "reflection problem" prevents the linear-in-means model from giving clear-cut indications for policy implementations.

In addition, social effects are also contaminated by nonsocial or correlated effects. Rather than representing any interaction effects, the estimated parameters could be the result of omitted variable bias or spurious social effect. People clearly choose their neighborhood and sort themselves into different neighborhoods based on various characteristics.[3] This implies that the error terms are likely to consist of two components as follows:

$$\epsilon_{ir} = \alpha_r + u_{ir}, \tag{3}$$

where α_r represents the characteristics that are common to all group members, usually unobserved, and u_{ir}'s are innovations. If the correlations between the group unobservables α_r and the outcome are positive, as they are in most cases, then the parameters representing social interaction effects will tend to be overestimated. For example, if high ability students are more likely to associate with high ability peers, then the social interaction coefficients estimated from the model will be positive even in the absence of social interaction effects, due to the effects of the omitted or unobserved variable for ability.

Another source of correlated effects is the common variables faced by people in the same group. The similarity in the outcome of the group members might reflect the effects of some common institutional envi-

[2] Note that it is implicitly assumed that the variables having direct effects on the individuals and those generating contextual effects are the same. If the two sets of variables were not identical, then the structural parameters would be identified. To see this, consider the extreme case where $x_{ir,1}$ and $x_{ir,2}$ do not have common elements—for example, $x_{ir,1}$ is income and $x_{ir,2}$ is race. Then the model becomes $y_{ir} = \lambda_0 E(y_r|r) + \beta_{10}x_{ir,1} + \beta_{20}E(x_{r,2}|r) + \epsilon_{ir}$, and the reduced form will be

$$y_{ir} = \beta_{10}x_{ir,1} + \frac{\lambda_0\beta_{10}}{1-\lambda_0}E(x_{r,1}|r) + \frac{\beta_{20}}{1-\lambda_0}E(x_{r,2}|r) + \epsilon_{ir}.$$

Thus all the parameters are identified. However, as discussed in Manski (1993) and Durlauf (2003), among others, without prior information it is difficult to find variables that are in $x_{ir,1}$ ($x_{ir,2}$) but not in $x_{ir,2}$ ($x_{ir,1}$). Therefore, we will consider the general case in which the sets of variables in $x_{ir,1}$ and $x_{ir,2}$ are identical, without making any arbitrary exclusion assumption.

[3] See, e.g., Massey, Gross, and Shibuya (1993) and Jargowsky (1997) for detailed discussions. This is referred to as the endogenous neighborhood membership problem.

ronments or shocks instead of social interaction effects. For instance, a good teacher tends to improve the performance of the whole class. If a student's outcome is regressed on her own characteristics and her classmates' outcome and characteristics, without controlling for teacher quality, positive estimates of the social interaction coefficients will be obtained even without social interaction effects.

B. Existing Solutions

To the best of my knowledge, the "reflection problem" has seldom been confronted directly. Most studies simply avoid it by focusing on either endogenous effects or contextual effects, assuming that the other type of social effects are not present.[4] It is also common to replace the contemporary peer outcomes in the model with lagged peer variables.[5]

Most attention has been devoted to the endogenous membership or the omitted variable bias problem. The most popular strategy is the instrumental variable method. Evans et al. (1992) is one of the first papers that seriously consider the neighborhood selection problem. The authors use metropolitan aggregate data as instrumental variables to control for the endogeneity of peer group formation and show that peer effects on teenage pregnancy and school dropout decisions become insignificant. This poses a big challenge to studies that treat peer groups as exogenous, but the validity of this study is also open to debate. The biggest concern is about the validity of the instrumental variables because it is hard to guarantee that they are correlated with the peer variables while uncorrelated with the structural errors. In a more recent paper, Rivkin (2001) demonstrates how using aggregate level data as instrumental variables leads to exaggeration of the specification errors.

In the broader neighborhood effects literature, a family fixed effect strategy is often used to isolate the effects of neighborhood on a child's outcome from those of unobserved family characteristics (e.g., Aaronson 1998; Plotnick and Hoffman 1999). This approach looks at outcomes for sibling pairs in families that moved across neighborhoods. One potential problem, as discussed in Aaronson (1998), is that the unobserved family characteristics faced by the siblings may not be the same. In fact, the change of neighborhood is often associated with some other change in the family, like divorce or unemployment. As a result, the estimation

[4] Fertig (2003) is one of the few studies that incorporate both endogenous and contextual effects in a model. It specifies a nonlinear form of the peer variables and relies on this nonlinearity to identify both effects. However, the specification of the form of the nonlinearity is ad hoc. Another notable exception is Ioannides and Zabel (2008), which uses an SAR model incorporating both endogenous and contextual effects in a random effect setting to analyze housing demand.

[5] See Dietz (2002) for a comprehensive review of the neighborhood effects literature.

obtained by this strategy is likely to be contaminated by the omitted variables.

Recent attempts have used more enriched data sets where the formation of reference group is, or is believed to be, free of individuals' choices. For instance, Sacerdote (2001) and Zimmerman (2003) specify peer groups as roommates and rely on the random assignment of roommates to identify peer effects on student outcomes. Moffitt (2001) argues that randomized "partial-population intervention" experiments that directly affect only a subset of the population can help to identify social interaction effects. This quasi-experimental strategy has the potential to solve the endogenous sorting problem and is suggested by several researchers, including Manski (1993) and Rivkin (2001). However, the validity of this strategy depends heavily on the design and implementation of the experiment because it is possible that what is supposed to be random is actually a result of self-selection.[6]

Some studies, including Hoxby (2000) and Hanushek et al. (2003), employ a similar strategy but instead of using experiments, they exploit credible exogenous perturbations in the peer groups to identify peer effects. One potential problem is that the exogenous perturbations might be too small to provide reliable information for identification. Graham (2008) assumes that α_r is a random effect instead of a fixed effect and identifies social interactions using conditional variance restrictions. This will be plausible if one can defend the random effect assumption and has data satisfying the assumptions in Graham (2008).

III. An SAR Model with Individual Specific Social Interactions

To avoid the limitations of the linear-in-means model, we consider an SAR model with individual specific social interactions in a group interaction setting. The classic SAR model departs from the Manski (1993) model by measuring peer variables as the weighted averages of observed peer outcomes and characteristics instead of group expectations.[7] Unlike the Manski (1993) model, which focuses only on the deterministic part of the model, the SAR model also incorporates the information contained in the stochastic part, that is, the spatially correlated error terms. This additional source of information could be valuable for identification or improving estimation efficiency. In several papers (e.g., Lee 2003, 2004, and 2007a), Lee demonstrates that when the exogenous regressors do not

[6] See Heckman and Smith (1995) and Heckman et al. (1998) for discussion on difficulties in using experiments in social studies. Moffitt (2001) systematically discusses issues in analyzing social interactions by policy interventions.

[7] The estimation and properties of the SAR model have been studied in Anselin (1988), Kelejian and Prucha (1998, 1999), and Lee (2003, 2004, 2007a), among others.

provide enough information for identifying the model—for instance, in the pure SAR model where there are no exogenous regressors or in the SAR model where the exogenous regressors are irrelevant—one can exploit the information contained in the error terms to identify the endogenous social interaction coefficient λ_0. Thus, the SAR model could be an appealing approach in social interaction estimations. In fact, the SAR model with group interactions has been employed in several empirical studies of social interactions. However, to the best of my knowledge, few of the studies that use the SAR model have seriously exploited the available information to identify both endogenous and contextual social effects. Instead, most of them focus only on endogenous social effects.[8] Lee (2007b) is among the first to seriously consider these issues and extend the SAR model to incorporate both endogenous and contextual effects and group fixed effects in a group interaction setting. This study further extends Lee's (2007b) model to incorporate the more realistic asymmetric social interactions patterns under group interaction scenario.

A. The Model

The model is given by:

$$Y_r = \lambda_0 W_r Y_r + X_r \beta_{10} + W_r X_r \beta_{20} + l_r \alpha_r + \epsilon_r, \quad r = 1, \ldots, R, \tag{4}$$

where Y_r and X_r are the vector and matrix of outcomes and characteristics for the m_r members in group r. The term W_r is an $m_r \times m_r$ row-normalized spatial weights matrix of known constants with zero diagonal elements, and l_r is the m_r-dimensional vector of ones. The term λ_0 represents endogenous effects, and contextual effects are captured by β_{20}. The term α_r captures the group fixed effects, which are the observable and, more importantly, unobservable effects of the common factors faced by the same group members. Finally, the ϵ_{ir}'s are assumed to be innovations with $\mathrm{Var}(\epsilon_r) = \sigma_0^2 I_{m_r}$.

The key difference between this model and Lee's (2007b) is in the specification of the spatial weights matrix, W_r. In Lee (2007b), an individual is assumed to be equally affected by all the other members in the same group. This is a convenient and necessary assumption, especially when one does not have more information regarding the social interaction patterns within the group. In contrast, I specify W_r based on the actual friendship network within each group. Friendship networks are believed to be an important social context for adolescents. The sociology literature,

[8] See, e.g., Case and Katz (1991). As pointed out in n. 4, Ioannides and Zabel (2008) use an SAR model incorporating both endogenous and contextual effects. But the information contained in the disturbances is not explored in their random effects setting.

including Haynie (2001, 2002), Duncan, Boisjoly, and Harris (2001), and Maxwell (2000), discusses how friendship networks build information flow, social norms, social acceptances, social expectations, and so on. Their studies show that adolescents are significantly influenced by their friends. The importance of friends is also recognized by many economists. For example, in an illustration of his social capital theory, Becker (1996, 13) mentions that "a teenager may begin to smoke, join a gang, and neglect his studies mainly because his friends smoke, are gang members, and do not pay attention to school." The unique design of Add Health makes it possible to identify an adolescent's peers as her friends.

Similar to Lee (2007b), the term α_r captures the group fixed effects, which are the observable or, more importantly, unobservable effects of the common factors faced by the same group members. The identification for the linear-in-means model with fixed effects is impossible with cross-sectional data.[9] However, in this SAR model, because the peer measurements are not constant for individuals in the same group, the model can be identified with cross-sectional data in the presence of unobserved group fixed effects. Note that in this model, although the spatial weights matrix is exogenously given, it does not follow that the formation of reference groups is random, due to the presence of group fixed effects.[10]

B. The Identification Sources

Lee (2007b) demonstrates that both endogenous and contextual effects are identifiable in his model as long as group sizes are not constant so that the social interaction patterns across groups are different, and that the identification will be weak when group sizes are large. In the model considered here, W_r could have any arbitrary structure based on the actual friendship networks, and it may not be symmetric if the friend nominations are not reciprocal. Therefore, the interaction patterns depend on

[9] This can be easily seen by considering a model without social interactions but with unobserved fixed effects, $\alpha_r = \lambda_0 E(y_r|r) + \beta_{20} E(x_r|r)$. Then the process generated by this model will be identical to that generated by the linear-in-means model with social interactions but without fixed effects.

[10] As pointed out by a referee, the group fixed effects model cannot deal with the possible unobservable factors in common within groups; therefore the endogeneity of friendship formation has been only partially addressed. Our results could still be contaminated by correlated effects if people tend to make friends with others who have similar ability levels within the school grade. Meanwhile, there are lots of other reasons, such as common interests in sports, etc., for people to make friends. As a matter of fact, the results with SAR disturbance in Sec. IV.D partly confirm that friendship formation is not necessarily motivated by academic purpose. In a future project, we will consider a simultaneous system to model friendship formation and social interactions at the same time. That will provide an alternative way to address the endogeneity of friendship formation, although the identification and estimation of that system could be difficult.

the specific structure of the spatial weights matrix for each group, which introduces additional nonlinearity into the system. In this setting, even when all the group sizes are the same, there could still be sufficient variations in the social interaction patterns across groups due to different structures within each group. Therefore the model is identifiable even when all the group sizes are identical. Moreover, large group sizes should not pose any difficulty in identification as long as the spatial weights matrix W_r is sparse, so that each individual has only a few neighbors within each group. Some recent papers, including Calvó-Armengol et al. (2005), De Giorgi, Pellizzari, and Redaelli (2007), Bramoullé et al. (2009), and Laschever (2009), among others, systematically analyze the identification issues for network models which have a structure similar to that of the model considered here. Most discussions below closely follow these works.

To see where the identifications come from, let us first consider the model without fixed effects:

$$Y_r = \lambda_0 W_r Y_r + X_r \beta_{10} + W_r X_r \beta_{20} + \epsilon_r, \; r = 1, \ldots, R. \tag{5}$$

As discussed in Bramoullé et al. (2009), this model is identified if and only if $E(W_r Y_r | X_r)$ is not perfectly collinear with the regressors $(X_r, W_r X_r)$ so that some instruments can be found for the endogenous vector $W_r Y_r$. This condition is equivalent to that in which the matrices, I_r, W_r, W_r^2 are linearly independent. This will be true as long as the networks are partially overlapping, that is, some individual may not be friends with her friends' friends. Then some peers of peers do not directly affect an individual but only indirectly through her peers, and the characteristics of peers of peers may serve as instruments for peers' behavior of an individual. In particular, consider three individuals i, j, k in group r such that j is i's friend, k is j's friend, but k is not i's friend. It follows that for individual i, $W_r^2 X_r$ (i.e., x_{kr}) is a valid instrument for $W_r Y_r$ (i.e., y_{jr}), since x_{kr} affects y_{ir} but only indirectly, through its effect on y_{jr} (i.e., i and k are separated by a link of distance 2). As shown in Bramoullé et al. (2009), the presence of an intransitive triad guarantees the linear independency among I_r, W_r, W_r^2. Most networks have intransitive triads, and the natural exclusion restrictions induced by the network structure guarantee identification of the model.

Now consider the model with group fixed effect as in equation (4). Following Lee, Liu, and Lin (forthcoming), I eliminate the group fixed effect by the de-group-mean transformation, that is, by multiplying equation (4) with the matrix

$$J_r = I_{m_r} - \frac{1}{m_r} l_r l_r',$$

where I_{m_r} is the identity matrix of dimension m_r, and l_r is the m_r-dimensional vector of ones. Then, the model becomes

$$J_r Y_r = \lambda_0 J_r W_r Y_r + J_r X_r \beta_{10} + J_r W_r X_r \beta_{20} + J_r \epsilon_r. \tag{6}$$

Denoting $\hat{Y}_r = J_r Y_r$, $\hat{X}_r = J_r X_r$, and $\hat{\epsilon}_r = J_r \epsilon_r$, the model in (6) can be rewritten as

$$\hat{Y}_r = \lambda_0 J_r W_r \hat{Y}_r + \hat{X}_r \beta_{10} + J_r W_r \hat{X}_r \beta_{20} + \hat{\epsilon}_r. \tag{7}$$

Similarly, this model is identified if and only if $E(J_r W_r \hat{Y}_r | X_r)$ is not perfectly collinear with the regressors $(\hat{X}_r, J_r W_r \hat{X}_r)$, which is equivalent to that in which the matrices, I_r, W_r, W_r^2, W_r^3, are linearly independent. The condition is more demanding because some information has been used to deal with the fixed effects. Bramoullé et al. (2009) show that if there are two agents i and j in the group separated by a link of distance 3, then I_r, W_r, W_r^2, W_r^3 are linearly independent and the model is identified. Consider four individuals i, j, k, l, such that j is i's friend, k is j's friend, l is k's friend, but l is not i's friend. Then x_{lr} can serve as an instrument for y_{jr} in individual i's equation, since x_{lr} affects y_{ir} but only indirectly, through its effect on y_{kr}. Thus the model is identified. Bramoullé et al. (2009) show that identification holds for most networks and fails only for a few networks that have very restrictive structures, such as complete bipartite networks.

Therefore, the endogenous and contextual effects are identified as soon as there are some variations in the reference groups across individuals. By allowing the peer measurements to vary across individuals in the same group, we introduce an additional source of nonlinearity and break down the linear dependency between the endogenous and contextual effects, thus resolving the "reflection problem."

C. The Maximum Likelihood Estimation

Following Lee et al. (forthcoming), we estimate the model with group fixed effects by maximum likelihood estimation. Consider equation (7). Note that the error terms $\hat{\epsilon}_r$ are not innovations. Instead, they are correlated within the group and heteroskedastic. Furthermore, the variance-covariance matrix for the error terms is given by

$$\mathrm{Var}(\hat{\epsilon}_r) = J_r J_r' \sigma_0^2 = J_r \sigma_0^2 = \sigma_0^2 \Sigma_r, \tag{8}$$

which is singular and with rank $(m_r - 1)$, so there is a set of $(m_r - 1)$ linearly independent relationships among the elements of $\hat{\epsilon}_r$.

As in the theory of generalized inverses for the estimation of linear regression models, the eigenvalues eigenvectors decomposition technique is used to eliminate the linear dependency among observations to obtain an effectively independent sample. Let $[F_r, H_r]$ be the orthogonal matrix of J_r, where F_r corresponds to the eigenvalues of one and H_r corresponds to the zero eigenvalues.

Then the model in (7) can be rewritten into the following with effective observations and independently and identically distributed (i.i.d.) disturbances:

$$Y_r^* = \lambda_0 W_r^* Y_r^* + X_r^* \beta_{10} + W_r^* X_r^* \beta_{20} + \epsilon_r^*, \tag{9}$$

where $Y_r^* = F_r' \hat{Y}_r$ and $\epsilon_r^* = F_r' \hat{\epsilon}_r$ are $(m_r - 1)$–dimensional vectors, $X_r^* = F_r' \hat{X}_r$ is an $(m_r - 1) \times k$–dimensional matrix, and $W_r^* = F_r' W_r F_r$ is an $(m_r - 1) \times (m_r - 1)$-dimensional matrix. Note that $\mathrm{Var}(\epsilon_r^*) = \sigma_0^2 I_{m_r^*}$ where $m_r^* = m_r - 1$.

The model in (9) has the same structure as a typical SAR model and thus can be estimated by various standard methods in the spatial econometrics literature. Although the diagonal elements of the spatial weights matrix W_r^* are not zero in (9), it turns out that the nonzero diagonal feature of W_r^* does not pose a threat to the consistency of various estimation procedures. Denote $X_r^* = (X_r^*, W_r^* X_r^*)$ and $\beta = (\beta_1', \beta_2')$. Under the normality assumption, the log likelihood for group r is given by

$$\ln L_r = -\frac{m_r^*}{2} \ln (2\pi\sigma^2) + \ln |I_{m_r^*} - \lambda W_r^*| - \frac{1}{2\sigma^2} [(I_{m_r^*} - \lambda W_r^*) Y_r^* - X_r^* \beta]' [(I_{m_r^*} - \lambda W_r^*) Y_r^* - X_r^* \beta], \tag{10}$$

which can be written in terms of the original variables as follows:

$$\ln L_r = -\frac{(m_r - 1)}{2} \ln (2\pi\sigma^2) - \ln (1 - \lambda) + \ln |I_{m_r} - \lambda W_r| - \frac{1}{2\sigma^2} [(I_{m_r} - \lambda W_r) Y_r - X_r \beta]' J_r [(I_{m_r} - \lambda W_r) Y_y - X_r \beta]. \tag{11}$$

Finally, the log likelihood for the whole sample is given by $\ln \ell_n = \sum_r^R \ln L_r$.[11]

[11] More details about the theoretical and methodological features of a similar model with SAR disturbances can be found in Lee et al. (2008). Lin (2005) uses a different method, i.e., deviation from friend mean method, to get rid of the fixed effect and estimate the model by a two-stage least squares (2SLS) procedure, with instrumental variables generated by some functions of the exogenous regressors $\hat{X}_r$, $\hat{W}_r \hat{X}_r$, and the spatial weights matrix $\hat{W}_r$. However, using 2SLS is not efficient, as it only exploits the information contained in the deterministic part of the mode, leaving the information in the stochastic part unused. To achieve more efficient and precise estimation results, we use the maximum likelihood method to estimate the model.

IV. Empirical Framework

A. The Add Health Survey

Add Health is a school-based study of adolescents in grades 7–12 in 132 schools that were selected based on a stratified nationally representative sample of all public and private high schools in the United States (Harris 2009). With the objective of examining adolescents' development in a social context, Add Health contains detailed information about the respondents' characteristics and social networks, providing ideal data for studying social interaction effects.

Three waves of data are available. In wave I, an in-school questionnaire was given to all the students attending the sampled schools from September 1994 to April 1995, resulting in a total sample of over 90,000 students. The questions covered the respondent's demographics; family background; academic outcome; health-related behaviors, including drug use; and teenage pregnancy. Most important, the adolescents were asked to nominate up to five male and five female friends. Friends' identification numbers make it possible to link a respondent's information to her friends' and construct friendship networks. One year later, in wave II, about 14,000 adolescents participated in an in-home survey. From August 2001 to April 2002, all the wave I respondents who could be located were interviewed for the wave III in-home survey, which contains a sample size of over 15,000.[12]

For this research, the desired data set is the wave I in-school survey. Because it covers all the students in the sampled schools, the respondent's friends are also likely to be in the sample. GPA can be calculated from the respondent's grades in several subjects, including English or language arts, history or social science, mathematics, and science, which provides a useful measure for academic attainment. Finally, the demographic and family questions permit a variety of control variables to be included in the study.

Thanks to the unique design of Add Health, we are able to identify an adolescent's peers as her friends. In this study, "group" refers to a grade level in the same school.[13] Within a school-grade group, only a respondent's friends are identified as her peers and are given equal weight, while all the others in the same group are assigned zero weight.[14] Therefore, the entry of the spatial weights matrix, $w_{r,ij}$, will be $1/n$ if respondent i identifies person j as one of her n friends in her grade. Otherwise, a zero

[12] Only a few students were randomly chosen to participate in the in-home survey, whereas the in-school survey covered all students in a school.

[13] It would be better if we could specify group at the class level. Unfortunately, the Add Health survey does not provide such information.

[14] The next subsection considers alternative ways to specify the elements of the spatial weights matrix.

will be assigned. The group fixed effects term α_r captures the effects of the common variables, observable or, more commonly, unobservable, that are faced by all the students in the same grade of the same school.

B. Sample Summary Statistics

The sample is based on several selection criteria. In order to be in this sample, a respondent must have valid information not only on her own characteristics but also on her friends'. Table 1 gives detailed information about sample selection procedure. Among the original 90,118 respondents, 70,639 have complete information on their own characteristics. Of these, 3.6% do not have valid information about their friends, so they are excluded from the sample. Table 1 also demonstrates the racial composition and family background of the sample, which changes as observations that fail to satisfy the sample construction criteria are excluded. We can see that the final sample has a slightly higher GPA, a higher fraction of whites, a larger percentage of respondents living with both parents, and also a higher proportion of respondents whose mothers have a high school education or work at professional occupations or do not work. On the other hand, we have slight decreases in the fractions of both blacks and Hispanics, and a decrease in the percentage of respondents whose mothers have less than a high school education. Overall, the differences are modest, and the sample selection bias is not a concern for our sample.

The final sample consists of 68,131 observations. Of these, 18,572 individuals are isolated—that is, they do not name a friend and are never nominated by others. The isolated sample consists of 482 school-grade groups, with a minimum, mean, and maximum group size of 2, 38.5, and 427, respectively. Conversely, there are 49,559 observations and 486 groups in the network sample. The minimum of the group sizes is 2, the mean is 102, and the maximum is 446. The numbers of nominated friends in the network sample range from 1 to 10, with the mean being 3.5. Only about 0.32% of these respondents nominated the maximum of 10 friends, so censoring at 10 does not appear to be a significant problem.

Table 2 provides the summaries and definitions of the variables used in this study. For the set of individual characteristics X_r, following previous studies (e.g., Duncan et al. 2001), and given data availability, the model uses age, years in school, gender, race, sports club membership, family structure, mother's education, and mother's occupation. Both endogenous and contextual effects are included. For the contextual effects, we use the same set of variables, allowing any element that directly determines an individual's outcome to also affect her peers'. As for the endogenous effects, unlike existing studies that use lagged peer outcome in order to avoid the "reflection problem," we use the contemporary measure of peer achievement, which will not cause any identification

Table 1
Sample Construction

Sample Selection Criteria	N	GPA	White	Black	Hispanic	Live with Both Parents	Mom Education Less than High School	Mom Education High School	Professional	Stay Home
Total respondents in Add Health Wave I In-School Survey	90,118	2.798 (78,277)	.532 (88,137)	.195 (88,137)	.151 (88,137)	.722 (86,540)	.100	.290	.246	.198
After deletion due to no valid ID information	85,627	2.799 (74,544)	.532 (83,812)	.195 (83,812)	.150 (83,812)	.723 (82,389)	.100	.291	.248	.198
After deletion due to no valid age information	85,267	2.800 (74,368)	.533 (83,582)	.195 (83,582)	.150 (83,582)	.723 (82,182)	.101	.291	.249	.199
After deletion due to no valid grade information	84,721	2.800 (73,999)	.534 (83,086)	.194 (83,086)	.149 (83,086)	.724 (81,713)	.100	.292	.249	.199
After deletion due to no valid gender information	84,292	2.801 (73,673)	.535 (82,684)	.194 (82,684)	.149 (82,684)	.724 (81,325)	.100	.292	.249	.199
After deletion due to no valid race information	82,684	2.803 (72,498)	.535	.194	.149	.725 (80,106)	.101	.293	.251	.200
After deletion due to no valid number of years in school information	82,461	2.803 (72,370)	.536	.194	.149	.725 (79,949)	.101	.294	.251	.200
After deletion due to no valid family structure information	79,949	2.809 (70,639)	.544	.189	.146	.725	.103	.300	.257	.205
After deletion due to no valid GPA information	70,639	2.809	.562	.178	.138	.736	.100	.305	.264	.205
After deletion due to no valid friend information	68,131	2.817	.569	.175	.135	.739	.097	.306	.267	.204

Note.—Values in parentheses denote the number of observations when different from the total sample size N. The definitions of the variables are given in table 2.

Table 2
Variable Definitions and Summary Statistics

Variable	Definition	Own Characteristics		Friends' Mean Characteristics	
		Mean	SD	Mean	SD
GPA	Average grade in mathematics, science, English or language arts, and history or social science	2.817	(.805)	2.914	(.601)
Age	Age	15.061	(1.674)	14.909	(1.614)
Years in school	No. of years in current school	2.501	(1.420)	2.585	(1.300)
Male	if male; 0 otherwise	.486	(.500)	.427	(.359)
(Female)	if female; 0 otherwise	.514	(.500)	.573	(.359)
(White)	if white; 0 otherwise	.569	(.495)	.609	(.421)
Black	if black; 0 otherwise	.175	(.380)	.161	(.342)
Asian	if Asian; 0 otherwise	.065	(.247)	.063	(.202)
Hispanic	if Hispanic; 0 otherwise	.135	(.341)	.115	(.270)
Other race	if American Indian or other race; 0 otherwise	.056	(.229)	.052	(.155)
Sports club member	if sports club member; 0 else	.527	(.499)	.579	(.368)
(Not sports club member)	if not sports club member; 0 otherwise	.473	(.499)	.421	(.368)
Live with both parents	if living with both parents; 0 otherwise	.739	(.439)	.771	(.300)
(Not live with both parents)	if not living with both parents; 0 otherwise	.261	(.439)	.229	(.300)
Mom education less than HS	if mom's education less than high school; 0 otherwise	.097	(.296)	.086	(.206)
(Mom education HS)	if mom's education high school; 0 otherwise	.306	(.461)	.310	(.320)
Mom education more than HS	if mom's education beyond high school; 0 otherwise	.422	(.494)	.459	(.361)
Mom education missing	if the information about mom's education is missing; 0 otherwise	.101	(.301)	.086	(.198)
Professional	if mom is a doctor, lawyer, scientist, teacher, executive, director, and the like; 0 otherwise	.267	(.442)	.296	(.311)
(Stay home)	if mom is a homemaker, retired, or does not work; 0 otherwise	.204	(.403)	.194	(.270)
Other jobs	if mom's occupation is not among the "Professional" or "Stay home"; 0 otherwise	.362	(.481)	.373	(.326)
Welfare	if mom receives public assistance, such as welfare; 0 else	.008	(.087)	.005	(.052)
Mom job missing	if the information about mom's job is missing; 0 otherwise	.085	(.279)	.073	(.182)

Note.—The variables in brackets are the omitted categories in the following estimations.

problem in the SAR model. As noted by Hanushek et al. (2003), using lagged peer outcome may leave out some effects of concurrent peer behavior that cannot be captured by the lagged variables, leading to an underestimate of peer influences.

The mean age of the respondents is about 15, on average, and they have attended their current school for 2.5 years. Among the 68,131 observations, 48.6% are male, 13.5% are Hispanic, 56.9% are white, and 17.5% are black. Asians account for 6.5%, and 5.6% of the sample are of other races. About 73.9% of the respondents live with both parents. The highest education level achieved is high school for about 30% of the respondents' mothers, is beyond high school for 42.2% of the mothers, and is less than 12 years for 9.7% of the mothers. As for mother's occupation, Add Health provides a detailed list with more than 15 categories. We combine these occupations into four broader categories, along with a missing indicator.[15] Specifically, 26.7% of the respondents' mothers work at professional occupations, such as teachers, doctors, lawyers, and executives. Twenty percent of the mothers are homemakers or do not work. Less than 1% of the mothers receive welfare,[16] and 36.2% of the mothers hold other jobs. The dependent variable GPA is the average grade of four subjects: mathematics, science, English or language arts, and history or social science.[17] The mean GPA of the sample is 2.817 out of 4, with a standard deviation of 0.805. As we can see from table 2, the means of the friend's mean characteristics are similar to those of own characteristics.

These variables cover almost all the background information available in this in-school survey.[18] One limitation is that we do not have information on family income, which is arguably an important determinant in adolescents' academic achievement. But the education and occupations of the mothers should provide some information about family income and thus should capture most, if not all, of the effects of family income. On the other hand, information at the school level, such as school quality and teacher/pupil ratio, is unnecessary thanks to the fixed effects estimation strategy.

C. Empirical Results

We first estimate models consisting of either endogenous or contextual effects, which are similar to the specifications of many existing studies,

[15] Different combinations of the occupations have been considered, and the estimation results are robust to these changes.

[16] This is lower than the whole population of mothers of school-age children, where the fraction is about 5%. The underreporting may be due to the adolescents' ignorance about their mothers' receipt of welfare.

[17] Since only letter grades are reported, we assign A 4, B 3, C 2, and D 1.

[18] The in-home survey provides richer background information but contains only a subset of the sample.

with and without school-grade fixed effects. Then, we consider the full model with both social effects, with and without controlling for group unobservables. Some discussions regarding the robustness of the results are also provided.

1. Models with Either Endogenous or Contextual Effects

Table 3 presents the results for the models with only endogenous or contextual effects. Models (1) and (2) are the models with only endogenous effects, without and with controlling for the school-grade fixed effects, respectively. The differences between the results from these two models are huge. In particular, the estimated endogenous effect coefficient in model (1) is 0.143 and in model (2) is 0.302, and both are significant. Further, the estimation results for most respondent characteristics, including age, years in school, other race, mother's occupation of "other," mother's welfare status, and so on, also exhibit significant change for the two models. The significant increase in the log likelihood value suggests great improvement of the model with fixed effects over the one without. Therefore, the omitted variables seriously contaminate the peer effect estimation. For these two specifications, one cannot identify a clear interpretation for the estimated endogenous interaction coefficients since they capture the mixture of endogenous and contextual effects, not the pure endogenous effects.

Models (3) and (4) are the models with contextual effects only, without and with fixed effects. Both models show strong, yet different, contextual effects. Some estimated contextual effect coefficients, such as years in school, other race, mother's education less than high school, and mother's occupation of "other" even switch signs. Other contextual variables, including black, mother's welfare status, and mother's occupation of "professional jobs" also exhibit great differences. The own characteristics whose estimation coefficients are sensitive to the inclusion of fixed effects include age, mother's occupation of "other," whether the respondent lives with both parents, and black. Similar to the models with only endogenous effects, the inclusion of fixed effects greatly improves the goodness of fit, as shown from the significant change in the log likelihood values. However, we still cannot draw a clear inference from these estimates, since they represent the mixed effects of endogenous and contextual variables. Unfortunately, these specifications are common practice in the social interaction literature, due in great part to the convenience of circumventing the "reflection problem."

2. Models with Both Endogenous and Contextual Effects

In order to separate endogenous effects from contextual effects, we estimate the full model with both effects, presenting the results in table 4.

Table 3
Results for the Models with Endogenous Effects or Contextual Effects Only

Explanatory Variable	(1)	(2)	(3)	(4)
Endogenous effect	.143***	.302***		
	(.002)	(.005)		
Contextual effects (peer means):				
Age			−.029***	−.095***
			(.002)	(.009)
Years in school			−.009**	.005
			(.004)	(.006)
Male			.079***	.020*
			(.011)	(.010)
Black			−.012	−.101***
			(.016)	(.017)
Asian			.092***	.138***
			(.021)	(.022)
Hispanic			−.072***	−.071***
			(.017)	(.018)
Other race			.097***	−.102***
			(.024)	(.022)
Sport			.250***	.098***
			(.011)	(.010)
Live with both parents			.349***	.127***
			(.013)	(.012)
Mom education less than HS			.085***	−.053***
			(.020)	(.018)
Mom education more than HS			.212***	.174***
			(.013)	(.012)
Mom education missing			.212***	−.022
			(.020)	(.018)
Mom's job is professional			.155***	.019
			(.016)	(.015)
Mom other jobs			.095***	−.058***
			(.014)	(.013)
Mom on welfare			−.004	−.158**
			(.070)	(.064)
Mom job missing			.135***	−.108***
			(.022)	(.020)
Own X:				
Age	.140***	−.127***	.152***	−.129***
	(.001)	(.004)	(.001)	(.005)
Years in school	−.003	.011***	.009***	.010***
	(.002)	(.003)	(.003)	(.003)
Male	−.127***	−.158***	−.188***	−.176***
	(.006)	(.006)	(.007)	(.006)
Black	−.116***	−.190***	−.093***	−.170***
	(.009)	(.010)	(.012)	(.012)
Asian	.275***	.227***	.236***	.232***
	(.013)	(.013)	(.015)	(.014)
Hispanic	−.135***	−.141***	−.126***	−.129***
	(.010)	(.010)	(.011)	(.011)
Other race	.029**	−.079***	.019	−.082***
	(.014)	(.012)	(.014)	(.012)
Sport	.221***	.112***	.194***	.111***
	(.006)	(.006)	(.007)	(.006)
Live with both parents	.269***	.129***	.265***	.134***
	(.007)	(.007)	(.007)	(.007)
Mom education less than HS	−.057***	−.086***	−.047***	−.087***
	(.011)	(.010)	(.011)	(.010)
Mom education more than HS	.243***	.172***	.230***	.181***
	(.007)	(.007)	(.008)	(.007)
Mom education missing	.112***	−.006	.098***	−.006
	(.011)	(.010)	(.011)	(.010)

Table 3 ***(Continued)***

Explanatory Variable	(1)	(2)	(3)	(4)
Mom's job is professional	.102***	.023***	.101***	.021**
	(.009)	(.008)	(.009)	(.008)
Mom other jobs	.030***	−.047***	.035***	−.052***
	(.008)	(.007)	(.008)	(.007)
Mom on welfare	.012	−.048	.018	−.046
	(.036)	(.032)	(.036)	(.032)
Mom job missing	.035***	−.085***	.032***	−.090***
	(.012)	(.011)	(.012)	(.011)
Fixed effect	No	Yes	No	Yes
Likelihood value	−81,395	−71,369	−81,627	−72,372

Note.—The dependent variable is GPA. The sample size is 68,131. Standard errors are in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Again, large differences occur between model (5), the full model without controlling for the group fixed effects, and model (6), the full model with fixed effects. In particular, the estimated endogenous effect coefficient shrinks by about one-third for model (6) as compared to model (5). The contextual variables that are sensitive to the inclusion of fixed effects are years in school, black, Asian, other race, mother's education less than high school, and mother's occupation of "other." The introduction of the fixed effects also greatly affects the estimation results for several own characteristics, such as age, other race, mother's occupation of "other," mother's occupation of "professional jobs," and whether the respondent lives with both parents. Finally, the log likelihood value also exhibits great improvement for model (6).

Among the six models presented in tables 3 and 4, the log likelihood values clearly indicate that model (6), the model with both endogenous and contextual effects and with group fixed effects provides the best goodness of fit of the data. The results from this model show strong evidence for both endogenous effects and contextual effects. In particular, the estimated endogenous effect coefficient is 0.274, which is about 34% of a standard deviation in the average GPA, suggesting that a 1 standard deviation increase in peer average achievement raises own GPA by 0.221 points or by 7.85% of its mean of 2.817.[19] These results show that the social multiplier effect does exist in school learning. As for the contextual effects, most variables show significant impact. In particular, the contextual variables that show negative effects include age, black (at 5% level), other race (at 5% level), mother's welfare status (at 10% level), and mother's occupation of "other." The estimated coefficients range from

[19] The magnitude of the estimate is similar to those in other studies. For example, Hanushek et al. (2003) find that the estimated peer effect coefficients range from 0.15 to 0.24 in student academic achievement. And note that they specify peer group by grade, instead of friendship as here.

Table 4
Results for the Models with Endogenous and Contextual Effects

Explanatory Variable	(5)	(6)
Endogenous effect	.404***	.274***
	(.005)	(.005)
Contextual effects (peer means):		
Age	−.085***	−.046***
	(.001)	(.009)
Years in school	−.012***	.000
	(.004)	(.006)
Male	.122***	.060***
	(.011)	(.010)
Black	.025*	−.037**
	(.015)	(.017)
Asian	−.027	.056**
	(.020)	(.022)
Hispanic	.015	−.011
	(.016)	(.018)
Other race	.075***	−.055**
	(.022)	(.022)
Sport	.104***	.051***
	(.010)	(.010)
Live with both parents	.170***	.080***
	(.012)	(.012)
Mom education less than HS	.077***	−.022
	(.019)	(.018)
Mom education more than HS	.080***	.099***
	(.012)	(.012)
Mom education missing	.125***	−.015
	(.019)	(.018)
Mom's job is professional	.071***	.009
	(.015)	(.014)
Mom other jobs	.049***	−.038***
	(.013)	(.013)
Mom on welfare	−.013	−.112*
	(.066)	(.063)
Mom job missing	.070***	−.079***
	(.021)	(.020)
Own *X*:		
Age	.156***	−.122***
	(.001)	(.004)
Years in school	.011***	.010***
	(.003)	(.003)
Male	−.188***	−.175***
	(.007)	(.006)
Black	−.091***	−.161***
	(.011)	(.012)
Asian	.227***	.221***
	(.014)	(.014)
Hispanic	−.122***	−.126***
	(.010)	(.011)
Other race	.007	−.072***
	(.013)	(.012)
Sport	.164***	.102***
	(.006)	(.006)
Live with both parents	.223***	.124***
	(.007)	(.007)
Mom education less than HS	−.054***	−.078***
	(.011)	(.010)

Table 4 ***(Continued)***

Explanatory Variable	(5)	(6)
Mom education more than HS	.203***	.165***
	(.007)	(.007)
Mom education missing	.071***	−.004
	(.011)	(.010)
Mom's job is professional	.084***	.020**
	(.009)	(.008)
Mom other jobs	.024***	−.047***
	(.007)	(.007)
Mom on welfare	−.003	−.043
	(.034)	(.032)
Mom job missing	.014	−.082***
	(.011)	(.011)
Fixed effect	No	Yes
Likelihood value	−78,308	−71,174

NOTE.—The dependent variable is GPA. The sample size is 68,131. Standard errors are in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

−0.037 (for black) to −0.112 (for mother's welfare status). On the other hand, having peers who are male, Asian (at 5% level), who participate in a sports club, who live with both parents, or who have mothers with more than a high school education will help improve an individual's GPA. The contextual variables of mothers with more than a high school education and whether the respondent lives with both parents show the largest effects, with coefficients of 0.099 and 0.080, respectively. All the other variables have an estimated coefficient around 0.05.

As for the respondent's own characteristics, most estimated coefficients have the expected signs. In general, older students tend to do worse, and students who have been in the current school for longer periods tend to do better. Male students score 0.175 grade points lower than female students. Hispanic, black, and students of other races score 0.126, 0.161, and 0.072 points lower than white students, respectively, while Asian students score 0.221 points higher than white students. Students who participate in a sports club and who live in an intact family of both parents score 0.102 and 0.124 points higher, respectively. Mother's education is an important determinant of grades, and the relationship is positive. Specifically, children of mothers with a less than high school education score 0.078 points lower than those of high school graduate mothers, while children of mothers who have education beyond high school achieve 0.165 points higher. Compared to students whose mothers do not work, children of teachers, lawyers, or other professional job holders are more successful (at 5% level), while children of other job holders have lower grades. From these results, we can see that the variables of Asian race, participation in a sports club, living with both parents, and mother with more than high school education have positive effects on own GPA as well as friends' GPA. Meanwhile, variables of age, black, other race, and mother's occupation of "other" have negative effects on both own GPA and friends' GPA.

As discussed in many studies, including Amato (1993), Pong and Ju (2000), and Sun and Li (2001), adolescents living with both parents enjoy more parental warmth, emotional stability, and parental involvement and support, lower levels of parent-child conflict, and so forth, which helps improve an adolescent's academic performance. This analysis contributes to the literature by showing that an adolescent's academic achievement is significantly affected not only by her own family background but also by the family backgrounds of her peers. Although this study cannot indicate the mechanism by which this influence occurs, there are several reasonable possibilities. Children who live with both parents tend to be more self-disciplined and do not disrupt the classroom; they may have more learning resources available at home and share them with their friends, and the like. Positive associations between mother's education and child academic attainment are among the most replicated results in the literature (e.g., Bee et al. 1982; Haveman and Wolfe 1995). Children with educated mothers enjoy an enriched home learning environment, more readily available family resources, and probably higher family expectations and higher innate ability inherited from their parents. Further, better-educated mothers may value education more and spend more time in monitoring the schools and teachers. The results of this study indicate that these benefits clearly generate spillover effects to their friends. A similar effect may occur for other groups of students. Those children in the study who attain higher grades—Asians and those participating in sports clubs—may have more developed organization and reasoning skills, which are passed to their friends. On the other hand, students who are less disciplined and experience less success in school—in this study those who are older students, black, of other race, or with a mother's occupation of "other"—may have fewer family resources available and may tend to disrupt their friends in learning. Since these estimates are obtained after controlling for the school-grade fixed effects, they provide sensible evidence for the existence of peer effects, both endogenous and contextual, in student academic attainment.

D. Some Alternative Specifications

In this subsection, we shall consider several alternative specifications of the spatial weights matrix and the model and see how the results change. All models considered in this section include unobserved group fixed effects.

1. Exclusion of the Isolated Sample

In the final sample of 68,131 observations, 18,572 respondents (or 27.3%) are isolated, which may affect the estimation results. Table 5 summarizes the estimation results for three models, that is, the model with endogenous effects only, the one with contextual effects only, and the model with both effects, based on the network sample only. Comparing model (7) with model (2), model (8) with model (4), and model (9) with model (6), we can see that the estimation results are quite robust to the change of the sample selection criterion. Since the isolated observations do not interact with other people, they do not appear to affect the estimations of the social interaction effect coefficients. However, dropping these observations leads to slight inflations of the standard errors of the estimated coefficients on own characteristics for all three models. At the same time, the great increases in the likelihood values may indicate the efficiency gain associated with the use of the more effective sample.

2. An Alternative Specification of the Weighting Matrix

In this study, the spatial weights matrix is defined by the real friendship nominations, and it will be asymmetric if the friendship nominations are not symmetric. For example, if Tom (person 1) nominated Bob (person 2) as his friend, while Bob did not name Tom as his friend, then the (1, 2) entry of the weighting matrix will have some positive weight, while the (2, 1) element will have zero weight. Therefore, this study does not treat friendships as reciprocal in nature, which means that even if Bob is Tom's friend, it does not necessarily follow that Tom is also Bob's friend. This is consistent with several studies, including Haynie (2001) and Vaquera and Kao (2008), which find that friendships are not always reciprocal and that reciprocated friendships, unlike the nonreciprocated ones, are likely to be more intimate and important. Other studies, such as Laursen (1993) and Erwin (1998), on the other hand, argue that friendships are reciprocal by definition.

To examine how the friendship reciprocity assumption might affect the empirical estimation, we refine the weighing matrix by assuming that friendships are reciprocal, that is, if Bob nominated Tom as his friend, then Tom will be treated as Bob's friend by definition, regardless of his own nominations, and vice versa. Then the resulting weighting matrix will be symmetric, and those who did not nominate a friend but are nominated by others will be included in the network sample. The network sample resulting from the reciprocity assumption consists of 60,495 observations, 10,936 more compared to the network sample based on real friendship nominations. In the new network sample, there are 489 groups, and the minimum, mean, and maximum of the group sizes are 2, 123.7, and 484, respectively. Both nominations received and numbers of friends

Table 5
Results from Network Sample Only

Explanatory Variable	(7)	(8)	(9)
Endogenous effect	.287***		.259***
	(.005)		(.005)
Contextual effects (peer means):			
Age		−.093***	−.049***
		(.008)	(.008)
Years in school		.005	.001
		(.006)	(.006)
Male		.029***	.067***
		(.011)	(.010)
Black		−.124***	−.073***
		(.019)	(.019)
Asian		.153***	.079***
		(.022)	(.022)
Hispanic		−.070***	−.017
		(.018)	(.017)
Other race		−.102***	−.060***
		(.022)	(.021)
Sport		.099***	.057***
		(.010)	(.010)
Live with both parents		.127***	.084***
		(.012)	(.011)
Mom education less than HS		−.052***	−.024
		(.018)	(.017)
Mom education more than HS		.173***	.104***
		(.012)	(.011)
Mom education missing		−.021	−.015
		(.018)	(.018)
Mom's job is professional		.018	.010
		(.014)	(.014)
Mom other jobs		−.059***	−.040***
		(.012)	(.012)
Mom on welfare		−.156**	−.114*
		(.062)	(.061)
Mom job missing		−.109***	−.082***
		(.020)	(.019)
Own *X*:			
Age	−.136***	−.138***	−.127***
	(.005)	(.006)	(.005)
Years in school	.012***	.011***	.011***
	(.003)	(.004)	(.003)
Male	−.160***	−.187***	−.187***
	(.006)	(.007)	(.007)
Black	−.186***	−.139***	−.122***
	(.011)	(.017)	(.016)
Asian	.205***	.205***	.188***
	(.015)	(.017)	(.016)
Hispanic	−.147***	−.128***	−.122***
	(.012)	(.013)	(.013)
Other race	−.084***	−.087***	−.074***
	(.014)	(.014)	(.014)
Sport	.105***	.102***	.091***
	(.007)	(.007)	(.007)
Live with both parents	.138***	.144***	.131***
	(.008)	(.008)	(.008)
Mom education less than HS	−.096***	−.095***	−.084***
	(.012)	(.012)	(.012)
Mom education more than HS	.170***	.181***	.161***
	(.008)	(.008)	(.008)

Table 5 ***(Continued)***

Explanatory Variable	(7)	(8)	(9)
Mom education missing	−.027**	−.027**	−.024**
	(.012)	(.012)	(.012)
Mom's job is professional	.021**	.017*	.016*
	(.009)	(.009)	(.009)
Mom other jobs	−.048***	−.054***	−.049***
	(.008)	(.008)	(.008)
Mom on welfare	−.088**	−.081**	−.077**
	(.039)	(.040)	(.039)
Mom job missing	−.088***	−.094***	−.083***
	(.013)	(.013)	(.013)
Fixed effect	Yes	Yes	Yes
Likelihood value	−50,687	−51,577	−50,450

NOTE.—The dependent variable is GPA. The sample size is 49,559. Standard errors are in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

nominated range from 1 to 36, with a mean of 5.1; more than 93% of the sample has fewer than 10 friends. Table 6 presents the estimation results based on friendship reciprocity assumption for three models, that is, the model with endogenous effects only, the one with contextual effects only, and the model with both effects. Comparing model (10) with model (2), model (11) with model (4), and model (12) with model (6), imposing the friendship reciprocity assumption does not change estimation results much, although the standard errors of the estimated coefficients on own characteristics for all three models have become slightly smaller. The great drops in the likelihood values for all three models indicate the advantage of the weighting matrix based on real friendship nomination over the one based on reciprocity assumption.

3. An Alternative Model with SAR Disturbances

The model we have considered so far is the model given in (4), where the error terms are assumed to be innovations. However, another equally plausible model is a spatial model with an SAR error structure given as follows:

$$Y_r = \lambda_0 W_r Y_r + X_r\beta_{10} + W_r X_r \beta_{20} + l_r\alpha_r + \epsilon_r,$$

$$\epsilon_r = \rho_0 W_r \epsilon_r + u_r, r = 1, \ldots, R. \quad (12)$$

The model in (4) is a special case with $\rho_0 = 0$. This model implies that in addition to the endogenous effects and contextual effects, some unobserved characteristics of the friends are also interdependent. This could follow, for instance, if people choose their friends based on unobservables that are not constant over the group members, that is, some unobservables that are not captured by the fixed effect term α_r. Lee et al. (forthcoming) find that if SAR disturbances are mis-specified as a model with innovation

Table 6
Results from Sample with Reciprocal Friendships

Explanatory Variable	(10)	(11)	(12)
Endogenous effect	.303***		.270***
	(.005)		(.005)
Contextual effects (peer means):			
Age		−.116***	−.067***
		(.008)	(.008)
Years in school		.010*	.005
		(.006)	(.006)
Male		.000	.045***
		(.009)	(.010)
Black		−.104***	−.038**
		(.017)	(.016)
Asian		.179***	.098***
		(.022)	(.021)
Hispanic		−.050***	.000
		(.018)	(.019)
Other race		−.112***	−.059***
		(.023)	(.023)
Sport		.112***	.065***
		(.010)	(.010)
Live with both parents		.153***	.101***
		(.012)	(.012)
Mom education less than HS		−.042**	−.015
		(.018)	(.018)
Mom education more than HS		.231***	.153***
		(.013)	(.012)
Mom education missing		.016	.021
		(.019)	(.018)
Mom's job is professional		−.006	−.011
		(.015)	(.015)
Mom other jobs		−.084***	−.058***
		(.013)	(.013)
Mom on welfare		−.134**	−.107*
		(.058)	(.057)
Mom job missing		−.111***	−.071***
		(.020)	(.020)
Own *X*:			
Age	−.125***	−.126***	−.118***
	(.004)	(.004)	(.004)
Years in school	.009***	.009***	.008***
	(.003)	(.003)	(.003)
Male	−.161***	−.174***	−.173***
	(.005)	(.006)	(.006)
Black	−.188***	−.162***	−.155***
	(.009)	(.011)	(.011)
Asian	.223***	.223***	.209***
	(.012)	(.013)	(.013)
Hispanic	−.137***	−.120***	−.118***
	(.010)	(.010)	(.010)
Other race	−.080***	−.077***	−.071***
	(.011)	(.012)	(.011)
Sport	.118***	.115***	.107***
	(.005)	(.006)	(.005)
Live with both parents	.121***	.126***	.115***
	(.006)	(.006)	(.006)
Mom education less than HS	−.077***	−.073***	−.067***
	(.009)	(.009)	(.009)
Mom education more than HS	.171***	.177***	.162***
	(.006)	(.006)	(.006)

Table 6 ***(Continued)***

Explanatory Variable	(10)	(11)	(12)
Mom education missing	−.003	−.003	−.002
	(.009)	(.009)	(.009)
Mom's job is professional	.030***	.025***	.025***
	(.008)	(.008)	(.008)
Mom other jobs	−.044***	−.048***	−.043***
	(.007)	(.007)	(.007)
Mom on welfare	−.049*	−.056*	−.049*
	(.029)	(.029)	(.029)
Mom job missing	−.084***	−.087***	−.080***
	(.010)	(.010)	(.010)
Fixed effect	Yes	Yes	Yes
Likelihood value	−83,543	−84,579	−83,238

Note.—The dependent variable is GPA. The sample size is 79,067. Standard errors are in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

errors, the results will be seriously biased. To check the sensitivity of the estimation results to this model specification, we can estimate three models with SAR disturbances: the model with endogenous effects only, the one with contextual effects only, and the model with both effects.

The results are reported in table 7. A careful comparison between model (13) and model (2), model (14) and model (4), and model (15) and model (6) reveals that the estimated coefficients on own characteristics are robust to this specification, but several of the social interaction coefficients are not. In particular, the endogenous effect coefficients increase from 0.302 to 0.495 for the model with endogenous effect only, and change from 0.274 to 0.473 for the model with both effects. The spatial coefficients in the error terms for both models are both around −0.25. Furthermore, some contextual effect coefficients also exhibit great change after the introduction of SAR errors. For instance, the contextual coefficient on black is negative and significant at 5% level in the model with i.i.d. disturbance, and the one on Asian is positive and significant at 5% level. Both of these contextual variables switch sign and become insignificant in the model with SAR errors. The contextual coefficients on Hispanic and mother's education less than high school both change from negative to positive, although neither is significant in either specifications. As for the model without endogenous effects, the spatial coefficient in the error term is significant at 0.271, and the other coefficients do not change much. Therefore, the estimation results, especially the estimated social interaction coefficients, are not robust in this regard. The imposing of the condition $\rho_0 = 0$ results in slight decreases in the likelihood values for all three models.[20]

[20] The results for the model with SAR disturbances and for the model with friendship reciprocity assumption based on the network sample are similar to

Table 7
Results for Models with SAR Disturbances

Explanatory Variable	(13)	(14)	(15)
Endogenous effect	.495***		.473***
	(.008)		(.009)
rho	−.257***	.271***	−.237***
	(.010)	(.024)	(.011)
Contextual effects (peer means):			
Age		−.076***	−.017**
		(.009)	(.008)
Years in school		.005	−.003
		(.006)	(.005)
Male		.010	.094***
		(.011)	(.010)
Black		−.105***	.016
		(.019)	(.016)
Asian		.123***	−.004
		(.024)	(.021)
Hispanic		−.054***	.026
		(.019)	(.017)
Other race		−.069***	−.036*
		(.023)	(.021)
Sport		.084***	.019**
		(.011)	(.010)
Live with both parents		.113***	.045***
		(.013)	(.012)
Mom education less than HS		−.054***	.008
		(.018)	(.017)
Mom education more than HS		.138***	.056***
		(.013)	(.012)
Mom education missing		−.018	−.012
		(.019)	(.018)
Mom's job is professional		.012	.008
		(.015)	(.014)
Mom other jobs		−.051***	−.022*
		(.013)	(.012)
Mom on welfare		−.132**	−.082
		(.064)	(.061)
Mom job missing		−.108***	−.047**
		(.020)	(.019)
Own X:			
Age	−.122***	−.125***	−.119***
	(.004)	(.005)	(.004)
Years in school	.010***	.010***	.010***
	(.003)	(.003)	(.003)
Male	−.143***	−.174***	−.176***
	(.005)	(.006)	(.006)
Black	−.156***	−.169***	−.158***
	(.009)	(.012)	(.012)
Asian	.211***	.227***	.216***
	(.012)	(.013)	(.014)
Hispanic	−.118***	−.132***	−.122***
	(.010)	(.011)	(.011)
Other race	−.073***	−.075***	−.067***
	(.012)	(.012)	(.012)
Sport	.105***	.107***	.099***
	(.005)	(.006)	(.006)
Live with both parents	.126***	.129***	.120***
	(.006)	(.007)	(.007)
Mom education less than HS	−.078***	−.083***	−.076***
	(.010)	(.010)	(.010)

Table 7 ***(Continued)***

Explanatory Variable	(13)	(14)	(15)
Mom education more than HS	.168***	.171***	.159***
	(.007)	(.007)	(.007)
Mom education missing	−.005	−.005	−.004
	(.010)	(.010)	(.010)
Mom's job is professional	.021***	.020**	.018**
	(.008)	(.008)	(.008)
Mom other jobs	−.047***	−.049***	−.045***
	(.007)	(.007)	(.007)
Mom on welfare	−.049	−.048	−.040
	(.031)	(.032)	(.031)
Mom job missing	−.086***	−.087***	−.079***
	(.011)	(.011)	(.011)
Fixed effect	Yes	Yes	Yes
Likelihood value	−71,169	−71,243	−71,070

Note.—The dependent variable is GPA. The sample size is 68,131. Standard errors are in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

V. Conclusion

The estimation of peer effects is plagued by the "reflection problem," the omitted variable bias problem, and limitations of the data. In the linear-in-means model in Manski (1993), it is impossible to separately identify endogenous and contextual social interaction effects. In contrast, the SAR model is free of the "reflection problem," due to the nonlinearity introduced by the variations in the measurements of the peer variables. By incorporating group fixed effects, Lee (2007b) extends the SAR model to better fit in the social interaction setting. This research formulates a generalized version of Lee's (2007b) model to account for asymmetric or individual specific social interactions. By applying the proposed model to the Add Health data, we obtain estimates for peer effects that are less contaminated by confounding factors. Unlike previous studies that have data only at the school or grade level, Add Health offers detailed information about an individual's social network, which enables us to specify peer groups as friendship networks.

The estimation results from the SAR models with and without fixed effects are distinct. The estimation results from the model with and without school-grade fixed effects differ greatly, which indicates the severity of the omitted variable bias and/or the selection bias issues. The results from the full SAR model with both endogenous and contextual effects and the group fixed effects suggest that both endogenous and contextual effects exist in student academic attainment. In particular, the estimated endogenous effect coefficient is 0.274, which is about 34% of a standard deviation in the average GPA, suggesting that a 1 standard deviation

those based on the whole sample, and therefore are not reported. These results are available from the author upon request.

increase in peer average achievement raises own GPA by 0.221 points. Being surrounded by peers who are male, Asian (at 5% level), participate in sports clubs, living with both parents, or have mothers with more than high school education will help improve an individual's GPA. Conversely, having friends who are older, black (at 5% level), of other race (at 5% level), have a mother on welfare (at 10% level), or whose mother's occupation is "other" has a negative effect. This study provides justification for policies relying on the existence of peer effects, such as school voucher programs. The results show some robustness to the sample selection criteria and the friendship reciprocity assumption. However, introducing SAR disturbances into the model greatly inflates the endogenous coefficient, along with a negative interaction coefficient in the error terms.

One limitation of this study is that the social interactions considered are in a group interaction setting only, although we allow for asymmetric and flexible interaction structure within a group. We need to specify ex ante some group like class or school and assume that people only interact with the group members. In real life, social networks are not confined to any "artificial" boundary such as a class or school group. Therefore, it would be desirable to extend the model to a non-group-interaction framework and model the formation of peer groups and social interactions simultaneously. It would also be a promising research to apply the model to analyze other adolescent development outcomes, such as school dropout, drug use, or teenage pregnancy.

References

Aaronson, Daniel. 1998. Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources* 33, no. 4:915–46.

Amato, Paul R. 1993. Children's adjustment to divorce: Theories, hypotheses, and empirical support. *Journal of Marriage and Family* 55: 23–38.

Anselin, Luc. 1988. *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic.

Becker, Gary S. 1996. *Accounting for tastes*. Cambridge, MA: Harvard University Press.

Bee, Helen L., Kathryn E. Barnard, Sandra J. Eyres, Carol A. Gray, Mary A. Hammond, Anita L. Spietz, Charlene Snyder, and Barbara Clark. 1982. Prediction of IQ and language skill from perinatal status, child performance, family characteristics, and mother-infant interaction. *Child Development* 53, no. 5:1134–56.

Benabou, Roland. 1996. Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance. *American Economic Review* 86, no. 3:584–609.

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. *Journal of Human Resources* 34, no. 2:268–93.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics* 150:41–55.

Brock, William A., and Steven N. Durlauf. 2001. Interactions-based models. In *Handbook of econometrics*, ed. James J. Heckman and Edward Leamer. Amsterdam: Elsevier Science.

Brook, Judith S., and Michael D. Newcomb. 1995. Childhood aggression and unconventionality: Impact on later academic achievement, drug use, and workforce involvement. *Journal of Genetic Psychology* 156, no. 4: 393–410.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2005. Peer effects and social networks in education and crime. Discussion Paper no. 5244, CEPR, London.

Case, Anne C., and Lawrence F. Katz. 1991. The company you keep: The effects of family and neighborhood on disadvantaged youths. Working Paper no. 3705, National Bureau of Economic Research, Cambridge, MA.

Coleman, James S., et al. 1966. *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2007. Be as careful of the books you read as of the company you keep: Evidence on peer effects in educational choices. Discussion Paper no. 2833, Institute for the Study of Labor (IZA), Bonn.

Dietz, Robert D. 2002. The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Science Research* 31, no. 4:539–75.

Duncan, Greg J., Johanne Boisjoly, and Kathleen M. Harris. 2001. Sibling, peer, neighbor, and schoolmate correlations as indicators of the importance of context for adolescent development. *Demography* 38, no. 3:437–47.

Durlauf, Steven N. 2003. Neighborhood effects. Working Paper no. 17, University of Wisconsin–Madison, Social System Research Institute.

Epple, Dennis, and Richard E. Romano. 1998. Competition between private and public schools: Vouchers and peer group effects. *American Economic Review* 88, no. 1:33–62.

Erwin, Phil. 1998. *Friendship in childhood and adolescence*. New York: Routledge.

Evans, William N., Wallace E. Oates, and Robert M. Schwab. 1992. Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy* 100, no. 5:966–91.

Fertig, Michael. 2003. Educational production, endogenous peer group formation and class composition—evidence from the PISA 2000 study. Discussion Paper no. 714, Institute for the Study of Labor (IZA), Bonn.

Gaviria, Alejandro, and Steven Raphael. 2001. School based peer effects and juvenile behavior. *Review of Economics and Statistics* 83:257–68.

Graham, Bryan S. 2008. Identifying social interactions through conditional variance restrictions. *Econometrica* 76, no. 3:643–60.

Grodner, Andrew, and Thomas J. Kniesner. 2006. Social interactions in labor supply. *Journal of the European Economic Association* 4, no. 6: 1226–48.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* 18, no. 5:527–44.

Harris, Kathleen M. 2009. The National Longitudinal Study of Adolescent Health (Add Health), Waves I and II, 1994–96; Wave III, 2001–2; Wave IV, 2007–9. Machine-readable data file and documentation. Carolina Population Center, University of North Carolina at Chapel Hill.

Haveman, Robert, and Barbara Wolfe. 1995. The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature* 23:1829–78.

Haynie, Dana L. 2001. Delinquent peers revisited: Does network structure matter? *American Journal of Sociology* 106:1013–57.

———. 2002. Friendship networks and delinquency: The relative nature of peer delinquency. *Journal of Quantitative Criminology* 18, no. 2:99–134.

Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. Characterizing selection bias using experimental data. *Econometrica* 66, no. 5:1017–98.

Heckman, James J., and Jeffrey A. Smith. 1995. Assessing the case for social experiments. *Journal of Economic Perspectives* 9, no. 2:85–110.

Henderson, J. Vernon, Peter Mieszkowski, and Yvon Sauvageau. 1978. Peer group effects and educational production function. *Journal of Public Economics* 10:97–106.

Hoxby, Caroline. 2000. Peer effects in the classroom, learning from gender and race variation. Working Paper no. 7867, National Bureau of Economic Research, Cambridge, MA.

Ioannides, Yannis M., and Jeffrey E. Zabel. 2008. Interactions, neighborhood selection and housing demand. *Journal of Urban Economics* 63, no. 1:229–52.

Jargowsky, Paul A. 1997. *Poverty and place: Ghettos, barrios, and the American city*. New York: Russell Sage.

Jones, Ethel B., and John D. Jackson. 1990. College grades and labor market rewards. *Journal of Human Resources* 25, no. 2:252–66.

Kelejian, Harry H., and Ingmar R. Prucha. 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. *Journal of Real Estate Finance and Economics* 17:99–121.

———. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40:509–33.

Kremer, Michael. 1993. Population growth and technological change: One million B.C. to 1990. *Quarterly Journal of Economics* 108, no. 3:681–716.

Laschever, Ron. 2009. The doughboys network: Social interactions and the employment of World War I veterans. Working paper, Department of Economics, University of Illinois at Urbana-Champaign.

Laursen, Brett. 1993. *Close friendships in adolescence*. San Francisco: Jossey-Bass.

Lee, Lung-fei. 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22, no. 4:307–35.

———. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* 72:1899–1926.

———. 2007a. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137:489–514.

———. 2007b. Identification and estimation of spatial econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140, no. 2:333–74.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin. Forthcoming. Specification and estimation of social interaction models with network structures. *Econometrics Journal.*

Lin, Xu. 2005 Peer effects and student academic achievement: An application of spatial autoregressive model with group unobservables. Unpublished manuscript, Department of Economics, Ohio State University.

Loury, Linda D., and David Garman. 1995. College selectivity and earnings. *Journal of Labor Economics* 13, no. 2:289–308.

Manski, Charles F. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60, no. 3:531–42.

———. 2000. Economic analysis of social interactions. *Journal of Economic Perspectives* 14, no. 3:115–36.

Massey, Douglas S., Andrew B. Gross, and Kumiko Shibuya. 1993. Migration, segregation, and the geographic concentration of poverty. *American Sociological Review* 59, no. 3:425–45.

Maxwell, Kimberly A. 2000. Do friends matter? The role of peer influence on adolescent risk behavior. PhD diss., University of Pennsylvania.

Miguel, Edward, and Michael Kremer. 2004. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72, no. 1:159–217.

Moffitt, Robert A. 2001. Policy interventions, low-level equilibria and social interactions. In *Social dynamics*, ed. Steven N. Durlauf and H. Peyton Young. Cambridge, MA: MIT Press.

Morell, Darlene. 1993. Admission criteria, enrollment management, and the probability of college completion: A validity study. Unpublished manuscript, Student Research and Information, Student Affairs, University of California, San Diego.

Nechyba, Thomas J. 2000. Mobility, targeting, and private-school vouchers. *American Economic Review* 90, no. 1:130–46.

Newcomb, Michael D., and Peter M. Bentler. 1986. Drug use, education aspirations, and work force involvement: The transition from adolescence to young adulthood. *American Journal of Community Psychology* 14, no. 4:302–21.

Plotnick, Robert D., and Saul D. Hoffman. 1999. The effect of neighborhood characteristics on young adult outcomes: Alternative estimates. *Social Science Quarterly* 80:1–18.

Pong, Suet-Ling, and Dong-Beom Ju. 2000. The effects of change in family structure and income on dropping out of middle and high school. *Journal of Family Issues* 21:147–69.

Rivkin, Steven G. 2001. Tiebout sorting, aggregation and the estimation of peer group effects. *Economics of Education Review* 20, no. 3:201–9.

Sacerdote, Bruce. 2001. Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics* 116, no. 2:681–704.

Sun, Yongmin, and Yuanzhang Li. 2001. Marital disruption, parental investment, and children's academic achievement. *Journal of Family Issues* 22:27–62.

Vaquera, Elizabeth, and Grace Kao. 2008. Do you like me as much as I like you? Friendship reciprocity and its effects on school outcomes among adolescents. *Social Science Research* 37:55–72.

Woittiez, Isolde, and Arie Kapteyn. 1998. Social interactions and habit formation in a model of female labour supply. *Journal of Public Economics* 70, no. 2:185–205.

Zimmerman, David J. 2003. Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics* 85, no. 1:9–23.