Quantitative Economics 3 (2012), 421–470 1759-7331/20120421

Estimating spillovers using panel data, with an application to the classroom

Peter Arcidiacono
Duke University

Gigi Foster
University of New South Wales

Natalie Goodpaster
Analysis Group, Inc.

Josh Kinsler
University of Rochester

Obtaining consistent estimates of spillovers in an educational context is hampered by at least two issues: selection into peer groups and peer effects emanating from unobservable characteristics. We develop an algorithm for estimating spillovers using panel data that addresses both of these problems. The key innovation is to allow the spillover to operate through the fixed effects of a student's peers. The only data requirements are multiple outcomes per student and heterogeneity in the peer group over time. We first show that the nonlinear least squares estimate of the spillover parameter is consistent and asymptotically normal for a fixed T. We then provide an iterative estimation algorithm that is easy to implement and converges to the nonlinear least squares solution. Using University of Maryland transcript data, we find statistically significant peer effects on course grades, particularly in courses of a collaborative nature. We compare our method with traditional approaches to the estimation of peer effects, and quantify separately the biases associated with selection and spillovers through peer unobservables.

Keywords. Panel data, fixed effects, peer effects, education.

JEL classification. C18, C23, I21.

Peter Arcidiacono: psarcidi@econ.duke.edu
Gigi Foster: gigi.foster@unsw.edu.au
Natalie Goodpaster: natalie.goodpaster@gmail.com
Josh Kinsler: joshua.kinsler@rochester.edu
We thank Joe Altonji, Pat Bayer, Jane Cooley, Paul Frijters, John Haltiwanger, Caroline Hoxby, Tom Nechyba, Bruce Sacerdote, seminar participants at Duke University, University of Georgia, Yale University, and participants at the 2004 Labour Econometrics Workshop held at the University of Auckland and the NBER's Higher Education Workshop for helpful comments, as well as the University of Maryland administrators who have provided us with the data used in this paper. Shakeeb Khan and Nese Yildiz were particularly helpful. Special thanks go to Chris Giordano and Bill Spann of the Maryland Office of Institutional Research and Planning. We also thank Graham Gower for superior research assistance.

DOI: 10.3982/QE145

1. Introduction

The question of how peers affect student achievement underlies many debates in applied economics. Peer effects are relevant to the estimation of the impact of affirmative action, school quality, and public school improvement initiatives such as school vouchers, and are central to more immediate concerns such as how best to group students to maximize learning.[1] However, despite this wide field of potential relevance, the empirical estimation of spillovers—whether in the education context or elsewhere— is not straightforward.

There are at least two barriers that must be overcome when estimating spillovers on student achievement.[2] The first is the selection problem. When individuals choose their peer groups, high-ability[3] students may sort themselves into peer groups with other high-ability students. With ability only partially observable, positive estimates of peer effects may result even when no peer effects are present because of a positive correlation between the student's unobserved ability and the observed ability of his peers. Researchers have undertaken a variety of estimation strategies to try to overcome the selection problem,[4] but significant empirical problems linger, both because researchers only have access to incomplete measures of ability and because peer effects may operate differently when peers are chosen rather than assigned.

A second barrier is that spillovers may work in part through characteristics or actions that are not observed by the econometrician. The importance of peer effects may be significantly understated if the primary channel through which they operate is unobserved. Peer effects through unobservables in education have received little attention outside of Altonji, Huang, and Taber (2004) and Graham (2008).[5]

We introduce a new algorithm for estimating spillovers using panel data that overcomes both these obstacles. Our key innovation is that the peer effects are captured through a linear combination of individual fixed effects. Utilizing fixed effects to capture the impact of peers is well suited to environments where time-varying peer unobservables do not affect individual choices. This can hold when the outcome of interest is a

[1] Epple, Romano, and Sieg (2003) showed that prices colleges charge differ by ability, suggesting the importance of peer effects.

[2] A third barrier is measurement of the peer group. See Chandrasekhar and Lewis (2011) for a discussion of circumventing measurement error in the peer group.

[3] For ease of exposition, we refer to the bundle of individuals' performance-relevant characteristics as ability.

[4] One set of papers uses proxy variables to break the link between unobserved and peer ability (Arcidiacono and Nicholson (2005), Hanushek, Kain, Markman, and Rivkin (2003), and Betts and Morell (1999)). Another set of papers relies on some form of random assignment (Sacerdote (2001), Zimmerman (2003), Winston and Zimmerman (2003), Foster (2006), Lehrer and Ding (2007), Carrell and Hoekstra (2010), Carrell, Fullerton, and West (2009), Carrell, West, and Malmstrom (2008), and Hoxby (2001)). Finally, researchers have tried to circumvent the endogeneity problem with instrumental variables (Evans, Oates, and Schwab (1992)). See Epple and Romano (2011) for a review of the different approaches.

[5] Graham (2008) required random assignment into classes and because his analysis is cross sectional, must make stronger assumptions on the variance of the idiosyncratic term across classes. Mas and Moretti (2009) estimated spillovers in the workplace through both observables and unobservables. They used a two-stage method that yields inconsistent and downward-biased estimates of the spillover parameters when T is fixed. This is likely to be unimportant in their setting because their data contain workers with many observations over time.

choice and individuals only have expectations of peer choices. The assumption on peer unobservables is also valid when individuals only have partial control over their outcomes, as is the case with test scores and grades, and either have expectations of peer choices or peer choices are made based solely on observables.

Constructing the spillover as a linear combination of individual fixed effects results in a nonlinear optimization problem. Estimating individual unobserved heterogeneity in non-linear panel data models often results in biased estimates of the key parameters of interest—the incidental parameters problem.[6] As N goes to infinity for a fixed T, the estimation error for the fixed effects often does not vanish as the sample size grows, contaminating the estimates of the parameters of interest.[7] We show, however, that the nonlinear least squares estimate of the spillover parameter is consistent and asymptotically normal as $N \to \infty$ with T fixed, even though the fixed effects themselves are not consistent.[8] These consistency arguments also extend to the case where peer effects persist over time.

While nonlinear least squares yields consistent estimates of the spillover parameters, the dimensionality of the problem renders nonlinear least squares infeasible. We develop an iterative algorithm that, under certain conditions, produces the same estimates as nonlinear least squares. The algorithm toggles between estimating the individual fixed effects and the spillover parameters. Each iteration lowers the sum of squared errors, with a fixed point reached at the nonlinear least squares solution to the full problem.

We customize the model for application to peer effects in education, using student-level data from the University of Maryland. Six semesters of transcript data are available, covering the semesters from the spring of 1999 to the fall of 2001. We observe grades for every class each student took over the course of this period as long as the student lived on campus during any one of the six semesters.

We estimate the model separately for each of three types of courses, finding significant peer effects that vary by course type. A 1 standard deviation increase in peer ability yields average returns similar to those from between a 3 percent and an 11 percent of a standard deviation increase in own ability, depending on the course type and specification. The lowest returns are found in math and science, and the highest returns are found in the social sciences.

Our model allows us to quantify selection both within and across course types. Within course types, we compare the amount of selection with respect to observed and unobserved student ability. To arrive at these measures, we decompose each of the estimated student fixed effects, or what we label total ability, into an observed and an unobserved component using typical observed ability measures, such as Scholastic Aptitude

[6]Neyman and Scott (1948) were the first to document the incidental parameters problem.

[7]Hahn and Newey (2004) provided two methods of bias correction: a panel jackknife and an analytical correction. Woutersen (2002) and Fernandez-Val (forthcoming) considered estimators from bias-corrected moment conditions. In a similar vein, Arellano and Hahn (2006) and Bester and Hansen (forthcoming) considered bias-correcting the initial objective function.

[8]Other special cases where the incidental parameters problem does not require a bias correction are Manski (1987), Honore (1992), and Horowitz and Lee (2004).

Test (SAT) scores and high school performance. For all course types, we find greater selection on unobserved ability than observed ability. However, selection is highest when measured using total ability, a reflection of the significant correlation between peer observed and unobserved ability within a section.[9] Because we estimate our model separately for each course type, we can compare student ability in their primary course of study to their ability in other fields, thereby quantifying selection across course types. We find strong evidence of both comparative and absolute advantage. On average, students select course types for which they are best suited. However, math and science students show greater aptitude overall in every course type.

Finally, we compare our peer effect estimates to those that would be obtained using more conventional methods. In particular, we examine separately the two obstacles present in traditional peer effects estimation: selection into peer groups and the effect of peer unobservables. Controlling for selection only, which is what is accomplished using random assignment, we show that the estimated peer effects are lower than the peer effects obtained using our method. This is because random assignment techniques rely on incomplete measures of peer ability. We then reintroduce selection into the model and show that the bias in the peer effect estimate can be either positive or negative when both issues are present. For humanities courses, the peer effect estimate continues to be biased downward since the peer unobservables problem dominates the selection problem. The opposite is true for math and science courses, where the peer effect estimate using conventional methods is more than four times our original estimate. The differences across course types are due in part to the much higher correlation between individual observed ability and peer unobserved ability in math and science relative to the humanities.

The remainder of the paper proceeds as follows. Section 2 presents the baseline model, the identification result, and the solution algorithm. Section 3 extends the model to incorporate correlated effects, endogenous effects, and heterogeneity in peer spillovers. Monte Carlo evidence on the performance of the algorithm is presented in Section 4. Section 5 describes the University of Maryland data and Section 6 presents the results. Section 7 explores selection within and across course types, and Section 8 illustrates the biases associated with traditional peer effect measures. Section 9 concludes.

2. Estimating spillovers with panel data

In this section, we present a model and estimation strategy for measuring achievement spillovers using student fixed effects. The model is constructed keeping in mind that our application is measuring peer effects in college, where we are interested in the interactions that occur within discussion sections in large classes.

We first consider a case where one's outcome depends only on one's own fixed effect and the fixed effects of the other individuals in a predefined peer group. We show that it is possible to obtain consistent estimates of the spillover and that there is a computationally cheap way to obtain the solution. All proofs appear in the Appendix A.

[9] By construction, observed and unobserved ability are orthogonal in the population.

2.1 *Identifying spillovers using panel data*

Our baseline model has individual i's outcome at time t in peer group n, Y_{itn}, depending on his own observed and unobserved characteristics, X_{it} and u_{it}, a linear function of the observed and unobserved characteristics of each of the other students in his peer group, and a transitory error, ε_{itn}. Denote as $M_{tn}+1$ the total number of individuals in peer group n at time t. Each member of peer group n at time t then has M_{tn} peers. Denote as $\mathbb{M}_{tn\sim i}$ the set of individuals (numbering M_{tn}) in peer group n at time t with individual i removed. Our baseline specification can then be written

$$Y_{itn} = X_{it}\beta_1 + u_{it}\beta_2 + \frac{1}{M_{tn}} \sum_{j\in\mathbb{M}_{tn\sim i}} (X_{jt}\gamma_1 + u_{jt}\gamma_2) + \varepsilon_{itn}. \tag{1}$$

In addition to the assumption of linearity, the specification in (1) is restrictive along a number of dimensions. There are no endogenous effects, as peer choices do not enter the outcome equation. There are also no correlated effects, as there are no variables to capture the commonality of the environment faced by all members of student i's time-t peer group. Finally, this specification does not allow for heterogeneity in the susceptibility to peer influence.

While each of these restrictions is relaxed in the next section, even in this special case, estimation is problematic when peer groups are chosen. In particular, there may be correlation between u_{it} and the sum of observed peer characteristics, leading to biased estimates of γ_1. Also, we are not able to capture the peer influence through unobservables, meaning that γ_2 is inestimable without further assumptions. While random assignment can remove the correlation between u_{it} and observed peer characteristics, the inability to capture spillovers through unobservables remains.[10]

We now make an additional assumption: the relevance to outcomes of peer characteristics is proportional to that of own characteristics, meaning that we can write[11]

$$\gamma_1 = \gamma_o\beta_1,$$
$$\gamma_2 = \gamma_o\beta_2.$$

This implies, for example, that if two dimensions of an individual's ability are equally important in their effect on Y_{itn}, then those two dimensions of peer ability are also equally important in determining Y_{itn}. This same assumption is used in Altonji, Huang, and Taber (2004).

Now define

$$\alpha_{ito} = X_{it}\beta_1 + u_{it}\beta_2.$$

[10] Using random assignment to identify the spillover also disregards the possibility that spillovers operate differently in selected versus randomized contexts.

[11] For the remainder of the paper, we designate population parameters with an o subscript (γ_o) and designate estimates of the population parameters with a caret ($\hat{\gamma}$).

We can then rewrite equation (1) as

$$Y_{itn} = \alpha_{ito} + \frac{\gamma_o}{M_{tn}} \sum_{j \in \mathbb{M}_{tn\sim i}} \alpha_{jto} + \varepsilon_{itn}. \tag{2}$$

An individual's outcome is then a function of the individual's own ability at t plus the mean ability of the other students in the peer group at t.

We are then interested in solutions to the nonlinear least squares problem

$$\min_{\alpha,\gamma} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{itn} - \alpha_{it} - \frac{\gamma}{M_{tn}} \sum_{j \in \mathbb{M}_{tn\sim i}} \alpha_{jt} \right)^2. \tag{3}$$

If individual ability varies over time, as it does in the above specification, γ_o is not identified unless multiple observations per student are available in each time period. However, additional structure can be placed on the evolution of the α_o's to ensure that the spillover parameter is identified even if multiple observations per time period are not available. In the following discussion, we investigate the properties of our estimator of γ_o under two common assumptions about how ability evolves:

A1 Static. *Individual ability is assumed fixed over time,* $\alpha_{ito} = \alpha_{io}$. *An individual's outcome is then a function of his own fixed effect plus the mean of the fixed effects of the other students in the peer group.*

A2 Cumulative. *Ability accumulates according to the interactions between the individual and his peers. At the initial time period, ability is given by* α_{io}. *When there is no depreciation, from* $t = 2$ *onward ability accumulates according to*

$$\alpha_{ito} = \alpha_{io} + \sum_{t'=1}^{t-1} \frac{\gamma_o}{M_{t'n}} \sum_{j \in M_{t'n\sim i}} \alpha_{jt'o}. \tag{4}$$

Note that with only one observation per student available in each time period, it would be difficult to allow for more flexible forms of time-varying ability.

The two approaches to the evolution of ability require estimating the same number of parameters—an α_{io} for each individual plus an estimate of the spillover parameter, γ_o. Maintaining the assumptions of linearity and proportionality previously discussed, we prove that the least squares solution to equation (3) is a consistent estimator of γ_o under the following set of assumptions.

Theorem 1. *Let* $\mathcal{N}$ *denote the number of individuals who are observed at least two times and satisfy* $\sum_{j \in \mathbb{M}_{tn\sim i}} (\frac{\alpha_{jto}}{M_{tn}}) \neq \sum_{j \in \mathbb{M}_{t'n\sim i}} (\frac{\alpha_{jto}}{M_{t'n}})$ *for some* t, t'. *Suppose either* A1 *or* A2 *holds. Additionally, suppose the following statements*:

(i) *We have* $E(\varepsilon_{itn}\varepsilon_{jsk}) = 0 \; \forall j \neq i, t \neq s, n \neq k$.

(ii) *We have* $E(\varepsilon_{itn}\alpha_{jto}) = 0 \; \forall i, j, t, n$.

(iii) *We have* $E(\alpha_{ito}^4) < \infty \; \forall i, t, n$.

(iv) *We have* $E(\varepsilon_{itn}) = 0$ *and* $E(\varepsilon_{itn}^4) < \infty \; \forall i, t, n$.

(v) *Either* $E(\varepsilon_{itn}^2|n,t) = E(\varepsilon_{jtn}^2|n,t)\ \forall i,j,t,n$ *or* $\mathrm{Cov}(\varepsilon_{itn}^2, N_i) = 0$, *where* N_i *is the number of observations for individual* i. *Further, under A2,* $E(\varepsilon_{itn}^2|t) = E(\varepsilon_{it'n}^2|t')$.

(vi) *The parameter* γ_o *lies in the interior of a compact parameter space* Γ, *where the largest element of* Γ *is given by* $\overline{\gamma}$. *Furthermore,* $\overline{\gamma} < \underline{M}$, *where* $\underline{M}$ *is the smallest class size.*

If (i)–(vi) *hold, then* $\hat{\gamma}$ *is* $\sqrt{N}$ *consistent and an asymptotically normal estimator of* γ_o *for fixed* T.

While most of the above assumptions are standard, a few nonstandard assumptions deserve closer inspection. Assumption (i) requires that the residuals across any two observations be uncorrelated. Any correlation across outcomes for the same individual is captured by the individual fixed effect, while correlation in outcomes across individuals in the same peer group is entirely captured by the peer effect.[12] Assumption (v) requires that either the residuals within a peer group have equal variance, implying that only heteroskedasticity across peer groups can be accommodated, or, if heteroskedasticity operates at the individual level, it is uncorrelated with the number of times the individual is observed in the data. This assumption is generally applied in virtually all papers in the peer effects literature, as standard errors are typically clustered at the class level. Assumption (v) does, however, have to be strengthened in the accumulation case, as the variance in the error term cannot be correlated with time.

Because the estimates of the individual effects are inconsistent for fixed T, one would expect the estimator of γ_o to be downward biased as a result of measurement error. Indeed, two-step approaches, such as that taken in Mas and Moretti (2009), in which estimates of the individual effects are obtained in a first step and then taken as given in the second-step estimation of the spillover parameter, do suffer from attenuation bias.[13] Similarly, an approach that utilizes the average of the peer average grades over time to measure peer ability also is downward biased as a result of measurement error.[14] The intuition for why measurement error does not lead to attenuation bias in our case follows directly from the structure of the consistency proof. The proof relies on solving for each of the individual effects as a function of the data and γ, and then substituting these functions for the individual effects in (3). The key is that when solving for α_i as a function of γ, we account for the direct effect of α_i on own outcomes and the indirect effect of α_i on all the individuals who happen to be paired with i. The strength of the indirect effect is determined by the spillover parameter, and as the spillover parameter moves, so too do the implied estimates of the individual effects. Accounting for the relationship be-

[12] In the next section, we introduce a procedure for capturing correlated effects that do not work through the peer effect.

[13] In Mas and Moretti (2009), measurement error is likely less of an issue due to the large number of time periods per individual in their data. However, in other settings, particularly those in education, the number of time periods per individual is likely to be small.

[14] Using the average of the peer average grades as a proxy for peer ability induces two types of measurement error. The first source of measurement error is related to the fact that grades are a noisy signal of student ability. The second source of measurement error arises from misspecification of the production function, since there are now extra peer terms included in the average grades. When we estimate the model taking this approach, the estimated spillover effect declines by approximately 25 percent.

tween the estimate of the spillover parameter and the estimates of the individual effects removes the measurement error issue. Minimizing the concentrated objective function with respect to γ alone yields the desired result.

The structure of the proof suggests an obvious estimation strategy: minimize the concentrated least squares problem with respect to γ. However, while concentrating out the α's is useful for proving consistency, the resulting formulas are quite cumbersome and difficult to calculate. Directly solving (3) is also generally not possible because of the dimensionality of the problem. Instead, we consider an iterative estimation strategy that both circumvents the dimensionality problem and yields the same solution as direct maximization. The next section introduces the computational procedure and discusses how it relates to the broader literature regarding estimation of high-dimensional problems.

For the remainder of the paper, we focus on the case where α_{ito} does not vary over time, since our empirical application investigates the existence of peer effects in a college setting—a setting where there is heterogeneity in both the number of courses taken and the course topic. The accumulation of skill is likely to be less important here relative to primary or secondary schooling, since college courses generally do not build directly on one another. However, the methods we describe below can easily be adapted to the accumulation case as well.

2.2 *Computing spillovers with panel data*

Before moving directly to the computation of the spillover model outlined in the previous section, we illustrate how our proposed procedure can ameliorate a somewhat simpler computational problem prominently discussed in the literature. An outstanding problem in applied microeconomics is how to estimate models that contain multiple types of fixed effects where each set of fixed effects is of a large dimension.[15] We begin with this econometric problem, since the iterative method we employ solves the issue of multiple fixed effects en route to estimating spillovers. We focus on two papers in particular—Rivkin, Hanushek, and Kain (2005) and Abowd, Kramarz, and Margolis (1999)—to illustrate the difficulties in estimating large numbers of fixed effects.

Rivkin, Hanushek, and Kain (2005) modeled gains in test scores as a function of the observed characteristics of the students, X_i, and teacher fixed effects, π_{jo}, where i indexes individuals and j indexes teachers.[16] The change in test scores from time $t-1$ to t, given that the individual has teacher j at time t, ΔY, is then modeled as

$$\Delta Y = \beta_o X_i + \pi_{jo} + \varepsilon_{it}. \tag{5}$$

[15] Harris and Sass (2006, 2011) used our method in estimation of models with multiple classes of high-dimensioned fixed effects. Burke and Sass (forthcoming) used our method to estimate peer effects in Florida public schools.

[16] Our model is presented in levels, since we are working with collegiate data, where defining a baseline of achievement is somewhat difficult. However, the model can be applied exactly as written if gains are the outcome of interest. The key differences are that the outcome is now a gain and that the individual fixed effects reflect heterogeneity in ability to improve. Peer effects in this case would also work through an individual's ability to improve. To the extent that gains are employed to eliminate time invariant unobserved heterogeneity, our model can handle this directly by estimating fixed effects at multiple levels.

Note that X_i includes characteristics of the students that do not vary over time. However, X_i may not include the full set of individual characteristics that are relevant for achievement gains, and the omitted variables may be correlated with the π_{jo}'s due to streaming of students and/or systematic selection of certain teachers into classrooms with higher- or lower-ability students. As an alternative, we could estimate the model with both student and teacher fixed effects:

$$\Delta Y = \theta_{io} + \pi_{jo} + \varepsilon_{it}. \tag{6}$$

However, estimating both sets of fixed effects simultaneously would be infeasible given the large number of students and teachers in their data.

Abowd, Kramarz, and Margolis (1999) were interested in modeling wages as a function of both firm and worker fixed effects. The most basic model they were interested in estimating contains just individual and firm-specific effects in a regression of log earnings. A more interesting case occurs when there are tenure effects that vary across firms. For simplicity, assume that the effects of tenure are linear. Labeling X_{ijt} as the amount of tenure individual i has in firm j at time t, the outcome equation is

$$Y_{ijt} = \theta_{io} + \pi_{jo} + \phi_{jo} X_{ijt} + \varepsilon_{ijt}, \tag{7}$$

where ϕ_{jo} is the firm-specific return to tenure. Abowd and Kramarz (1999) recognized that with over 1 million workers and 500,000 firms, they could not estimate the above equation directly. Instead, they considered a number of estimation techniques, none of which results in least squares estimates of the firm and worker fixed effects without imposing additional assumptions on the data generating process.[17]

Our approach yields least squares estimates of both firm and worker effects in a computationally feasible way without imposing any extraneous orthogonality conditions. Estimating the firm–worker model by ordinary least squares (OLS) solves

$$\min_{\theta,\pi,\phi} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{ijt} - \theta_i - \pi_j - \phi_j X_{ijt})^2. \tag{8}$$

Minimizing this function in one step remains infeasible as a result of the large number of firms and workers. Instead, we propose an iterative method that yields OLS estimates of the parameters of interest while circumventing the dimensionality problem. Given starting values for the θ's, the algorithm iterates on two steps; the qth iteration uses the following steps:

Step 1. Conditional on θ^{q-1}, estimate π^q and ϕ^q by OLS.

Step 2. Conditional on π^q and ϕ^q, estimate θ^q by OLS.

The process continues until the parameters converge. Because the sum of squared errors

[17] Abowd, Creecy, and Kramarz (2002) provided one way to recover the exact least squares estimates of the firm and worker effects when both vectors are of a high dimension. Using the code provided on the authors' website, we compared the performance of our estimator to the new estimator in Abowd, Creecy, and Kramarz (2002). With 500,000 firms, 1 million workers, and a linear returns-to-tenure parameter, our algorithm produced the same parameter estimate and reduced the required computational time by 25 percent.

is decreased at each step, we eventually converge to the parameter values that minimize the least squares problem in (8), regardless of which pair of parameters we guess first to start the algorithm. The primary advantage of our method in applications such as those described above is that it is capable of estimating extremely large sets of fixed effects in a reasonable amount of time.

The model becomes slightly more complicated when the outcomes are allowed to depend on functions of the individual effects themselves. The iterative estimation strategy we employ involves toggling between estimating the spillover parameter by OLS and estimating the individual effects. The additional complexity arises in the second step. In the firm–worker example, the qth iteration estimate for θ_i does not depend on the qth iteration estimate of θ_j. However, in the spillover model, i's outcome is a function of α_{io} and α_{jo} for all $j \in \mathbb{M}_{tn}$. This suggests that we need to minimize the conditional likelihood function over all of the α's directly. We are able to avoid this by instead repeatedly updating α_i using the first-order condition from the least squares problem.

Consider the first-order condition of the nonlinear least squares problem with respect to α_i:

$$\begin{aligned} 0 = \sum_{t=1}^{T} \Bigg[&\left(Y_{itn} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j \in \mathbb{M}_{tn\sim i}} \alpha_j \right) \\ &+ \sum_{j \in \mathbb{M}_{tn\sim i}} \frac{\gamma}{M_{tn}} \left(Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k \in \mathbb{M}_{tn\sim j}} \alpha_k \right) \Bigg]. \end{aligned} \tag{9}$$

Solving for α_i yields

$$\begin{aligned} \alpha_i = \sum_t \Bigg[&Y_{itn} - \frac{\gamma}{M_{tn}} \sum_{j \in \mathbb{M}_{tn\sim i}} \alpha_j \\ &+ \sum_{j \in \mathbb{M}_{tn\sim i}} \frac{\gamma}{M_{tn}} \left(Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k \in \mathbb{M}_{tn\sim i\sim j}} \alpha_k \right) \Bigg] \Bigg/ \left(T + \sum_t \frac{\gamma^2}{M_{tn}} \right). \end{aligned} \tag{10}$$

Note that we have extracted the α_i terms from the last term in (9) to derive (10). We establish in Theorem 2 the conditions under which, given any initial set of α's, repeatedly updating the α's using (10) yields a fixed point.

Theorem 2. *Denote $f(\alpha)$ as a function that maps from $\mathcal{R}^N \to \mathcal{R}^N$, where the ith element of $f(\alpha)$ is given by the right-hand side of* (10) *for all $i \in N$. A sufficient condition for $f(\alpha)$ to be a contraction mapping is that the maximum value of γ is less than* 0.4.

The restriction on the maximum value of γ is needed to ensure that the feedback effects are not too strong. With Theorem 2 giving a solution method for the α's conditional on the γ's, our algorithm iterates on estimating the α's using $f(\alpha)$ (taking the γ's as given) and then estimating the γ's taking the α's as given. Each of these two steps lowers the sum of squared errors and, analogous to the estimator in Section 2, converges to the nonlinear least squares solution. In practice, we have found that the algorithm performs substantially faster if the α's are only updated until the sum of squared errors falls before moving on to reestimating γ.[18] To summarize, the algorithm is started with an initial guess for the α's and iterates on two steps until convergence; the qth iteration is given by the following steps.[19]

Step 1. Conditional on α^{q-1}, estimate γ^q by OLS.

Step 2. Conditional on γ^q, update α^q according to (10).

3. Model extensions

The baseline model makes a number of simplifying assumptions regarding the channel through which the spillover operates, the shared group environment, and the form of the spillover effect. The following sections discuss extensions of the model to address these complications.

3.1 *Endogenous effects*

Up to this point, we have ignored how individual and peer choices may affect outcomes: endogenous effects.[20] The peer effects literature that allows for endogenous effects can be broken into two classes of models. In the first class, the outcome of interest is itself a choice, and this choice is directly affected by the actual or expected choices of an individual's peers. In the second class of models, the outcome of interest is not completely within the individual's control. However, choices by both the individual and the individual's peers directly affect the outcome. In this case, it is own effort and the effort that other individuals exert that affect own outcomes, but others' outcomes per se do not appear in the own-outcome equation. Cooley (2009a) showed that identification is much more complicated in this second class of models.[21]

In Appendix B, we consider the complications introduced by endogenous effects by setting up a structural representation of each of these classes of models and showing what our estimator is able to recover using reduced-form estimation in each case. The key result is that in our empirical application, as well as a wide variety of endogenous-effects settings, the fixed-effects-based approach itself does not restrict one's ability to separately identify the various peer effect channels when compared with the standard observables-based approaches. However, one setting where the fixed effects approach does not consistently identify the reduced-form peer effect occurs when individuals have complete control over their own outcomes and peer outcomes enter directly into

[18] For most iterations of our models, updating the α's just once led to a decrease in the sum of squared errors.

[19] Note that since the model is nonlinear, local minima may be a possibility. We have not run into this issue in practice, either in the Monte Carlo exercises or when using the real data.

[20] An emerging literature examines identification and estimation of endogenous and exogenous peer effects in network models. See Bramoulle, Djebbari, and Fortin (2009), Giorgi, Pelizzari, and Redaellil (2007), and Lin (2010).

[21] See Durlauf and Ioannides (2001) for a review of the identification problems in these models as well as the different approaches to overcoming them.

the choice equation. An example of this setting is smoking choices, where peer smoking behavior is directly observed. The threat to identification in this case, namely correlation in the reduced-form errors within a peer group, is not present in our application, since own grades are not fully controlled and peer grades are unobserved.[22]

3.2 *Correlated effects*

We next discuss an extension to our baseline model in which common shocks—correlated effects—influence outcomes. If each individual peer group is exposed to a different environment, it is impossible to separate the correlated effects from the exogenous effects without further parameterizations. A restriction that is easily imposed for correlated effects in our context is a component to one's grade that is course-specific. This is important as grade inflation and where the curve lies affect one's final grade irrespective of own and peer ability. Denote individual i's outcome at time t, in peer group n, and course c as Y_{itnc}, and denote the set of students in course c at time t by $\mathbb{M}_{tc}$. Placing course fixed effects into the achievement equation yields

$$Y_{itnc} = \alpha_{io} + \frac{\gamma_o}{M_{tn}} \sum_{j \in \mathbb{M}_{tn\sim i}} \alpha_{jo} + \delta_{tco} + \varepsilon_{itnc}. \tag{11}$$

The δ_{tco}'s are then the course fixed effects, used to capture correlated effects, since all peer groups are formed within courses. The nonlinear least squares problem we are now interested in solving takes the form[23]

$$\min_{\alpha,\gamma,\delta} \sum_{i}^{N} \sum_{t=1}^{T} \left(Y_{itnc} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j \in \mathbb{M}_{tn\sim i}} \alpha_j - \delta_{tc} \right)^2. \tag{12}$$

The estimation strategy is identical to the baseline model, with one additional step for updating the course fixed effects. The updating equation for the course fixed effects is derived from the first-order condition of equation (12) with respect to δ. Because each outcome is associated with only one δ and because δ always enters linearly, the updating equation boils down to a simple average.[24] Adding other types of fixed effects simply requires calculating an additional average.

3.3 *Heterogeneous effects*

The assumption that each student is affected in the same manner by their classmates is restrictive and, as pointed out by Hoxby and Weignarth (2005), not particularly interesting from a policy perspective. In particular, the linear-in-means model implies that,

[22] We do not pursue the endogenous effects extension further, since the conditions necessary to separately identify exogenous and endogenous effects are not satisfied in our empirical application. As a result, the estimates of our model can be interpreted as a combination of both exogenous and endogenous effects.

[23] While the consistency of $\hat{\gamma}$ is unchanged regardless of whether we include other time-varying regressors, it is particularly clear here since we can rewrite equation (12) without δ_{tc} by demeaning the dependent variable at the course level.

[24] Note that the updating equations for α and γ need to be altered to account for the presence of the course fixed effects.

in terms of grades, any winners from reshuffling peers are perfectly balanced by those who lose from the reshuffling. We now relax this assumption by extending our spillover framework to allow for either heterogeneity in the *response* to peers or heterogeneity in the *influence* of peers.

The first extension allows the effect of peers to vary with an individual's own characteristics, a model we refer to as heterogeneity in responsiveness to peers. A simple example would be if female students are influenced more by peer ability than male students. We can express a spillover model that incorporates heterogeneity in the responsiveness to peers as

$$Y_{itn} = \alpha_{io} + \frac{\sum_{j \in \mathbb{M}_{tn \sim i}} \alpha_{jo}}{M_{tn}} (X_i * \gamma_o) + \delta_{tco} + \varepsilon_{itn},$$

where X_i denotes the observable characteristics of individual i.

The second model, which we refer to as heterogeneity in peer influence, allows the strength of the peer effect to depend on the interaction between own and peer characteristics. For example, male students may be affected more by other male students than they are by female students. For ease of exposition, assume that students can be assigned to one of two groups, such as male or female, or black or white. Heterogeneity in peer influence can then be easily incorporated as

$$Y_{itn} = \alpha_{io} + \frac{1}{M_{tn}} \left(\gamma_{1o} \sum_{j \in \mathbb{M}_{tng \sim i}} \alpha_{jo} + \gamma_{2o} \sum_{j \in \mathbb{M}_{tng'}} \alpha_{jo} \right) + \delta_{tco} + \varepsilon_{itn},$$

where $\mathbb{M}_{tng \sim i}$ is the set of all students in peer group n at time t who are in the same group as i, excluding individual i, and $\mathbb{M}_{tng'}$ are all individuals in peer group n at time t who are not in the same group as i.

The steps required to estimate either of the above models are identical to those outlined in the previous section, although each step becomes slightly more complicated. Rather than estimating a single γ by OLS in Step 1, multiple γ's need to be estimated. Computationally, Step 2 is also more cumbersome, since the first-order condition for α_i likely depends on i's type and the type of peers with whom i is grouped.

4. Monte Carlo simulations

To investigate the properties of our iterative estimator, we now run simulations using different assumptions about the composition of and selection into the peer groups. To mirror our empirical application, we focus on the case where peer effects are transitory, although the performance of our estimators is quite good in the value added case as well.[25] In each setting, the model is simulated using 10,000 students. We simulate the model 100 times under various states of the world constructed by varying four dimensions of the problem:

[25]Results are available from the authors upon request.

1. *Observations per student.* The number of outcomes observed per student varies across simulations between 2, 5, and 10. Five observations is the maximum number a researcher may have when analyzing grade school or high school test score data, and 10 observations is more likely when analyzing grades achieved in university-level courses. More observations per student implies more accurate measurement of the α_o's.

2. *Students per peer group.* The number of students per peer group varies across simulations between 2 and 15. Two observations is the minimum number required to identify a spillover in this type of model; 15 students per peer group is in the range of what might be observed in typical classroom-based data sets.

3. *Selection into classes.* To show that our estimator solves the selection problem, we simulate the model under alternative assignment rules. Under random assignment, the average standard deviation of the α_o's within a peer group equals the standard deviation of α_o in the population. We also simulate the model with selection such that the average standard deviation of the α_o's within a peer group is 75 percent of the population standard deviation.

4. *Transitory component.* The noisier is the outcome measure, the noisier are the estimates of the α_o's. The distribution of the α_o's is set at $N(0,1)$. The ε's are distributed with mean zero and standard deviation (σ_ε) equal to 1.15 or 1.95. We also investigate the impact of heteroskedasticity at the peer group level.

The common group-level shock used to model the correlated effect is not statistically associated with the abilities of students in the classes. However, students are sorted into classes based on ability. Thus, the average standard deviation of abilities within a class is smaller than the standard deviation of abilities in the population.

Table 1 documents the model's performance when the true value of γ_o is 0.15. Regardless of assignment procedure or section size, $\hat{\gamma}$ is centered around the truth. However, two interesting patterns emerge in the estimates and standard errors of $\hat{\gamma}$. First,

Table 1. Baseline model, $\gamma_o = 0.15$.[a]

Obs. per Student	Peer Group Size		Random Assignment $\sigma_\varepsilon = 1.95$	Random Assignment $\sigma_\varepsilon = 1.15$	Selection $\sigma_\varepsilon = 1.95$	Selection $\sigma_\varepsilon = 1.15$
2	2	$\hat{\gamma}$	0.151	0.151	0.140	0.151
			(0.034)	(0.016)	(0.060)	(0.024)
		R^2	0.706	0.822	0.713	0.828
5	10	$\hat{\gamma}$	0.150	0.150	0.146	0.149
			(0.041)	(0.021)	(0.059)	(0.033)
		R^2	0.482	0.686	0.494	0.697
10	10	$\hat{\gamma}$	0.150	0.148	0.148	0.152
			(0.025)	(0.012)	(0.036)	(0.018)
		R^2	0.415	0.644	0.429	0.659

[a]The R^2 values reported in this table pertain to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student so as to manipulate the amount of variation in performance that is explained by the ability measure. Parameter values are averages over 100 simulations on a population of 10,000 students.

Table 2. Varying class size and heteroskedasticity, $\gamma_o = 0.15$.[a]

Obs. per Student	Peer Group Size		Random Assignment		Selection	
			σ_ε	σ_ε	σ_ε	σ_ε
5	$U[5,15]$		1.95	1.15	1.95	1.15
		$\hat{\gamma}$	0.158	0.150	0.160	0.145
			(0.041)	(0.022)	(0.059)	(0.030)
		R^2	0.484	0.686	0.502	0.704
5	$U[5,15]$		$N(1.95, 0.09)$	$N(1.15, 0.09)$	$N(1.95, 0.09)$	$N(1.15, 0.09)$
		$\hat{\gamma}$	0.148	0.151	0.146	0.149
			(0.037)	(0.022)	(0.059)	(0.026)
		R^2	0.480	0.67	0.497	0.692
5	$U[5,15]$		$N(0.95 + \frac{\text{size}}{10}, 0.09)$	$N(0.15 + \frac{\text{size}}{10}, 0.09)$	$N(0.95 + \frac{\text{size}}{10}, 0.09)$	$N(0.15 + \frac{\text{size}}{10}, 0.09)$
		$\hat{\gamma}$	0.153	0.149	0.150	0.149
			(0.049)	(0.029)	(0.061)	(0.027)
		R^2	0.459	0.633	0.474	0.649

[a]The R^2 values reported in this table pertain to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student so as to manipulate the amount of variation in performance that is explained by the ability measure. For the bottom two-thirds of the table, the standard deviation of the random error varies across peer groups according to the distributions listed. Parameter values are averages over 100 simulations on a population of 10,000 students.

$\hat{\gamma}$ is more precisely measured when students are randomly assigned to classes. Selection in this case can be thought of as occurring at two levels: the classroom level and the teacher level. Teacher-level sorting refers to the idea that teachers are often assigned students of similar ability over time. These results reflect the fact that sorting at the teacher level confounds the estimate of the correlated effect and reduces the precision of the classroom-level peer effect estimate. In fact, if selection occurred only at the classroom level, the peer effect estimates would be more precise than in the random assignment case (ceteris paribus), since there would be greater variation in peer ability. Second, as the peer group size increases, the precision of $\hat{\gamma}$ decreases.[26] This is again related to the variation in peer ability across classes. With smaller class sizes, other things equal, there is greater variation in peer ability across classes, which yields more precise estimates of the spillover.

Many applications involving peer effects involve possible heteroskedasticity at the class level. Table 2 shows the performance of the algorithm in the presence of heteroskedasticity, including the case when heteroskedasticity is a function of the size of the class. The first panel of results illustrates that heterogenous class size does not affect the performance of the peer effects estimator: $\hat{\gamma}$ remains centered around 0.15 as

[26]This can be seen in Table 1, since as the number of observations per student increases, we should naturally see an increase in precision. Yet we do not see that increase between the first and second rows, because peer group size increases as well, driving standard errors up. We only see the increase in precision between the second and third rows, when the number of observations per student rises as peer group size is held constant. The negative association of peer group size and precision, as well as all other relationships discussed here, have also been verified in numerous additional Monte Carlo exercises; results are available upon request from the authors.

Table 3. Heterogenous gamma models.[a]

Model		Random Assignment	Selection
Heterogeneity in responsiveness to peers	$\gamma_{1o} = 0.15$	0.151	0.146
		(0.025)	(0.029)
	$\gamma_{2o} = 0.1$	0.100	0.094
		(0.033)	(0.032)
	R^2	0.683	0.699
Heterogeneity in peer influence	$\gamma_{1o} = 0.15$	0.150	0.151
		(0.029)	(0.037)
	$\gamma_{2o} = 0.1$	0.102	0.098
		(0.030)	(0.039)
	R^2	0.687	0.684

[a]The R^2 values reported in this table pertain to the regression of grades onto the constructed fixed effect values. Parameter values are averages over 100 simulations on a population of 10,000 students. Each student is observed 5 times with a total group size of 10 students.

peer group size varies uniformly between 5 and 15. The second and third panels add heteroskedasticity to the heterogenous class size case. In the second panel, σ_ε is drawn from a Normal distribution with a mean of 1.15 or 1.95 and a standard deviation of 0.3. It is assumed that each peer group draws from the same distribution. In the third panel, the mean of the distribution of σ_ε shifts according to the size of the peer group. The peer effects estimator continues to perform quite well regardless of the type of heteroskedasticity. Across the various distributions of σ_ε and sorting scenarios, we estimate a peer effect centered on the truth.

As noted previously, the linear-in-means model may not be the most interesting case from the policy maker's perspective. We suggested two extensions to the baseline framework that would relax this assumption. Table 3 illustrates the performance of the heterogeneous effect models, where the basic structure of the Monte Carlo experiments is kept intact. In each case, we assume students are characterized by one binary variable. The results indicate that the estimation framework previously outlined is amenable to heterogenous peer effects.

5. Data

With the model producing consistent estimates of the spillover parameter and performing well in our Monte Carlo simulations, we now turn to the data used in estimation. The administrative data set used in this paper covers all undergraduates observed residing in University of Maryland on-campus housing during any of the following six academic semesters: Spring 1999 (S99), Fall 1999 (F99), Spring 2000 (S00), Fall 2000 (F00), Spring 2001 (S01), or Fall 2001 (F01). The data set includes students living off-campus in a given semester as long as they were observed living on-campus during at least one of the six semesters. Ninety percent of University of Maryland entering freshmen live on campus in their first semester,[27] so the data set includes at least 90 percent of the University

[27]This number is taken from publicly available statistics posted on the university's web page.

of Maryland undergraduate population who began study sometime in the six-semester period.[28] A "section" is a subset of students from an entire course that meets together formally at least once a week. In these smaller groups, greater communication and interaction is expected of students. Our section-level spillover captures how the influence of the ability of other students in the same section impacts on own course grade.

To generate the student–section-level sample, we first placed two major restrictions on the data set: students had to have valid A–F grade information for the given section and they could not be the only student observed in the section that semester.[29] Students who withdrew from a course, audited, or received a nonletter grade (such as pass) were excluded from the sample due to concerns that they might not have been present during sections and classes.[30] Dropping students with no letter grade reduced the sample from 351,940 to 324,181. We then deleted all observations on sections that were not in one of three well defined academic subgroups: (i) humanities (86,844 observations), (ii) social sciences (77,312 observations), and (iii) hard sciences and mathematics (82,675 observations).[31] This left a combined sample of 246,831 student–section observations, representing 18,511 individual students. Sample sizes are provided in Table 4.

Table 4. Sample sizes.[a]

	S99	F99	S00	F00	S01	F01	Total
1. Student–sections	27,900	37,231	37,109	45,991	45,054	53,546	246,831
2. Students	7126	9646	9458	11,760	11,393	13,662	63,045
3. Unique sections	3095	3388	3408	3628	3543	3754	20,816
4. Unique courses	1030	1079	1172	1189	1246	1252	6968
5. Single-section courses	632	682	733	752	778	795	4372
6. Student–sections (only multisection courses)	23,206	31,627	30,599	38,056	36,260	43,853	203,601

[a]Figures represent the data set after applying the restrictions noted in the text. The unrestricted data set contained 351,940 student–section observations. Rows 3 and 4 show the total number of unique sections and courses, respectively, in which anyone in the sample during the given semester was observed.

[28]There is a less complete representation for upperclassmen, some of whom entered before our observation period and may not have lived on campus during the period. However, our identification of peer effects comes from large, multisection courses in which freshmen predominate. In tests of whether classes that were underrepresented had lower estimated peer effects than those with a complete representation, we found no meaningful differences.

[29]Numeric grade equivalents were assigned as follows: A = 4, B = 3, C = 2, D = 1, and F = 0.

[30]If two separate grades were recorded for the student for a given section, the highest grade was used. We dropped students who did not receive a final grade in a given course, which assumes that these students did not affect the outcomes of their peers. We could have treated those who attrit as full members of the class provided we observed a grade for these individuals in another course. The other course would then pin down the individual's fixed effect.

[31]Excluded courses include those that are generally more vocationally oriented, but very diverse; for example, journalism, nutrition and food science, landscape architecture, and library science. Because our model estimates a homogeneous underlying ability for each course type, we did not include these courses in a separate category due to our concern that the underlying ability necessary to succeed in them is not sufficiently homogeneous across the category.

Finally, while our method does not require the presence of observable characteristics about individuals, the data set to which we apply it does offer an array of observable measures about each student. We examine later in the paper how characteristics such as SAT scores correlate with our estimates of student ability.

6. Estimates of classroom spillovers

We now turn to our model specifications and estimates. We first describe and estimate a model that restricts the peer effect such that the spillover depends only on the mean ability in the section. The second specification allows the size of the spillover to depend on one's own characteristics. For example, those who have high SAT verbal scores may receive higher benefits from their peers than those who have low SAT verbal scores. With the results of the two models in hand, we then show how predictable ability is given observable measures such as SAT scores, high school grade point average, and demographics.

6.1 *Homogeneous gamma model*

With n, c, and t indexing sections, courses, and semesters, we have the same specification as in equation (11) except that now, with the number of individuals in each section varying, we restrict the spillover to depend on the mean fixed effect of the other individuals in the same section of a course:

$$Y_{itnc} = \alpha_{io} + \frac{\gamma_o}{M_{tn}} \sum_{j \in \mathbb{M}_{tn\sim i}} \alpha_{jo} + \delta_{tco} + \varepsilon_{itnc}. \tag{13}$$

Because grades are assigned at the course level, there is a relationship between students who share a course but are not in the same section that cannot be captured by the section peer effect, γ_o. We might expect, for example, that if the course is graded on a curve and the entire class is extremely able, a mediocre student's grade may suffer. By including fixed effects at the course level, the δ_{tco}'s, we can make the outcome measure comparable across classes.

Consistent with the data section, we split courses into three types: humanities, social sciences, and math and science. A student's performance in each type of course will differ according to the particular student's strengths and weaknesses. Therefore, instead of encapsulating all the attributes of a student into one ability measure, we allow students to have separate ability measures for each course type in which they are enrolled. As noted above, all courses used in our analysis are classified as belonging to one of the following course types: humanities, social sciences, or math and science. We estimate an independent ability measure for each type of course for each student, conditional on the student's enrollment in at least one class within that course type.[32] Another supporting rationale for the empirical division into course types is that the amount of interaction, and therefore the size of the peer effect, may differ by course type. The algorithm is then

[32]Information as to which courses were assigned to which course types is available from the authors upon request.

run separately for each type of course, yielding three sets of peer and class effects estimates, as well as separate student ability measures for each course type taken.[33]

Table 5 shows the results from estimating equation (13) for each of the three types of courses. Standard errors are calculated using a wild bootstrap procedure.[34] The results indicate positive and significant section peer effects for all course types. The magnitudes of the section-level peer effects suggest that peer effects are most important in the social sciences and least important in math and science. This pattern may reflect the amount of collaborative work required in each course type as well as the differing amounts of discussion that occur in the sections.

To understand the importance of peer ability relative to own ability, we need to take into account the differences in variation of peer and own ability. There is likely to be less variation in peer ability than in own ability, as peer ability averages over a cross section of students, leading to some heterogeneity canceling out. The second and third columns of Table 6 show the standard deviation of mean peer ability and the standard deviation of individual ability, respectively. The fourth column then shows the fraction of a standard deviation of own ability that is equivalent, in terms of its effect on grades, to a 1 standard deviation increase in peer ability. This is calculated by dividing the standard deviation

Table 5. Peer effects results by course type: homogeneous gamma model.[a]

	Humanities	Soc. Sci.	Math/Sci.
Section peer ability	0.1613	0.1960	0.0483
	(0.016)	(0.019)	(0.013)
N	86,844	77,312	82,675
R^2	0.6373	0.6321	0.6861

[a]The dependent variable is the grade in the class. Class fixed effects are estimated in all specifications. Standard errors are obtained using a wild bootstrap.

Table 6. Standard deviations of estimated ability and marginal effects: homogeneous gamma model.[a]

Course Type	Section SD	Population SD	Marginal Effect Ratio
Humanities	0.2823	0.6804	0.0669
Social science	0.3103	0.7125	0.0853
Math and science	0.5952	0.9498	0.0302

[a]Section SD is the standard deviation of average peer ability across the sample of student–section observations of the given course type. Population SD is the standard deviation of ability across the sample of student–section observations in courses of the given course type. These calculations are both based on the fixed effects estimated by the homogeneous gamma model. Marginal Effect Ratio shows the ratio of the effect on grades from a 1 standard deviation increase in peer ability to the effect on grades from a 1 standard deviation increase in own ability.

[33]Note that while classes with only one section do not help in the estimation of the spillover parameter directly, these classes are still useful in pinning down the student fixed effects.

[34]The wild bootstrap is advantageous in this setting since it allows for heteroskedasticity of an unknown form and does not require any resampling. For additional details on the wild bootstrap, see Davidson and MacKinnon (2006).

of mean peer ability by the standard deviation of individual ability and multiplying this number by the estimated γ.

The gap evident in the raw marginal effects between math and science and the other course types is somewhat mitigated because there is relatively more heterogeneity in peer ability in math and science courses than in humanities or social science courses. A 1 standard deviation increase in peer ability is shown to be equivalent to a maximum of 9 percent of the effect of a 1 standard deviation increase in individual ability in the social sciences, and to a minimum of 3 percent of the effect of a 1 standard deviation increase in individual ability in math and science courses.

6.2 *Heterogeneous gamma model*

Table 7 shows the results of a peer effect model that allows for heterogeneity in the response to peers.[35] Response to peer ability is allowed to vary according to an individual's gender, race, and SAT scores. The qualitative results for humanities and social sciences are the same. Relative to white males, Asians see less of a return to peer ability, while females and other nonwhite students see higher returns. Both SAT math and SAT verbal scores are associated with higher returns to peer ability. While Asians in math and science again see lower returns to peer ability, females and other nonwhites also see lower returns relative to their white male counterparts. The interaction of the peer effect with SAT verbal score is once again positive, but the sign on the SAT math interaction is now negative.

The differences in the SAT interactions across fields suggest that two competing forces may be at play. First, those who have higher test scores may have skills that make

Table 7. Peer effects results by course type: heterogeneous gamma model.[a]

Section Peer Ability	Humanities	Soc. Sci.	Math/Sci.
Overall	0.2058	0.2227	0.0940
	(0.020)	(0.025)	(0.018)
Female	0.0970	0.0584	−0.0517
	(0.019)	(0.021)	(0.018)
Asian	−0.0098	−0.0347	−0.0346
	(0.035)	(0.030)	(0.023)
Other nonwhite	0.0375	0.0252	−0.0420
	(0.028)	(0.031)	(0.025)
SAT math	0.0410	0.0507	−0.0560
	(0.012)	(0.016)	(0.011)
SAT verbal	0.0222	0.0147	0.0635
	(0.011)	(0.014)	(0.009)
N	86,844	77,312	82,675
R^2	0.6376	0.6323	0.6864

[a]The dependent variable is the grade in the class. Class fixed effects are estimated in all specifications. Standard errors are obtained using a wild bootstrap.

[35]We also allowed for peer effects to be stronger for those of similar races and genders, with little change in the results. Standard errors were obtained using the wild bootstrap.

Table 8. Standard deviations of estimated ability and marginal effects: heterogeneous gamma model.[a]

Course Type	Section SD	Population SD	Avg. Marginal Effect	Marginal Effect Ratio
Humanities	0.2383	0.6368	0.2595	0.0970
			(0.0622)	
Social science	0.2849	0.6747	0.2480	0.1048
			(0.0554)	
Math and science	0.6028	0.9660	0.0555	0.0347
			(0.0636)	

[a]Section SD is the standard deviation of average peer ability across the sample of student–section observations of the given course type. Population SD is the standard deviation of ability across the sample of student–section observations in courses of the given course type. These calculations are both based on the fixed effects estimated by the heterogeneous gamma model. The fourth column shows the marginal effect on grades from a 1-point increase in peer ability. Marginal Effect Ratio shows the ratio of the effect on grades from a 1 standard deviation increase in peer ability to the effect on grades from a 1 standard deviation increase in own ability. The numbers in parentheses are standard deviations of the marginal effects of a 1-point increase in peer ability that are estimated to occur in the sample.

them better able to benefit from their peers. Working against this, however, is that there is more scope for students to benefit the lower they are in the ability distribution. SAT verbal and math scores may not be highly correlated with the ability to perform well in the humanities and social sciences, implying that the first effect dominates in these course types. However, the SAT math score may be highly correlated with the ability to perform well in math and science classes, leading to the second effect dominating.

Averaging across all individuals within a course type, the relative magnitude of the peer effects is unchanged from the homogeneous gamma model. Peer ability is most important in social science courses and least important in math and science courses. However, the overall magnitude of the peer effects is significantly higher, increasing by an average of over 40 percent across course type. Relative to a 1 standard deviation increase in own ability, the effects of a 1 standard deviation increase in peer ability are also higher in the heterogeneous gamma model as shown in the final column of Table 8. The ratio of the effects of a 1 standard deviation increase in mean peer ability to a 1 standard deviation increase in own ability range from a low of 3.5 percent for math and science to a high of 10.5 percent in the social sciences.

6.3 *Analysis of ability*

Next, we explore the extent to which the fixed effects from our iterative algorithm are predictable using the observed proxies for ability that are consistently used in related literature, and the extent to which the fixed effects estimated using the two methods differ. To facilitate this comparison, we use SAT scores, high school performance, and a host of other observable student attributes as regressors to construct a conglomerate observable measure of ability. This approach is analogous to the creation of an academic index (as employed by Sacerdote (2001)). For each course type, we regress our estimated student fixed effects on an array of previous performance measures and demographics. These results are presented Table 9.

TABLE 9. Regression of fixed effects on observed ability.[a]

	Homogeneous Gamma Model			Heterogenous Gamma Model		
	Hum.	Soc. Sci.	Math/Sci.	Hum.	Soc. Sci.	Math/Sci.
SAT math	0.00	0.11	0.36	−0.20	−0.04	0.51
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SAT verbal	0.08	0.11	0.00	−0.04	0.06	−0.19
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
HS GPA	0.49	0.49	0.73	0.49	0.49	0.72
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Honors	0.14	0.15	0.17	0.14	0.15	0.17
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Sports	−0.07	−0.14	0.01	−0.08	−0.14	0.00
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
In state	0.05	0.08	0.12	0.05	0.08	0.12
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
N	17,332	15,264	16,077	17,332	15,264	16,077
R^2	0.22	0.24	0.34	0.13	0.17	0.36
Race/gender dummies	Y	Y	Y	Y	Y	Y

[a]The dependent variable in the second through fourth columns is the student-level fixed effects estimated in the homogeneous gamma model; the dependent variable in the last three columns is the student-level fixed effects estimated in the heterogeneous gamma model. All regressions also include a female dummy variable and dummy variables for Black, Hispanic, Asian, and American Indian. Standard errors are robust to heteroskedasticity.

The second through fourth of Table 9 show results when we use the fixed effects estimated in our homogeneous gamma model as the dependent variable. The statistical significance and magnitudes of the coefficients on SAT math and SAT verbal scores vary across the four different course types in predictable ways. For example, SAT math scores are insignificant when explaining ability in humanities courses, but are a better proxy for math and science ability. The opposite is true for SAT verbal scores, with higher SAT verbal scores associated with higher ability in the humanities, but uncorrelated with ability in math and science.

The last three columns in Table 9 show the corresponding results for the heterogeneous gamma model. Recalling the results found in Table 7 regarding the positive association of SAT verbal scores with stronger peer effects across all course types, it is unsurprising that the coefficient on own SAT verbal score is smaller and even negative for some course types when predicting own ability.[36] Combining these findings with the results presented here suggests that SAT verbal scores have very little to do with ability in the absolute, but rather reflect how capable an individual is at extracting rents from others.

The R^2 for these regressions ranges from 0.13 to 0.36, depending on the course type and whether we use the homogeneous or heterogeneous gamma model to generate the individual fixed effects. That these observable characteristics only explain a small portion of our estimated ability measures suggests the possibility of large biases associated

[36]Previous works such as Arcidiacono (2004), Arcidiacono and Vigdor (forthcoming), and Arcidiacono, Cooley, and Hussey (2008) have all found no returns to verbal test scores in the labor market.

with the unobserved ability problem when following a selection-on-observables, or random assignment, approach. At the same time, because our estimated fixed effects are noisy estimates of unobserved ability, the R^2 of these regressions is biased downward. As a result, we likely overstate the role of unobserved ability.

7. Quantifying selection

In this section we quantify how much selection is taking place within course types with respect to both observed and unobserved ability.[37] For ease of exposition, we refer to the predicted values of the regressions in Table 9 as observed ability and to the residuals of those regressions as unobserved ability. Thus, by construction, observed and unobserved ability are uncorrelated at the individual level. However, unobserved individual ability and observed peer ability will be correlated if students sort by total ability. By decomposing ability into its observed and unobserved components, we can calculate the correlation between unobserved individual ability and observed peer ability, which is the crux of the selection problem. We also examine selection across course types. Because we estimate separate abilities for each course type, we are able to determine whether students choose to take more courses in areas where they are comparatively more able.

7.1 *Selection within course types*

Table 10 provides information by course type on the selection evident with respect to both observed and unobserved ability. We use the underlying ability as estimated by the homogeneous gamma model and the heterogeneous gamma model, as well as the observed and unobserved portions of this ability. The first row for every course type shows the section-size-weighted average of the sectionwide standard deviation of the variable in question, across all sections in the particular course type; the second row for every course type shows the simple standard deviation of the variable in question across the sample of students taking courses of the given course type. The third row provides the ratio of the two. The smaller are the numbers in the third row, the tighter is the distribution of the variable within sections relative to the unsorted distribution and, therefore, the more selection is evident with respect to that variable.

For all three course types, there is more selection on unobserved ability than on observed ability. In the social sciences and, in particular, for the humanities, there is more selection on the estimated α's as a whole than on either observed ability or unobserved ability separately, with the highest levels of selection found in math and science. These patterns are driven by the correlation between peer observed and unobserved ability. For the homogeneous gamma specification, the correlation coefficients between peer observed and unobserved ability are 0.03, 0.07, and 0.35 in the humanities, social sciences, and math and science, respectively. The selection on the estimated α's in math and science is particularly strong relative to selection on either observed or unobserved

[37] See Altonji, Elder, and Taber (2005) for more discussion of selection on observed and unobserved factors.

Table 10. Selection based on observables and estimated ability.[a]

Course Type	Homogeneous Gamma			Heterogenous Gamma		
	α	$\hat{\alpha}$	α_u	α	$\hat{\alpha}$	α_u
Humanities						
Avg. section-level SD	0.6154	0.3372	0.5385	0.5849	0.2487	0.5415
Population SD	0.8046	0.3773	0.7106	0.7619	0.2711	0.7121
Ratio	0.7649	0.8937	0.7578	0.7677	0.9174	0.7604
Social science						
Avg. section-level SD	0.6365	0.3750	0.5667	0.6056	0.2994	0.5688
Population SD	0.8618	0.4228	0.7510	0.8226	0.3343	0.7517
Ratio	0.7386	0.8869	0.7546	0.7362	0.8956	0.7567
Math and science						
Avg. section-level SD	0.7475	0.5036	0.6666	0.7636	0.5323	0.6649
Population SD	1.0567	0.6165	0.8583	1.0719	0.6432	0.8575
Ratio	0.7074	0.8169	0.7767	0.7124	0.8276	0.7754

[a]Each set of rows corresponds to sections in one course type; each set of columns corresponds to one version of the model (homogeneous gamma versus heterogeneous gamma). Variables under analysis appear in the heading row: α is ability as estimated by our model, $\hat{\alpha}$ is observed ability, and α_u is unobserved ability. Avg. section-level SD is the average (across all sections, and weighted by section size) of the standard deviation of the variable within a section. Population SD is the standard deviation of the variable in the population of students taking courses of the given course type. Ratio is the ratio of the first of these to the second, and shows the degree of selection into sections with respect to each variable displayed.

ability, which is consistent with a high correlation between peer observed and unobserved ability.

With the observed and unobserved ability measures, it is also to possible to estimate the correlation between unobserved individual ability and observed peer ability, which feeds directly into the bias associated with the selection problem. The correlation coefficients for unobserved individual ability and observed peer ability are 0.03, 0.06, and 0.20 for humanities, social sciences, and math and science, respectively. The high correlation coefficient for math and science suggests that the upward bias associated with peer effect estimation using a selection-on-observables approach might be quite large.

7.2 *Selection across course types*

Because the vast majority of students are observed in courses of multiple types during their tenure at Maryland, we obtain multiple estimates of ability for most students. Calculating the correlations between estimated ability levels illuminates the extent to which good performance in each of the three course types is driven by similar student attributes as performance in the other course types and, therefore, provides an empirical index of the academic similarity of course types.[38]

Panel A of Table 11 shows the correlation coefficients among estimated ability levels across the three course types from the homogeneous gamma model. These correlations are created using estimated abilities from students observed in all course types.[39] The

[38]For ease of exposition, we focus on the homogeneous gamma model for the rest of the paper.

[39]Correlations among estimated abilities were also calculated for all students who were observed in each pair of course types. Similar coefficients resulted.

Table 11. Correlations of estimated abilities across course types.[a]

Course Type	Humanities	Social Science
	A. Abilities	
Humanities	1.0000	
Social science	0.6875	1.0000
Math and science	0.6469	0.6776
	B. Predicted Abilities	
Humanities	1.0000	
Social science	0.9643	1.0000
Math and science	0.8808	0.9593

[a]The abilities used in these correlation matrices are those of the 12,715 students who took classes in all three course types; they are estimated by the homogeneous gamma model. Panel A displays correlations among the full abilities, while panel B displays correlations among the predicted values from regressing the estimated abilities from the homogeneous gamma model on observable variables (as shown in the second through fourth columns of Table 9).

correlation coefficients are all quite large and close together, ranging from 0.65 to 0.69.[40] Panel B of Table 11 displays analogous results using only the portion of estimated abilities for each student that is predictable using our observable variables. While the relative relationships among observable abilities by course type are the same, the strength of the relationships is much stronger. The differences across the panels suggest that ability is much more heterogeneous than can be captured by observed ability measures.

With this information as background, Table 12 illustrates the degree to which individual students are observed to be sorting into the types of courses for which they appear, based on our model, to be best suited. In particular, we label a student as specializing in a particular course type if the number of courses taken in that course type is higher than the number of courses taken in either of the other two course types. Panel A of this table displays results using the estimated abilities from our homogeneous gamma model standardized to a $N(0,1)$ distribution, and panel B displays results using only the portion of those estimated abilities that could be predicted based on observable characteristics, also standardized to $N(0,1)$. The rows correspond to the sets of students who specialize in humanities, social sciences, and math and science, respectively. The numbers along the rows give the mean (normalized) fixed effect for each of the different course types.

We see two striking patterns in panel A. First, students who specialize in humanities courses are estimated to be less able across the board than those who specialize in either the social sciences or math and science. Students specializing in math and science are the most able across the board, with higher average fixed effects in each course type. Even more interesting, students in each specialization group appear to have specialized in the area for which they are most suited.[41] Panel B of Table 12, where we use only

[40]The pattern of correlations for the heterogenous gamma model is similar, although the correlation coefficients are slightly lower: 0.64 for social sciences and humanities, 0.62 for social sciences and math and science, and 0.54 for math and science and humanities.

[41]Paglin and Rufolo (1990) found similar sorting patterns using Graduate Record Exam data and transcript data from the University of Oregon and Oregon State University. They found that students with high math ability tend to take courses in which there is a high return to this skill.

Table 12. Specialization of students into course types by relative aptitude.[a]

	Humanities	Soc. Sci.	Math/Sci.	N
	A. Abilities			
Humanities specializers	−0.08	−0.19	−0.26	3978
Social science specializers	0.04	0.10	−0.07	3547
Math and science specializers	0.14	0.21	0.43	3745
	B. Predicted Abilities			
Humanities specializers	−0.04	−0.13	−0.21	3977
Social science specializers	−0.11	−0.10	−0.11	3547
Math and science specializers	0.18	0.27	0.36	3745

[a]In panel A, each cell shows the mean of the deviations of students' ability to perform in the course type of that column (as estimated by our homogeneous gamma model, and standardized to a normal $(0, 1)$ distribution) from the sample standardized mean of estimated ability across the course type, for the population of that row. "Specializers" are students who are observed to take more courses in the given course type than in either of the other two course types.

observed ability to examine selection, also illustrates both absolute and comparative advantage, but reflects more attenuated distributions of abilities and less heterogeneity than panel A.

8. Comparing the method to conventional methods

To compare our estimated peer effects to those that would be obtained using conventional techniques, we first conceptualize the estimation problem as follows. Given a model where the ability of each student can be decomposed into observed versus unobserved portions, there are two econometric obstacles to the accurate estimation of the spillover. The first obstacle is a positive correlation between the student's own unobserved ability and his peer group's observed ability, which also leads in any given sample to a correlation between the peer group's observed ability and the peer group's unobserved ability. This problem leads to an upward bias of the spillover parameter. The second obstacle is that when only observables are used to form the peer ability measures, the underlying distribution of peer ability is attenuated, leading to downward pressure on the estimated impact of a 1 standard deviation change in peer ability.

To examine the quantitative impact of these two problems separately, we first artificially eliminate the correlation between the student's own unobserved ability and the peer group's observed ability by differencing out our estimated individual fixed effects and course effects from student grades. We regress these adjusted grades on observed peer ability, rather than total peer ability. This enables us to examine the consequences for estimation of using an incomplete measure of peer ability in a case where the link between individual unobservables and peer observables is broken.

The second row of Table 13 for each course type presents the results of this first exercise, where we use total ability (our estimated α's) for the individual and observed ability for peers. For comparison, the first row of the table for each course type gives the original spillover estimate produced using our algorithm. Looking at the fourth column of the table, we can see that the estimated effects of a 1 standard deviation increase in observed peer ability are at most two-thirds the size of a 1 standard deviation increase in

TABLE 13. Comparing the method to conventional results: homogeneous gamma model.[a]

	Own Ability	Section Peers' Ability	Effect of 1 SD Change in Peer Ability	Own Ability	Peers' Ability
Humanities	1	0.1613	0.0455	Total	Total
	(—)	(0.0007)			
	1	0.1642	0.0285	Total	Observed
	(—)	(0.0008)			
	0.9349	0.2502	0.0435	Observed	Observed
	(0.0076)	(0.0077)			
Social science	1	0.1960	0.0608	Total	Total
	(—)	(0.0008)			
	1	0.2061	0.0370	Total	Observed
	(—)	(0.0009)			
	0.8518	0.4085	0.0733	Observed	Observed
	(0.0077)	(0.0077)			
Math and science	1	0.0483	0.0287	Total	Total
	(—)	(0.0008)			
	1	0.0516	0.0193	Total	Observed
	(—)	(0.0009)			
	0.7784	0.3519	0.1313	Observed	Observed
	(0.0062)	(0.0063)			

[a]The dependent variable is grade in the class. When own ability is restricted to equal 1, class effects are removed before estimation by demeaning using the "true" values of the class effect as estimated by our homogeneous gamma model. When own ability is unrestricted, class fixed effects are absorbed in estimation. For the purposes of this table, we ignore the sampling variation in the parameter estimates used to construct our observed ability measures.

total peer ability. This first-order decrease in effect magnitude is evident because the impact of unobserved peer ability is not captured in the second row except through the correlation between observed and unobserved peer ability.[42]

We next investigate what happens when the econometrician additionally assumes that students select into peer groups based only on observable characteristics (the selection-on-observables approach). In the third row of Table 13, we again use observed peer ability, but now we use observed ability for the individuals as well rather than the estimated fixed effects. The positive correlation between unobserved individual ability and observed peer ability biases the selection-on-observables estimate of the spillover parameter upward. While the estimate of the spillover parameter is biased upward, the effect of a 1 standard deviation increase in peer ability may still be smaller because the variance in observed peer ability is smaller than the variance in peer ability as a whole. As can be seen in the fourth column, this is indeed the case for humanities, the course type with the smallest correlation between unobserved individual ability and observed

[42]Under random assignment, a 1 standard deviation increase in peer ability would produce even smaller effects than those shown in the second row of Table 13. There are two reasons for this. First, peer observed ability and peer unobserved ability are positively correlated in our data, but would not be correlated under random assignment. The second reason why random assignment leads to even lower estimates of a 1 standard deviation increase in peer ability is that random assignment itself leads to less heterogeneity in mean peer ability across sections than when higher ability students choose sections with other high ability students.

peer ability. For the social sciences, the selection-on-observables estimate of a 1 standard deviation increase, although higher than our original estimate, is still closer than the estimates given in the second row that mitigate the selection problem. However, for math and science, the estimated effect of a 1 standard deviation increase in peer ability is significantly higher using the selection-on-observables approach than using our algorithm. This is driven by (i) the high correlation between unobserved individual ability and observed peer ability in math and science, (ii) the fact that observed ability is a greater fraction of total ability in math and science, and (iii) the fact that the underlying peer effect estimate from our method is smallest in math and science, which mitigates the underestimation of a 1 standard deviation increase in peer ability.[43]

9. Conclusion

Accurate estimation of peer effects in the classroom is plagued by at least two issues, both of which have to do with ability not being fully observed. First, there is selection into the peer group, which leads to a positive correlation between unobserved individual ability and observed peer ability. If ignored, this correlation leads to upward-biased estimates of peer effect parameters. On the other hand, underestimation of the effects of peers may result from ignoring peer effects that operate through unobservables.

We present a new iterative method for estimating educational peer effects that overcomes both these obstacles. Our methodology rests on two primary assumptions: outcomes are linear in peer characteristics and the relevance to outcomes of peer characteristics is proportional to that of own characteristics. Under these assumptions, all that is required from a data standpoint is that there are multiple observations per student, with the peer group changing over time. We control for individual effects and allow the peer effect to operate through a linear combination of the other individual effects. We show that our estimator is consistent and asymptotically normal for fixed T as N goes to infinity. Consistency is shown in the case where the effects are transitory and when they are persistent. We also develop an iterative algorithm that is computationally much cheaper than direct nonlinear least squares minimization, yet produces the nonlinear squares results upon convergence. Monte Carlo results suggest that the model performs quite well, even when the number of observations per student is small.

We estimate the model on transcript data from the University of Maryland. Small but significant peer effects are found, with evidence of heterogeneity by course type. Social science courses show the largest peer effects, whereas grades in math and science courses rely least on peer ability and most heavily on a student's own ability.

Previous efforts to estimate spillover effects in education that do not rely on random assignment are often plagued by concerns regarding selection on unobservables. Our data suggest that this is a valid issue. Students select into sections based more on unobservable factors than on observable factors. This leads to correlation in unobserved own ability and both observed and unobserved peer ability that, if ignored, biases the spillover parameter upward. There is also much selection across course type. Students

[43] Indeed, if the spillover parameter were zero, there would be no scope for the attenuation effect.

sort into course types where their relative abilities are highest, suggesting comparative advantage is important in the selection of courses. However, absolute advantage is also present, as those who primarily choose math and science course are more able in both humanities and social sciences than those who choose to specialize in one of the other areas.

Our method allows us to quantify the effects of both the selection problem and the problem of not being able to estimate peer effects through unobservables. Estimation using data on different course types illustrates how the setting dictates which of these problems is more important. For math and science courses, the estimated spillover parameter from our model is small. This, coupled with abundant selection into math and science courses, leads to estimates from a selection-on-observables approach that significantly overstate the importance of peers. However, for humanities courses, the estimated spillover parameter is larger than in math and science, and this fact coupled with much less selection than in math and science makes the selection-on-observables approach yield a peer effects estimate similar to that estimated by our model. We also show that a random assignment approach, which removes the correlation between individual unobserved ability and peer observed ability, but ignores peer effects through unobservables, significantly understates the impact of peers on achievement.

Appendix A: Proofs

A.1 *Proof of Theorem* 1

For ease of exposition, we illustrate the proof assuming that students are grouped with at most one other student at any point in time. We also focus only on case A1 where there is no accumulation. The proof for general class sizes and accumulation (case A2) are straightforward extensions and are provided in a supplementary file on the journal website.[44] Keeping with the literature, we also assume a homogeneous peer effect that is proportional to the ability of a student's peer. The proof can be readily expanded to multiple γ's.

We consider the following limiting case:

1. We observe students for at most two time periods.
2. Within each class, there is only one student who is observed for two periods. The other student is observed for only one time period.

Remark 1. Clearly if the estimator is consistent for $T = 2$, it is also consistent for $T > 2$. The second simplification is equivalent to allowing all but one of the individual effects in a class to vary over time. For example, suppose there were $2\mathcal{N}$ students observed for two periods, implying that $2\mathcal{N}$ individual effects would be estimated. We could, however, allow the individual effect to vary over time for one student in each group, making sure to choose these students in such a way that they are matched with someone in both periods whose individual effect does not vary over time.[45] Then $3\mathcal{N}$ individual effects

[44] See http://qeconomics.org/supp/145/supplement.pdf.

[45] To see how these assignments work, consider a two period model where the groups in period 1 are $\{A, B\}$ and $\{C, D\}$, and the groups in period 2 are $\{A, C\}$ and $\{B, D\}$. We could let the individual effects vary

would be estimated. Having one individual whose effect varies over time is equivalent to estimating two individual effects—it is the same as having two different individuals who were each observed once. If the estimator is consistent in this case, then it is also consistent under the restricted case when all of the individual effects are time invariant (fixed effects).

Consider the set of students who are observed for two time periods. Each of these students has one peer in period 1 and one peer in period 2. Denote a student block as one student observed for two periods plus his two peers. There are then $\mathcal{N}$ blocks of students, one block for each student observed twice, with three students in each block. Denote the first student in each block as the student who is observed twice, where α_{1n} is the individual effect. The individual effect for the first classmate in block n is α_{2n}, while the individual effect for the second classmate in block n is α_{3n}.

The optimization problem is then

$$\begin{aligned} \min_{\alpha,\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \Bigg(& (y_{11n} - \alpha_{1n} - \gamma\alpha_{2n})^2 + (y_{12n} - \alpha_{1n} - \gamma\alpha_{3n})^2 \\ & + \sum_{i=2}^{3} (y_{in} - \alpha_{in} - \gamma\alpha_{1n})^2 \Bigg). \end{aligned} \tag{14}$$

Within each block, there are four terms: two residuals for the student observed twice and a residual for the peer in each period.

Remark 2. Note that, conditional on γ, the estimates of individual effects in one block will not affect the estimates of the individual effects in another block. Hence, we are able to focus on individual blocks in isolation from one another when concentrating out the α's as a function of γ.

Our proof then consists of the following five lemmas, each of which is proven later in this appendix.

We first show that the α's can be written as closed form expressions of γ and the data.

Lemma 1. *The vector of unobserved student abilities, α, can be concentrated out of the least squares problem and written strictly as a function of γ and y. Ability for the student in block n observed in both periods is given by*

$$\alpha_{1n} = \frac{y_{11n} + y_{12n} - \gamma(y_{2n} + y_{3n})}{2(1-\gamma^2)},$$

while the abilities for the peers in block n are given by

$$\alpha_{2n} = \frac{y_{2n} + \gamma^2 y_{3n} - \gamma y_{12n} - \gamma^3 y_{11n}}{1-\gamma^4}$$

for either $\{A, D\}$ or $\{B, C\}$. In both these cases, each group in each time period would have one student observed twice and one student observed once. The number of individual effects would then increase from four to six.

and

$$\alpha_{3n} = \frac{y_{3n} + \gamma^2 y_{2n} - \gamma y_{11n} - \gamma^3 y_{12n}}{1 - \gamma^4}.$$

We then show the form of the minimization problem when the α's are concentrated out.

Lemma 2. *Concentrating the α's out of the original least squares problem results in an optimization problem over γ that takes the form*

$$\min_{\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \frac{(y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n}))^2}{2(1+\gamma^2)}.$$

Our nonlinear least squares problem now has only one parameter, γ. We are now in a position to investigate the properties of our estimator of γ_o. For ease of notation, define $q(w, \gamma)$ as

$$q(w, \gamma) = \frac{(y_{11} - y_{12} + \gamma(y_3 - y_2))^2}{2(1+\gamma^2)},$$

where $w \equiv \mathbf{y}$. We let $\mathcal{W}$ denote the subset of $\mathbb{R}^4$ representing the possible values of w. Our key result is then Lemma 3, which establishes identification.

Lemma 3. *We have*

$$E\big[q(w, \gamma_o)\big] < E\big[q(w, \gamma)\big] \quad \forall \gamma \in \Gamma, \gamma \neq \gamma_o.$$

Theorem 12.2 of Wooldridge (2002) establishes that sufficient conditions for consistency are identification and uniform convergence. Having already established identification, Lemma 4 shows uniform convergence.

Lemma 4. *We have*

$$\max_{\gamma \in \Gamma} \left| \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} q(w_n, \gamma) - E\big[q(w, \gamma)\big] \right| \xrightarrow{p} 0.$$

Consistency then follows from Theorem 12.2 of Wooldridge: $\hat{\gamma} \xrightarrow{p} \gamma_o$.

Finally, we establish asymptotic normality of $\hat{\gamma}$. Denote $s(w, \gamma_o)$ and $H(w, \gamma_o)$ as the first and second derivatives of $q(w, \gamma)$ evaluated at γ_o. Then Lemma 5 completes the proof.

Lemma 5. *We have*

$$\sqrt{\mathcal{N}}(\hat{\gamma} - \gamma_o) \xrightarrow{d} N\big(0, A_o^{-1} B_o A_o^{-1}\big),$$

where

$$A_o \equiv E\big[H(w, \gamma_o)\big]$$

and

$$B_o \equiv E\big[s(w, \gamma_o)^2\big] = \mathrm{Var}\big[s(w, \gamma_o)\big].$$

Proof of Lemma 1. Our objective is to show that the system of equations obtained by differentiating equation (14) with respect to α can be expressed as a series of equations in terms of γ and y, and that these expressions are as given in Lemma 1. Again, conditional on γ, the estimates of individual effects in one block will not affect the estimates of the individual effects in another block. Thus, we can work with the system of first-order conditions within one block and then generalize the results to the full system of equations. The first-order condition for α_{1n} (student in each block who is observed in both time periods) is given by

$$0 = \frac{-2}{\mathcal{N}}\left[(y_{11n} - \alpha_{1n} - \gamma\alpha_{2n}) + (y_{12n} - \alpha_{1n} - \gamma\alpha_{3n}) + \gamma\sum_{i=2}^{3}(y_{in} - \alpha_{in} - \gamma\alpha_{1n})\right],$$

while the first-order condition for α_{2n} and α_{3n} are, respectively, given by

$$0 = \frac{-2}{\mathcal{N}}\big[(y_{2n} - \alpha_{2n} - \gamma\alpha_{1n}) + \gamma(y_{11n} - \alpha_{1n} - \gamma\alpha_{2n})\big]$$

and

$$0 = \frac{-2}{\mathcal{N}}\big[(y_{3n} - \alpha_{3n} - \gamma\alpha_{1n}) + \gamma(y_{12n} - \alpha_{1n} - \gamma\alpha_{3n})\big].$$

Within each block, this yields a relatively simple system of three equations and three unknown abilities. The first-order conditions for α_{2n} and α_{3n} can be rearranged such that

$$\alpha_{2n} = \frac{y_{2n} + \gamma y_{11n} - 2\gamma\alpha_{1n}}{1+\gamma^2}$$

and

$$\alpha_{3n} = \frac{y_{3n} + \gamma y_{12n} - 2\gamma\alpha_{1n}}{1+\gamma^2}.$$

Notice that the equation for α_{2n} depends only on the own outcome, the outcome of individual 1 when grouped with individual 2, and the ability of individual 1. A similar result occurs for α_{3n}. Thus, the only thing linking individuals 2 and 3 within a block is the ability of individual 1.

Rearranging the first-order condition for α_{1n} such that the α_{1n} are grouped on the left-hand side of the equation results in

$$\alpha_{1n}\big(2+2\gamma^2\big) = y_{11n} + y_{12n} + \gamma(y_{2n} + y_{3n}) - 2\gamma(\alpha_{2n} + \alpha_{3n}),$$

and substituting for α_{2n} and α_{3n} using the previously derived formulas yields

$$\begin{aligned}\alpha_{1n}\bigl(2+2\gamma^2\bigr) &= y_{11n}+y_{12n}+\gamma(y_{2n}+y_{3n})\\ &\quad -\frac{2\gamma}{1+\gamma^2}\bigl(y_{2n}+y_{3n}+\gamma(y_{11n}+y_{12n})-4\gamma\alpha_{1n}\bigr).\end{aligned}$$

Moving all the α_{1n} terms to the left side and finding common denominators on both sides of the equation results in

$$\begin{aligned}&\frac{\alpha_{1n}((2+2\gamma^2)(1+\gamma^2)-8\gamma^2)}{1+\gamma^2}\\ &\quad=\frac{(1+\gamma^2)(y_{11n}+y_{12n}+\gamma(y_{2n}+y_{3n}))-2\gamma(y_{2n}+y_{3n}+\gamma(y_{11n}+y_{12n}))}{1+\gamma^2}.\end{aligned}$$

Canceling out the denominators and simplifying both sides of the equation yields

$$\alpha_{1n}\bigl(2\bigl(1-\gamma^2\bigr)^2\bigr)=\bigl(1-\gamma^2\bigr)(y_{11n}+y_{12n})-\gamma\bigl(1-\gamma^2\bigr)(y_{2n}+y_{3n}).$$

Dividing both sides of the equation by $2(1-\gamma^2)^2$ yields the desired result that

$$\alpha_{1n}=\frac{y_{11n}+y_{12n}-\gamma(y_{2n}+y_{3n})}{2(1-\gamma^2)}.$$

The solution for α_{1n} can now be substituted back into the first-order conditions for α_{2n} and α_{3n} to yield solutions strictly as functions of γ and y. Substituting α_{1n} into the equation for α_{2n} and finding a common denominator yields

$$\alpha_{2n}=\frac{2(1-\gamma^2)(y_{2n}+\gamma y_{11n})-2\gamma(y_{11n}+y_{12n}-\gamma(y_{2n}+y_{3n}))}{2(1-\gamma^2)(1+\gamma^2)}.$$

Factoring out the 2 in the numerator and expanding the resulting expression yields

$$\alpha_{2n}=\frac{(1-\gamma^2+\gamma^2)y_{2n}+(\gamma(1-\gamma^2)-\gamma)y_{11n}-\gamma y_{12n}+\gamma^2 y_{3n}}{(1-\gamma^2)(1+\gamma^2)}.$$

Some simple manipulation leads to the final result that

$$\alpha_{2n}=\frac{y_{2n}+\gamma^2 y_{3n}-\gamma y_{12n}-\gamma^3 y_{11n}}{1-\gamma^4}.$$

Obtaining the solution for α_{3n} proceeds in exactly the same way, and yields a formula that mirrors the solution for α_{2n} with the appropriate indices changed to reflect when individual 3 is grouped with individual 1. The result is

$$\alpha_{3n}=\frac{y_{3n}+\gamma^2 y_{2n}-\gamma y_{11n}-\gamma^3 y_{12n}}{1-\gamma^4}. \qquad \square$$

Proof of Lemma 2. Lemma 1 provides a solution for α strictly as a function of y and γ. We can substitute this solution back into the original optimization problem to derive the result in Lemma 2.

Consider minimizing the sum of squared residuals within a particular block n. There are four residuals within each block: two for the student observed twice and one each for the corresponding peer. We begin by simplifying the residual for the first observation of the student observed twice, which is given by the expression

$$e_{11n} = y_{11n} - \alpha_{1n} - \gamma\alpha_{2n}.$$

Substituting for α_{1n} and α_{2n} in e_{11n} with the results from Lemma 1 results in

$$e_{11n} = y_{11n} - \frac{y_{11n} + y_{12n} - \gamma(y_{2n} + y_{3n})}{2(1-\gamma^2)} - \frac{\gamma(y_{2n} + \gamma^2 y_{3n} - \gamma y_{12n} - \gamma^3 y_{11n})}{1-\gamma^4}.$$

Finding a common denominator and combining like terms in the numerator yields

$$\begin{aligned} e_{11n} = {} & \big(\big(2\big(1-\gamma^4\big) - \big(1+\gamma^2\big) + 2\gamma^4\big)y_{11n} - \big(\big(1+\gamma^2\big) - 2\gamma^2\big)y_{12n} \\ & + \big(\gamma\big(1+\gamma^2\big) - 2\gamma\big)y_{2n} + \big(\gamma\big(1+\gamma^2\big) - 2\gamma^3\big)y_{3n}\big) / \big(2\big(1-\gamma^4\big)\big). \end{aligned}$$

Simplifying the numerators on each of the y terms and factoring the denominator yields

$$e_{11n} = \frac{(1-\gamma^2)y_{11n} - (1-\gamma^2)y_{12n} - \gamma(1-\gamma^2)y_{2n} + \gamma(1-\gamma^2)y_{3n}}{2(1-\gamma^2)(1+\gamma^2)}.$$

Finally, we can cancel all the $(1-\gamma^2)$ terms to arrive at

$$e_{11n} = \frac{y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n})}{2(1+\gamma^2)}.$$

The expression for e_{12n} as a function of γ and y can be similarly derived by substituting in α_{1n} and α_{3n}. However, the expressions for e_{12n} and α_{3n} are mirror images of the expressions for e_{11n} and α_{2n}. Thus, e_{12n} will take the exact same form as e_{11n} except the subscripts denoting the period or classmate are swapped. The expression is

$$e_{12n} = \frac{y_{12n} - y_{11n} + \gamma(y_{2n} - y_{3n})}{2(1+\gamma^2)}.$$

The residuals for the one observation individuals in each block, e_{2n} and e_{3n}, are given by

$$e_{2n} = y_{2n} - \alpha_{2n} - \gamma\alpha_{1n}$$

and

$$e_{3n} = y_{3n} - \alpha_{3n} - \gamma\alpha_{1n}.$$

To write these strictly as functions of γ and y, we again use the results of Lemma 1. Substituting for α_{1n} and α_{2n} in e_{2n} yields

$$e_{2n} = y_{2n} - \frac{y_{2n} + \gamma^2 y_{3n} - \gamma y_{12n} - \gamma^3 y_{11n}}{1-\gamma^4} - \frac{\gamma(y_{11n} + y_{12n} - \gamma(y_{2n} + y_{3n}))}{2(1-\gamma^2)}.$$

Finding a common denominator and simplifying the resulting expressions yields

$$e_{2n} = \frac{\gamma(y_{12n} - y_{11n} + \gamma(y_{2n} - y_{3n}))}{2(1+\gamma^2)}.$$

The expression for e_{3n} is similar to that of e_{2n}, except the subscripts differ to reflect the time period in which individual 3 is grouped with 1. Thus the solution for e_{3n} will mirror the solution for e_{2n}, except that the appropriate subscripts are swapped across terms. The final expression for e_{3n} is

$$e_{3n} = \frac{\gamma(y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n}))}{2(1+\gamma^2)}.$$

The original optimization problem written as a function of the residuals in each block n takes the form

$$\min_{\alpha,\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} (e_{11n}^2 + e_{12n}^2 + e_{2n}^2 + e_{3n}^2).$$

Now we can substitute in for each residual using the formulas previously derived. However, a cursory glance at the formulas for e_{11n}, e_{12n}, e_{2n}, and e_{3n} reveals that

$$e_{11n} = -e_{12n} = -\gamma e_{2n} = \gamma e_{3n}.$$

Using these relationships, we can rewrite the least squares problem as

$$\min_{\alpha,\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \bigl((2+2\gamma^2)e_{11n}^2\bigr).$$

Substituting in with our solution for e_{11n} yields

$$\min_{\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left((2+2\gamma^2) \frac{(y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n}))^2}{(2(1+\gamma^2))^2} \right).$$

Canceling terms results in the optimization problem

$$\min_{\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \frac{(y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n}))^2}{2(1+\gamma^2)}. \qquad \square$$

Proof of Lemma 3. The population objective function as a function of γ is given by

$$E\bigl[q(w,\gamma)\bigr] = E\left[\frac{(y_{11} - y_{12} + \gamma(y_3 - y_2))^2}{2(1+\gamma^2)}\right].$$

Substituting for y with the data generating process yields

$$\begin{aligned} E\bigl[q(w,\gamma)\bigr] = E\bigl[&\bigl(\alpha_{1o} + \gamma_o\alpha_{2o} + \varepsilon_{11} - (\alpha_{1o} + \gamma_o\alpha_{3o} + \varepsilon_{12}) \\ &+ \gamma\bigl(\alpha_{3o} + \gamma_o\alpha_{1o} + \varepsilon_3 - (\alpha_{2o} + \gamma_o\alpha_{1o} + \varepsilon_2)\bigr)\bigr)^2 / \bigl(2(1+\gamma^2)\bigr)\bigr]. \end{aligned}$$

Canceling the appropriate terms and combining like terms in the numerator leaves

$$E\big[q(w,\gamma)\big]=E\bigg[\frac{((\gamma_o-\gamma)(\alpha_{2o}-\alpha_{3o})+(\varepsilon_{11}-\varepsilon_{12})+\gamma(\varepsilon_3-\varepsilon_2))^2}{2(1+\gamma^2)}\bigg].$$

Opening up the square term leaves

$$\begin{aligned}E\big[q(w,\gamma)\big]=E\bigg[&\frac{1}{2(1+\gamma^2)}\big((\gamma_o-\gamma)^2(\alpha_{2o}-\alpha_{3o})^2+(\varepsilon_{11}-\varepsilon_{12})^2+\gamma^2(\varepsilon_3-\varepsilon_2)^2\\&+2(\gamma_o-\gamma)(\alpha_{2o}-\alpha_{3o})(\varepsilon_{11}-\varepsilon_{12})\\&+2\gamma(\gamma_o-\gamma)(\alpha_{2o}-\alpha_{3o})(\varepsilon_3-\varepsilon_2)\\&+2\gamma(\varepsilon_{11}-\varepsilon_{12})(\varepsilon_3-\varepsilon_2)\big)\bigg].\end{aligned}$$

By Theorem 1(i) and (ii), the final three terms in the numerator all have expectation 0. Similarly, any covariance terms associated with the first three terms in the numerator will have expectation 0. The final simplified expression is given by

$$E\big[q(w,\gamma)\big]=\frac{(\gamma_o-\gamma)^2E[(\alpha_{2o}-\alpha_{3o})^2]+E[\varepsilon_{11}^2]+E[\varepsilon_{12}^2]+\gamma^2(E[\varepsilon_3^2]+E[\varepsilon_2^2])}{2(1+\gamma^2)},$$

which we can rewrite in the manner

$$E\big[q(w,\gamma)\big]=\frac{(\gamma_o-\gamma)^2E[(\alpha_{2o}-\alpha_{3o})^2]}{2(1+\gamma^2)}+\frac{E[\varepsilon_{11}^2]+E[\varepsilon_{12}^2]+\gamma^2(E[\varepsilon_3^2]+E[\varepsilon_2^2])}{2(1+\gamma^2)}.$$

Note that by Theorem 1(v) $E[\varepsilon_{11}^2]=E[\varepsilon_2^2]$ and $E[\varepsilon_{12}^2]=E[\varepsilon_3^2]$, implying that we can rewrite the above equation as

$$E\big[q(w,\gamma)\big]=\frac{(\gamma_o-\gamma)^2E[(\alpha_{2o}-\alpha_{3o})^2]}{2(1+\gamma^2)}+\big(E\big[\varepsilon_{11}^2\big]+E\big[\varepsilon_{12}^2\big]\big)/2.$$

The first term in the above expression is strictly greater than 0 for all $\gamma\neq\gamma_o$ and the second term does not depend upon γ. As a result, $E[q(w,\gamma_o)]<E[q(w,\gamma)]$ for all $\gamma\in\Gamma$ when $\gamma\neq\gamma_o$. □

Proof of Lemma 4. Uniform convergence, requires that

$$\max_{\gamma\in\Gamma}\bigg|\frac{1}{\mathcal{N}}\sum_{n=1}^{\mathcal{N}}q(w_n,\gamma)-E\big[q(w,\gamma)\big]\bigg|\xrightarrow{p}0.$$

Theorem 12.1 in Wooldridge states four conditions that the data and q must satisfy so that the above condition holds.

1. *The parameter Γ is compact*. This condition is satisfied by Theorem 1(vi).
2. *For each $\gamma\in\Gamma$, $q(\cdot,\gamma)$ is Borel measurable on $\mathcal{W}$*. Since $q(\cdot,\gamma)$ is a continuous function of w, it is also Borel measurable.

3. *For each* $w \in \mathcal{W}$, $q(w, \cdot)$ *is continuous on* Γ. Our concentrated objective function is continuous in γ.

4. *For all* $\gamma \in \Gamma$, $|q(w, \gamma)| \leq b(w)$, where b is a nonnegative function on $\mathcal{W}$ such that $E[b(w)] < \infty$. Note that $q(w, \gamma)$ is always positive, so we can ignore the absolute value. We derive a bounding function $b(w)$ in the manner

$$\begin{aligned} q(w, \gamma) &= \frac{(y_{11} - y_{12} + \gamma(y_3 - y_2))^2}{2(1+\gamma^2)} \\ &= \frac{(y_{11} - y_{12})^2 + \gamma^2(y_3 - y_2)^2 + 2\gamma(y_3 - y_2)(y_{11} - y_{12})}{2(1+\gamma^2)} \\ &\leq \frac{2(y_{11} - y_{12})^2}{2(1+\gamma^2)} + \frac{2\gamma^2(y_3 - y_2)^2}{2(1+\gamma^2)} \\ &\leq (y_{11} - y_{12})^2 + (y_3 - y_2)^2, \end{aligned}$$

where the third line follows from the triangle inequality. Our bounding function is then

$$b(w) = (y_{11} - y_{12})^2 + (y_3 - y_2)^2,$$

where we have shown that $b(w) \geq q(w, \gamma)$ for all y.

We now show that $E[b(w)] < \infty$, completing the proof. Note that $E[b(w)]$ is given by

$$E\big[b(w)\big] = E\big[(y_{11} - y_{12})^2 + (y_3 - y_2)^2\big].$$

Using the triangle inequality, we can rewrite the above expression as

$$\begin{aligned} E\big[b(w)\big] &\leq E\big[2y_{11}^2 + 2y_{12}^2 + 2y_3^2 + 2y_2^2\big] \\ &\leq 2\big(E\big[y_{11}^2\big] + E\big[y_{12}^2\big] + E\big[y_3^2\big] + E\big[y_2^2\big]\big). \end{aligned}$$

Next we substitute in for y using the data generating process. Consider $E[y_{11}^2]$, which is given by

$$E\big[y_{11}^2\big] = E\big[(\alpha_{1o} + \gamma_o \alpha_{2o} + \varepsilon_{11})^2\big].$$

Applying the triangle inequality again yields

$$E\big[y_{11}^2\big] \leq 3\big(E\big[\alpha_{1o}^2\big] + \gamma_o^2 E\big[\alpha_{2o}^2\big] + E\big[\varepsilon_{11}^2\big]\big).$$

Theorem 1(iii) and (iv) ensure that all of the terms on the right-hand side of the inequality in the above equation are finite. Thus, $E[y_{11}^2]$ is finite. By a similar argument, it can be shown that all the terms in $E[b(w)]$ are finite. □

Proof of Lemma 5. Theorem 12.3 in Wooldridge (2002) states six conditions that must hold for $\hat{\gamma}$ to be distributed asymptotically normal. Many of these conditions involve the first and second derivatives of $q(w, \gamma)$. We begin our proof of asymptotic normality by deriving the first and second derivatives of the objective function.

The first derivative of the objective function, or the score, is given by

$$
\begin{aligned}
s(w,\gamma) = {} & \frac{1}{4(1+\gamma^2)^2}\bigl[2\bigl(1+\gamma^2\bigr)\bigl(2(y_3-y_2)\bigl(y_{11}-y_{12}+\gamma(y_3-y_2)\bigr)\bigr) \\
& - 4\gamma\bigl(\bigl(y_{11}-y_{12}+\gamma(y_3-y_2)\bigr)^2\bigr)\bigr].
\end{aligned}
$$

Expanding the square and grouping on the γ terms yields

$$
s(w,\gamma) = \frac{1}{(1+\gamma^2)^2}\bigl[\bigl(1-\gamma^2\bigr)(y_{11}-y_{12})(y_3-y_2) + \gamma\bigl(-(y_{11}-y_{12})^2+(y_3-y_2)^2\bigr)\bigr].
$$

The Hessian of the objective function is simply the derivative of the score, $\frac{\partial s(y,\gamma)}{\partial \gamma}$, and is written

$$
\begin{aligned}
H(w,\gamma) = {} & \frac{1}{(1+\gamma^2)^4} \\
& \times \bigl(\bigl(1+\gamma^2\bigr)^2\bigl((y_3-y_2)^2-(y_{11}-y_{12})^2-2\gamma(y_3-y_2)(y_{11}-y_{12})\bigr) \\
& - 4\gamma\bigl(1+\gamma^2\bigr) \\
& \times \bigl(\gamma\bigl((y_3-y_2)^2-(y_{11}-y_{12})^2\bigr)+\bigl(1-\gamma^2\bigr)(y_3-y_2)(y_{11}-y_{12})\bigr)\bigr).
\end{aligned}
$$

Factoring out a $(1+\gamma^2)$ and combining like terms greatly simplifies the above expression, leaving

$$
\begin{aligned}
H(w,\gamma) = {} & \frac{1}{(1+\gamma^2)^3}\bigl(\bigl(1-3\gamma^2\bigr)\bigl((y_3-y_2)^2-(y_{11}-y_{12})^2\bigr) \\
& - 2\gamma\bigl(3-\gamma^2\bigr)(y_3-y_2)(y_{11}-y_{12})\bigr).
\end{aligned}
$$

We now show that the six conditions of Theorem 12.3 in Wooldridge (2002) are satisfied. We will refer to the above formulations of the score and Hessian throughout.

1. *The parameter γ_o must be in the interior of Γ*. This condition is satisfied by Theorem 1(vi).

2. *The function $s(w,\cdot)$ is continuously differentiable on the interior of Γ for all $w \in \mathcal{W}$*. Since $H(w,\gamma)$ is continuous in γ, $s(w,\cdot)$ is continuously differentiable.

3. *Each element of $H(w,\gamma)$ is bounded in absolute value by a function $b(w)$, where $E[b(w)] < \infty$*. We derive a bounding function $b(w)$ in the manner

$$
\begin{aligned}
H(w,\gamma) &= \frac{(1-3\gamma^2)((y_3-y_2)^2-(y_{11}-y_{12})^2)-2\gamma(3-\gamma^2)(y_3-y_2)(y_{11}-y_{12})}{(1+\gamma^2)^3}, \\
\bigl|H(w,\gamma)\bigr| &\le \bigl|\bigl(1-3\gamma^2\bigr)\bigl((y_3-y_2)^2-(y_{11}-y_{12})^2\bigr) \\
&\quad - 2\gamma\bigl(3-\gamma^2\bigr)(y_3-y_2)(y_{11}-y_{12})\bigr|, \\
\bigl|H(w,\gamma)\bigr| &\le \bigl|\bigl(1-3\gamma^2\bigr)\bigl((y_3-y_2)^2-(y_{11}-y_{12})^2\bigr)\bigr| \\
&\quad + \bigl|2\gamma\bigl(3-\gamma^2\bigr)(y_3-y_2)(y_{11}-y_{12})\bigr|,
\end{aligned}
$$

$$
\begin{aligned}
\left|H(w,\gamma)\right| &\le \left(1+3\gamma^2\right)\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right)\\
&\quad+\left(3+\gamma^2\right)\left|2\gamma(y_3-y_2)(y_{11}-y_{12})\right|,\\
\left|H(w,\gamma)\right| &\le \left(1+3\gamma^2\right)\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right)\\
&\quad+\left(3+\gamma^2\right)\left|\gamma\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right)\right|,\\
\left|H(w,\gamma)\right| &\le \left(1+3\gamma^2\right)\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right)\\
&\quad+\left(3+\gamma^2\right)|\gamma|\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right),
\end{aligned}
$$

where the second to last line utilizes the fact that $(y_3-y_2)^2+(y_{11}-y_{12})^2 > 2(y_3-y_2)(y_{11}-y_{12})$ as $((y_3-y_2)-((y_{11}-y_{12}))^2 > 0$. Let $\overline{\gamma}$ and $\underline{\gamma}$ denote the largest and smallest elements of the set Γ. The γ that maximizes the right-hand side is given by $\gamma^* = \max\{\overline{\gamma}, -\underline{\gamma}\} < \infty$. Our bounding function is then

$$
\begin{aligned}
b(w) &= \left(1+3\gamma^{*2}\right)\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right)\\
&\quad+\gamma^*\left(3+\gamma^{*2}\right)\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right)\\
&= \left(1+\gamma^*\left(\gamma^{*2}+3\gamma^*+3\right)\right)\left((y_3-y_2)^2+(y_{11}-y_{12})^2\right),
\end{aligned}
$$

where we have shown that $b(w) \ge H(w,\gamma)$ for all w. Notice that the absolute value of γ is no longer necessary, since by definition γ^* is always positive.

We now show that $E[b(w)] < \infty$, completing the proof:

$$
E\left[b(w)\right] = \left(1+\gamma^*\left(\gamma^{*2}+3\gamma^*+3\right)\right)E\left[(y_3-y_2)^2+(y_{11}-y_{12})^2\right].
$$

When deriving the bounding function for $q(w,\gamma)$, we showed that $E[(y_3-y_2)^2+(y_{11}-y_{12})^2] < \infty$. Since γ^* is also finite, $E[b(w)] < \infty$.

4. *The equality* $A_o \equiv E[H(w,\gamma_o)]$ *is positive definite*. We first note that we can interchange the expectations and the partial derivatives: $E[H(w,\gamma)] = \partial^2 E[q(w,\gamma)]/\partial\gamma^2$. From Lemma 3, we know that we can write

$$
E\left[q(w,\gamma)\right] = \frac{(\gamma-\gamma_o)^2 E[(\alpha_{2o}-\alpha_{3o})^2]}{2(1+\gamma^2)} + \left(E\left[\varepsilon_{11}^2\right]+E\left[\varepsilon_{12}^2\right]\right)/2.
$$

Note that γ affects two terms: $(\gamma-\gamma_o)^2$ and the denominator. However, because we are going to evaluate the expected Hessian at γ_o, we only need the second derivative of the first term, $(\gamma-\gamma_o)^2$. All of the other partial derivatives will either be multiplied by $(\gamma-\gamma_o)^2$ or $(\gamma-\gamma_o)$, both of which are zero when $\gamma=\gamma_o$. The second derivative of $(\gamma-\gamma_o)^2$ with respect to γ is positive. This second derivative is then multiplied by the expectation of a squared object in the numerator and divided by the sum of squared objects in the denominator. Thus, the expectation of the Hessian evaluated at γ_o is strictly positive.

5. *We have* $E[s(w,\gamma_o)] = 0$. Note that $E[s(w,\gamma)] = \partial E[q(w,\gamma)]/\partial\gamma$. Differentiating $E[q(w,\gamma)]$ with respect to γ leaves terms that are multiplied by $(\gamma-\gamma_o)$ or by $(\gamma-\gamma_o)^2$, implying that if we evaluate the derivative at $\gamma=\gamma_o$, then the expected score is zero.

6. *Each element of* $s(w, \gamma_o)$ *has finite second moment.* Given that the score has only one element, this condition boils down to $E[s(w, \gamma_o)^2] < \infty$. To show this, we square the score function, repeatedly apply the triangle equality, and evaluate the expected value at the true γ:

$$
\begin{aligned}
E\bigl[s(w, \gamma_o)^2\bigr] = E\biggl(&\frac{1}{(1+\gamma_o^2)^4}\bigl[\bigl(1-\gamma_o^2\bigr)(y_{11}-y_{12})(y_3-y_2) \\
&+ \gamma_o\bigl(-(y_{11}-y_{12})^2 + (y_3-y_2)^2\bigr)\bigr]^2\biggr).
\end{aligned}
$$

Repeatedly applying the triangle inequality yields

$$
\begin{aligned}
E\bigl[s(w, \gamma_o)^2\bigr] &\le E\biggl(\frac{1}{(1+\gamma_o^2)^4}\bigl[2\bigl(1-\gamma_o^2\bigr)^2(y_{11}-y_{12})^2(y_3-y_2)^2 \\
&\qquad + 2\gamma_o^2\bigl(-(y_{11}-y_{12})^2 + (y_3-y_2)^2\bigr)^2\bigr]\biggr) \\
&\le E\biggl(\frac{4}{(1+\gamma_o^2)^4}\bigl[2\bigl(1-\gamma_o^2\bigr)^2\bigl(y_{11}^2+y_{12}^2\bigr)\bigl(y_3^2+y_2^2\bigr) \\
&\qquad + \gamma_o^2\bigl((y_{11}-y_{12})^4 + (y_3-y_2)^4\bigr)\bigr]\biggr) \\
&\le E\biggl(\frac{4}{(1+\gamma_o^2)^4}\bigl[2\bigl(1-\gamma_o^2\bigr)^2\bigl(y_{11}^4+y_{12}^4+y_3^4+y_2^4\bigr) \\
&\qquad + 4\gamma_o^2\bigl(\bigl(y_{11}^2+y_{12}^2\bigr)^2 + \bigl(y_3^2+y_2^2\bigr)^2\bigr)\bigr]\biggr) \\
&\le E\biggl(\frac{8}{(1+\gamma_o^2)^4}\bigl[\bigl(1-\gamma_o^2\bigr)^2\bigl(y_{11}^4+y_{12}^4+y_3^4+y_2^4\bigr) \\
&\qquad + 4\gamma_o^2\bigl(y_{11}^4+y_{12}^4+y_3^4+y_2^4\bigr)\bigr]\biggr) \\
&\le E\biggl(\frac{8}{(1+\gamma_o^2)^2}\bigl[y_{11}^4+y_{12}^4+y_3^4+y_2^4\bigr]\biggr) \\
&\le \frac{8}{(1+\gamma_o^2)^2}E\bigl(y_{11}^4+y_{12}^4+y_3^4+y_2^4\bigr).
\end{aligned}
$$

Now we substitute for y with the data generating process. Consider $E[y_{11}^4]$, which is given by

$$
E\bigl[y_{11}^4\bigr] = E\bigl[(\alpha_{1o} + \gamma_o\alpha_{2o} + \varepsilon_{11})^4\bigr].
$$

Repeatedly applying the triangle inequality yields

$$
\begin{aligned}
E\bigl[y_{11}^4\bigr] &\le 9E\bigl[\bigl(\alpha_{1o}^2 + \gamma_o^2\alpha_{2o}^2 + \varepsilon_{11}^2\bigr)^2\bigr] \\
&\le 27\bigl(E\bigl[\alpha_{1o}^4\bigr] + \gamma_o^4E\bigl[\alpha_{2o}^4\bigr] + E\bigl[\varepsilon_{11}^4\bigr]\bigr).
\end{aligned}
$$

Theorem 1(iii) and (iv) ensure that all of the terms on the right-hand side of the inequality in the above equation are finite. Thus, $E[y_{11}^4]$ is finite. By a similar argument, it can be shown that all the terms in the expectation of the squared score are finite. □

A.2 *Proof of Theorem* 2

The first-order condition for α_i can be written as

$$
\begin{aligned}
0 = {} & \sum_{t=1}^{T}\left(Y_{itn} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j\in\mathbb{M}_{tn\sim i}} \alpha_j\right) \\
& + \sum_{t=1}^{T} \sum_{j\in\mathbb{M}_{tn\sim i}} \frac{\gamma}{M_{tn}}\left(Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k\in\mathbb{M}_{tn\sim j}} \alpha_k\right).
\end{aligned}
\tag{15}
$$

Solving for α_i and collecting terms, we have

$$
\begin{aligned}
\alpha_i = {} & \sum_{t=1}^{T}\left[Y_{itn} - \frac{\gamma}{M_{tn}} \sum_{j\in\mathbb{M}_{tn\sim i}} \alpha_j\right. \\
& \left. + \frac{\gamma}{M_{tn}} \sum_{j\in\mathbb{M}_{tn\sim i}} \left(Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k\in\mathbb{M}_{tn\sim j\sim i}} \alpha_k\right)\right] \Big/ \left(T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}}\right).
\end{aligned}
\tag{16}
$$

Now we stack these equations, such that the $N \times 1$ vector of α's runs down the left-hand side of the stack. To apply our iterative method, we make a first guess at this vector and then use this guess to generate OLS-derived estimates of the other parameters appearing in the model. Once obtained, these estimates are then plugged into the right-hand side of these equations and we update our guess of the α vector. Let the first of any two consecutive guesses of the α vector be called simply α, and let the second (updated) guess be called α'. We would like to show that our mapping, call it f, from $\alpha \to \alpha'$ is a contraction mapping; that is, $\rho(f(\alpha), f(\alpha')) < \beta\rho(\alpha, \alpha')$ for some $\beta < 1$, where ρ is a valid distance function. Using a Euclidean distance function for ρ, our task is then to show the conditions under which, for a chosen $\beta < 1$,

$$
\begin{aligned}
& \left(\sum_{i=1}^{N}\left(-\sum_{t=1}^{T}\left[\frac{\gamma}{M_{tn}} \sum_{j\in\mathbb{M}_{tn\sim i}} \tilde{\alpha}_j\right.\right.\right. \\
& \left.\left.\left. + \frac{\gamma}{M_{tn}} \sum_{j\in\mathbb{M}_{tn\sim i}} \left(\tilde{\alpha}_j + \frac{\gamma}{M_{tn}} \sum_{k\in\mathbb{M}_{tn\sim j\sim i}} \tilde{\alpha}_k\right)\right] \Big/ \left(T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}}\right)\right)^2\right)^{1/2}
\end{aligned}
\tag{17}
$$

will be less than

$$
\beta\left(\sum_{i=1}^{N} \tilde{\alpha}_i^2\right)^{1/2},
\tag{18}
$$

where $\tilde{\alpha} = \alpha - \alpha'$ and N again refers to the total student population. Factoring out the α's, this requirement can be rewritten as

$$\left(\sum_{i=1}^{N}\left(-\frac{\sum_{t=1}^{T}\left[\left(\frac{2\gamma}{M_{tn}}+\frac{\gamma^2(M_{tn}-1)}{M_{tn}^2}\right)\sum_{j\in\mathbb{M}_{nt\sim i}}\tilde{\alpha}_j\right]}{T+\sum_{t=1}^{T}\frac{\gamma^2}{M_{tn}}}\right)^2\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{19}$$

Expanding the inner square on the left-hand side of the inequality and repeatedly applying the triangle inequality yields

$$\left(\sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{T}T\left[\left(\frac{2\gamma}{M_{tn}}+\frac{\gamma^2(M_{tn}-1)}{M_{tn}^2}\right)^2\left(\sum_{j\in\mathbb{M}_{nt\sim i}}\tilde{\alpha}_j\right)^2\right]}{\left(T+\sum_{t=1}^{T}\frac{\gamma^2}{M_{tn}}\right)^2}\right)\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{20}$$

Expanding the square on the sum of the $\tilde{\alpha}_j$'s and applying the triangle inequality leaves

$$\left(\sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{T}T\left[\left(\frac{2\gamma}{M_{tn}}+\frac{\gamma^2(M_{tn}-1)}{M_{tn}^2}\right)^2 M_{tn}\sum_{j\in\mathbb{M}_{nt\sim i}}\tilde{\alpha}_j^2\right]}{\left(T+\sum_{t=1}^{T}\frac{\gamma^2}{M_{tn}}\right)^2}\right)\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{21}$$

Inside the square brackets of equation (22), there are no $\tilde{\alpha}_i$, since this term reflects the purged first-order condition from individual i. However, $\tilde{\alpha}_i$ will be present in the first-order condition from all of i's classmates over time. Because the M_{tn} in the denominator reflect the peer group sizes experienced by individual i over time, all the terms on the left-hand side of the inequality containing an $\tilde{\alpha}_i$ will have different denominators. Substituting $\overline{M}$ for M_{tn} in the denominator ensures a common denominator across the terms containing an $\tilde{\alpha}_i$:

$$\left(\sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{T}T\left[\left(\frac{2\gamma}{M_{tn}}+\frac{\gamma^2(M_{tn}-1)}{M_{tn}^2}\right)^2 M_{tn}\sum_{j\in\mathbb{M}_{tn\sim i}}\tilde{\alpha}_j^2\right]}{T^2\left(1+\frac{\gamma^2}{\overline{M}}\right)^2}\right)\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{22}$$

This substitution is valid, since it shrinks the denominator for every term on the left-hand side of the inequality, making it less likely to hold. Now we can easily collect all the terms containing an $\tilde{\alpha}_i$, yielding

$$\left(\sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{T} T\left[\left(\frac{2\gamma}{M_{tn}}+\frac{\gamma^2(M_{tn}-1)}{M_{tn}^2}\right)^2 M_{tn}^2\tilde{\alpha}_i^2\right]}{T^2\left(1+\frac{\gamma^2}{\overline{\overline{M}}}\right)^2}\right)\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{23}$$

The additional M_{tn} term in the numerator comes from the fact that $\tilde{\alpha}_i$ will show up once for each of the M_{tn} peers at time t. Bringing the M_{tn}^2 inside the parentheses in the numerator yields

$$\left(\sum_{i=1}^{N}\left(\frac{\sum_{t=1}^{T} T\left[\left(2\gamma+\gamma^2-\frac{\gamma^2}{M_{tn}}\right)^2\right]}{T^2\left(1+\frac{\gamma^2}{\overline{\overline{M}}}\right)^2}\right)\tilde{\alpha}_i^2\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{24}$$

Notice that we can again substitute for M_{tn} with $\overline{M}$, since this will strictly increase the coefficient on $\tilde{\alpha}_i$, making it less likely that the inequality is satisfied. Making this substitution and canceling the T^2 terms leaves

$$\left(\sum_{i=1}^{N}\left(\frac{\left(2\gamma+\gamma^2-\frac{\gamma^2}{\overline{\overline{M}}}\right)^2}{\left(1+\frac{\gamma^2}{\overline{\overline{M}}}\right)^2}\right)\tilde{\alpha}_i^2\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{25}$$

which can be rewritten as

$$\frac{2\gamma+\gamma^2-\frac{\gamma^2}{\overline{\overline{M}}}}{1+\frac{\gamma^2}{\overline{\overline{M}}}}\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2} < \beta\left(\sum_{i=1}^{N}\tilde{\alpha}_i^2\right)^{1/2}. \tag{26}$$

As long as the γ's are such that (26) is satisfied, we have a contraction mapping. The denominator of the leading term is strictly greater than 1, implying that if the numerator is strictly less than 1 then the contraction holds for some $\beta < 1$. If $2\gamma+\gamma^2 < 1$, the numerator will be strictly less than 1, which is true for $\gamma \le 0.4$.[46]

[46] An identical restriction on γ is required in the case of an unbalanced panel. To derive this, simply define ρ as a weighted Euclidean distance where the individual weights are given by the number of observations for student i, T_i.

Appendix B: Endogenous effects

In this section, we show how our framework can be incorporated to allow for endogenous effects. We introduce a new variable, Z_{itn}, that affects the choices of the individual but affects his peers only through the individual's choice. Throughout, we assume that Z_{itn} is uncorrelated with all the ε's. For ease of notation, we also focus on the case where peer groups consist of only two individuals.[47] We first consider the case where individuals have total control of the outcome: the outcome of interest is a choice. We then consider the case that is most relevant to our empirical work, where individuals only have partial control over the outcome.

B.1 *Total control*

We first consider the case where Y_{itn} is directly affected by Y_{jtn}. In this case, the linear model is

$$Y_{itn} = \alpha_{io} + \gamma_o \alpha_{jo} + \phi_o Y_{jtn} + \theta_o Z_{itn} + \varepsilon_{itn}. \tag{27}$$

Substituting the expression for Y_{jtn} into (27) and solving for Y_{itn} yields

$$\begin{aligned} Y_{itn} = {} & \left(\frac{1+\phi_o\gamma_o}{1-\phi_o^2}\right)\alpha_{io} + \left(\frac{\gamma_0+\phi_o}{1-\phi_o^2}\right)\alpha_{jo} \\ & + \frac{\theta_o Z_{itn}}{1-\phi_o^2} + \frac{\phi_o\theta_o Z_{jtn}}{1-\phi_o^2} + \frac{\varepsilon_{itn}+\phi_o\varepsilon_{jtn}}{1-\phi_o^2}. \end{aligned} \tag{28}$$

Note that the last term—the reduced-form error—has both ε_{itn} and ε_{jtn}. The reduced-form errors will then be correlated between individuals who share a peer group, violating Theorem 1(iii). In estimation, this correlation is partially absorbed by the peer fixed effects, which in turn prohibits consistent estimation of the coefficient on α_{jo}. Our conclusion is that when the outcome variable is a choice that is affected by the actual choices of one's peers, we cannot obtain a consistent estimate of the parameter on the peer fixed effects for fixed T. Note, however, that if the spillovers operated only through observables, which would imply replacing the α_{io}'s with $X\beta$, then all of the structural parameters would be identified.

We now consider the case where individuals only have expectations about what their peers will choose. This situation maps well to a wide variety of outcomes where the behavior of others either is not perfectly observed or occurs at exactly the same time as own behavior and, therefore, cannot be a direct input to own behavior. In particular, suppose that ε_{jt} is unknown to individual i and has mean zero. The outcome equation is then

$$Y_{itn} = \alpha_{io} + \gamma_o \alpha_{jo} + \phi_o E(Y_{jtn}) + \theta_o Z_{itn} + \varepsilon_{itn}. \tag{29}$$

Again substituting in for Y_{jtn} and solving for Y_{itn} yields

$$Y_{itn} = \left(\frac{1+\phi_o\gamma_o}{1-\phi_o^2}\right)\alpha_{io} + \left(\frac{\gamma_0+\phi_o}{1-\phi_o^2}\right)\alpha_{jo} + \frac{\theta_o Z_{itn}}{1-\phi_o^2} + \frac{\phi_o\theta_o Z_{jtn}}{1-\phi_o^2} + \frac{\varepsilon_{itn}}{1-\phi_o^2}. \tag{30}$$

[47] Results for larger peer groups are available upon request.

Theorem 1(iii), which states that the reduced-form error is uncorrelated between peer group members, is no longer violated by the model. We can then write (30) as

$$Y_{itn} = \alpha^*_{io} + \gamma^*_o \alpha_{jo} + \theta^*_o Z_{itn} + \phi^*_o Z_{jtn} + \varepsilon^*_{itn}, \tag{31}$$

where

$$\alpha^*_{io} = \left(\frac{1+\phi_o\gamma_o}{1-\phi_o^2}\right)\alpha_{io},$$

$$\gamma^*_o = \left(\frac{\gamma_0+\phi_o}{1+\phi_o\gamma_o}\right),$$

$$\theta^*_o = \frac{\theta_o}{1-\phi_o^2}, \qquad \phi^*_o = \phi_o\theta^*_o, \qquad \varepsilon^* = \frac{\varepsilon_{itn}}{1-\phi_o^2}.$$

Estimating the reduced form then makes it possible to recover all the structural parameters, as would also hold in the standard case where the α_{io}'s were replaced by a set of observables multiplied by a vector of coefficients. We can recover $\hat{\phi}$ and $\hat{\theta}$ from $\hat{\phi}^*$ and $\hat{\theta}^*$. Next, given $\hat{\phi}$, we can obtain $\hat{\gamma}$ using $\hat{\gamma}^*$ as $\hat{\gamma} = (\hat{\gamma}^* - \hat{\phi})/(1 - \hat{\gamma}^*\hat{\phi})$.

One key identifying assumption in this case is that the expected choices of the individual's peers are formed on the basis of observed characteristics and the peer fixed effects, both of which are uncorrelated with the structural errors. Identification of the underlying parameters using our fixed-effects approach also requires Z_{itn} to be time-varying. If it is not, then Z_{itn} would be absorbed into the reduced-form individual effect and we would be back to using two coefficients to recover three parameters. We would be left with the same estimating equation as the baseline model, and the reduced form would be a linear combination of own and peer fixed effects plus the Z values of the peers, but we could not separate out the endogenous effects from the exogenous effects. Note that in the case that spillovers operate only through observable characteristics, Z_{itn} is only required to vary across individuals, not within person.

B.2 *Partial control*

As pointed out by Cooley (2009a, 2009b), the estimation issues become much more complicated when individuals only have partial control over their outcomes. For example, in educational settings where grades are the outcome of interest, it is not the grades of the other students in the class that affect the student's grades, but the effort the other students exert. Moreover, students cannot directly choose their grades, but can only choose effort levels, which in turn combine with other forces (including peer effort) to determine their grades. Separating out endogenous and exogenous effects is much harder in this case.

We now show what we can identify when individuals make choices that only partially affect their outcome, and where the choices of others influence both own choices and own outcomes. Let e_{itn} indicate the continuous choice individuals make to affect outcome Y_{itn}. Adding e_{itn} and e_{jtn} to the baseline model as direct influences on outcomes yields

$$Y_{itn} = \alpha_{io} + \phi_{1o} e_{itn} + \gamma_o \alpha_{jo} + \phi_{2o} e_{jtn} + \varepsilon_{itn}. \tag{32}$$

The utility associated with choosing a particular value of e_{itn} depends on the individual's fixed effect, α_{io}, as well as on the choices of the other individual and their individual effect. Similar to the previous case, we assume that there is an additional variable, Z_{itn}, that affects the choice of effort. We assume that the utility function takes the form

$$\begin{aligned}&U\big(e_{itn}, E(Y_{itn})\big)\\&\quad = E(Y_{itn}) + e_{itn}(\lambda_{1o}\alpha_{io} + \lambda_{2o}Z_{itn} + \lambda_{3o}e_{jtn} + \lambda_{4o}\alpha_{jo}) - e_{itn}^2/2,\end{aligned} \tag{33}$$

where we have normalized the coefficient on the squared term. The first-order condition from maximizing (33) with respect to e_{itn} and solving for e_{itn} implies that own optimal effort can be written as

$$e_{itn} = \phi_{1o} + \lambda_{1o}\alpha_{io} + \lambda_{2o}Z_{itn} + \lambda_{3o}e_{jtn} + \lambda_{4o}\alpha_{jo}. \tag{34}$$

Substituting for e_{jtn} from j's maximization problem then yields:

$$\begin{aligned}e_{itn} &= \big((1+\lambda_{3o})\phi_{1o} + (\lambda_{1o} + \lambda_{3o}\lambda_{4o})\alpha_{io}\\&\quad + (\lambda_{4o} + \lambda_{3o}\lambda_{1o})\alpha_{jo} + \lambda_{2o}Z_{itn} + \lambda_{3o}\lambda_{2o}Z_{jtn}\big)\big/\big(1-\lambda_{3o}^2\big).\end{aligned} \tag{35}$$

Substituting for e_{itn} and e_{jtn} in equation (32) and collecting terms implies we can rewrite (32) as

$$Y_{itn} = \alpha_{io}^* + \phi_{1o}^* Z_{itn} + \gamma_o^*\alpha_{jo}^* + \phi_{2o}^* Z_{jtn} + \varepsilon_{it}^*, \tag{36}$$

where

$$\alpha_{io}^* = C + \left(1 + \frac{\phi_{1o}(\lambda_{1o} + \lambda_{3o}\lambda_{4o}) + \phi_{2o}(\lambda_{4o} + \lambda_{3o}\lambda_{1o})}{1-\lambda_{3o}^2}\right)\alpha_i,$$

$$\gamma_o^* = \frac{((1-\lambda_{3o}^2)\gamma_o + \phi_{2o}(\lambda_{1o} + \lambda_{3o}\lambda_{4o}) + \phi_{1o}(\lambda_{4o} + \lambda_{3o}\lambda_{1o}))}{(1-\lambda_{3o}^2 + \phi_{1o}(\lambda_{1o} + \lambda_{3o}\lambda_{4o}) + \phi_{2o}(\lambda_{4o} + \lambda_{3o}\lambda_{1o}))},$$

$$\phi_{1o}^* = \frac{\lambda_2(\phi_{1o} + \lambda_{3o}\phi_{2o})}{1-\lambda_{3o}^2}, \qquad \phi_{2o}^* = \frac{\lambda_2(\phi_{2o} + \lambda_{3o}\phi_{1o})}{1-\lambda_{3o}^2}, \qquad \varepsilon_{itn}^* = \frac{\varepsilon_{itn}}{1-\lambda_{3o}^2},$$

and where C is the adjustment to α_{io}^* coming from the ϕ_o terms that are not multiplying a regressor.

Reduced-form estimation then yield estimates of three coefficients, $\hat{\phi}_1^*$, $\hat{\phi}_2^*$, and $\hat{\gamma}^*$, that are functions of six underlying parameters. What we can say is that $\hat{\phi}_1^* > 0$ implies that individual effort either directly affects the outcome or affects the outcome through the other individual's effort, which in turn affects the individual's outcome. Similarly, if the coefficient on Z_{jtn}, $\hat{\phi}_2^*$, is greater than zero, we can conclude that peer effort matters in some form, either directly or through affecting the individual's own effort. Once again, these results are essentially identical to those in Cooley (2009b), subject to replacing observable characteristics with individual effects.

References

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002), "Computing person and firm effects using linked longitudinal employer–employee data." Technical report, Center for Economic Studies, U.S. Census Bureau. [429]

Abowd, J. M. and F. Kramarz (1999), "The analysis of labor markets using matched employer–employee data." In *Handbook of Labor Economics*, Vol. 3B (O. C. Ashenfelter and D. Card, eds.), chapter 40, 2629–2710, Elsevier Science B.V., Amsterdam. [429]

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999), "High wage workers and high wage firms." *Econometrica*, 67 (2), 251–333. [428, 429]

Altonji, J. G., T. E. Elder, and C. R. Taber (2005), "Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools." *Journal of Political Economy*, 113 (1), 151–184. [443]

Altonji, J. G., C.-I. Huang and C. R. Taber (2004), "Estimating the cream skimming effect of private school vouchers on public school students." Working paper 16579, NBER. [422, 425]

Arcidiacono, P. (2004), "Ability sorting and the returns to college major." *Journal of Econometrics*, 121 (1–2), 343–375. [442]

Arcidiacono, P., J. Cooley, and A. Hussey (2008), "The economic returns to an MBA." *International Economic Review*, 49 (3), 873–899. [442]

Arcidiacono, P. and S. Nicholson (2005), "Peer effects in medical school." *Journal of Public Economics*, 89 (2–3), 327–350. [422]

Arcidiacono, P. and J. Vigdor (forthcoming), "Does the river spillover? Estimating the economic returns to attending a racially diverse college." *Economic Inquiry*. [442]

Arellano, M. and J. Hahn (2006), "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects." Working paper wp2006_0613, CEMFI. [423]

Bester, C. A. and C. Hansen (forthcoming), "A penalty function approach to bias reduction in non-linear panel models with fixed effects." *Journal of Business & Economics Statistics*. [423]

Betts, J. R. and D. Morell (1999), "The determinants of undergraduate grade point average: The relative importance of family background, high school resources, and peer effects." *Journal of Human Resources*, 32 (2), 268–293. [422]

Bramoulle, Y., H. Djebbari, and B. Fortin (2009), "Identification of peer effects through Social Networks." *Journal of Econometrics*, 150 (1), 41–55. [431]

Burke, M. and T. Sass (forthcoming), "Classroom peer effects and student achievement." *Journal of Labor Economics*. [428]

Carrell, S., R. Fullerton, and J. West (2009), "Does your cohort matter? Measuring peer effects in academic achievement." *Journal of Labor Economics*, 27, 439–464. [422]

Carrell, S. and M. Hoekstra (2010), "Externalities in the classroom: How domestic violence harms everyone's kids." *American Economic Journal: Applied Economics*, 2, 211–228. [422]

Carrell, S., J. West, and F. Malmstrom (2008), "Peer effects in academic cheating." *Journal of Human Resources*, 43, 173–207. [422]

Chandrasekhar, A. G. and R. Lewis (2011), "Econometrics of sampled networks." Unpublished manuscript, MIT. [422]

Cooley, J. (2009a), "Alternative mechanisms of peer spillovers in the classroom: Implications for identification and policy." Working paper, University of Cambridge. [431, 465]

Cooley, J. (2009b), "Desegregation and the achievement gap: Do diverse peers help?" Working paper, University of Cambridge. [465, 466]

Davidson, R. and J. G. MacKinnon (2006), "Bootstrap methods in econometrics." In *Palgrave Handbook of Econometrics*, Vol. 1 (T. Mills and K. Patterson, eds.), 812–838, Palgrave Macmillan, Basingstoke. [439]

Durlauf, S. N. and Y. M. Ioannides (2001), "Social interactions." *Annual Review of Economics*, 2, 451–478. [431]

Epple, D. and R. Romano (2011), "Peer effects in education: A survey of the theory and evidence." In *Handbook of Social Economics* (J. Benhabib, M. Jackson, and A. Bisen, eds.), 1053–1164, Elsevier, Amsterdam. [422]

Epple, D., R. Romano, and H. Sieg (2003), "Peer effects, financial aid and selection into colleges and universities: An empirical analysis." *Journal of Applied Econometrics*, 18, 501–525. [422]

Evans, W. N., W. E. Oates, and R. M. Schwab (1992), "Measuring peer group effects: A study of teenage behavior." *Journal of Political Economy*, 100 (5), 966–991. [422]

Fernandez-Val, I. (forthcoming), "Fixed effects estimation of structural parameters and marginal effects in panel probit models." *Journal of Econometrics*. [423]

Foster, G. (2006), "It's not your peers, and it's not your friends: Some progress toward understanding the peer effect mechanism." *Journal of Public Economics*, 90, 1455–1475. [422]

Giorgi, G. D., M. Pelizzari, and S. Redaellil (2007), "Be as careful of the books you read as of the company you keep: Evidence on peer effects in educational choices." Working paper 14948, NBER. [431]

Graham, B. (2008), "Identifying social interactions through conditional variance restrictions." *Econometrica*, 76, 643–660. [422]

Hahn, J. and W. Newey (2004), "Jackknife and analytical bias reduction for nonlinear panel models." *Econometrica*, 72, 1295–1319. [423]

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003), "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18 (5), 527–544. [422]

Harris, D. and T. Sass (2006), "Value-added models and the measurement of teacher quality." Unpublished manuscript, Urban Institute. [428]

Harris, D. and T. Sass (2011), "Teacher training, teacher quality and student achievement." *Journal of Public Economics*, 95, 798–812. [428]

Honore, B. (1992), "Trimmed LAD and least squares estimation of truncated and censored models with fixed effects." *Econometrica*, 60, 533–565. [423]

Horowitz, J. L. and S. Lee (2004), "Semiparametric estimation of a panel data proportional hazard model with fixed effects." *Journal of Econometrics*, 119, 155–198. [423]

Hoxby, C. (2001), "Peer effects in the classroom: Learning from gender and race variation." Working paper 7867, NBER. [422]

Hoxby, C. and G. Weignarth (2005), "Taking race out of the equation: School reassignment and the structure of peer effects." Working paper, Harvard University. [432]

Lehrer, S. and W. Ding (2007), "Do peers affect student achievement in China's secondary schools?" *Review of Economics and Statistics*, 89 (2), 300–312. [422]

Lin, X. (2010), "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables." *Journal of Labor Economics*, 28 (4), 825–860. [431]

Manski, C. F. (1987), "Semiparametric analysis of random effects linear models from binary panel data." *Econometrica*, 55, 357–362. [423]

Mas, A. and E. Moretti (2009), "Peers at work." *American Economic Review*, 99, 112–145. [422, 427]

Neyman, J. and E. Scott (1948), "Consistent estimates based on partially consistent observations." *Econometrica*, 16, 1–32. [423]

Paglin, M. and A. Rufolo (1990), "Heterogeneous human capital, occupational choice, and male–female earnings differences." *Journal of Labor Economics*, 8 (1), 123–144. [445]

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005), "Teachers, schools, and academic achievement." *Econometrica*, 73 (2), 417–458. [428]

Sacerdote, B. I. (2001), "Peer effects with random assignment: Results for dartmouth roommates." *Quarterly Journal of Economics*, 116 (2), 681–704. [422, 441]

Winston, G. C. and D. J. Zimmerman (2003), "Peer effects in higher education." In *College Decisions: How Students Actually Make Them and How They Could* (C. Hoxby, ed.), University of Chicago Press, Chicago. [422]

Wooldridge, J. M. (2002), *Econometric analysis of cross-section and panel data*. MIT Press, Cambridge. [451, 457, 458]

Woutersen, T. (2002), "Robustness against incidental parameters." Unpublished manuscript, University of Arizona. [423]

Zimmerman, D. J. (2003), "Peer effects in higher education: Evidence from a natural experiment." *Review of Economics and Statistics*, 85 (1), 9–23. [422]

Submitted March, 2011. Final version accepted March, 2012.