

Estimating Spillovers using Panel Data, with an Application to the Classroom*

Peter Arcidiacono, Duke University
Gigi Foster, University of South Australia
Natalie Goodpaster, Analysis Group, Inc.
Josh Kinsler, University of Rochester

February 12, 2009

Abstract

Obtaining consistent estimates of spillovers in an educational context is hampered by 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 non-linear least squares estimate of the spillover is consistent and asymptotically normal as $N \rightarrow \infty$ with T fixed. We then provide an iterative estimation algorithm that is easy to implement and that converges to the non-linear 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.

*We thank Joe Altonji, Pat Bayer, 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 was 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.

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. 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. The first is the selection problem. When individuals choose their peer groups, high ability¹ 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,² 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 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 has received little attention outside of Altonji et al. (2004). Random assignment is able to circumvent the selection problem but the obstacle to estimation posed by peer effects through unobservables remains.

We introduce a new algorithm for estimating spillovers using panel data that overcomes both these obstacles. The key innovation is that the peer effects are captured through a linear combination of individual fixed effects. Additional types of fixed effects, such as institutional

¹For ease of exposition we refer to the bundle of individuals' performance-relevant characteristics as 'ability.'

²One set of papers uses proxy variables to break the link between unobserved and peer ability (Arcidiacono and Nicholson (2005), Hanushek et al. (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), and Hoxby (2001)). Finally, researchers have tried to circumvent the endogeneity problem with instrumental variables (Evans et al. (1992)).

effects, can also be incorporated into the algorithm. Thus our method, while originally developed for use in social effects research, can also be used in applications where fixed effects of large dimension and of potentially multiple types must be estimated.

Constructing the spillover as a linear combination of individual fixed effects results in a non-linear 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.³ As N goes to infinity for a fixed T , the estimation error for the individual effects often does not vanish as the sample size grows, contaminating the estimates of the parameters of interest.⁴ We show, however, that the nonlinear least squares estimate of the spillover parameter is consistent and asymptotically normal as $N \rightarrow \infty$ with T fixed, even though the fixed effects themselves are not consistent.⁵

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.

As a precursor to the exposition of our full spillover model, we show how our iterative estimator can be used to identify multiple types of fixed effects in contexts where no spillovers are being estimated. Then, we customize the model for application to peer effects in education, and estimate the customized model 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 where ability is allowed

³Neyman and Scott (1948) were the first to document the incidental parameters problem.

⁴Hahn and Newey (2004) provide two methods of bias correction: a panel jackknife and an analytical correction. Woutersen (2002) and Fernandez-Val (2005) consider estimators from bias-corrected moment conditions. In a similar vein, Arellano and Hahn (2006) and Bester and Hansen (2005) consider bias-correcting the initial objective function.

⁵Other special cases where the incidental parameters problem does not require a bias correction are Manski (1987), Honore (1992), and Horowitz and Lee (2004).

to vary by course type. We find significant peer effects which vary by course type. A one standard deviation increase in peer ability yields average returns similar to those from between a 3 percent and a 11 percent of a standard deviation increase in own ability, depending upon the course type and specification. The lowest returns are found in math and science with the highest returns 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 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.⁶ Because we estimate our model separately for each course type, we can compare student ability in the primary course of study to 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 re-introduce 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 downwards since the peer unobservables problem dominates the selection issue. 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 is due in part to the much higher correlation

⁶By construction, observed and unobserved ability are orthogonal in the population.

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 discusses our iterative estimation algorithm in the context of some outstanding problems in the applied microeconomics literature. Section 3 presents the model, its properties, our estimation strategy, and evidence from Monte Carlo simulations. Section 4 describes the University of Maryland data, and Section 5 presents the results. Section 6 explores selection within and across course types and Section 7 illustrates the biases associated with traditional peer effect measures. Section 8 concludes.

2 Estimating Models with Multiple Types of Fixed Effects

An outstanding problem in the applied microeconomic literature is how to estimate models containing multiple types of fixed effects where each set of fixed effects is of a large dimension. 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 then present our full spillover model in the next section. We focus on two papers in particular: Rivkin et al. (2005) and Abowd et al. (1999), to illustrate the difficulties in estimating large numbers of fixed effects.

Rivkin et al. (2005) model gains in test scores as a function of the observed characteristics of the students, X_i , and teacher fixed effects, δ_j , where i indexes individuals and j indexes teachers. 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:

$$\Delta Y = \alpha X_i + \delta_j + \epsilon_{it} \quad (1)$$

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 δ_j '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:

$$\Delta Y = \alpha_i + \delta_j + \epsilon_{it} \quad (2)$$

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

Abowd et al. (1999) are interested in estimating models that include both firm and worker fixed effects. The most basic model they are interested in estimating contains just individual and firm-specific effects regressed on log earnings. Labeling log earnings for individual i in firm j at time t as Y , the outcome equation is:

$$Y = \alpha_i + \delta_j + \epsilon_{it} \quad (3)$$

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 as the amount of tenure individual i has in firm j at time t , the outcome equation is:

$$Y = \alpha_i + \delta_j + \phi_j X + \epsilon_{it} \quad (4)$$

where ϕ_j is the firm-specific return to tenure. Abowd and Kramarz (1999) recognize that with over 1 million workers and 500,000 firms, they cannot estimate the above equation directly. Instead, they consider 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.

Our approach yields least squares estimates of both firm and worker effects in a computationally feasible way without imposing any extraneous orthogonality conditions. Consider the following regression equation which nests the models from both Rivkin et al. (2005) and Abowd and Kramarz (1999):

$$Y_{ijt} = \alpha_i + \beta X_{it} + \delta_j + \gamma_j Z + \epsilon_{ijt} \quad (5)$$

Estimating this equation by OLS solves:

$$\min_{\alpha, \gamma} \sum_{i=1}^N \sum_{t=1}^T (Y_{ijt} - \alpha_i - \beta X_{it} - \delta_j - \gamma_j Z)^2 \quad (6)$$

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 α 's and the β 's, the algorithm iterates on two steps with the q th iteration following:

- Step 1: Conditional on α^{q-1} and β^{q-1} , estimate δ^q and γ^q by OLS.
- Step 2: Conditional on δ^q and γ^q , estimate α^q and β^q by OLS.

The process continues until the parameters converge. Because the sum of squared errors is decreased at each step, we will eventually converge to the parameter values that minimize the least squares problem in (6), 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 in this section is that it is capable of estimating extremely large sets of fixed effects in a reasonable amount of time. The model becomes significantly more complicated when the outcomes are allowed to depend on functions of the fixed effects themselves. We discuss this enhancement now.

3 Estimating Spillovers with Panel Data

In this section we present a model and estimation strategy for measuring achievement spillovers using student fixed effects, building upon the iterative technique outlined in the prior section. The model is constructed keeping in mind that our application will be 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 as well as the fixed effects of the other individuals in a defined peer group. We show that it is possible to obtain consistent estimates of the spillover and that there is a computationally cheap way of obtaining the solution. Although we focus on the case where the spillover effect is homogeneous across individuals, all proofs go through when the spillover is allowed to vary provided the number of parameters governing the spillover effect does not vary with the sample size. We then show how the model is easily expanded to incorporate correlated effects, highlighting again the advantages of iterative estimation in solving the dimensionality problem. Finally, we provide Monte Carlo evidence that the small-sample behavior of our iterative estimator is also appealing. All proofs are in the appendix.

3.1 Baseline Model

Our baseline model has individual i 's outcome at time t , Y_{it} , depending upon his own observed and unobserved permanent characteristics, X_i and u_i , the observed and unobserved permanent characteristics of each of the other students in his peer group, and a transitory error ϵ_{it} . Define I_{ijt} as an indicator function that takes on a value of one if individuals i and j are in the same peer group at time t , and zero otherwise, and let N be the number of students in the population. Our baseline specification is then:

$$Y_{it} = X_i\beta_1 + u_i\beta_2 + \sum_{j \neq i}^N I_{ijt} (X_j\gamma_1 + u_j\gamma_2) + \epsilon_{it} \quad (7)$$

The specification in (7) is restrictive along a number of dimensions. In the language of Manski (1995), there are no endogenous effects as Y_{jt} has no direct effect on Y_{it} for any j , and there are 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. Further, there are no accumulation effects as spillovers today do not affect outcomes tomorrow. However, even in this special case estimation is still problematic when peer groups are chosen. In particular, there may be correlation between u_i and the sum of observed peer characteristics, leading to biased estimates of γ_1 . Also, we will not be 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_i and observed peer characteristics, the inability to capture spillovers through unobservables remains.⁷

We next make one additional assumption: the relevance to outcomes of peer characteristics is proportional to that of own characteristics, meaning that we can write γ_1 and γ_2 as

$$\begin{aligned} \gamma_1 &= \gamma\beta_1 \\ \gamma_2 &= \gamma\beta_2 \end{aligned}$$

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

⁷Using random assignment to identify the spillover also disregards the possibility that spillovers operate differently in selected versus randomized contexts.

Now define α_i as:

$$\alpha_i = X_i\beta_1 + u_i\beta_2$$

We can then rewrite equation (7) as:

$$Y_{it} = \alpha_i + \sum_{j \neq i}^N I_{ijt}(\gamma\alpha_j) + \epsilon_{it} \quad (8)$$

An individual's outcome is then a function of the individual's fixed effect plus a linear function of the fixed effects of the other students in the peer group. Using fixed effects in this way allows us to abstract from many other covariates that may affect student outcomes. All of the heterogeneity in fixed student characteristics that might affect student outcomes, such as birth cohort, sex, IQ, or race is captured with this one measure.

Our goal is then to show the properties of the solution to the non-linear least squares problem:

$$\min_{\alpha, \gamma} \sum_{i=1}^N \sum_{t=1}^T \left(Y_{it} - \alpha_i - \sum_{j \neq i}^N I_{ijt}(\gamma\alpha_j) \right)^2 \quad (9)$$

Given the non-linearities present in the problem, one may suspect that for fixed T , the estimates of γ 's will be biased as a result of the incidental parameters problem. However, we show that under standard assumptions this is not the case.

Theorem 1. *Under the following assumptions, the estimates of γ from solving the minimization problem in (9) are $\sqrt{N}$ consistent and asymptotically normal for fixed T as $N \rightarrow \infty$ where N is the number of peer groups:*

1. $\epsilon_{it} \perp \epsilon_{js} \ \forall \ j \neq i, \ t \neq s$
2. $\epsilon_{it} \perp \alpha_j \ \forall \ i, \ j$
3. $Var(\epsilon_{it})$ is finite and constant across i and t
4. $Var(\alpha_i)$ is finite

The structure of the proof relies on solving each of the α 's as a function of the data and γ and then substituting these functions in for the α 's in (9). Minimizing with respect to γ alone then yields the result.

While concentrating out the α 's is useful for proving consistency, the formulas are quite cumbersome and difficult to calculate. Directly solving (9) 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.

To implement the algorithm, consider the first order condition of the nonlinear least squares problem with respect to α_i :

$$0 = \sum_{t=1}^T \left[\left(Y_{it} - \alpha_i - \sum_{j \neq i}^N I_{ijt}(\gamma \alpha_j) \right) + \sum_{j \neq i}^N I_{ijt} \gamma \left(Y_{jt} - \alpha_j - \sum_{k \neq j}^K I_{jkt}(\gamma \alpha_k) \right) \right] \quad (10)$$

Solving for α_i yields:

$$\alpha_i = \frac{\sum_t \left[Y_{it} - \sum_{j \neq i} I_{ijt} \gamma \alpha_j + \sum_{j \neq i} \gamma \left(Y_{jt} - \alpha_j - \sum_{k \neq i, j} I_{jkt} \gamma \alpha_k \right) \right]}{T + \sum_t \sum_{j \neq i} I_{ijt} \gamma^2} \quad (11)$$

Note that we have pulled out the α_i terms from the last term in (10) to derive (11). We establish in Theorem 2 the conditions under which, given any initial set of α 's, repeatedly updating the α 's using (11) yields a fixed point.

Theorem 2. *Denote $f(\alpha)$ as a function mapping from $\mathcal{R}^N \rightarrow \mathcal{R}^N$ where the i th element of $f(\alpha)$ is given by the right hand side of (11) $\forall i \in N$. A sufficient condition for $f(\alpha)$ to be a contraction mapping is that the maximum of γ/M is less than 0.4, where M is the number of other individuals in each student's peer group.*

The restriction on the maximum value 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 re-estimating γ .⁸ To summarize, the algorithm is started with an initial guess for the α 's and iterates on two steps until convergence, with the q th iteration given by:

⁸For most iterations of our models updating the α 's just once led to a decrease in the sum of squared errors.

- Step 1: Conditional on α^{q-1} , estimate γ^q by OLS.
- Step 2: Conditional on γ^q , update α^q according to (11).

3.2 Correlated Effects

We now extend the model to allow for correlated effects. 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 will affect one's final grade irrespective of own and peer ability. Denote I_{ict} as an indicator for whether individual i is in course c at time t where $c \in [1, C]$. Placing course fixed effects into the achievement equation yields:

$$Y_{it} = \alpha_i + \sum_{j \neq i}^N I_{ijt}(\gamma \alpha_j) + \sum_{c=1}^C I_{ict} \delta_c + \epsilon_{it} \quad (12)$$

The δ_c 's are then the course fixed effects used to capture correlated effects since all peer groups are formed within courses.

The non-linear least squares problem we are now interested in solving takes the following form:

$$\min_{\alpha, \gamma, \delta} \sum_i^N \sum_{t=1}^T \left(Y_{it} - \alpha_i - \sum_{j \neq i}^N I_{ijt}(\gamma \alpha_j) - \sum_{c=1}^C I_{ict} \delta_c \right)^2 \quad (13)$$

While the consistency of γ is unchanged regardless of whether we include other time-varying regressors, it is particularly clear here since we can re-write the above expression without δ_c by demeaning the dependent variable at the course level.

The first order condition with respect to α_i changes to reflect the presence of the course fixed effects:

$$\alpha_i = \frac{\sum_t \left[Y_{it} - \sum_c I_{ict} \delta_c - \sum_{j \neq i} \gamma \alpha_j + \sum_{j \neq i} I_{ijt} \gamma \left(Y_{jt} - \alpha_j - \sum_c I_{jct} \delta_c - \sum_{k \neq i, j} I_{jkt} \gamma \alpha_k \right) \right]}{T + \sum_t \sum_{j \neq i} I_{ijt} \gamma^2} \quad (14)$$

The updating rule for the α 's then follows directly from (14), once α terms are collected for each student.

For a given set of α 's and γ 's, the course fixed effects can be calculated according to:

$$\delta_c = \frac{\sum_i \sum_t I_{ict} (Y_{it} - \alpha_i - \sum_{j \neq i} I_{ijt} \gamma \alpha_j)}{\sum_i \sum_t I_{ict}} \quad (15)$$

The estimation strategy is then the same as without correlated effects, with one additional step. As before, each step of the estimation decreases the sum of squared errors. The algorithm is set by an initial guess of the α 's and δ 's and then iterates, with the q th iteration following:

- Step 1: Conditional on α^{q-1} and δ^{q-1} , estimate γ^q by OLS.
- Step 2: Conditional on δ^{q-1} and γ^q , update α^q according to (14).
- Step 3: Conditional on α^q and γ^q , update δ^q according to (15).

Adding other types of fixed effects simply adds additional steps to the estimation, with each set of fixed effects updated in an additional step. For example, had data been available on teaching assistants, it would have been possible for us to control for teaching assistant effects as well.

3.3 Monte Carlo Simulations

To investigate the properties of our iterative estimator in the presence of correlated effects, we now run simulations using different assumptions about the empirical setting. 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:

1. *Observations per student*- The number of outcomes observed per student varies across simulations between 2, 5, and 10. 5 is the maximum number of observations 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 's.
2. *Students per peer group*- The number of students per peer group varies across simulations between 2 and 10. 2 is the minimum number of observations required to identify a spillover in this type of model. 10 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 's within a peer group equals the standard deviation of α in the population. We also simulate the model with selection such that the average standard deviation of the α 's within a peer group is 75% of the population standard deviation. The selection level in the non-random assignment simulation corresponds with the sorting observed in the Maryland transcript data.⁹

4. *Transitory component*- The noisier the outcome measure, the noisier the estimates of the α 's will be. The distribution of the α 's is set at $N(0,1)$. The ϵ 's are distributed with mean zero and standard deviation equal to 1.15 or 1.95.

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 upon ability. Thus, the average standard deviation of abilities within a class is smaller than standard deviation of abilities in the population.

To verify first that the estimator does not give positive peer effects when no peer effects are present, Table 1 shows simulation results when the true peer effect, γ , is zero. In all cases, regardless of the number of observations per student, the level of selection, and the noisiness of the outcome measure, $\hat{\gamma}$ is precisely estimated around zero. There is no evidence of bias when the true peer effect is zero.

Table 2 documents the model's performance when the true value of γ is 0.15. Again, regardless of assignment procedure or section size, γ is centered around the truth. However, two interesting patterns emerge in the estimates and standard errors of γ . First, γ is more precisely measured when students are randomly assigned to sections. Selection in this case can be thought of as occurring at two levels: the class level and the section level. These results reflect the fact that selection at the class level confounds the estimate of the correlated effect and reduces the precision of the section-level peer effect estimate. In fact, if selection occurred only at the section 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 γ decreases.¹⁰ This is again related

⁹In our data, the standard deviation of ability within peer groups is between 70% and 77% of the standard deviation of ability in the population depending upon the course type.

¹⁰This can be seen in Table 2 since as the number of observations per student increases, we should naturally see an increase in precision—yet we do not, because peer group size increases as well, driving standard errors

to the variation in peer ability across sections. With smaller section sizes, other things equal, there is greater variation in peer ability across sections which yields more precise estimates of the spillover.

4 Data

With the model producing consistent estimates of the spillover parameter and performing well in our Monte Carlos, 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, Fall 1999, Spring 2000, Fall 2000, Spring 2001, or Fall 2001. 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. 90% of University of Maryland entering freshmen live on campus in their first semester,¹¹ so the data set includes at least 90% of the University of Maryland undergraduate population who began study sometime in the six-semester period. 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, multi-section courses in which freshmen predominate.¹² A “section” in the American undergraduate context 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 ability of other students in the same section impacts any given student’s eventual grade in a course.

To generate the student-section-level sample used to estimate our models, we first placed two major restrictions on the data set: students had to have valid A through F grade information for the given section, and they could not be the only student observed in the section that semester.¹³ These restrictions yielded a sample of 300,640 student-section observations. We

up. 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.

¹¹This number is taken from publicly-available statistics posted on the University’s web page.

¹²In tests of whether classes that were under-represented had lower estimated peer effects than those with a complete representation, we found no meaningful differences.

¹³Numeric grade equivalents were assigned as follows: A = 4; B = 3; C = 2; D = 1; and F = 0. Students who

then deleted all observations on sections that were not in one of three well-defined academic subgroups: (1) humanities (86,844 observations), (2) social sciences (77,312 observations), and (3) hard sciences and mathematics (82,675 observations).¹⁴ This left a combined sample of 246,831 student-section observations, representing 18,511 individual students. Sample sizes are provided in Table 3.

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. Later in the paper, we use the following information about students in further analysis of our models' results: SAT math and SAT verbal scores; high school grade point average; sex; race; whether the student was included in the honors program; whether the student participated in sports; and whether the student was an in-state student (as opposed to out-of-state or international).

5 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 only depends upon the mean ability in the section. The second specification allows the size of the spillover to depend upon one's own characteristics. For example, those with high SAT verbal scores may receive higher benefits from their peers than those with 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.

withdrew from a course, audited, or received a non-letter grade (such as Pass) were excluded from the sample due to concerns that they might not have been present during sections and classes. If two separate grades were recorded for the student for a given section, the highest grade was used.

¹⁴Excluded 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.

5.1 Homogeneous Gamma Model

With c and t indexing courses and sections, we have the same specification as in equation (12) except that now, with the number of individuals in each section varying, we restrict the spillover to depend upon the mean fixed effect of the other individuals in the same section of a course:

$$Y_{it} = \alpha_i + \sum_{j \neq i}^N I_{ijt} \left(\frac{\gamma \alpha_j}{M_{it}} \right) + \sum_{c=1}^C I_{ict} \delta_c + \epsilon_{it} \quad (16)$$

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 . 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. Similarly, the teacher of a course may respond to higher average class ability by teaching in a way that enhances learning and therefore raises grades for at least some portion of the class. By including fixed effects at the course level, the δ_c '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.¹⁵ 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.

Adjusting equation (16) to incorporate abilities and peer influence that vary by course

¹⁵Information as to which courses were assigned to which course types is available from the authors upon request.

type yields:

$$Y_{ikt} = \alpha_{ik} + \sum_{j \neq i}^N I_{ijt} \left(\frac{\gamma \alpha_{jk}}{M_{it}} \right) + \sum_{c=1}^C I_{ict} \delta_c + \epsilon_{ikt} \quad (17)$$

where k indexes course type. This model is run through our algorithm 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. Note that while classes with only one section do not help in the identification of γ directly, these classes are still useful in identifying the α 's.

Table 4 shows the results from estimating equation (16) for each of the three types of courses. 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.

In order 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 first two columns of Table 5 show the standard deviation of mean peer ability and the standard deviation of individual ability, respectively. The third column then shows the fraction of a standard deviation of own ability that is equivalent, in terms of its effect on grades, to a one-standard-deviation increase in peer ability. This is calculated by dividing the standard deviation of mean peer ability by the standard deviation of individual ability and multiplying this number by the estimated gamma.

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 one-standard-deviation increase in peer ability is shown to be equivalent to a maximum of 9% of the effect of a one-standard-deviation increase in individual ability in the social sciences, and to a minimum of 3% of the effect of a one standard deviation increase in individual ability in math and science courses.

5.2 Heterogeneous Gamma Model

The assumption that each student is affected in the same manner by their classmates is restrictive and not as interesting from a policy perspective. In particular, the linear-in-means model implies that, in terms of grades, any winners from reshuffling peers are perfectly balanced by those who lose from the reshuffling.¹⁶ To relax this assumption, we apply our iterative estimator to an enhanced spillover model in which peer effects are allowed to vary with the observable characteristics of the student:

$$Y_{ikt} = \alpha_{ik} + \sum_{j \neq i}^N \frac{I_{ijt} \alpha_{jk}}{M_{it}} (X_i \gamma) + \sum_{c=1}^C I_{ict} \delta_c + \epsilon_{ikt} \quad (18)$$

where X includes whether the student is female, whether the student is Asian or another race, as well as the student's SAT math and SAT verbal scores.

Table 6 shows the results from this heterogeneous peer effects model. 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 non-white 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 non-whites 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 with higher test scores may have skills that make 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 homogenous gamma model. Peer ability is most important in

¹⁶See Hoxby and Weignarth (2005) for a discussion.

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% across course type. Relative to a one-standard-deviation increase in own ability, the effects of a one-standard-deviation increase in peer ability are also higher in the heterogeneous gamma model as shown in the fourth column of Table 7. The ratio of the effects of a one-standard-deviation increase in mean peer ability to a one-standard-deviation increase in own ability range from a low of 3.5% for math and science to a high of 10.5% in the social sciences.

5.3 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 consistently used in related literature, and to what extent the fixed effects estimated using the two methods differ. In order 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 8.

Columns 1 through 3 of Table 8 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. Consistent with Arcidiacono (2004), females outperform males across all course types. Conditional on SAT scores and high school GPA, whites perform better than all other ethnic groups including Asians.

The second set of columns in Table 8 shows the corresponding results for the heterogeneous gamma model. Recalling the results found in Table 6 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. Previous work such as Arcidiacono (2004), Arcidiacono and Vigdor (2007), and Arcidiacono et al. (forthcoming) have all found no returns to verbal test scores in the labor market. 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.

We find that 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 with the unobserved ability problem when following a selection-on-observables, or random assignment, approach. With much of peer ability unobserved, we would expect to underestimate the extent to which outcomes are affected by one's peers. However, selection into sections may occur based upon one's own unobserved ability, which in turn may be correlated with observed peer ability. This latter effect characterizes the selection problem, and is likely to bias peer effects estimates upward. While random assignment circumvents this latter problem, a downward bias remains from not taking into account the impact of unobservable peer attributes on outcomes.

6 Quantifying Selection

In this section we quantify how much selection is taking place within course types with respect to both observed and unobserved ability.¹⁷ For ease of exposition, we refer to the predicted values of the regressions in Table 8 as 'observed ability' and 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. In particular, we can split our student sample by which course type was chosen the most. Because we estimate separate abilities for each course type, we

¹⁷See Altonji et al. (2005) for more discussion of selection on observed and unobserved factors.

are able to determine whether students choose to take more courses in areas where they are comparatively more able. In addition to examining comparative advantage, we can also see whether students who choose particular course types enjoy an absolute performance advantage by determining if their mean fixed effects are higher on average across all course types.

6.1 Selection Within Course Types

Table 9 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 section-wide 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 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.

This table allows a direct analysis of the comparative degree of selection evident with respect to observed versus unobserved ability. The patterns for both the homogeneous specification and the heterogeneous specification are almost exactly the same. For all three course types there is more selection on unobserved ability than on observed ability, with higher ratios of section standard deviations to population standard deviations found for 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 ability, which is consistent with the high correlation between peer observed and unobserved ability.

With the observed and unobserved ability measures it is also 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. We test this hypothesis in section 7.

6.2 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. Each estimated ability measure is specific to courses of a particular type, by virtue of the estimation procedure, and as such each can be assumed to reflect a skill set that is particularly in demand in that course type. 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. For ease of exposition, we focus on the homogeneous gamma model for the rest of the paper.

Panel A of Table 10 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.¹⁸ The correlation coefficients are all quite large and close together, ranging from 0.65 to 0.69.¹⁹ Panel B of Table 10 displays analogous results using only the portion of estimated abilities for each student that is predictable using our observable variables. The relative relationships amongst observable abilities by course type are the same, with observable abilities to perform in the humanities and in math and science being related to a lesser extent than those of the other two course type pairings (humanities and social sciences, and social sciences and math and science). However, the strength of the relationships is much stronger across the board with correlation coefficients

¹⁸Correlations among estimated abilities were also calculated for all students who were observed in each pair of course types. Similar coefficients resulted.

¹⁹The 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.

above 0.95 for observed abilities for two out of three pairings. 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 11 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 gives 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. This can be seen from the fact that the highest number in each row is that corresponding to the course type specialized in by that sample of students—the highest number in each row is on the diagonal.²⁰ Panel B of Table 11, where we use only observed ability to examine selection, also illustrates both absolute and comparative advantage, but reflects more attenuated distributions of abilities and less heterogeneity than shown in Panel A.

²⁰Paglin and Rufolo (1990) find similar sorting patterns using GRE exam data and transcript data from the University of Oregon and Oregon State University. They find that students with high math ability tend to take courses in which there is a high return to this skill.

7 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 one-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.

Row 2 of Table 12 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, Row 1 of the table for each course type gives the original spillover estimate produced using our algorithm. Looking at Column 3 of the table, we can see that the estimated effects of a one-standard-deviation increase in observed peer ability are at most two-thirds the size of a one-standard-deviation increase in total peer ability. This first-order decrease in effect magnitude is evident because the impact of unobserved peer ability is not captured in Row 2 except through the correlation between observed and unobserved peer ability.

However, under random assignment, a one-standard-deviation increase in peer ability would produce even smaller effects than those shown in Row 2 of Table 12 for each course type. 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.²¹ This leads to an upward bias on our estimates of the spillover parameters: the estimated coefficients in the second row for each course type are all higher than those in the first row. The largest percentage increase in the peer ability coefficient is for courses in math and science, where the correlation between observed and unobserved peer ability is the highest. The second reason why random assignment will lead to even lower estimates of a one-standard-deviation increase in peer ability is that random assignment itself leads to less heterogeneity in mean peer ability across sections when higher ability students choose sections with other high ability students. This attenuation in the distribution of peer ability means that a one-standard-deviation increase in peer ability will be smaller under random assignment than in a self-selected context.

We next investigate what happens when the econometrician additionally assumes that students select into peer groups only based on observable characteristics (the “selection on observables” approach). In Row 3 of Table 12, 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 one-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 third column, this is indeed the case for humanities, the course type with the smallest correlation between unobserved individual ability and observed peer ability. For the social sciences, the selection-on-observables estimate of a one-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 one-standard-deviation increase in peer ability is significantly higher using the selection-on-observables approach than using our algorithm. This is driven by (1) the high correlation between unobserved individual ability and observed peer ability in math and science; (2) the fact that observed ability is a greater fraction of

²¹Recall that observed and unobserved ability are uncorrelated by construction at the individual level. Under random assignment, observed and unobserved peer ability will also be uncorrelated. The positive correlation in observed and unobserved peer ability in our data stems from student sorting based upon total ability.

total ability in math and science; and (3) the fact that the underlying peer effect estimate from our method is smallest in math and science, which mitigates the underestimation of a one-standard-deviation increase in peer ability.²²

8 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 biased upward estimates of peer effects parameters. On the other hand, underestimation of the effects of peers may result from ignoring peer effects through unobservables.

We present a new iterative method for estimating educational peer effects that overcomes both these obstacles. All that is required is that there are multiple observations per student with the peer group changing over time. We then control for individual effects and allow the peer effect to operate through a linear combination of the other individual fixed effects. We show that our estimator is consistent and asymptotically normal for fixed T as N goes to infinity. We also develop an iterative algorithm that is computationally much cheaper than direct non-linear least squares minimization yet produces the non-linear 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. The 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 observed and unobserved peer ability that, if ignored, biases the spillover parameter upward.

²²Indeed, if the spillover parameter were zero, there would be scope for the attenuation effect.

There is also much selection across course type. Students 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. Our different course types illustrate 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 much selection into math and science courses, leads to estimates from a selection-on-observables approach that significantly overstates the importance of peers. However, for humanities courses the estimated spillover parameter is larger than in math and science. Coupled with much less selection than in math and science, the selection-on-observables approach places similar importance on peers to what is estimated from our model. This is in contrast to random assignment, which removes the correlation between individual unobserved ability and peer observed ability but ignores peer effects through unobservables, leading to random assignment significantly understating the impact of peers on achievement.

There are many avenues to be explored in future research. First, future research will relax the assumption that an individual's ability to help others is proportional to an individual's own ability to perform well. The individual who asks the clarifying question may be more useful to others than a smarter individual who remains quiet. It is possible to extend the model such that spillover ability is treated differently from the ability to perform well for oneself.

Second, peer effects here are purely transitory. Future work will relax this assumption by allowing the effect of peer ability in a particular course to decay over time, as well as to influence performance in other contemporaneous classes. This will help us determine the long-run impact of peer groups on educational achievement, and may result in higher peer effect estimates as we account for spillovers beyond the classroom.

Finally, rather than separately estimating ability in each course type, the model could be expanded to allow a factor structure on ability and allow the returns to the different abilities to vary by course type. This expansion would allow a better exploitation of large data sets,

such as ours, with outcomes that have heterogeneous determinants, since performance on all outcome measures collectively would be used to estimate individual abilities.

References

- Abowd, J. M. and Kramarz, F.: 1999, “The Analysis of Labor Markets using Matched Employer-Employee Data”, in O. C. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3B, Elsevier Science B.V., chapter 40, pp. 2629–2710.
- Abowd, J. M., Kramarz, F. and Margolis, D. N.: 1999, “High Wage Workers and High Wage Firms”, *Econometrica* **67**(2), 251–333.
- Altonji, J. G., Elder, T. E. and Taber, C. R.: 2005, “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools”, *Journal of Political Economy* **113**(1), 151–184.
- Altonji, J. G., Huang, C.-I. and Taber, C. R.: 2004, “Estimating the Cream Skimming Effect of Private School Vouchers on Public School Students”. Working Paper.
- Arcidiacono, P.: 2004, “Ability Sorting and the Returns to College Major”, *Journal of Econometrics* **121**(1-2), 343–375.
- Arcidiacono, P., Cooley, J. and Hussey, A.: forthcoming, “The Economic Returns to an MBA”, *International Economic Review*.
- Arcidiacono, P. and Nicholson, S.: 2005, “Peer Effects in Medical School”, *Journal of Public Economics* **89**(2-3), 327–350.
- Arcidiacono, P. and Vigdor, J.: 2007, “Does the River Spillover? Estimating the Economic Returns to Attending a Racially Diverse College”. Working Paper.
- Arellano, M. and Hahn, J.: 2006, “A Likelihood-based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects”. Unpublished Manuscript.
- Bester, C. A. and Hansen, C.: 2005, “A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects”. Unpublished Manuscript.
- Betts, J. R. and Morell, D.: 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.

- Evans, W. N., Oates, W. E. and Schwab, R. M.: 1992, “Measuring Peer Group Effects: A Study of Teenage Behavior”, *Journal of Political Economy* **100**(5), 966–991.
- Fernandez-Val, I.: 2005, “Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects”. Unpublished Manuscript.
- 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.
- Hahn, J. and Newey, W.: 2004, “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models”, *Econometrica* **72**, 1295–1319.
- Hanushek, E. A., Kain, J. F., Markman, J. M. and Rivkin, S. G.: 2003, “Does Peer Ability Affect Student Achievement?”, *Journal of Applied Econometrics* **18**(5), 527–544.
- Honore, B.: 1992, “Trimmed LAD and Least Squares Estimation of Truncated and Censored Models with Fixed Effects”, *Econometrica* **60**, 533–565.
- Horowitz, J. L. and Lee, S.: 2004, “Semiparametric Estimation of a Panel Data Proportional Hazard Model with Fixed Effects”, *Journal of Econometrics* **119**, 155–198.
- Hoxby, C.: 2001, “Peer Effects in the Classroom: Learning from Gender and Race Variation”. NBER Working Paper 7867.
- Hoxby, C. and Weignarth, G.: 2005, “Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects”. Working Paper.
- Lehrer, S. and Ding, W.: 2007, “Do Peers Affect Student Achievement in China’s Secondary Schools?”, *Review of Economics and Statistics* **89**(2), 300–312.
- Manski, C.: 1995, *“Identification Problems in the Social Sciences”*, Harvard University Press, Cambridge, Massachusetts.
- Manski, C. F.: 1987, “Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data”, *Econometrica* **55**, 357–362.
- Neyman, J. and Scott, E.: 1948, “Consistent Estimates Based on Partially Consistent Observations”, *Econometrica* **16**, 1–32.

- Paglin, M. and Rufolo, A.: 1990, “Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences”, *Journal of Labor Economics* **8**(1), 123–144.
- Rivkin, S. G., Hanushek, E. A. and Kain, J. F.: 2005, “Teachers, Schools, and Academic Achievement”, *Econometrica* **73**(2), 417–458.
- Sacerdote, B. I.: 2001, “Peer Effects with Random Assignment: Results for Dartmouth Roommates”, *Quarterly Journal of Economics* **116**(2), 681–704.
- Winston, G. C. and Zimmerman, D. J.: 2003, “Peer Effects in Higher Education”, in C. Hoxby (ed.), *College Decisions: How Students Actually Make Them and How They Could*, University of Chicago Press.
- Woutersen, T.: 2002, “Robustness Against Incidental Parameters”. Unpublished Manuscript.
- Zimmerman, D. J.: 2003, “Peer Effects in Higher Education: Evidence from a Natural Experiment”, *Review of Economics and Statistics* **85**(1), 9–23.

Table 1: Monte Carlo Simulations: True Gamma = 0

Obs. Per	Peer Group		Random Assignment		Selection	
Student	Size		$\sigma_\epsilon=1.95$	$\sigma_\epsilon=1.15$	$\sigma_\epsilon=1.95$	$\sigma_\epsilon=1.15$
2	2	γ	0.0011	-0.0002	-0.0094	-0.0042
			(0.035)	(0.015)	(0.060)	(0.023)
		R ²	0.7044	0.8209	0.7055	0.82
5	10	γ	0.0005	0.0011	0.0039	-0.0055
			(0.037)	(0.021)	(0.061)	(0.032)
		R ²	0.4831	0.6845	0.4815	0.6852
10	10	γ	0.0049	-0.0018	0.0013	-0.0003
			(0.026)	(0.014)	(0.036)	(0.015)
		R ²	0.415	0.644	0.4148	0.6453

Note: The R-squared values reported in this table are those pertaining 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 in order to manipulate the amount of variation in performance that is explained by the ability measure.

Table 2: Monte Carlo Simulations: True Gamma = .15						
Obs. Per	Peer Group		Random Assignment		Selection	
Student	Size		$\sigma_\epsilon=1.95$	$\sigma_\epsilon=1.15$	$\sigma_\epsilon=1.95$	$\sigma_\epsilon=1.15$
2	2	γ	0.1507	0.1511	0.14	0.1508
			(0.034)	(0.016)	(0.060)	(0.024)
		R^2	0.7055	0.8219	0.7125	0.8277
5	10	γ	0.1499	0.1498	0.1463	0.1489
			(0.041)	(0.021)	(0.059)	(0.033)
		R^2	0.4822	0.6862	0.4943	0.6967
10	10	γ	0.1502	0.148	0.148	0.1524
			(0.025)	(0.012)	(0.036)	(0.018)
		R^2	0.4152	0.644	0.4288	0.6594

Note: The R-squared values reported in this table are those pertaining 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 in order to manipulate the amount of variation in performance that is explained by the ability measure.

Table 3: Sample Sizes

	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 Multi-Section Courses)	23,206	31,627	30,599	38,056	36,260	43,853	203,601

Note: 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.

Table 4: Peer Effects Results by Course Type: Homogeneous Gamma Model

	Humanities	Soc.Sci.	Math/Sci.
Section peer ability	0.1613 (0.0007)	0.1960 (0.0008)	0.0483 (0.008)
N	86,844	77,312	82,675
R^2	0.6373	0.6321	0.6861

Note: Dependent variable is the grade in the class. Class fixed effects are estimated in all specifications.

Table 5: Standard Deviations of Estimated Ability and Marginal Effects: Homogeneous Gamma Model

Course Type	Section Std Dev	Population Std Dev	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

Note: “Section Std Dev” is the standard deviation of average peer ability across the sample of student-section observations of the given course type. “Population Std Dev” 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 (1) the effect on grades from a one-standard-deviation increase in peer ability to (2) the effect on grades from a one-standard-deviation increase in own ability.

Table 6: Peer Effects Results by Course Type: Heterogeneous Gamma Model

	Humanities	Soc.Sci.	Math/Sci.
Section peer ability	0.2058 (0.0008)	0.2227 (0.0013)	0.0940 (0.0013)
Section peer ability*Female	0.0970 (0.0010)	0.0584 (0.0018)	-0.0517 (0.0018)
Section peer ability*Asian	-0.0098 (0.0016)	-0.0347 (0.0026)	-0.0346 (0.0022)
Section peer ability*Other Nonwhite	0.0375 (0.0014)	0.0252 (0.0026)	-0.0420 (0.0026)
Section peer ability*SATm	0.0410 (0.0006)	0.0507 (0.0012)	-0.0560 (0.0011)
Section peer ability*SATv	0.0222 (0.0005)	0.0147 (0.0010)	0.0635 (0.0009)
N	86,844	77,312	82,675
R^2	0.6376	0.6323	0.6864

Note: Dependent variable is the grade in the class. Class fixed effects are estimated in all specifications.

Table 7: Standard Deviations of Estimated Ability and Marginal Effects: Heterogeneous Gamma Model

Course Type	Section Std Dev	Population Std Dev	Avg Marginal Effect	Marginal Effect Ratio
Humanities	0.2383	0.6368	0.2595 (.0622)	0.0970
Social Science	0.2849	0.6747	0.2480 (.0554)	0.1048
Math and Science	0.6028	0.9660	0.0555 (.0636)	0.0347

Note: "Section Std Dev" is the standard deviation of average peer ability across the sample of student-section observations of the given course type. "Population Std Dev" 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. Column 3 shows the marginal effect on grades from a one-point increase in peer ability. "Marginal effect ratio" shows the ratio of (1) the effect on grades from a one-standard-deviation increase in peer ability to (2) the effect on grades from a one-standard-deviation increase in own ability. The numbers in parentheses are standard deviations of the marginal effects of a one-point increase in peer ability that are estimated to occur in the sample.

Table 8: Regression of Fixed Effects on Observed Ability

	Homogeneous Gamma Model			Heterogenous Gamma Model		
	Hum.	Soc.Sci.	Math/Sci.	Hum.	Soc.Sci.	Math/Sci.
SATm	0.00	0.11	0.36	-0.20	-0.04	0.51
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SATv	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)
Female	0.26	0.17	0.15	-0.15	0.03	0.27
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Black	-0.28	-0.23	-0.18	-0.44	-0.29	-0.10
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Hispanic	-0.15	-0.19	-0.17	-0.31	-0.25	-0.08
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
Asian	-0.12	-0.13	-0.10	-0.08	-0.05	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Amer. Ind.	-0.39	-0.35	-0.31	-0.55	-0.41	-0.22
	(0.12)	(0.12)	(0.16)	(0.12)	(0.12)	(0.16)
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

Note: The dependent variable in columns 1 through 3 is the student-level fixed effects estimated in the homogeneous gamma model; the dependent variable in columns 4 through 6 is the student-level fixed effects estimated in the heterogeneous gamma model. The excluded racial/ethnic category is white; racial/ethnic categories are mutually exclusive. Standard errors are robust to heteroskedasticity.

Table 9: Selection based on Observables and Estimated Ability

Course Type	Homogeneous Gamma			Heterogeneous Gamma		
	α	$\hat{\alpha}$	α_u	α	$\hat{\alpha}$	α_u
Humanities						
Avg Section-level SD	.6154	.3372	.5385	.5849	.2487	.5415
Population SD	.8046	.3773	.7106	.7619	.2711	.7121
Ratio	.7649	.8937	.7578	.7677	.9174	.7604
Social Science						
Avg Section-level SD	.6365	.3750	.5667	.6056	.2994	.5688
Population SD	.8618	.4228	.7510	.8226	.3343	.7517
Ratio	.7386	.8869	.7546	.7362	.8956	.7567
Math and Science						
Avg Section-level SD	.7475	.5036	.6666	.7636	.5323	.6649
Population SD	1.0567	.6165	.8583	1.0719	.6432	.8575
Ratio	.7074	.8169	.7767	.7124	.8276	.7754

Note: Each set of rows corresponds to sections in one coursetype; 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.

Table 10: Correlations of Estimated Abilities Across Course Types

<i>Panel A: Abilities</i>		
Course type	Humanities	Social Science
Humanities	1.0000	
Social Science	0.6875	1.0000
Math and Science	0.6469	0.6776

<i>Panel B: Predicted abilities</i>		
Course type	Humanities	Social Science
Humanities	1.0000	
Social Science	0.9643	1.0000
Math and Science	0.8808	0.9593

The abilities used in these correlation matrices are those of the 12,715 students who took classes in all three course types, and they are estimated by the homogeneous gamma model. Panel A displays correlations amongst the full abilities, while Panel B displays correlations amongst the predicted values from regressing the estimated abilities from the homogeneous gamma model on observable variables (as shown in the first three columns of Table 8).

Table 11: Specialization of Students into Course Types By Relative Aptitude

Panel A: Abilities

	Humanities	Soc.Sci.	Math/Sci.	N
Humanities-specializers	-.08	-.19	-.26	3978
Social Science-specializers	.04	.10	-.07	3547
Math and Science-specializers	.14	.21	.43	3745

Panel B: Predicted abilities

	Humanities	Soc.Sci.	Math/Sci.	N
Humanities-specializers	-.04	-.13	-.21	3977
Social Science-specializers	-.11	-.10	-.11	3547
Math and Science-specializers	.18	.27	.36	3745

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.

Table 12: Comparing the Method to Conventional Results: Homogeneous Gamma Model

	Own ability	Section peers' ability	Effect of 1-stddev chg 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
	(-)	(.0009)			
	0.7784	0.3519	0.1313	Observed	Observed
	(0.0062)	(0.0063)			

Note: Dependent variable is grade in the class. For models shown in Rows 1 and 2 for each coursetype, class effects are removed before estimation by demeaning using the ‘true’ values of the class effect as estimated by our homogeneous gamma model. For Row 3, 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, which may impact the standard errors reported here. Observations are as in Table 4, although creating the second and third rows involved dropping observations for which observables were missing. The maximum number of observations dropped for a course type was 6.