NBER WORKING PAPER SERIES

ESTIMATING THE CREAM SKIMMING EFFECT OF SCHOOL CHOICE

Joseph G. Altonji
Ching-I Huang
Christopher R. Taber

Working Paper 16579
http://www.nber.org/papers/w16579

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2010

This research was supported by a grant from the Searle Foundation, the Institute for Policy Research, Northwestern University, and the Economic Growth Center, Yale University. We have received valuable comments from seminar and conference participants at a number of institutions. We also thank Steven Berry, Charles Manski, Derek Neal, and Jesse Rothstein for helpful discussions. Earlier drafts were circulated under the title "Estimating the Cream Skimming Effect of Private School Vouchers on Public School Students". Mistakes are our responsibility. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating the Cream Skimming Effect of School Choice
Joseph G. Altonji, Ching-I Huang, and Christopher R. Taber
NBER Working Paper No. 16579
December 2010
JEL No. I21

ABSTRACT

We develop a framework that may be used to determine the degree to which a school choice program may harm public school stayers by luring the best students to other schools. This framework results in a simple formula showing that the "cream-skimming" effect is increasing in the degree of heterogeneity within schools, the school choice takeup rate of strong students relative to weak students, and the importance of peers. We use the formula to investigate the effects of a voucher program on the high school graduation rate of the students who would remain in public school. We employ NELS:88 data to measure the characteristics of public school students, to estimate a model of the private school entrance decision, and to estimate peer group effects on graduation. We supplement the econometric estimates with a wide range of alternative assumptions about school choice and peer effects. We find that the cream skimming effect is negative but small and that this result is robust across our specifications.

Joseph G. Altonji
Department of Economics
Yale University
Box 208264
New Haven, CT 06520-8264
and NBER
joseph.altonji@yale.edu

Ching-I Huang
Department of Economics
National Taiwan University
21, Syujhou Road, Taipei 100, Taiwan
chingihuang@ntu.edu.tw

Christopher R. Taber
Department of Economics
University of Wisconsin -Madison
1180 Observatory Dr
Social Sciences Building #6448
Madison, WI 53706-1320
and NBER
ctaber@ssc.wisc.edu

1 Introduction

Dissatisfaction with the performance of the U.S. educational system, particularly in minority urban school districts, has led to a surge in interest in and experimentation with a variety of school choice programs. To assess the overall effect of any school choice program on educational outcomes, one must address (at least) three questions. First, by how much do children who exercise their option to switch schools benefit? Second, does increased competition lead schools to improve? Third, will a choice program lure the best students away from current schools, and if so, how large is the negative "cream skimming effect" on those who remain behind? The same three issues are central to assessing the major types of choice programs: vouchers, charter schools, and choice within the public school system.

There is a lot of research on the first question and some work on the second, but virtually no work on measuring the cream skimming effect, which is the subject of our paper.[1] Quantifying this effect is difficult. Starting from knowledge of who currently attends public school, one must assess who will move in response to the program, estimate the consequences of peers for outcomes, and then weight the change in the peers of public school students by the estimates of the peer effect parameters. In fact, it is hard to conceive of a controlled experiment, let alone a natural experiment, that could directly identify this effect. Even if one could randomize eligibility for a choice plan across schools, one still would not know the counterfactual of who in the control school would remain if they were eligible for the plan. One must combine experimental or non-experimental variation with a model of school choice and outcomes.

The first contribution of the paper is to show that for a broad class of models of school choice and peer effects, the cream skimming effect is determined by the covariance between the relative probability that a student will move in response to the choice program with an index of the differences between the student's characteristics and the average characteristics of his or her classmates. The weights of the index are the peer effect coefficients. The covariance is increasing in the amount of heterogeneity within schools, the relative response of advantaged students to the voucher, and the magnitude of the peer effect coefficients. To see the intuition, note that the cream skimming effect will be zero under three separate conditions. The first is zero heterogeneity within public schools, because then the leavers will be the same as the stayers.

[1]Much of the U.S. research on the direct benefits from private school attendance is in the context of Catholic schools. The results are mixed, but studies by Evans and Schwab (1995), Neal (1997), Grogger and Neal (2000), and Altonji, Elder and Taber (2005) suggests that students who attend Catholic schools perform substantially better than they when he would have done in a public school. The evidence is strongest for urban minority students, and the main effects appear to be on high school graduation and college attendance rates rather than on achievement on standardized tests. Rouse and Barrow (2009) survey the evidence from a number of voucher experiments on the achievement effects of private school attendance and conclude that they are small. They also summarize the literature on effects of attending a charter school and on whether competition improves public school performance. See also Hoxby (2003).

The second is the absense of a relationship between the probability that a student leaves and student characteristics that affect classmates, because then the voucher will not change the average values of relevant peer measure. Finally, the cream skimming effect will be zero if peer effects on outcomes are zero.

Although our formula for the cream skimming effect could be applied to other school choice programs, we focus on voucher programs because our data are much more informative about selection into private schools. We consider both broad based voucher programs and programs that are targeted to low income students or urban students and use high school graduation as our main outcome of interest.

The cream skimming formula provides the structure for what amounts to a mix of formal econometric analysis and sensitivity analysis. The formal econometric analysis uses data from the National Education Longitudinal Survey of 1988 (NELS:88) and proceeds in three stages. The first stage is to estimate the effect of a voucher on who attends public school. The second is to estimate the effects of peers on outcomes. The final stage is to apply our formula and compute the cream skimming effect.

In the first stage of our base case model we estimate a probit model for public school choice that identifies the conditional probability that a student with specified observed characteristics chooses public school, given the status quo of no voucher program. We model the voucher as a shift in the index determining school choice for a given group of students. This allows us to compute the relative probability that public school students will remain in public school given the level of the voucher. Furthermore, the assumption implies the natural result that students who are currently in private schools continue to attend them after a voucher is put into place. This permits us to obtain the distribution of observed and unobserved characteristics of students who will remain in public school by using the relative probabilities of continued public school attendance to reweight the distribution for public school students under the status quo. By comparing reweighted means to the means of public school students under the status quo one obtains estimates of how mean family income, mean parental education, mean eighth grade test scores and other characteristics of high school peers will change for those who remain in public high schools. A big advantage of our approach is that we do not need variation in tuition or voucher levels to estimate which students are likely to respond to the voucher program.

We work with several alternative specifications of the demand for public schools, including a nested logit specification in which Catholic and non-Catholic private schools are treated as separate alternatives, a case in which we fit the school choice model to the Milwaukee school voucher experience, and the extreme assumption that the students who move in response to the voucher come entirely from the top of the achievement distribution. The most interesting and difficult case is a model in which peer quality influences school choice. To handle this case, we develop and implement a new methodology for estimating binary consumer choice models when demand depends on the characteristics of the other buyers and information on the characteristics of the other buyers is only available to the researchers for those who choose the product. In our case, the attractiveness of public school depends on the characteristics of the other public school students. We have information on the potential public high school classmates of those who choose public school but not of those who choose private schools. We show that with data on a set of variables that are correlated with the characteristics of students in a high school district, such as average family income and racial composition, one can simultaneously estimate the distribution of the characteristics of the students who choose public high school, with selection accounted for, while estimating the effects of student body characteristics on the choice.[2] Our methodology could be applied in other situations in which consumer demand depends the characteristics of other consumers, and consumer characertitics are correlated with location, and sampling is choice based because of the costs of a random sample.

The second stage is to estimate the extent of peer effects on high school graduation. We start with the standard procedure of regressing high school fixed effects for the outcome on observed student body characteristics and address a number of econometric issues. The most difficult is accounting for the effects of the voucher on *unobserved* peer characteristics. For example, the usual school level parental background measures, such as average family income and average parental education are only crude measures of the resources that parents provide to their children. Roughly speaking, we use the school choice model to infer the mean for each high school of the index of unobserved student characteristics that determine school choice. Observed and unobserved characteristics that influence school choice need not be the only student characteristics that influence peer outcomes, but, as we show below, student characteristics influence the cream skimming effect only to the extent that they are related to school choice. In one case, we calculate the cream skimming effect under a range of assumptions about the relative size of the effect on school outcomes of the school averages of unobserved and observed student body characteristics influence school choice. Our method of handling unobserved characteristics might be useful in other contexts.

In the final stage, we use our estimates of the peer effects models and estimates of the shift in both the observed and unobserved characteristics of the peers of students who remain in public school to estimate the effect of the voucher program on the high school graduation rate of the public school stayers. We report results for a number of different model specifications and also investigate voucher programs that are targeted to low income families, low income schools, low achieving schools, and urban areas. Overall, our results are robust and indicate that a large scale

[2] The problem is simpler if information is available for all students about the student body of the public high school option. We do not know of a US data set that is nationally representative and provides such information.

voucher program would have small effects on the high school graduation probabilities of those who remain in public school. One might expect the cream skimming problem to be even less severe in other choice programs such as charter schools given the available evidence on selection into charter schools. If this is the case we would conclude that the magnitude of this effect is likely very small in those cases as well.

We do not know of another study that is directly comparable to ours, but the literature on voucher programs in the US and elsewhere provides some evidence about who would take up vouchers. Howell and Peterson (2002) examine several programs targeted to low income students and conclude the the degree of positive selection in such programs is relatively small. There are no universal voucher programs to study in the U.S., but Hsieh and Urquiola (2002) find that Chile's universal voucher program induced higher income and higher ability children to move to private sector schools. Ladd (2002) finds that selection in New Zealand's choice program worked in a direction similar to the Chilean program and that "the expansion of choice in that country exacerbated the problems of the schools at the bottom of the distribution and the reduced the ability of those schools to provide an adequate education." Given that the composition of charter schools is heavily influenced by the specific areas in which they were introduced and the missions of the schools, one cannot easily draw conclusions about a universal or a targeted voucher program from aggregate statistics on the composition of charter schools. Nevertheless, there is little indication that charter schools lead to a large exodus of the most advantaged children from regular public schools, particularly when compared to private schools.[3]

Our work is also related to a few papers on the general equilibrium effects of voucher programs with peer effects, including Manski (1992), Epple and Romano (1998, 2002, 2003), Epple, Newlon, and Romano (2002), and Caucutt (2002). Simulations of calibrated versions of the models usually show cream skimming, although the magnitude varies with the details of the model specification and assumed parameter values. Nechyba (1999, 2000, 2003) shows that migration can have a countervailing effect on low income households who remain in public school. To our knowledge, the only paper that explicitly estimates and simulates the extent of peer group effects with vouchers is Ferreyra (2007). She estimates her model using school district data from several large metropolitan areas. She then simulates the effects of vouchers.

Ferreyra (2007) does not use data on school quality, but infers the production function for it based on location and schooling decisions. This is an important limitation because peer characteristics could influence school and location choices for a number of reasons, including: a) the possibility that they affect school outcomes and parents care about the school outcomes, b) the possibility that parent's care about outcomes and think that peer effects are important even though they may not be and c) the possibility that parents care about peer quality per se.

[3]See RPP International (2001), and Fisler (2002), Lacireno-Paquet et al (2002), and Zimmer et al (2009).

In contrast, we are the first to directly estimate the effects of peers on particular outcomes and then simulate the effects of cream skimming on these outcome for public school students. Our analysis is complementary with the general equilibrium papers.

The paper continues in Section 2, where we present our school choice model, define the cream skimming parameter, and derive the key equation that determines the effect. In section 3 we discuss estimation in situations in which peer effects depend only on observed variables. Sections 4 and 5 we discuss the NELS:88 data and provide descriptive statistics. In Section 6 we briefly discuss the estimates of our basic school choice model, estimates of the effects of student characteristics and peer characteristics on high school graduation, and estimate the cream skimming effect. In Section 7 we allow peer characteristics to affect school choice and in Section 8 we allow for unobserved school characteristics that influence choice and for unobserved peer characteristics. Sections 9 and 10 consider additional alternative assumptions about peer effects and about school choice. We conclude in Section 11.

2 A Model of School Choice and Outcomes and a Definition of the Cream Skimming Effect

In this section, we begin by presenting the basic model for school choice and classmate effects on school outcomes that underlies our analysis of the effects of school choice on students who remain in their original school. We then define our parameter of interest—the cream skimming effect—and analyze the factors that determine it.

2.1 School Choice

Let $\mathcal{S}$ be the set of all public schools in the country. Each student i is assigned to a particular school district and denote the public school in that district as S_i. School choice programs are indexed by τ. This could represent a program dictating a certain level of a voucher or entry of a particular class of charter schools. The indicator variable $P_i^\tau = 1$ if the student chooses their default public school and 0 if the student chooses an alternative school under program τ. In general, the choice P_i^τ depends τ both through direct effect of the program on the student and through the effect of the program on the choices of other students. For most of the empirical results in the paper, we focus on a universal voucher program that provides the same voucher amount to all students. We denote the status quo as $\tau = 0$ and the corresponding choice as P_i^0. When we use the expression $P_i^\tau = 1$ as a conditioning argument, we are conditioning on i choosing public school when the school choice program is τ. The conditioning argument $P_i^0 = 1$ means that i chose a public school in the status quo (and this variable will be observable in the data set).

Let $V(X_i^*, S_i, \tau)$ be the difference in the utility of attending public school S_i under choice program τ and the utility of the best alternative school option given $i's$ location, where X_i^* is a vector that includes all variables, both observed and unobserved, that influence the relative utility of private school. Thus P_i^τ is determined by

$$P_i^\tau = \mathbf{1}(V(X_i^*, S_i, \tau) > 0). \tag{1}$$

Note that the function V allows school choice to depend on the specific school S_i. This dependence may arise through fixed school characteristics that would not be altered by τ, as well as through peer effects that would be altered.

2.2 School Outcomes, School Quality, and Peer Effects

Let $Y_i(\tau)$ be an outcome that individual i would achieve if he or she attended S_i under choice program τ. Examples of outcomes are test scores, high school completion, college attendance, and earnings. $Y_i(\tau)$ is determined by

$$Y_i(\tau) = X_i'\gamma + \theta(S_i, \tau) + \varepsilon_i, \tag{2}$$

where $\theta(S_i, \tau)$ is a school quality component that is common to all individuals who attend S_i, X_i is a vector of observed characteristics of i that influence the outcome and ε_i is an index of unobservable individual factors that is uncorrelated with X_i and $\theta(S_i, \tau)$. The school effect $\theta(S_i, \tau)$ depends on τ through peer effects that change as different students attend public school. Note that (2) rules out interactions between X_i and , which could be added.

For any public school $s \in \mathcal{S}$ and any voucher program τ, peer effects are a function of the vector

$$\overline{Z}(s, \tau) = E(Z_i \mid S_i = s, P_i^\tau = 1),$$

where Z_i is the vector of observed and unobserved student characteristics that influence other students and $\overline{Z}(s, \tau)$ is the average of Z_i for students who choose S_i given the voucher program τ. An important special case is $Z_i = X_i$.

Without much loss of generality given that Z_i may include known nonlinear functions, we assume that the school fixed-effect can be expressed as

$$\theta(s, \tau) = \overline{Z}(s, \tau)'\delta + Q_s' B_Q + \xi_s, \tag{3}$$

where the observed variables Q_s and the error component ξ_s capture other determinants of school quality that are not influenced by the voucher, such as the characteristics of the building, the principal, and the teachers.[4]

[4] We use the term "peer effects" to refer to the influence of the average values in a school of a variety of student

2.3 The Cream Skimming Effect $\pi^p(\tau)$

For individual i we define the treatment effect of vouchers conditional on staying in public school as

$$\pi_i(\tau) \equiv Y_i(\tau) - Y_i(0) = \theta(S_i, \tau) - \theta(S_i, 0) = [\bar{Z}(S_i, \tau) - \overline{Z}(S_i, 0)]'\delta.$$

Our parameter of interest is the average value of this "cream skimming" effect for public school stayers under a school choice program:

$$\begin{aligned} (4) \qquad \pi^p(\tau) &\equiv E\left(\pi_i(\tau) \mid P_i^\tau = 1, P_i^0 = 1\right) \\ &= \left[E\left(\overline{Z}(S_i, \tau) \mid P_i^\tau = P_i^0 = 1\right) - E\left(\overline{Z}(S_i, 0) \mid P_i^\tau = P_i^0 = 1\right)\right]'\delta. \end{aligned}$$

Thus we need to identify δ, $E\left(\overline{Z}(S_i, \tau) \mid P_i^\tau = P_i^0 = 1\right)$, and $E\left(\overline{Z}(S_i, 0) \mid P_i^\tau = P_i^0 = 1\right)$ to identify the cream skimming effect. So far we have not made assumptions about what is observable to the econometrician.

We now derive the formula for $\pi^p(\tau)$ that underlies our empirical investigation. The term $E\left(\overline{Z}(S_i, 0) \mid P_i^\tau = P_i^0 = 1\right)$ involves P_i^τ, which is not observed. However, P_i^0 is observed, and so we can condition on $P_i^0 = 1$, the set of people currently in public school. Define $\chi_i \equiv \{X_i^*, Z_i, S_i\}$. Letting G represent a generic distribution, we can write

$$E\left(\overline{Z}(S_i, 0) \mid P_i^\tau = P_i^0 = 1\right) = \int \overline{Z}(S_i, 0)\, dG\left(\chi_i \mid P_i^\tau = P_i^0 = 1\right).$$

Bayes theorem implies that

$$dG\left(\chi_i \mid P_i^\tau = P_i^0 = 1\right) = \frac{\Pr(P_i^\tau = 1 \mid P_i^0 = 1, \chi_i) dG\left(\chi_i \mid P_i^0 = 1\right)}{\Pr(P_i^\tau = 1 \mid P_i^0 = 1)} = \psi_i(\tau) dG\left(\chi_i \mid P_i^0 = 1\right),$$

where $\psi_i(\tau) \equiv \frac{\Pr(P_i^\tau = 1 \mid P_i^0 = 1, \chi_i)}{\Pr(P_i^\tau = 1 \mid P_i^0 = 1)}$ is the relative probability of remaining in public school after the voucher program τ is put into effect, conditional on χ_i. Consequently,

$$E\left(\overline{Z}(S_i, 0) \mid P_i^\tau = P_i^0 = 1\right) = \int \overline{Z}(S_i, 0)\, \psi_i(\tau) dG\left(\chi_i \mid P_i^0 = 1\right) = E(\overline{Z}(S_i, 0)\, \psi_i(\tau) \mid P_i^0 = 1).$$

By the same line of reasoning,

$$\begin{aligned} E\left(\overline{Z}(S_i, \tau) \mid P_i^\tau = P_i^0 = 1\right) &= \int \overline{Z}(S_i, \tau)\, dG\left(\chi_i \mid P_i^\tau = 1, P_i^0 = 1\right) \\ &= \int \overline{Z}(S_i, \tau)\, \psi_i(\tau) dG\left(\chi_i \mid P_i^0 = 1\right) = E(\overline{Z}(S_i, \tau)\, \psi_i(\tau) \mid P_i^0 = 1) \end{aligned}$$

body characteristics that are determined prior to high school. These include parental education and income, race, gender, and performance in lower grades. We are side stepping the reflection problem discussed by Manski (1993). It does not matter for our simulations whether peer effects operate through covariates or outcomes. One can interpret δ in (3) as the reduced form coefficients of a model with reflection. The reduced form is all that is needed.

Using the above two equations and (4), we can write $\pi^p(\tau)$ as

(5)
$$\pi^p(\tau) = E(\psi_i(\tau)[\overline{Z}(S_i,\tau) - \overline{Z}(S_i,0)]'\delta | P_i^0 = 1).$$

The cream skimming formula simplifies further if the gain from choosing the default school, $V(X_i^*, S_i, \tau)$, is monotone in τ, so that no student chooses to attend the default school under the proposed choice policy if they would not under the status quo. (That is under monontonicity, if $P_i^0 = 0$ then $P_i^\tau = 0$). This is a very natural assumption in the case expansion of charter schools or introduction of a voucher program although it would not make sense for an expansion of choices within public schools.[5]

Under monotonicity, using the law of iterated expectations,

$$\begin{aligned} E\left(\overline{Z}(S_i,\tau) \mid P_i^\tau = 1, P_i^0 = 1\right) &= E\left(\overline{Z}(S_i,\tau) \mid P_i^\tau = 1\right) = E(E\left(Z_j \mid S_j = S_i, P_i^\tau = 1\right) | P_i^\tau = 1) \\ &= E\left(Z_i \mid P_i^\tau = 1\right) = E\left(\psi_i(\tau) Z_i \mid P_i^0 = 1\right). \end{aligned}$$

Consequently, under monotonicity, (5) reduces to[6]

(6)
$$\pi^p(\tau) = cov(\psi_i(\tau), [Z_i - \overline{Z}(S_i,0)]'\delta | P_i^0 = 1).$$

Equation (6) shows that the cream skimming effect $\pi^p(\tau)$ is the covariance between $\psi_i(\tau)$ and $[Z_i - \overline{Z}(S_i,0)]'\delta$. It easy to see that $\pi^p(\tau)$ depends on three factors. The first is the extent and the nature of the variation in $\psi_i(\tau)$ If $\psi_i(\tau)$ does not vary across i, then students who move in response to τ are more or less a random sample and the characteristics of the peers of students who remain in public school do not change. If $\psi_i(\tau)$ does vary but is unrelated to Z_i, there would be no cream skimming and the cream skimming effect would be 0. By the same token, $\pi^p(\tau)$ is more negative the greater degree to which $\psi_i(\tau)$ declines with characteristics that benefit other students (i.e, characteristics that increase $[Z_i - \overline{Z}(S_i,0)]'\delta$).

The second determinant of $\pi^p(\tau)$ is the extent of heterogeneity in peer characteristics within a school. $\pi^p(\tau)$ will be zero if there is no heterogeneity in Z_i within a school (eg., parental background is identical). In this case $Z_i - \overline{Z}(S_i,0)$ would zero for all i and once again the treatment effect would be identically zero. The more heterogeneity within a school, the more negative the cream skimming effect could potentially be. Components of Z_i matter for the cream skimming effect only to the extent that they are correlated with $\psi_i(\tau)$. This provides an

[5]While we estimate some models in which montonicity does not have to hold, in every case we present below it does hold. The simplified version of (5) eases interpretation of the empirical work.

[6]Getting from the expectation to the covariance uses the additional fact that $E(Z_i \mid P_i^0 = 1) = E\left(\overline{Z}(S_i,0) \mid P_i^0 = 1\right)$.

important justification for the restrictions that we impose on the peer effects function $\overline{Z}(s,\tau)'\delta$ below.

The third determinant is the magnitude of the peer effect coefficients δ. $\pi^p(\tau)$ will be identically zero if there are no peer group effects ($\delta = 0$). More generally, the more peers matter for school outcomes, the more important the cream skimming effect will be (assuming that better and more advantaged students from a given school are more likely to move).

Thus one can see the importance of all three factors. The cream skimming effect will be 0 if *any* of the three channels are 0, not just if *all* three are 0. By the same logic, to get a large value of $\pi^p(\tau)$ all of the channels must be sizeable. We obtain small values for $\pi^p(\tau)$ below because of the combination of factors, not a single one. When one plugs reasonable estimates of the three channels into the formula, small estimates of $\pi^p(\tau)$ come out.

3 Estimating the Cream Skim Effect

In the next subsections we present the econometric model of school choice that we use to estimate $\psi_i(\tau)$, discuss estimation of the school outcome model (2)-(3), and then discuss the construction of $\hat{\pi}^p(\tau)$. In formula (6) note that $\psi_i(\tau)$ is the only parameter that depends on the particular choice program. The difference $[Z_i - \overline{Z}(S_i, 0)]$ is just raw data that can be observed (or approximated) directly. The peer effect parameter δ is difficult to estimate, but there is much previous work on this. Thus the novel and challenging aspect of estimating $\pi^p(\tau)$ is estimation of $\psi_i(\tau)$. If one were to investigate the cream skimming effect of a charter school program or other types of school choice, one could use estimate $\psi_i(\tau)$ for that program and then plug it into (6).[7] In this paper, we consider the voucher case using observational data on how people select into private school to understand how people might react to different types of voucher programs. We focus on a model in which students either remain in public school or attend a private school.

3.1 An Econometric Model of Public School Attendance

In our baseline case, we estimate $\psi_i(\tau)$ from a probit model of public school attendance. Partition X_i^* as $\{X_i, u_i\}$,where X_i is a vector of observed characteristics and the scalar u_i is an index of unobserved characteristics that influence the attractiveness of the public school S_i relative to private schools. u_i is assumed to be independent of X_i. We assume that the utility gain from attending public school S_i under voucher program τ is

$$V(X_i^*, S_i, \tau) = X_i'\beta + \varphi\bar{Z}(S_i,\tau)'\delta - t_i(\tau) + u_i \tag{7}$$

[7] Note that without monotonicity that is no longer true. One would have to use the model to simulate $\bar{Z}(S_i,\tau)$ as well, and then apply (5).

where $t_i(\tau)$ is the voucher that individual i is eligible for under voucher program τ. Note that we are implicitly assuming that the voucher affects choices only through the direct effect and through the public school peers $\left(\bar{Z}(S_i,\tau)'\delta\right)$, not by changing the characteristics of the private schools options available to i.[8]

The decision rule of student i facing voucher program τ boils down to

$$P_i^{\tau} = 1(X_i'\beta + \varphi\bar{Z}(S_i,\tau)\delta - t_i(\tau) + u_i > 0) \tag{8}$$

Because we do not have data on t_i, we normalize $var(u_i) = 1$. This implicitly defines the scale of t_i such that a unit change in t_i has the same effect on school choice as a one standard deviation change in the u_i. We do not need to know the price elasticity of demand for private schools.[9] We instead define the "size" of the voucher in terms of the number of people induced to attend private school by the voucher. For our base case we choose τ so that $t_i(\tau) = t$ induces 10% of public school students to move.

If u_i is $N(0,1)$ then

$$\Pr(P_i^{\tau} = 1, X_i, S_i) = \Phi(X_i'\beta + \varphi\bar{Z}(S_i,\tau)'\delta - t_i(\tau)),$$

where Φ is the standard normal CDF. The ratio $\psi_i(\tau)$ of the probability that i will stay in public school if a voucher τ is introduced relative to the probability that the student is in public school under that status quo in which τ is 0 may be written as

$$\psi_i(\tau) = \frac{\left[\frac{\Phi(X_i'\beta+\varphi\bar{Z}(S_i,\tau)'\delta-t_i(\tau))}{\Phi(X_i'\beta-\varphi\bar{Z}(S_i,0)'\delta)}\right]}{\int\frac{\Phi(X_i'\beta+\varphi\bar{Z}(S_i,\tau)'\delta-t_i(\tau))}{\Phi(X_i'\beta-\varphi\bar{Z}(S_i,0)'\delta)}dG(\chi_i|P_i^0=1)}. \tag{9}$$

We construct $\hat{\psi}_i(\tau)$ using many different versions of the school choice model. In our base case specification, we ignore feedback effects of peers on school choice, and assume that $\varphi = 0$. In this case, our estimator is

[8] Our assumption is consistent with an expansion of the private school sector to accommodate increased demand provided that attributes that influence choice do not change. We are assuming that any feedback from voucher induced changes in the peer characteristics of public and private schools to school choice is of a second order of importance. One can accommodate the likely possibility that average distance from private schools would decrease in the wake of a large scale voucher program, with an effect on the demand for private schools. To do so, one could redefine $t_i(\tau) = t$ to be an index capturing both the effect of the tuition subsidy and of a uniform reduction in distance resulting from the private school expansion. In some sense, the reduction in distance associated with private school entry acts as a multiplier on the effect of a voucher on demand for private school. In practice however, one might expect the size of the distance reduction to depend on the existing stock of private schools and to vary across households depending on precisely where they live. One could model this but we have not done so.

[9] Dynarski, Gruber and Li (2008) use tuition discounts for number of children to estimate the price elasticity of demand for Catholic primary schools. Their results suggest that lower SES families are more responsive to the tuition. This would suggest that our estimates are an over-estimate of the cream skimming effect.

$$\hat{\psi}_i(\tau) \equiv \left[\frac{\Phi(X_i'\hat{\beta} - t_i(\tau))}{\Phi(X_i'\hat{\beta})}\right] \Big/ \left[\frac{1}{N_{P^0}} \sum_{j, P_j^0 = 1} \frac{\Phi(X_j'\widehat{\beta} - t_j(\tau))}{\Phi(X_j'\widehat{\beta})}\right] \tag{10}$$

where N_{p^0} is the number of sample members in public school.

In one set of cases we relax the assumption that peer effects have a negligible effect on the public school decision by freely estimating φ at a major cost in terms of computational complexity. In another case, we disaggregate Catholic and non-Catholic private schools using a nested logit specification.[10] We also experiment with alternative assumptions about unobservables. In our base case we assume that there are no unobservable peer effects and no fixed unobservables common to students assigned to $S_i = s$ that influence school choice. Relaxing this assumption is very difficult, as we discuss in Section 8 and Appendix 3, but has little effect on our basic findings.

3.2 Estimation of the School Outcome Parameters

We estimate the coefficient vector γ from (2) by OLS regression of $Y_i(0)$ on X_i with public school fixed effects. Although the estimates of γ are not our main focus, bias in the estimator for γ could spill over into bias in the estimation of the link between $\theta(s, \tau)$ and $\bar{Z}(s, \tau)$. Consequently, a discussion is in order even though we do not have a way to address the issue. First, measurement error is likely to lead to underestimation of γ (in absolute value) and bias $\hat{\delta}$ in the opposite direction to the extent that school level averages are less affected by measurement error and a substantial component of the true variation in X_i is across school. This is likely to lead to an overestimate of the importance for high school performance of the average level of eighth grade test scores (which have a random component to them), parental education, income, etc. Second, within school variation in omitted factors that influence education outcomes and income and are correlated with the within school variation in X_i will also influence $\hat{\gamma}$. The effect of this latter source of bias on $\hat{\delta}$ is harder to determine.

3.2.1 Estimation of the School Quality Parameter δ

In our base case, we assume that all elements of Z_i are observed. Because we only observe a sample of students from each public school and thus only observe $\overline{Z}(S_i, 0)$ with error, we use the JIVE estimator discussed in Devereux (2005) to estimate δ. Specifically, one may rewrite (3) as

$$\hat{\theta}(S_i, 0) = Z_i'\delta + Q_{S_i}'B_Q + \varepsilon_{\theta i} \tag{11}$$

[10] The parameters of a trinomial probit are poorly identified in our data.

where $\hat{\theta}(S_i, 0)$ is the estimate of the school fixed effect for the school S_i attended by person i and the error term $\varepsilon_{\theta i} = [\overline{Z}(S_i, 0)' - Z_i']\delta + \xi_s + [\hat{\theta}(S_i, 0) - \theta(S_i, 0)]$. We estimate δ and B_Q by IV regression. The instruments are Q_{S_i} and the "i left out means" $\overline{Z}_{S_{-i}}$ consisting of the average value of Z for sample members who attended $i's$ public school S_i, with i excluded.[11]

There are several sources of bias in $\hat{\delta}$. Advantaged students tend to go to better schools, suggesting a positive correlation between elements of $\overline{Z}_{S_{-i}}$ representing high socioeconomic status and $\epsilon_{\theta i}$ (through ξ_s). This would lead the estimates of δ to be biased upward (away from zero). We have already noted that bias in the γ estimates could spill over into δ, but the sign is ambiguous. The fact that some high schools have more than one feeder school will bias our estimates of δ to the extent that the mean of Z_i varies across feeder schools for a given high school if the sample is not representative of the mix of students from the various feeder schools. In practice, we usually only have students from one feeder school. In this situation, the component $[\overline{Z}(S_i, 0)' - Z_i]$ will be negatively correlated with $\overline{Z}_S$, the sample average in the high school. This effect biases the estimate of δ downward. On the other hand, these students were peers during eighth grade as well. Since δ is defined to be the effect of high school peers, δ will be biased upward to the extent that eighth grade peer effects continue to matter for high school. Bias in δ will bias the magnitude of $\pi^p(\tau)$ in the same direction. We suspect that the net effect of the various biases in δ is positive, which would mean that our estimates can be viewed as upper bounds for how large the treatment effect could be. However, we also experiment with alternative values of δ.

As it turns out, when δ is unrestricted,δ is poorly determined in the sample because there is substantial collinearity among the $\overline{Z}_{S_{-i}}$. Consequently, we focus on models that impose index restrictions on $\overline{Z}(s,\tau)'\delta$. The first is

$$\overline{Z}(s,\tau)'\delta = c + \delta_{X'\beta}\overline{X}(s,\tau)'\beta. \tag{12}$$

The above equation states that up to a factor of proportionality peer effects depend on average student characteristics in the same way that the school choice does. Even if this restriction is false, the fact that the cream skimming effect of the voucher has to work through $X_i'\beta + t_i(\tau)$ implies that one can think of (12) as a "reduced form" that is a first order approximation to $\overline{Z}(s,\tau)'\delta$.

The second and perhaps more natural assumption is that

$$\overline{Z}(s,\tau)'\delta = \delta_{X'\gamma}\overline{X}(s,\tau)'\gamma. \tag{13}$$

The restriction states that up to a factor of proportionality, peer effects depend on the school mean of X_i in the same way the outcome Y_i depends on X_i.

[11] Because we must deal with sampling error in our measures of peer characteristics, the use of nonlinear alternatives to the linear probability model with fixed effects would greatly complicate that analysis.

We estimate $\delta_{X'\beta}$ by JIVE which involves regressing $\hat{\theta}(S_i,0)$ on $X_i'\hat{\beta}$ and Q using $\overline{X}'_{S_{-i}}\hat{\beta}$ as an excluded instrumental variable for $X_i'\hat{\beta}$. We estimate $\delta_{X'\gamma}$ analogously by regressing $\hat{\theta}(S_i,0)$ on $X_i'\hat{\gamma}$ and Q_{S_i} using $\overline{X}'_{S_i-i}\hat{\gamma}$ as the excluded instrumental variable for $X_i'\hat{\gamma}$.

To evaluate (6), we use the fact that $\overline{Z}_{S_{-i}}$ is an unbiased estimator of $\overline{Z}\,(S_i,0)$ to rewrite (6) as[12]

$$\pi^p(\tau) = E(\psi_i(\tau)[Z_i - \overline{Z}\,(S_i,0)]'\delta|P_i^0 = 1) = E(\psi_i(\tau)[Z_i - \overline{Z}_{S_{-i}}]'\delta|P_i^0 = 1). \tag{14}$$

In the case of the $X'\gamma$ index model, we obtain a consistent estimator $\widehat{\pi}^p\,(\tau)$ by replacing the right hand side of (14) with its sample analogue

$$\widehat{\pi}^p\,(\tau) = \frac{1}{N_{P^0}} \sum_{\{i:\,P_i^0=1\}} \hat{\psi}_i(\tau)(Z_i - \overline{Z}_{S_{-i}})'\hat{\delta}. \tag{15}$$

$$= \frac{1}{N_{P^0}} \sum_{\{i:\,P_i^0=1\}} \hat{\psi}_i(\tau)(X_i - \overline{X}_{S_{-i}})\hat{\gamma}\hat{\delta}_{X'\gamma}. \tag{16}$$

In the case of the model with index $X_i'\beta$,

$$\widehat{\pi}^p\,(\tau) = \frac{1}{N_{P^0}} \sum_{\{i:\,P_i^0=1\}} \hat{\psi}_i(\tau)(X_i - \overline{X}_{S_{-i}})'\hat{\beta}\hat{\gamma}_{X'\beta}. \tag{17}$$

4 Data

NELS:88 is a National Center for Education Statistics (NCES) survey that began in the Spring of 1988. A total of 1032 schools contributed as many as 26 eighth grade students to the base year survey, resulting in 24,599 eighth graders participating.[13] Subsamples of these individuals were reinterviewed in 1990, 1992, 1994 and 2000. The NCES only attempted to contact 20,062 base-year respondents in the first and second follow-ups, and only 14,041 in the 1994 survey. Additional observations are lost due to attrition. A subsample consisting of 15,623 individuals were re-interviewed in 2000, when most respondents were 26 years old. We use information on income from this wave. Our analysis is based primarily on the restricted use version of NELS:88, to which we have merged characteristics of the geographic area and school district.

Parent, student, and teacher surveys in the base year provide information on family and individual background and as well as a very rich set of pre-high school achievement and behaviors. Each student was also administered a series of cognitive tests in the 1988, 1990, and 1992

[12] We leave Z_i out because random variation in Z across students from the same high school makes the correlation between Z_i and the mean including Z_i stronger than the correlation between Z_i and $\bar{Z}(P_i,0)$.

[13] This description draws heavily from Altonji, Elder and Taber (2003).

surveys to ascertain aptitude and achievement in math, science, reading, and history. We use the behavior measures and 8th grade test scores as person specific control variables and peer measures. They have the advantage of being determined prior to high school.

Our main outcome measure is high school graduation (HS_i). HS_i is one if the respondent graduated high school by the date of the 1994 survey, and zero otherwise. The school choice variables P_i, CH_i, NC_i are mutually exclusive indicators for whether the current or last school in which the individual was enrolled as of 1990 (two years after the eighth grade year) was a public high school, and Catholic high school, or a non-Catholic private high school, respectively.[14] Unless noted otherwise, the results in the paper are weighted.[15] Definitions of variables are provided in Appendix 1. In the empirical analysis missing values for key explanatory variables are replaced by their respective unweighted average values and we include missing value dummies in the school choice and outcome models for a few variables as indicated in the tables.[16] Observations with missing values of the school ID or the school type are dropped. The school choice sample contains 16,483 observations, of whom 14,193 chose a public high school, 1,354 chose a Catholic school, and 936 chose a non-Catholic private school.

Because of the complexity of the estimator of $\pi^p(\tau)$ and its components, we use a block bootstrap method to compute standard errors, confidence intervals, and bias corrections for most of the parameters. The blocks allow for correlation in the error terms among students who attend the same eighth grade and among the students who attend the same high school. The blocks consist of students from each set of eighth grades who sent at least one student to a common high school. See the Appendix 1 for more detail. The distribution of N_s, the number of sample members in each high school, is concentrated between 6 and 18 observations in our effective sample.

[14]A student who started in a private high school and transferred to a public school prior to the tenth grade survey would be coded as attending a public high school ($P = 0$). In the case of Catholic schools Altonji, Elder and Taber (2003) present evidence that is a minor issue.

[15]The sampling scheme in the NELS:88 is complicated. See the *NELS Base Year Sample Design Report* and the *National Education Longitudinal Study: 1988-1994 Methodology Report* for details. The weights depend in part on school choice and on outcomes, so it is important to weight. We use the 3rd follow-up panel weights (f3pnlwt) for all analyses involving high school graduation. For the school choice models we use the first follow-up panel weights (f1pnlwt). Weighting is particularly tricky when peer effects and/or unobserved school characteristics influence school choice, because the appropriate weights depend upon the behavior of the group NELS:88 eighth graders who are assigned to a particular public high school. We discuss the issues in Appendix 2 and 3, respectively, but do not provide a full solution.

[16]After the paper was essentially complete, we discovered a coding error in the "Lack of effort index" that we used in the analysis. This has little effect on the estimates of the choice and outcome equations and will be fixed in a later version. The descriptive statistics are for the correctly coded variable.

5 Descriptive Statistics

5.1 Student Outcome and Characteristics

Table 1 presents weighted means and standard deviations for the variables we use in the analysis, with imputed values excluded. The main point to be made from the table is that children who attend either Catholic high school or other private high schools are advantaged relative to students in public schools. For example, they come from families with substantially higher incomes, have better educated parents, are more likely to have both father and mother present, and have higher 8th grade achievement scores. They also have a 0.63 advantage in log income (10.84 versus 10.21). Using the estimates of the standard deviation of the school specific and student specific components of log family income (not reported) one may calculate that the income gap of 0.63 is equal to a 1.466 standard deviation (unweighted) shift in the component of parental income that varies across public high schools and to a 0.7037 standard deviation shift in the student specific component of family income. The gap in eighth grade math scores between private high school students and public high school students is 0.39 standard deviations. The gap amounts to 0.896 of standard deviation of the public high school specific component of this variable.

Table 1 also shows that private high school students tend to look stronger on a number of measures of 8th grade behavior. For example, they score lower on an index of delinquency, are less likely to have gotten into fights with other students, and are less likely to have behavior problems. The fraction of students who rarely complete homework is lower, they are much less likely to have repeated at least one grade between 4th and 8th grade, and they score much lower on a composite measure of the risk of dropping out. The large gap between private high school student and public school students on a broad range of observed characteristics that are relevant for school achievement is part of the cause for concern that vouchers will lead more advantaged students to leave public schools.

Turning to outcomes, students who attend private high schools are much more likely to graduate from high school than public high school students (0.942 versus 0.863) and more likely to be attending a four year college two years after the normal high school graduation year (0.584 versus 0.297).

6 Results for the Basic Model

In this section we present cream skimming estimates for our baseline specification. In this case, we exclude peer effects ($\varphi = 0$) and unobserved school district specific attributes from the school choice model (8). We also assume that only observed peer characteristics matter for school quality. We begin with a discussion of the school choice estimates and the effects of a student's own characteristics on outcomes. We then turn to the effects of student body characteristics on outcomes. Finally, we present estimates of the effects of a voucher program on the characteristics of those who remain in public school as well as estimates of $\pi^p(\tau)$.

6.1 Estimates of the Basic School Choice Model

Appendix Table A1 presents MLE-probit estimates of β for the basic school choice model (1) for the full sample. The dependent variable P_i^0 is one if the student attended public school and zero if the student attended a private high school.

We have chosen a rich set of regressors rather than a parsimonious specification to ensure that our estimates of peer effects are not contaminated by failure to control for the direct effects of a student's own characteristics and to ensure that the indices $X_i'\beta$ and $X_i'\gamma$ are strong predictors of preference for public school and of high school graduation. In addition to *Catholic*, which is 1 if the parents are Catholic, the equation contains gender, and race/ethnicity dummies, region dummies, and urban and suburban dummies. It also contains multiple measures of parental background (both parents present, father's education, mother's education, and log income), multiple measures of aptitude and achievement, including grades and 8th grade reading, math, science and history tests, and multiple behavioral and student performance measures. Consequently, individual variables are hard to interpret and most are statistically insignificant, although the variables are collectively highly significant and the pseudo R^2 of the model is 0.22

A few results are worth highlighting. Students with better-educated mothers and fathers are less likely to attend public school. Parental income is negative and significant. The effects are not large however. A four year increase in both father's education and mother's education accompanied by an increase in log family income of 0.4 would lead to a decline in the public school probability of about 0.084. Of course, these estimates hold the cognitive and behavior measures constant. Reading enters with a statistically signficant negative coefficient but coefficients on the other tests vary in sign and are not significant. All the behavioral measures that have a statistically significant negative association with high school graduation are positively associated with public school attendance, although typically not statistically significant. Students in urban areas are much less likely to attend public school. The same is true of suburban students. These results are probably influenced by the fact that private schools, particularly Catholic schools, are concentrated in urban and suburban areas.

Given the prevalence of Catholic high schools, the large negative coefficient on *Catholic* is not surprising. Because religious preference has very special role in the decision to attend a Catholic high school, *Catholic* is set to 0 when we evaluate the indices $X_i'\beta$ and $X_i'\gamma$ for the purpose of imposing index restrictions on the peer effect parameters δ. Under a voucher program, the fraction of private high schools that are Catholic would probably decline.

In results that are available from the authors on request, the marginal effects derivatives are typically higher for the urban subsample than the full sample, often by a factor of two or three. This is not surprising as families with high socioeconomic characteristics who live in the suburbs are more likely to send their children to public schools.

6.2 The Effect of a Student's Own Characteristics on High School Graduation

Appendix Table A2 presents estimates of γ, the effect of student's own characteristics on high school graduation, holding high school characteristics common to all students constant. Estimation is based on the public high school sample. The estimates are the coefficients from a linear probability model with high school fixed effects included. Block bootstrap confidence intervals are included in parentheses. We are well aware of the limitations of the linear probability model with fixed effects, but fixed effects probit or logit estimators are unattractive for a variety of reasons.

Because we work with a very rich model, the estimates for specific family background variables, 8th grade test score and achievement measures, and behavioral measures are hard to interpret and are not of central interest for our study. The results for most of the variables are consistent with the literature. We obtain positive coefficients on father's education and family income and they are significant. Not surprisingly, the math test score enters positively. The positive coefficient on Black is consistent with other studies of educational attainment that control for test scores and family background. The graduation probability is 0.148 lower for students who repeat a grade, 0.167 lower for students who are frequently absent, and .091 lower for students who rarely complete homework.

6.3 Effects of Student Body Characteristics on Outcomes

In Table 2 we report JIVE estimates of the effects of student body characteristics on outcomes with the restrictions (12) or (13) imposed on (11). The table reports $\hat{\delta}_{X'\beta}$ and $\hat{\delta}_{X'\gamma}$. 95% confidence intervals are in parentheses. These are based upon 1000 bootstrap replications. The vector Q_S consists of the region indicators, urban and suburban indicators, and a quadratic in distance from a Catholic school (coefficients not shown).

When we impose the restriction (12) that δ is proportional to β, our estimate of the coefficient $\delta_{X'\beta}$ on $\bar{X}'\hat{\beta}$ is -0.0453 with a confidence interval from -0.095 to 0.003. To give some sense of the magnitude, a change in the $X'\hat{\beta}$ of 0.3 would change the probability of attending a public school by roughly five percentage points. If we compare two high schools where $\bar{X}'\hat{\beta}$ differed by 0.3, the peers at the high school with a lower average propensity to attend public school would induce an increase in graduation of roughly $0.3 \times 0.0453 = 0.014$ in the graduation probability. If instead we impose the restriction (13), $\hat{\delta}_{X'\gamma}$ is 0.36. Given that $\overline{X}(s,\tau)'\hat{\gamma}$ and Y are in the same units, the point estimate says that the contribution of an increase in $X_i'\hat{\gamma}$ equal to $\Delta X_i'\hat{\gamma}$ for student i in a high school to the graduation rate of that high school is the sum of $\Delta X_i'\gamma$, the direct effect of $\bar{X}_i'\hat{\gamma}$ on i plus $0.36\Delta X_i'\gamma$. Consequently the fraction $0.265 = 0.36/(0.36+1)$ of the effect of $X_i'\hat{\gamma}$ on the graduation rate for a given high school operates through peer effects. This is a substantial externality. The estimate of $\hat{\delta}_{X'\gamma}$ is statistically distinct from zero, but the the 95% confidence interval is fairly wide: (0.029, 0.617).

6.4 The Effects of the Voucher Program on the Characteristics of Who Attends Public School

We begin by comparing the mean characteristics of public school stayers and movers for our base model and a universal voucher $t_i(\tau) = t$. For each bootstrap replication, the value of t is set to a level that is sufficient to induce 10% of the public school students to switch to private school. This point estimate of t is 0.5356, which can be interpreted as equivalent to a 0.5356 standard deviation change in the index of unobservables that determines school choice.

Point estimates and 95% confidence interval estimates of the means of selected elements of X_i are displayed for stayers and for movers in columns (1) and (2) of Table 3. The results show that the mean for movers is larger for two parents present, father's education, mother's education, log income, and all four test scores. Note that the sign, relative size and statistical significance of the differences between movers and stayers in the means of the elements of X_i are only weakly related to the sign, size and significance of the corresponding elements of $\hat{\beta}$ despite the key role of $\hat{\beta}$ in determining the relative odds that an individual will remain in public school in response to a voucher. For example, the mover-stayer difference in means is 1.26 for father's education and 0.94 for mother's education. The stayer-mover difference for a particular variable is affected by how it is correlated with other variables that influence school choice.

The fourth column of the table also reports the change in the average value of peer characteristics—the average value of Z of the peers of those who stay in public schools. This comes from the formula

$$\frac{1}{N_{P^0}} \sum_{i,\ P_i^0=1} \widehat{\psi}_i(\tau)(Z_i - \overline{Z}_{S_{-i}}).$$

Given the values in the fourth column, all we need to do to obtain the overall treatment effect is to multiply by δ. The changes are small. For example, there is little change in race/ethnic composition of peers. The prevalence of two-parent households drops by only -0.01. Father's and mother's education drop by -0.07 and -0.05, respectively and the log of parental income drops by -.025. The math test score declines by -0.12, which is only 0.012 standard deviations at the individual level.[17] Thus one can already see that it will take large peer group coefficients to lead to large overall cream skimming effect on outcomes.

A good way to summarize change in composition is compare the values of $X_i'\gamma$ and $X_i'\beta$ for movers and stayers. The bottom panel of Table 3 presents the point estimates and the 95% confidence intervals for the school means of the indices $X_i'\beta$, $\bar{X}_i{}'\beta$ $X_i'\gamma$, and $\bar{X}_i{}'\gamma$ by mobility group status. We exclude Catholic from X_i when computing $X_i'\gamma$ and $X_i'\beta$. Overall, the results suggest that a universal voucher program of the magnitude that we consider is unlikely to have a very large effect on the peers of the children who remain in public school. Consequently, unless outcomes are very sensitive to peers, the voucher program is not likely to have a substantial negative effect on how public school stayers do. However, in thinking about magnitudes one needs to compare the impact on the stayers to the gain of the movers. Since only 10% of the students move, the stayers are nine times more numerous than the movers. We will return to this issue momentarily.

6.5 Base Case Estimates of the Cream Skimming Effect $\pi^p(\tau)$

Table 4 presents a variety of estimates of the main parameter of interest— the cream skimming effect $\pi^p(\tau)$. Row 1 presents results for the base cases, for which the school choice model is column 1 of Table A1 and the estimates of γ are in Table A2. Keep in mind that the base case assumes that there is no school specific component in the school choice error term and that there is no peer interaction in the school choice model ($\varphi = 0$). It also assumes that peer effects depend on observed characteristics only. We consider alternative cases below.

In the column labeled "$X'\beta$ index" we impose the restriction (12). The point estimate of $\pi^p(\tau)$ is -0.0011 and the confidence interval is tight: -0.0017 to 0.0001. The lower bound estimate implies a small negative effect on stayers.[18]

In the column labeled "$X'\gamma$ Index" we impose (13). The point estimate of the change in the peer effect for stayers is -0.0013 and the lower bound to the confidence interval is -0.0023.[19] To put these numbers in perspective, it is helpful compare the direct benefits to students who are induced to move to the harm for students who are left behind after weighting by the size of the groups. Suppose that moving from public school to private school leads to an increase

[17]Not surprisingly, the means for movers of urban and suburban are larger than the means for stayers. In part, this reflects the fact that Catholic schools are much more prevalent in urban and suburban areas. Movers are more likely to be in the Northeast and somewhat less likely to be in the South, North Central and West regions. Movers also live closer to a Catholic school. Note that the "change in peers for stayers" for these variables reflect compositional shifts across schools. These variables are fixed for a given school, of course.

[18]Note that this effect can be calculated from numbers we discussed earlier. In particular $\widehat{\pi}^p(\tau) = \left[\frac{1}{N_{P^0}}\sum_{i,\ P_i^0=1}\widehat{\psi}_i(\tau)(X_i'\beta - \overline{X}'_{S_{-i}}\beta)\right]\widehat{\delta}_{X'\beta} = (0.0244)\times(-0.0453) = 0.0011$ where the first number comes from the second to last row of Table 3 and the second number comes from the first row of Table 2.

[19]As in the previous footnote, we can calculate this as $(-0.0037)\times(0.3614) = -0.0013$ where the numbers come from the last row of Table 3 and the first row of Table 4.

in the graduation rate by 0.06 for those who move. This estimate is in the range of what one obtains using single equation methods based on NELS:88 and is in the range of the lower bound estimates that Altonji, Elder and Taber (2005) obtain for Catholic high schools when they address the problem of selection on unobservables. The voucher program induces 10% of public school students to move, leaving 9 students in public school for everyone who moves. The point estimate of -0.0013 for the $X'\gamma$ index model implies that for each student who moves to private school the overall graduation rate for students who were in public school prior to the voucher rises by $0.06 - 0.0013 \times 9 = 0.048$. The gain of 0.06 for each student who moves is partially offset by a decline of 0.012 in the expected number of graduates among students who remain, an offset of about 20% of the direct benefit received by the child who switches to private school. Using the lower bound estimate of -0.0023, the negative impact on the number of stayers who graduates is 0.021 and the expected number of graduates among the pool of students who were in public school rises by 0.039 for each student who take up the voucher.

Although our focus is on high school graduation, during the process of our work we estimated the cream skimming effect using the alternative outcomes college enrollment in 1994 and the log of labor income in the year 2000 for full time workers. The estimates of $\pi^p(\tau)$ are small for these outcomes as well

7 Allowing School Choice to Depend on Peer Quality

In this section, we relax the assumption that the coefficient φ on the peer quality index $\bar{Z}(S_i, \tau)\delta$ in the school choice model (8) is 0. Doing so dramatically complicates that analysis for a number of reasons. The most important is that $\bar{Z}_{S_{-i}}$ is not observed for students who do not attend public school. Consequently, it must be simulated as part of the model estimation procedure. To the best of our knowledge, we are the first to estimate a demand model in which the choice of a consumer depends on characteristics of the other agents who choose it, and the relevant agent characteristics are only observed for those who choose the good. Our approach requires data on a set of observables at the district characteristic level, W_s, that shift the unconditional distribution of $X_i'\beta$ and $Z_i'\gamma$. The latter variables influence peer quality.[20]

Assume that

$$X_i'\beta = \mu^1_{s(i)} + \eta^1_i \, ; \quad X_i'\gamma = \mu^2_{s(i)} + \eta^2_i$$

[20]Bayer et al's (2007) model of housing and location demand is one of the few studies in which characteristics of other agents influence consumer choice. They have data on neighborhood characteristics for all consumers, which simplifies the analysis substantially. On the other hand, they consider location choice as well as housing choice and solve for equilibrium house prices, while we do not. Their model and estimation methodology builds on Berry et al (1994) and is very different from ours. Ferriera (2007) also estimates a model that allows for peers to affect choices. The style of estimation is also very different from ours.

where $s(i)$ is the school and i is the individual. We will usually suppress the dependence of s on i. Unconditionally, by which we mean prior to conditioning on the school choice decisions of those assigned to s, (η_i^1, η_i^2) is defined to have mean 0 and is uncorrelated with (μ_s^1, μ_s^2). We assume that W_s partially determines μ_s^1 and μ_s^2 through the equations

$$\begin{aligned} \mu_s^1 &= W_s'\alpha_1 + e_s^1 \\ \mu_s^2 &= W_s'\alpha_2 + e_s^2 \ , \end{aligned} \tag{18}$$

where (e_s^1, e_s^2) is $N(0, \Sigma_e)$. Similarly we assume that (η_i^1, η_i^2) is $N(0, \Sigma_\eta)$. Some but not all of the elements of W_s may be part of Q_s and affect school quality $\theta(s, \tau)$ and/or have a direct effect on school choice holding $X_i'\beta$ constant.

We estimate versions of the model analogous to the two cases discussed above. The case (13) assumes that effects of peers on school quality (δ) are proportional to γ while the case (12) assumes δ is proportional to β. To save space, we focus on the more complicated γ case in the text, but describe both the γ and the β case in Appendix 2. To simplify the expressions below define the conditional expectation

$$\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right) \equiv E\left(X_i'\gamma \mid P_i^\tau = 1, \mu_s^1, \mu_s^2\right).$$

Throughout this section we consider universal vouchers, so $t_i(\tau) = t(\tau)$ for all i. Keep in mind that we denote the absence of a voucher program as $\tau = 0$, with $t_i(0) = 0$.

Using the above equations and normality we may write $\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right)$ as

$$\begin{aligned} \overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right) &= \mu_s^2 + E\left(\eta_i^2 \mid \mu_s^1 + \eta_i^1 + \varphi\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right) - t(\tau) + u_i > 0\right) \\ &= \mu_s^2 + \frac{cov(\eta_i^1, \eta_i^2)}{\sqrt{1 + var(\eta_i^1)}}\lambda\left(\frac{\mu_s^1 - t(\tau) + \varphi\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right)}{\sqrt{1 + var(\eta_i^1)}}\right) \end{aligned} \tag{19}$$

where λ is the inverse Mill's ratio ($\phi(\cdot)/\Phi(\cdot)$). Since there is no closed form solution for $\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right)$, we have to solve for it numerically. Multiple equilibrium are possible because the demand for public school depends on the choices of other students. Fortunately, we have not found this to be a problem in practice.

We can estimate the treatment effect using our formula above with

$$\psi_i(\tau) = \frac{\left[\frac{\int \Phi(X_i'\beta + \varphi\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right) - t(\tau))dF\left(\mu_s^1, \mu_s^2 | X_i'\beta, X_i'\gamma, W_{s(i)}\right)}{\int \Phi(X_i'\beta + \varphi\overline{\mu}^2(0, \mu_s^1, \mu_s^2))dF\left(\mu_s^1, \mu_s^2 | X_i'\beta, X_i'\gamma, W_{s(i)}\right)}\right]}{\int \frac{\int \Phi(X_i'\beta + \varphi\overline{\mu}^2(\tau, \mu_s^1, \mu_s^2) - t(\tau))dF\left(\mu_s^1, \mu_s^2 | X_i'\beta, X_i'\gamma, W_{s(i)}\right)}{\int \Phi(X_i'\beta + \varphi\overline{\mu}^2(0, \mu_s^1, \mu_s^2))dF\left(\mu_s^1, \mu_s^2 | X_i'\beta, X_i'\gamma, W_{s(i)}\right)} dG(X_i, W_{s(i)} | P_i^0 = 1)}.$$

The estimation procedure is complicated, and so we relegate the details to Appendix 1. Here is a rough description. For the observed data, $\tau = 0$ and $t(0) = 0$, taking values of $(\alpha_1, \alpha_2, \Sigma_\eta, \Sigma_e)$ and $\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)$ as given, we estimate β and φ using the probit likelihood function for school choice and estimate γ from Eq. (2) using OLS. Given the estimates of β, γ, and φ, we use OLS and (18) to estimate α_1 and α_2. We then estimate Σ_η and Σ_e from the likelihood function for $X_i'\beta$ and $X_i'\gamma$ treating $\alpha_1, \alpha_2, \beta$, and γ as known. Finally, taking $\alpha_1, \alpha_2, \beta$, φ, Σ_η and Σ_e as given, for each s we update $\overline{\mu}^2(\tau, \mu_s^1, \mu_s^2)$ as the fixed point of (19). We iterate on this procedure until we find a fixed point.

The vector W_s consists of the average of the demographic characteristics for the zip code in which an eighth grade is located.[21] The characteristics are based on the 1990 Census and consist of percent black, percent Hispanic, an indicator for whether percent black is missing, median income, the percent of the population below the poverty line, and percent of the population with income more than double the poverty line. These variables explain an additional 11.96 percent of the cross high school variance in $X_i'\beta$ and 13.68 percent of the cross high school variance in $X_i'\gamma$ conditional on other area characteristics that we control for, which include census region, urbanicity and distance from the nearest Catholic high school.[22]

When (12) is imposed, $\hat{\varphi}$ is -0.3695 and the coefficients on the other variables in the choice equation are similar to those for the base model in Appendix Table A1. This says that a 0.1 increase in peer quality as measured by the index $\overline{X}'\beta$ governing the propensity to attend public school lowers the probability of attending by about 37% as much as a 0.1 increase in $X_i'\beta$ raises it. As can be seen in Table 2, the estimate of the school quality coefficient $\delta_{X'\beta}$ is -0.0412, which is very close to the estimate without φ. To give some sense of the magnitude, as in the base model, a change in the $\bar{X}'\hat{\beta}$ of 0.3 would change the proability of attending a public school by roughly five percentage points. If we compare two high schools where the index differs by 0.3, the peers at the high school with a lower propensity to attend public school would induce an increase of roughly $0.3 \times 0.0412 = 0.012$ in the graduation probability.

In the case of peer model γ, $\hat{\varphi}$ is 4.68 and $\delta_{X'\gamma}$ is 0.3614. $\hat{\varphi}$ is significant at the 10% level but not the 5% level.[23] The estimates of $\hat{\varphi}$ and $\hat{\delta_{X'\gamma}}$ imply that the peer effect of a shift of 0.1 in $\bar{X}_s\hat{\gamma}'$ raises the graduation rate by 0.036 and shifts the latent variable for public school attendance by about the same amount as the combined effect of an additional four more years of both mother's education and father's education. Of course, it is possible that peers affect the demand for public schools through other aspects of preferences in addition to the academic quality of the high school, as we mentioned earlier in our discussion of Ferreyra (2007). Note

[21] If public high school in NELS:88 receives sample members from NELS:88 eighth grades that are in different zip codes, we assign the average of the census characteristics for the different zip codes to students from all of eight grades. Our reasoning is that that the average will provide a better measure of the characteristics of the public high school option.

[22] These estimates are unweighted and based on an analysis of variance estimator. *Catholic* is excluded from both indices.

[23] The 90% confidence interval for this paramter is (0.2774,10.3294). The 90% confidence interval for the "$\bar{X}'\beta$" includes zero. It is (-0.6116,0.0486).

also that a difference across schools of 0.1 in $\bar{X}_s\hat{\gamma}'$ would be associated with a difference of 0.1 + 0.036 in the graduate rate for the high school. Families may choose schools based on average graduate rates rather than school quality per se.

The point estimates of the cream skimming effect for the two specifications are -0.0012 and -0.0014, respectively (Panel 2 of Table 4). These values are very close to the base case results with φ set to 0. We conclude that while peers play a role in school choice, they do not have much influence on the size of the cream skimming effect.[24] To provide intuition for this result, note first that our analysis already assumes that the students on the margin of attending are the most likely to to be affected by a voucher. We have already shown that the students who leave are advantaged relative to those who stay. The influence of peers on demand serves as a multiplier for the demand response to the voucher. This will influence the fraction of students who move as the result of a voucher of a given size. However, it does not change the mix of students who move by very much. If one were to allow for interactions between peer characteristics and student characteristics, then accounting for peers in the school choice model might make a bigger difference. We leave this extension to future research.

Given that our basic results are not sensitive to allowing peer characteristics to influence school choice and given the complexity of allowing for them, we set φ to 0 in what follows.

8 Unobservable Peer Characteristics

In this section and in Appendix 3 we present a methodology for estimating the cream skimming effect in the presence of both unobservable peer effects and unobservable school and community characteristics that influence school choice but do not depend on the voucher.

Using notation analogous to Section 7, define

$$\bar{u}\left(s,\tau\right) \;=\; E\left(u_i \mid S_i = s, P_i^{\tau} = 1\right); \qquad \bar{\varepsilon}\left(s,\tau\right)=\mathrm{E}(\varepsilon_i \mid S_i = s, P_i^{\tau} = 1)$$

where, as a reminder, u_i is the error term in the selection equation (1), ε_i is the error term in the outcome equation and $\overline{u}\left(s,\tau\right)$ and $\overline{\varepsilon}\left(s,\tau\right)$ are the means of u_i and ε_i among students assigned to school s who attend public school when the voucher program is τ. In what follows, partition Z_i into the observables X_i and an index of unobservables v_i and write the peer effect as

$$\bar{Z}(s,\tau)'\delta = \overline{X}\left(s,\tau\right)'\delta_x + \bar{v}(s,\tau)$$

where $\bar{v}(s,\tau) = E(v_i|S_i = s, P_i^{\tau} = 1)$. In the case of the β restriction (12), $\bar{v}(s,\tau) = 0$ and $\overline{X}\left(s,\tau\right)'\delta_x = \delta_{X'\beta}\overline{X}\left(s,\tau\right)'\beta$. A natural generalization of this model is to assume that peer

[24] We also estimated a model in which the peer effect on school choice operates through $X_i'\beta$ while the peer effect on the outcome operates through $X_i'\gamma$ and a model in which the peer effect in the decision operates through $X_i'\gamma$ but the peer effect in the outcome operates through $X_i'\beta$. These results give very similar results to those already provided.

quality depends upon the unobservable student characteristics that influence school choice, as in

$$\bar{Z}(s,\tau)'\delta = \delta_{X'\beta}\left[\overline{X}(s,\tau)'\beta + g\bar{u}(s,\tau)\right] \tag{20}$$

where $\bar{u}(s,\tau)$ enters with the coefficient $g\delta_{X'\beta}$.[25] Analogously the natural extension of the peer specification (13) is

$$\bar{Z}(s,\tau)'\delta = \delta_{X'\gamma}\left[\overline{X}(s,\tau)'\gamma + g\bar{\varepsilon}\,(s,\tau)\right]. \tag{21}$$

Recall that $\bar{Z}(s,\tau)'\delta$ represents the peer effect and so directly affects the output θ.

Conditional on β, which is identified from the school choice equation, g is identified under our functional form assumptions. However, we do not rely on the functional form assumptions and instead produce results for a range of values of g. The special case $g = 1$ corresponds to the assumption that "the unobservables are like observables" in that in (20), the coefficient relating $\theta(s,\tau)$ to the school mean of the index of observed variables that determines school choice, $\overline{X}(s,\tau)'\beta$, is the same as coefficient on the school mean of the error in the choice equation, $\overline{u}(s,\tau)$. Altonji, Elder, and Taber (2003, 2005) provide a model that can be used to justify $g = 1$ under some very strong assumptions. However, we do not impose only this value but instead consider the cases $g = .5$, $g = 1$, and $g = 1.5$.

To proceed, we first must define some new notation. As in the previous section, we decompose $X_i'\beta$ and $X_i'\gamma$, into

$$X_i'\beta = \mu^1_{s(i)} + \eta^1_i \;;\; X_i'\gamma = \mu^2_{s(i)} + \eta^2_i,$$

where s is the school level and i is the individual level. Analogously we decompose u_i and ε_i into

$$u_i = v^1_{s(i)} + \omega^1_i \;;\; \varepsilon_i = v^2_{s(i)} + \omega^2_i.$$

We assume that (η^1_i, η^2_i) and (ω^1_i, ω^2_i) are both jointly normal. The terms (μ^1_s, μ^2_s) may depend on W_s through (18), in which case we assume further that $(W_s'\alpha_1, W_s'\alpha_2)$ is jointly normal. However, we do not need to make use of the decomposition (18) and thus do not incorporate W_s and estimation of α_1 and α_2 into the analysis.

Let $\Sigma_\mu, \Sigma_v, \Sigma_\eta$, and Σ_ω be the variance covariance matrices of $(\mu^1_s, \mu^2_s), (v^1_s, v^2_s), (\eta^1_i, \eta^2_i)$ and (ω^1_i, ω^2_i) respectively. We use another "observables are like the unobservables" assumption which states that there is a single scalar a such that

[25] Equation (20) imposes the restrictions $\delta_x = \delta_{X'\beta}\beta$ and $\bar{v}(s,\tau) = g\delta_{X'\beta}\bar{u}(s,\tau)$.

$$\Sigma_v = a\Sigma_\mu$$

and

$$\begin{bmatrix} var(\omega_i^1) & cov(\omega_i^1,\omega_i^2) \\ cov(\omega_i^1,\omega_i^2) & var(\omega_i^2) \end{bmatrix} = a \begin{bmatrix} var(\eta_i^1) & cov(\eta_i^1,\eta_i^2) \\ cov(\eta_i^1,\eta_i^2) & var(\eta_i^2) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & V_{\omega^2} \end{bmatrix}.$$

Since $var(\omega_i^2)$ does not play a role in the analysis, we do not need to restrict this parameter.

The restrictions say that relative variances and the covariance of the unobservable cross district components on school choice and the outcome are the same as those of the observable components. The same is true of the student specific error components, with the exception of $var\left(\omega_i^2\right)$ which is not relevant in our approach. Estimation of Σ_μ and Σ_η is analogous to the endogenous peer quality model in the previous section and we discuss it in the appendix.

First consider the model in which the peer effect depends on the expected value of $X_i'\beta + gu_i$ of public school peers. As in the previous section we will consider the change from no vouchers ($t_i(0) = 0$) to universal vouchers ($t_i(\tau) = t$). The cream skimming effect of the new voucher program on those who remain in a public school from a school district with characteristics μ_s^1, v_s^1 is

$$\pi^p\left(\tau;\mu_s^1,v_s^1\right) = \delta_{X'\beta}\frac{\sigma_{\eta 11} + g\sigma_{\omega 11}}{\sqrt{\sigma_{\eta 11}+\sigma_{\omega 11}}}\left[\lambda\left(\frac{\mu_s^1+v_s^1-t}{\sqrt{\sigma_{\eta 11}+\sigma_{\omega 11}}}\right) - \lambda\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma_{\eta 11}+\sigma_{\omega 11}}}\right)\right] \tag{22}$$

Note that term $\mu_s^1 + gv_s^1$ drops out of the difference in conditional expectations because μ_s^1 and v_s^1 are school district specific rather than student specific. Consequently, they are not affected by the voucher. However, μ_s^1 and v_s^1 do influence the set of students who choose public school both before and after the voucher.[26]

When the peer effect is proportional to the expectation of $X_i'\gamma + g\varepsilon_i$ conditional on public school attendance, the cream-skimming effect is

$$\pi^p\left(\tau;\mu_s^1,v_s^1\right) = \delta_{X'\gamma}\frac{\sigma_{\eta 12} + g\sigma_{\omega 12}}{\sqrt{\sigma_{\eta 11}+\sigma_{\omega 11}}}\left[\lambda\left(\frac{\mu_s^1+v_s^1-t}{\sqrt{\sigma_{\eta 11}+\sigma_{\omega 11}}}\right) - \lambda\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma_{\eta 11}+\sigma_{\omega 11}}}\right)\right]. \tag{23}$$

To use (22) or (23) to identify the cream skimming effects for the β and γ specification of peer effects, we have to address a new issue. We do not observe (μ_s^1, v_s^1), which are arguments of cream skimming effect functions. We solve this problem by using Bayes theorem to infer the conditional distribution of (μ_s^1, v_s^1) given the characteristics of the other students who choose the

[26] For example, unobservables such as the skill of a district in marketing the public high school, the safety of the neighborhood immediately surrounding the high school, the attractiveness of the buildings, the characteristics of the local private school options, as well the district mean of unobserved family characteristics that influence school choice are all determinants of μ_s, the district average of u_i. These are not directly influenced by the voucher, and our analysis takes this fact into account. However, the fact most of the elements of X_i are student level might lead one to question the restriction "unobservables are like the observables restriction." One could explore sensitivity of the results to allowing the proportionality factor a to differ in the two restrictions, although we have not done so.

public school, with selection taken into account. We then integrate (μ_s^1, v_s^1) out and identify the average of of $\pi^p(\tau; \mu_s^1, v_s^1)$ over the distribution of μ_s^1, v_s^1 of those who stay in public school. The details are in Appendix 2.

In Table 2, $\hat{\delta}_{X'\beta}$ is 0.0169 when $g = 0.5$, 0.0157 when $g = 1$, and 0.0139 when $g = 1.5$,, which is close to the value of -0.045 in the base case. The estimates of $\delta_{X'\gamma}$ are 0.0314 when $g = 0.5$, 0.0321 when $g = 1$, and 0.0328 when $g = 1.5$, which are are substantially lower and more precise than the effect for the base model.

The estimates of the cream skimming effect are in the third panel of Table 4. Regardless of the assumption about g, the estimates of cream skimming effect are slightly positive when we use the $X'\beta$ index model. The estimates for the $X'\gamma$ model are about -0.0002. All of the estimates are close to zero and precisely estimated.

9 Sensitivity to Alternative Assumptions about the Effects of Peers on Outcomes

In the fifth panel of Table 4, we present estimates for alternative assumptions about the effects of peers on outcomes. In the first two rows of the panel we examine the sensitivity of our estimates of π to the coefficient $\delta_{X'\gamma}$ relating θ to $\overline{X}(s,\tau)'\gamma$. As noted earlier, our point estimate of $\delta_{X'\gamma}$ is 0.3614 when we exclude unobservable peer effects (Table 4, column 2, row 1) When we set $\delta_{X'\gamma}$ to 0.5, the estimate of the cream skimming effect is -0.0018. If we set $\delta_{X'\gamma}$ to 1, about 3 times our point estimate, the cream skimming effect is -0.0037, with a confidence interval from -0.0050 to -0.0023. On one hand, the decline of -0.0037 in the graduation rate is relative small in an absolute sense. However, since it is large enough to cancel out almost 2/3rds of the direct positive effect of moving 1 out of 10 students to private schools, assuming a private school effect of 0.06. On the other hand, a value of 1 seems extreme. It says that the indirect effect that $X_i'\gamma$ has on the graduate rate of a high school operating through $\overline{X}(s,\tau)'\gamma$ is as large as the direct effect on the graduate rate of person i.

Discussions of cream skimming often give special emphasis to negative consequences of isolation within public schools of children from low income families or racial minorities. In Table 4, Panel 5, we report results based on estimation of (3) restricting Z_i to consist only of average family income. This specification will tend to maximize the estimated impact of average family income on school outcomes. We obtain -0.0011, which is similar to our base case estimates. The next row of Table 4 is based on restricting Z_i to consist of only fraction African American. For this specification the estimate of the cream skimming effect is essentially 0. Note that from Table 3, the fraction of public school students who are African American increases by only 0.0035.

Finally, fifth panel also reports estimates under the assumption that peer effects are a linear combination of average test scores. In the "$X'\beta$ index" column we use test score coefficients from the school choice equation to form the index. In the "$X'\gamma$ index" column the test score coefficients are from outcome equation. Both of the estimates of π are very close to zero Finally, in the last row we assume peer effects operate only through father's education. The estimate of the cream skimming effect is only -0.0008.

10 Alternative School Choice Models

10.1 Treating Catholic and non-Catholic Private Schools as Distinct Options

Since preferences over Catholic and non-Catholic schools are likely to differ we explore the sensitivity to treating them as distinct choices. Table 4, Panel 4, reports estimates of the cream skimming effect using a nested logit specification. Catholic and non-Catholic private schools are in one nest and public school is in the other. Distinguishing private school type makes very little difference.[27]

We also consider a simpler case in which vouchers are only used for Catholic schools. The estimates in Panel 4 of Tables 6 are very similar to the base case and to the nested logit estimates.

10.2 School Choice Depends Only on $X'\gamma$

To guage the sensitivity of our results to possible mispecification of our models of school choice, we performed a simulation in which we assume that sorting into new voucher schools is determined *exclusively* by $X_i'\gamma$. To accomplish this, we rank public school students by $X_i'\hat{\gamma}$ and assume that the top 10% will move. Under this assumption, the average value of $X_i'\gamma$ for the movers is 0.5451 and the average value for the stayers is 0.3531. The average change in $\overline{X}(s,\tau)'\gamma$ for stayers is -0.0163. Using (13) as the peer effects specification, we obtain -0.0059=0.3614×(-0.0163) as the estimate of $\pi^p(\tau)$. This is a substantial effect relative to the private benefit of attending private school, but the assumption that the response to the voucher will be based entirely on the same index that determines peer effects is extreme. To see this point, contrast these results with the final row of table 3. When we freely estimate the school choice model the mean of $X_i'\gamma$ for movers is only 0.4144 and the change in the peers for stayers is only -0.0037.

[27] In the case of the peer effects specification (12) we use the index of coefficients that determine whether public school is chosen over the two alternatives. Our base model without unobservables and our models with unobservable peer effects generalize to the trinomial probit model in a natural way. However, we experienced numerical difficulties in estimating the trinomial choice model, which precluded the use of bootstrap methods to compute confidence intervals. The point estimates are generally consistent with the results that we obtained using a binomial choice model for a wide range of cases that we considered. Due to computational complexity, we have not tried the trivariate probit model when there are unobservable peer effects and unobserved school choice components that are specific to s or when peer quality influences demand.

Clearly the decision to attend public school depends on a lot more than just $X_i'\gamma$, but one can also see that with more selection one can obtain larger effects.

In the final panel of Table 4 we combine this experiment with setting $\delta_{X'\gamma} = 1$. This essentially makes two of the three factors in the cream skimming effect formula large–selection into private school is chosen to be as large as possible and the peer group effect is set to a very high value. This makes a large difference as we now obtain an estimate of $\pi^p(\tau)$ which is an order of magnitude higher than our original result.

10.3 Approximating School choice using the Milwaukee experience.

We also replace our school choice model with a model estimated with data from the Milwaukee voucher experience. We use data obtained from the Milwaukee Parental Choice Program data set (obtained from www.disc.wisc.edu). We restrict ourselves to the set of covariates that are similar in both the Milwaukee data sets and the NELS:88. This leaves us with Catholic religion, gender, race, a dummy variable for parent's present, father and mother's education, log of family income, and reading and math test scores (standardized). We estimate the choice parameters (β) using the data from Milwaukee only, but then estimate the outcome and peer effects from the NELS:88. In this case we obtain a small positive estimate of the cream skimming effect for both peer effect specifications. Thus when we compare our framework to an actual voucher program, it seems that if anything we are overstating the negative consequences.

10.4 Programs Targetted to Urban Students or to Low Income Students

We consider the cream-skimming effect of vouchers targetted to urban families after re-estimating our base model on the urban sample. The point estimates of the change in peer characteristics of stayers under an urban voucher program are very similar to those for a universal voucher and are small. The point estimates of $\pi^p(\tau)$ for the X'β index and X'γ index are 0.0071 and 0.0026 respectively (Table 4, Panel 4, row 3). However, the estimates are noisy for the (12) case, with a lower bound estimate of -0.0053, which is substantial.

Next we consider a program that limits eligibility to families whose incomes are in the lowest 20% of our sample. We again calculate the value of the voucher that would induce 10% of the eligible population to move to private school. The results are qualitatively similar to those for a universal voucher program in the sense that effects are small. However, because of the targeting, the peers of stayers become slightly *more* advantaged as a result of the voucher. For this reason, the point estimate of the cream skimming effect is 0.0001, which is positive but close to 0.[28]

[28]It is important to emphasize that we are estimating the effects on the full population of public school stayers of a voucher that moves 10% of the eligible population. This is only 2% of the full population. In this case if one

Programs that target all students in low income school districts regardless of income would have a different selection effect. We investigated this by re-estimating the base models on the subsample of schools in zipcodes with poverty rates above 16%, which is about 25% of the full sample. When peer effects depend on $X'\gamma$ the point estimate and is identical to the base case—-0.0013, but the confidence interval is considerably wider. When peer effects depend on $X'\beta$, the point estimate is -0.0018 but is very imprecise.[29]

11 Conclusion

The first contribution of the paper is to provide a simple formula showing that for a broad class of models of school choice and peer effects, the cream skimming effect is determined by the covariance between the relative probability that a student will move to private school in response to the voucher with an index of the differences between the student's characteristics and the average characteristics of his or her classmates. The index is weighted by the coefficients relating outcomes to peer characteristics. The formula for the cream skimming effect provides the structure for our empirical investigation.

We rely primarily on formal econometric analysis to estimate the school choice and school outcome parameters using several alternative models but we also perform a sensitivity analysis to alternative assumptions about school choice and the effects of peers on outcomes. We provide a method for allowing school choice to depend on peer characteristics even when peers are not directly observed for those who do not choose a public high school. We also provide a way to allow for both unobserved fixed characteristics of schools that influence school choice and for unobserved characteristics of peers that influence outcomes. Both methods may have other applications.

The specific parameter estimates vary with the details of the econometric specification and the voucher program specified. However, the point estimates and the lower bounds to the confidence interval estimates of the cream skimming effect of a voucher program on high school graduation rates are typically small in absolute value. The results suggest that the effects of vouchers on the *productivity* of public schools, either through a positive or negative response to competitive pressure or through an effect on the financial resources available in public schools,

---

wishes to compare the private gain in the high school graduation of those who take up the voucher to the losses of those who stay behind, one should multiply the cream skimming effect by 49 rather than the value of 9 that we used previously. However, given the small positive point estimate, even after taking this product, one is left with a small number, and it is positive rather than negative. One would like to know the effect of the targeted voucher on the targeted population. Unfortunately, it is not straightforward to use our approach to estimate the impact on members of the targeted subgroup, such as low income students, who remain in public schools unless there is no heterogeneity in the targeted group within schools or the samples of students from each public high school are large.

[29] The estimates are somewhat sensitive to the poverty rate cut off we choose, which is to be expected given the sampling error.

may be more important than the cream skimming effect.

Extrapolating to other types of choice programs is speculative since we do not examine them explicitly. However, our cream skimming effect formula would be identical. Under montonicity, the only difference would come from the relative probability of remaining in the default school, $\psi_i(\tau)$. In order for the overall effect to be sizeable, one would need the amount of cream skimming to be more severe than what we have simulated for voucher programs. The evidence to date on selection into Charter schools suggests that this is unlikely for the Charter school programs that have been introduced so far.

References

[1] Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," Northwestern University, revised April 2002.

[2] Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113(1), (2005): 151-184.

[3] Bayer, Patrick, Fernano Ferreira and Robert McMillan, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods", Journal of Political Economy 115(4) (2007): 588-638.

[4] Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium." *Econometrica 63 (July 1995):841-90.*

[5] Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland, *Catholic Schools and the Common Good*, Cambridge, Mass. : Harvard University Press, 1993.

[6] Bulkley, Katrina and Jennifer Fisler (2002), "An Overview of the Research on Charter Schools", CPRE Web Paper Series WP-01", June 2002,

[7] Caucutt, Elizabeth, "Educational Vouchers when there are Peer Group Effects-Size Matters," International Economic Review, Vol. 43 No. 1, February 2002, 195-222.

[8] Chubb, John E., and Terry M. Moe, *Politics, Markets, and America's Schools* (Washington, D.C.: The Brookings Institution, 1990).

[9] Coleman, James S., Thomas Hoffer, and Sally Kilgore, *High School Achievement: Public, Catholic, and Private Schools Compared* (New York, NY: Basic Books, Inc., 1982).

[10] Coleman, James S., and Thomas Hoffer, *Public and Private Schools: The Impact of Communities* (New York, NY: Basic Books, Inc., 1987).

[11] Cookson, Peter W., Jr., "Assessing Private School Effects: Implications for School Choice," in Edith Rasell and Richard Rothstein, eds., *School Choice: Examining the Evidence* (Washington, D.C.: Economic Policy Institute, 1993).

[12] Dynarski, Susan, Jonathan Gruber and Danielle Li, "Cheaper by the Dozen: Using Sibling Discounts at Catholic Schools to Estimate the Price Elasticity of Private School Attendance," NBER working Paper 15461, NBER, Inc.

[13] Evans, William N., and Robert M. Schwab, "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics,* 110 (1995), 947-974.

[14] Evans, William N., and Robert M. Schwab, "Who Benefits from Private Education: Evidence from Quantile Regressions," Department of Economics Working Paper, University of Maryland, August 1993.

[15] Epple, Dennis and Richard Romano,*"Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects",* 88(1), (March 1998): 33-62.

[16] ————, "Educational Vouchers and Cream Skimming", NBER Working Paper No. 9354, (November 2002).

[17] Epple, Dennis, Elizabeth Newlon, and Richard Romano,"The Effects of Educational Vouchers when Schools Track Students by Ability," *Public Economics*, 83, January 2002, 189-221.

[18] Epple, Dennis and Richard Romano,"Neighborhood Schools, Choice, and the Distribution of Educational Benefits," in *The Economics of School Choice,* ed. Caroline Hoxby, University of Chicago Press, Chicago, 2003, 227-286.

[19] Ferreyra, Maria Marta, "Estimating the Effects of Private School Vouchers in Multi-District Economies," American Economic Review, 97(3) (June 2007):789-817. .

[20] Figlio, David N., and Joe A. Stone, "Are Private Schools Really Better?," *Research in Labor Economics*, 18 JAI Press (2000): 115-140.

[21] Grogger, Jeff and Derek A. Neal, "Further Evidence on the Effects of Catholic Secondary Schooling," *Brookings-Wharton Papers on Urban Affairs, 2000*, 151-201.

[22] Heckman, James J., "Varieties of Selection Bias," *American Economic Review,* 80(1990).

[23] Howell, William and Paul Peterson, *The Education Gap,* Washington, DC, Brookings Press, (2002).

[24] Hoxby, Caroline M., "School Choice and School Productivity. Could School Choice be a Tide that Lifts All Boats", in *The Economics of School Choice*, C.M. Hoxby (editor), University of Chicago Press (2003): 287-342.

[25] Ladd, Helen F., "School Vouchers: A Critical View", *Journal of Economics Perspectives*, 16(4), (fall 2002):3-24.

[26] Manski, Charles F., "Educational Choice (Vouchers) and Social Mobility," *Economics of Education Review*, Vol. 1 No. 4, (1992), 351-369.

[27] Manski, Charles. F., "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, no3 (1993), 531-542.

[28] Murnane, Richard J., "A Review Essay - Comparisons of Public and Private Schools: Lessons from the Uproar." *Journal of Human Resources* 19 (1984), 263-77.

[29] Moffitt, Robert A., "Policy Intervensions, low-level equilibria, and social interactions," in *Social Dynamics*, Durlauf and Young eds., Washington D.C., Brookings Institution Press, 2001.

[30] Neal, Derek, "The Effects of Catholic Secondary Schooling on Educational Attainment," *Journal of Labor Economics* 15 (1997), 98-123.

[31] Neal, Derek, "How Vouchers Could Change the Market for Education", *Journal of Economic Perspectives*, 16(4), (Fall 2002):24-44.

[32] Nechyba, Thomas J. School Finance Induced Migration Patterns: The Impact of Private School Vouchers.", *Journal of Public Economic Theory*, 1(1), (1999): 5-50.

[33] ———, "Mobility, Targeting and Private School Vouchers." *American Economic Review*, 90(1) (March 2000): 130-146.

[34] Nechyba, Thomas, "Introducing School Choice into Multidistrict Public School Systems," n *The Economics of School Choice,* ed. Caroline Hoxby, University of Chicago Press, Chicago, 2003, 145-194.

[35] Rouse, Cecilia and Lisa Barrow, "School Vouchers and Student Achievement: Recent Evidence, Remaining Questions" *Annual Reviews of Economics, (September 2009):17-42.*

[36] Sander, William H. *Catholic Schools: Private and Social Effects*. Boston: Kluwer Academic Publishers, 2001.

[37] Witte, John F., "Private School Versus Public School Achievement: Are There Findings That Should Affect the Educational Choice Debate," *Economics of Education Review*, XI (1992), 371-394.

[38] Zimmer, Ron, Brian Gill, Kevin Booker, Stéphane Lavertu and John Witte, Do Charter Schools "Cream Skim" Students and Increase Racial-Ethnic Segregation?" http://www.vanderbilt.edu/schoolchoice/conference/papers/Zimmer_COMPLETE.pdf

Appendix 1: Data

A1.1 Sample Sizes By High School and Eighth Grade

Because of the complexity of the estimator of $\pi^p(\tau)$ and its components, we use a block bootstrap method to compute standard errors, confidence intervals, and bias corrections for most of the parameters. The methods accounts for correlation in the error terms among students who attend the same eighth grade and among the students who attend the same high school. The blocks consist of students from each set of eighth grades who sent at least one student to a common high school. For example, suppose that eighth grade A sent students to high school 1, 2, and 3, eighth grade B sent students to high school 1 and 3, and no other eighth grades represented in NELS:88 sent students to high school 1, 2, or 3. Then the students from eighth grade A and eight grade B constitute a block for purposes of constructing bootstrap replication samples.

About 86% of the high schools have students from only 1 eighth grade. This is to be expected because base year survey used eighth grade schools as strata. Among 39,000 schools containing the eighth grade in the U.S., 1,052 schools were selected. Since students usually go to a nearby high school, it is not very common in the sample for students from different eighth-grade schools to attend the same high school. About 58% of the eighth grades have sample students in only 1 high school. About 28% have sample students in 2 high schools and 10% in 3 high schools, with a small fraction sending sample members to 4 or more high schools. The distribution of observations per resampling block is concentrated between 6 and 30, but there are a few blocks with larger numbers of students. The largest block contains 965 students, and 7 blocks contain more than 100 students. We chose break up the blocks of more than 60 students into a separate block for each high school involved on pragmatic grounds, although doing so did not make much difference in the cases we checked. In practice, we also obtained similar confidence interval estimates if we treat students from each high school as a block.

The distribution of N_s, the number of sample students in each high school is concentrated between 6 and 18 observations.

A1.2 Description of Variables

NELS:88 variables used in the creation of the measures are shown in italics. This section draws upon Altonji, Elder, and Taber (2002)

Demographic Variables: These include indicators for female, hispanic, black, and whether catholic, which is created from parental responses concerning religion (*byp29*).

School Sector: Eighth Grade Sector (*g8ctrl1*) High School Sector (CH) (*g10ctrl1*)

Family Background Measures:

Catholic: 0-1 indicator for whether parents are Catholic (*byp29*).

Household composition: 0-1 indicator for whether the student lives with his/her mother and father in the base year. Created from (*byfcomp*).

Log family income : Continuous variable created using the midpoints of the ranges of the categorical variable base year variable *byfaminc* and $230,000 if families with income above $200,000 (the top category)

Missing value treatment: All family background variables are set equal to the sample mean when missing. 0-1 indicators for missing values are created for some of original variables as indicated in the tables.

Geographic Variables:

Region indicators and the Urban and Suburban indicators: Constructed from *g8region* and *g8urban* and refer to location of the 8th grade school the student attended. Missing values were dropped.

Distance to the nearest Catholic high school: This variable was constructed from the population weighted center of the zipcode of the 8th grade school and the population weighted centers of the zipcodes of all the Catholic high schools reported in Ganley's Catholic Schools in America, 1988 addition. See Altonji, Elder and Taber (2005). The units are 100,000 meters. A missing value indicator is included in the school choice equation.

Fraction black (*p008002*/*p001001*), fraction Hispanic (p0100001/p0010001), an indicator for whether fraction black is missing, median income, the fraction of the population below the poverty line ((*p1210001*+*p1210002*+*p1210003*)/*p0010001*), and the fraction of the population with income more than double the poverty line (*p1210009*/*p0010001*) are from the 1990 Census for the zipcode of the high school. Missing values were set to the sample mean.

Eighth Grade Test Score Measures:

We use the Item Response Theory scaled scores for reading, math, science, and history, civics and geography—*by2xrstd, by2xmstd, by2xsstd,* and *by2xhstd*. Missing values are set to the sample mean, and an indicator that is one when all of the tests are is included in the models. (With a few exceptions, the tests are either all missing or all available.)

Eighth Grade Behavioral and Performance-in-School Measures:

Delinquency Index: This variable is the sum of two variables and ranges from 0 to 4. The first is (*bys55a*), which is 1 if the student reports being sent to the office once or twice and 2 if sent more then 2 times. The second is *bys55e*, which is 1 if the student reports that his parents were contacted once or twice because of a behavior problem and 2 if they were contacted more than twice. It ranges from 0-4.

Student got in a fight: Created from student self-reported variable *bys55f*: 0 (never) 1 (once or twice) and 2 (more than twice) in the past semester.

Student performs below ability: 0-1 indicator variable taken from teacher surveys (*byt1_2* and *byt4_2*).

Student rarely completes homework: 0-1 indicator variable taken from teacher surveys (*byt1_3* and *byt4_3*).

Student frequently absent: 0-1 indicator variable taken from teacher surveys *(byt1_4* and *byt4_4*).

Student inattentive in class: 0-1 indicator variable taken from teacher surveys (*byt1_6* and *byt4_6*).

Student frequently disruptive in class: 0-1 indicator variable taken from teacher surveys (*byt1_8* and *byt4_8*).

Student Behavior Variables Missing: 0-1 indicator for whether any of the previous 5 variables are missing.

Trouble-Maker: 0-1 indicator variable created from *bys56e*, and coded as 1 if the student report indicates that other students see the respondent as a "very big" trouble-maker.

Behavior problem: 0-1 indicator variable created from *byp50*, regarding whether the parent considers their child to have a behavior problem in school.

Parents Contacted About Behavior: Created from *byp57e*, which measures the number of times parents report being contacted about behavior problems in the past school year. The values are 0 (never), 1 (once or twice), 2 (three or four time) and 3 (more than four times).

Limited English Proficiency Composite: 0-1 indicator variable (bylep). The NELS composite variable is based on student and teacher reports.

Repeated Grade: 0-1 indicator of whether a student repeated any grade 4-8, taken as the maximum of the student (bys74e-bys74i) and parent (*byp46e-byp46i*) reports.

Lack of Effort index: The base year student variable *bys75* measures "How many days of school did you miss over the past four weeks? The values are 0 (none) 1 (1 to 2) 2 (3 or 4) 3 (5 to 10) 4 (more than 10). *bys76* measures "How often do you cut or skip classes?" 0 (0) 1 (< once per week), 2 (at least once per week), 3 (daily). *bys77* is the response to "how many times were you late for school over the past four weeks?": 0 (0), 1 (1 or 2 days) 2 (3 or 4) 3 (5-10) 4 (more than 10). *bys78a, bys78b and bys78c* are responses to "How often do you come to class without pencil or paper when needed?", "How often do you come to class without books", and "How often do you come to class without homework. Each is coded as 3 (usually), 2 (often), 1 (seldom), 0 (never). The index is the sum of the 6 variables and ranges from 0 to 20.

Dropout risk index: This is NELS composite variable *byrisk*, ranging from 0-6. It is the sum of binary indicators for risk factors for dropout risk. The indicators are based on *byfcomp, bypared, byp6, bys41, bylep,* and *byfaminc.*

Grade Index: Based on *bygrads*, ranging from 0-4.

Gifted: 0-1 indicator for parent report of whether the student is currently enrolled in a gifted/talented program *(byp51)*.

Missing values of all variables were set to the sample mean.

Outcome Measures:

High School Graduation: 0-1 indicator for whether received high school diploma as of the third follow-up. One if *hsstat*=1.

College Attendance: 0-1 indicator for whether enrolled in a 4-year college as of April 1994. One if *enrl0494*=15 or 16.

Missing values are dropped.

Appendix 2: Estimation with Peer Effects in the School Choice Equation.

In this appendix we describe the estimation of the model when peer effects influence school choice, which is considered in Section 7. We first discuss estimation of the model when peers affect demand through the $E(X_i'\gamma)$ index, and then explain how the $E(X_i'\beta)$ index case differs. We estimate γ using the fixed effect for the base model. The data comes from the "no voucher" regime, and we suppress the indicator for the voucher program regime unless it is needed for clarity

A.2.1 The $X'\gamma$ Case

Partition the rest of the parameters that affect school choice into three subsets,

- β and φ
- α_1 and α_2
- Σ_η and Σ_e.

We estimate the model by iterating on the following procedure. In each iteration we update the parameters in the following steps, taking parameters from the previous step as given.

Step 1: Given an estimate of $(\alpha_1, \alpha_2, \Sigma_\eta, \Sigma_e, \beta)$ and $\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)$ from the previous iteration, we estimate β and φ using the likelihood for private and public schools. The log likelihood for individual i is

(24)

$$\begin{aligned}\mathcal{L}_i =& P_i w_i \log\left(\frac{\int \Phi\left(X_i'\beta + \varphi\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)\right)\phi\left(X_i'\beta^* - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right) d\Phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)}{\int \phi\left(X_i'\beta^* - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right) d\Phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)}\right) \\ &+ (1-P_i) w_i \log\left(1 - \frac{\int \Phi\left(X_i'\beta + \varphi\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)\right)\phi\left(X_i'\beta^* - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right) d\Phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)}{\int \phi\left(X_i'\beta^* - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right) d\Phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)}\right)\end{aligned}$$

where w_i is the sample weight for individual i and Φ is the cdf of a standard normal. We treat public and private school students symmetrically and do not use the data on peers to help update β in this step. Instead, we fix β^* in the above equation at the value $\hat{\beta}$ from the previous iteration rather than letting it change as we maximize the above likelihood function with respect to β and φ. This means that the update for $\hat{\beta}$ is chosen to maximize the likelihood of the school choice model rather than to make the $X_i'\beta$ distribution look approximately normal.

To see the intution behind Step 1, note that if we knew $\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)$ we would just run a probit of public school on X_i and $\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)$. Because we do not know it, we have to use the model to integrate out its conditional distribution given the data we have.

Step 2: Given β and γ, we estimate α_1 and α_2 by regressing $X_i'\beta$ and $X_i'\gamma$ on $W_{s(i)}$.

Step 3: Taking $(\beta,\gamma,\varphi,\alpha_1,\alpha_2)$ as given, we estimate Σ_η and Σ_e using the likelihood for public students only:

(25)

$$\mathcal{L}_s=\frac{\int\left[\prod_{\{i:S(i)=s\}}\left(\frac{\Phi\left(X_i'\beta+\varphi\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)\right)\phi\left(X_i'\beta-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)}{\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+var(\eta_i^1)}}\right)}\right)\right]\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+var(\eta_i^1)}}\right)d\Phi(\mu_s^1,\mu_s^2;W_s,\Sigma_e)}{\int\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+var(\eta_i^1)}}\right)d\Phi(\mu_s^1,\mu_s^2;W_s,\Sigma_e)}.$$

To see where the above equation comes from, one must consider the NELS sampling frame. In particular schools with larger values of $\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+var(\psi_i^1)}}\right)$ tend to be bigger. We make the simplifying assumption that the NELS sampling frame fixes the number of interviews at each public high school independent of the size of the school, but then oversamples bigger schools. In this case, $\mathcal{L}_s$ should be the likelihood of observing a particular realization of the vectors of values of $X_i'\beta$ and $X_i'\gamma$ for a particular sample of students from the school. Then, being loose with notation, the likelihood for a particular school takes the form

$$\mathcal{L}_s=\int\left[\prod_{\{i:S(i)=s\}}f(X_i'\beta,X_i'\gamma\mid P_i=1,\mu_s^1,\mu_s^2)\right]g_s\left(\mu_s^1,\mu_s^2\mid W_s\right)d\mu_s^1d\mu_s^2$$

where g_s is the probability density given the sampling scheme.

Then using Bayes theorem

$$\begin{aligned}f(X_i'\beta,X_i'\gamma\mid P_i=1,\mu_s^1,\mu_s^2)&=\frac{\Pr(P_i=1\mid X_i'\beta,X_i'\gamma,\mu_s^1,\mu_s^2)f(X_i'\beta,X_i'\gamma\mid\mu_s^1,\mu_s^2)}{\Pr(P_i=1\mid\mu_s^1,\mu_s^2)}\\&=\frac{\Phi\left(X_i'\beta+\varphi\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)\right)\phi\left(X_i'\beta-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)}{\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+var(\eta_i^1)}}\right)}\end{aligned}$$

and

$$\begin{aligned} g_s\left(\mu_s^1, \mu_s^2 \mid W_s\right) &= g\left(\mu_s^1, \mu_s^2 \mid W_s, P_i = 1\right) \\ &= \frac{Pr\left(P_i = 1 \mid W_s, \mu_s^1, \mu_s^2\right) g\left(\mu_s^1, \mu_s^2 \mid W_s\right)}{Pr\left(P_i = 1 \mid W_s\right)} \\ &= \frac{\Phi\left(\frac{\mu_s^1 + \varphi\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)}{\sqrt{1+var(\psi_i^1)}}\right) \phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)}{\int \Phi\left(\frac{\mu_s^1 + \varphi\overline{\mu}^2(0, \mu_s^1, \mu_s^2)}{\sqrt{1+var(\eta_i^1)}}\right) d\Phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)}. \end{aligned}$$

In reality, NELS:88 follows 8th grade sample members into high schools, and so the number of students sampled from high school s will depend on $\Phi\left(\frac{\mu_s^1 + \varphi\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)}{\sqrt{1+var(\psi_i^1)}}\right)$. Furthermore, up to a sample size of 11, the probability that s is included at all is increasing in the number of NELS:88 8th graders who start at the high school. Consequently, the probability that a particular student is followed depends on the $\Phi\left(\frac{\mu_s^1 + \varphi\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)}{\sqrt{1+var(\psi_i^1)}}\right)$ of other students. We do not address this. However, we use the average of the sample weights for the students who attend s to weight the value of $\mathcal{L}_s$.

Step 4: Taking all parameters as given, solve $\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)$ as a fixed point for the equation

$$\begin{aligned} \overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right) =& \mu_s^2 + E\left(\eta_i^2 \mid \mu_s^1 + \eta_i^1 + \varphi\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right) + \varepsilon_i > 0, \mu_s^1, \mu_s^2\right) \\ =& \mu_s^2 + \frac{cov(\eta_i^1, \eta_i^2)}{\sqrt{1 + var(\eta_i^1)}} \lambda\left(\frac{\mu_s^1 + \varphi\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)}{\sqrt{1 + var(\eta_i^1)}}\right) \end{aligned}$$

We iterate on this four-step procedure until we find a fixed point estimate of β. Each iteration is time consuming, in part because we must compute $\overline{\mu}^2\left(0, \mu_s^1, \mu_s^2\right)$ for each value of W_s in the data. However, the parameter estimates converge fairly quickly.

We then use the model estimates to simulate the effects of the voucher policy. The key to this is the construction of the weights ψ_i, which are a function of the probability that $P_i = 1$ under both the current regime and the alternative (τ) regime. The crucial element in this calculation is

$$\begin{aligned} \Pr(P_i = 1 \mid W_i, X_i'\beta, Z_i'\gamma; \tau) &= \int \Phi\left(X_i'\beta - t_i(\tau) + \varphi\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right)\right) dF(\mu_s^1, \mu_s^2 \mid W_{s(i)}, X_i'\beta, X_i'\gamma) \\ &= \frac{\int \Phi\left(X_i'\beta - t(\tau) + \varphi\overline{\mu}^2\left(\tau, \mu_s^1, \mu_s^2\right)\right) \phi\left(X_i'\beta - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right) d\Phi(\mu_s^1, \mu_s^2; W_{s(i)}, \Sigma_e)}{\int \phi\left(X_i'\beta - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right) d\Phi(\mu_s^1, \mu_s^2; W_{s(i)}, \Sigma_e)}. \end{aligned}$$

We then construct

$$\psi_i(\tau) = \frac{\frac{\Pr(P_i=1|W_{s(i)}, X_i'\beta, X_i'\gamma;\tau)}{\Pr(P_i=1|W_{s(i)}, X_i'\beta, X\prime\gamma;0)}}{\int \frac{\Pr(P_i=1|W_{s(i)}, X_i'\beta, X_i'\gamma;\tau)}{\Pr(P_i=1|W_{s(i)}, X_i'\beta, X_i'\gamma;0)} dG(W_{s(i)}, X_i'\beta, X_i'\gamma \mid P_i^0 = 1)}$$

A.2.2 The $X'\beta$ Case

We use the same iterative procedure for the $X_i'\beta$ model, but the equations are simpler. In this case define

$$\overline{\mu}^1\left(\tau,\mu_s^1\right) \equiv E\left(X_i'\beta \mid P_i^\tau = 1, \mu_s^1\right).$$

Given our model and distribution assumptions,

$$\begin{aligned}\overline{\mu}^1\left(\tau,\mu_s^1\right) &= \mu_s^1 + E\left(\eta_i^1 \mid \mu_s^1 + \eta_i^1 - t(\tau) + \varphi\overline{\mu}^1\left(\tau,\mu_s^1\right) + \varepsilon_i > 0\right) \\ &= \mu_s^1 + \frac{var(\eta_i^1)}{\sqrt{1+var(\eta_i^1)}}\lambda\left(\frac{\mu_s^1 - t(\tau) + \varphi\overline{\mu}^1\left(\tau,\mu_s^1\right)}{\sqrt{1+var(\eta_i^1)}}\right)\end{aligned}$$

The analogue of likelihoods (24) and (25) in the $X'\beta$ case are simpler than in the $X'\gamma$ case. They take on the forms

$$\begin{aligned}\mathcal{L}_i =& P_i w_i \log\left(\frac{\int \Phi\left(X_i'\beta + \varphi\overline{\mu}^1\left(0,\mu_s^1\right)\right)\phi\left(X_i'\beta^* - \mu_s^1; var(\eta_i^1)\right)d\Phi(\mu_s^1; W_s, Var(e_s^1))}{\int \phi\left(X_i'\beta^* - \mu_s^1; var(\eta_i^1)\right)d\Phi(\mu_s^1; W_s, Var(e_s^1))}\right) \\ &+ (1-P_i)w_i \log\left(1 - \frac{\int \Phi\left(X_i'\beta + \varphi\overline{\mu}^1\left(0,\mu_s^1\right)\right)\phi\left(X_i'\beta^* - \mu_s^1; var(\eta_i^1)\right)d\Phi(\mu_s^1; W_s, Var(e_s^1))}{\int \phi\left(X_i'\beta^* - \mu_s^1; var(\eta_i^1)\right)d\Phi(\mu_s^1; W_s, Var(e_s^1))}\right),\end{aligned}$$

and

$$\mathcal{L}_s = \frac{\int\left[\prod_{\{i:S(i)=s\}}\left(\frac{\Phi\left(X_i'\beta+\varphi\overline{\mu}^1\left(0,\mu_s^1\right)\right)\phi\left(X_i'\beta^*-\mu_s^1;var(\eta_i^1)\right)}{\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^1\left(0,\mu_s^1\right)}{\sqrt{1+var(\eta_i^1)}}\right)}\right)\right]\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^1\left(0,\mu_s^1\right)}{\sqrt{1+var(\eta_i^1)}}\right)d\Phi(\mu_s^1; W_s, Var(e_s^1))}{\int\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^1\left(0,\mu_s^1\right)}{\sqrt{1+var(\eta_i^1)}}\right)d\Phi(\mu_s^1; W_s, Var(e_s^1)).}$$

The final piece is

$$\begin{aligned}\Pr(P_i = 1 \mid W_i, X_i'\beta; \tau) &= \int \Phi\left(X_i'\beta - t(\tau) + \varphi\overline{\mu}^1\left(\tau,\mu_s^1\right)\right)dF(\mu_s^1 \mid W_{s(i)}, X_i'\beta) \\ &= \frac{\int \Phi\left(X_i'\beta - t(\tau) + \varphi\overline{\mu}^1\left(\tau,\mu_s^1\right)\right)\phi\left(X_i'\beta^* - \mu_s^1; var(\eta_i^1)\right)d\Phi(\mu_s^1; W_{s(i)}, Var(e_s^1))}{\int \phi\left(X_i'\beta^* - \mu_s^1; var(\eta_i^1)\right)d\Phi(\mu_s^1; W_{s(i)}, Var(e_s^1))},\end{aligned}$$

and again

$$\psi_i(\tau) = \frac{\frac{\Pr(P_i=1|W_{s(i)},X_i'\beta,Z_i'\gamma;\tau)}{\Pr(P_i=1|W_{s(i)},X_i'\beta,Z_i'\gamma;0)}}{\int\frac{\Pr(P_i=1|W_{s(i)},X_i'\beta,Z_i'\gamma;\tau)}{\Pr(P_i=1|W_{s(i)},X_i'\beta,Z_i'\gamma;0)}dG(W_{s(i)}, X_i'\beta, Z_i'\gamma \mid P_i^0 = 1)}.$$

Appendix 3: Unobservable School Effects and Unobservable Peer Characteristics

In this appendix we discuss estimation of the unobservable peer effect model we describe in the text. In practice there are two complicated issues: estimation of the variance/covariance components of the model and estimation of δ. Estimation of the variance/covariance model is quite similar to the case discussed in Appendix 2, so we do not go into great detail. However, estimation of δ is different than in the other models. The difference results from the need to infer the expected value of the unobservables from the observables. Basically, we do this using Bayes theorem, although the expression is quite involved. Our procedure is as follows.

1. We obtain β from a standard probit model for public school choice.

2. We estimate γ using fixed effects regression on the public school sample, but we correct for sample selection by including the inverse Mills-ratio term for public school choice in the regression.

3. Given β and taking the variance of $u^1_{s(i)}$ and e^1_i and $\sigma^2 = var\left(e^1_i + \eta^1_i\right)$ as known we use a likelihood function analogous to step 2 of Appendix 2. That is, we only use public schools and the likelihood is

$$\mathcal{L}_s = \frac{\int\int\int \left[\prod_{\{i:S(i)=s\}} \left(\frac{\Phi\left(\frac{X_i'\beta+v_s^1}{\sigma_{\omega^1}}\right)\phi\left(X_i'\beta-\mu_s^1, X_i'\gamma-\mu_s^2, \Sigma_\eta\right)}{\Phi\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma^2_{\eta^1}+\sigma^2_{\omega^1}}}\right)}\right)\right] \Phi\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma^2_{\eta^1}+\sigma^2_{\omega^1}}}\right) d\Phi\left(v_s^2\right) d\Phi(\mu_s^1, \mu_s^2; \Sigma_\mu)}{\int\int\int \Phi\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma^2_{\eta^1}+\sigma^2_{\omega^1}}}\right) d\Phi\left(v_s^2\right) d\Phi(\mu_s^1, \mu_s^2; \Sigma_\mu)}$$

4. Given Σ_μ and Σ_η, we calculate $a, \Sigma_v, Var(\omega_i^1)$, and $Cov(\omega_i^1, \omega_i^2)$ using the restrictions of the model.

5. Next we estimate $\delta_{X'\beta}$ and $\delta_{X'\gamma}$

   Let X_{-i} the matrix of values of X_j' for a random sample of individuals j who actually attend school $s(i)$, $j \neq i$. We will let $f_{\mu v^1}$ denote the density of $(\mu_s^1, \mu_s^2, v_s^1)$ conditional on $X_i^{\beta\gamma} \equiv (X_{-i}\beta, X_{-i}\gamma, P_i^0 = 1, X_i'\beta, X_i'\gamma)$ and let $f_{x^{\beta\gamma}}$ denote the likelihood of $X_i^{\beta\gamma}$ conditional on (μ_s^1, u_s^2, v_s^1). Then using Bayes theorem

$$f_{\mu v^1}(\mu^1_{s(i)}, \mu^2_{s(i)}, v^1_{s(i)} \mid X_i^{\beta\gamma}) = \frac{f_{x^{\beta\gamma}}(X_i^{\beta\gamma} \mid \mu^1_{s(i)}, v^1_{s(i)})\phi(\mu^1_{s(i)}, \mu^1_{s(i)}; \Sigma_\mu)\phi\left(v^1_{s(i)}; \sigma^2_{v^1}\right)}{\int\int\int f_{x^{\beta\gamma}}(X_i^{\beta\gamma} \mid \mu^1_{s(i)}, v^1_{s(i)})\phi(\mu^1_{s(i)}, \mu^2_{s(i)}; \Sigma_\mu)\phi\left(v^1_{s(i)}; \sigma^2_{v^1}\right) d\mu^1_{s(i)} d\mu^2_{s(i)} dv^1_{s(i)}},$$

   where the definition of $X_{-i}\beta$ and $X_{-i}\gamma$ implies that

$$f_{x^{\beta\gamma}}(X_i^{\beta\gamma} \mid \mu_s^1, \mu_s^2, v_s^1) = \frac{\prod_{j\in S_i} \Phi\left(\frac{X_j'\beta+v^1_{s(i)}}{\sigma_{\omega^1}}\right) \prod_{j\in S_i} \phi\left(X_j'\beta - \mu^1_{s(i)}, X_j'\gamma - \mu^2_{s(i)}, \Sigma_\eta\right)}{\prod_{\substack{j\in S_i \\ j\neq i}} \Phi\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma^2_\eta+\sigma^2_\omega}}\right)}.$$

We use $f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^\beta)$ to denote the conditional density of (μ_s^1, v_s^1) where $X_i^\beta \equiv (X_{-i}\beta, P_i^0 = 1, X_i'\beta)$.[30]

6. Using this density we know that

$$
\begin{aligned}
&E\left(X_i'\beta + gu_i \mid X_i^\beta\right) \\
&= \int\int E\left(\mu_s^1 + \eta_i^1 + g\left[v_s^1 + \omega_i^1\right] \mid \mu_s^1, v_s^1\right) f_{\mu^1 v^1}(\mu_s^1, , v_s^1 \mid X_i^\beta) d\mu_s^1 dv_s^1 \\
&= \int\int \left(\mu_s^1 + gv_s^1 + \frac{\sigma_{\eta 11} + g\sigma_{\omega 11}}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\lambda\left(\frac{\mu_s^1 + gv_s^1}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right)\right) f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^\beta) d\mu_s^1 dv_s^1.
\end{aligned}
$$

By running a regression of $Y - X'\gamma$ on this we can get a consistent estimate of $\delta_{X'\beta}$. In practice we use the sample analogue of this expression with Gauss-Quadrature to approximate the integrals.

In the case of the $X'\gamma$ index restriction (21), we use the same procedure to first estimate

$$
\begin{aligned}
&E\left(X_i'\gamma + g\varepsilon_i \mid X_i^{\beta\gamma}\right) \\
&= \int E\left(\mu_s^2 + \eta_i^2 + g\left[v_s^2 + \omega_i^2\right] \mid \mu_s^1, \mu_s^2, v_s^1\right) f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 d\mu_s^2 dv_s^1 \\
&= \int \left(\mu_s^2 + \frac{g\sigma_{v12} v_s^1}{\sigma_{v_s^1}^2} + \frac{\sigma_{\eta 12} + g\sigma_{\omega 12}}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\lambda\left(\frac{\mu_s^1 + gv_s^1}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right)\right) f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 d\mu_s^2 dv_s^1.
\end{aligned}
$$

and then obtain $\delta_{X'\gamma}$ from the analogous regression.

7. Finally we calculate the treatment effect. In the text we derive $\pi^p\left(\tau; \mu_s^1, v_s^1\right)$ for each model. We need to integrate across this. In either model the average value of the cream skimming effect may be written as

$$
\begin{aligned}
&E\left[\pi^p\left(\tau; \mu_s^1, v_s^1\right) \mid P_i^\tau = 1\right] \\
&= \frac{E\left[\pi^p\left(\tau; \mu_s^1, v_s^1\right) P_i^\tau\right]}{Pr\left[P_i^\tau = 1\right]} \\
&= \frac{\int \pi^p\left(\tau; \mu_s^1, v_s^1\right) \frac{\Phi\left(\left(X_i'\beta - t_i(\tau) + v_{s(i)}^1\right)/\sigma_{\omega 11}\right)}{\Phi\left(\left(X_i'\beta + v_{s(i)}^1\right)/\sigma_{\omega 11}\right)} f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_{-i}^\beta) d\mu_s^1 dv_s^1 dG\left(X_{-i}\beta, X_i'\beta \mid P_i^0 = 1\right)}{\int \frac{\Phi\left(\left(X_i'\beta - t_i(\tau) + v_{s(i)}^1\right)/\sigma_{\omega 11}\right)}{\Phi\left(\left(X_i'\beta + v_{s(i)}^1\right)/\sigma_{\omega 11}\right)} f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^\beta) d\mu_s^1 dv_s^1 dG\left(X_{-i}\beta, X_i'\beta \mid P_i^0 = 1\right)},
\end{aligned}
$$

where we use normality in the last line. Again we approximate with sample analogues.

[30] Our assumption that in the school attended by i, data on a random set of students are available and the number of students is not related to the school choice probability. We account for selection in who choses to attend the public school. However, in practice the number of students available depends on public school choice. This is because NELS:88 follows eighth graders into high schools and does not draw a random sample in high school. Furthermore, the probability that a high school attended by NELS:88 eighth graders is included in the first followup survey is increasing in the number who attend up until a school sample of 11. Consequently, the probability that student i is sampled depends on the choices of the other NELS:88 sample members from the same eight grade, which depend on $\mu_{s(i)}^1$. This is not accounted for in the expression for $f_{\mu v^1}(\mu_{s(i)}^1, \mu_{s(i)}^2, v_{s(i)}^1 \mid X_{-i}^\beta, P_i = 1, X_i)$. However, we use the average of the sample weights for the students who attend s to weight the value of $\mathcal{L}_s$.

Table 1: Descriptive Statistics: Full Sample By High School Type

Variables	All Schools	Public Schools	All Private	Catholic Private	Other Private
Demographics					
male	0.50	0.50	0.53	0.54	0.50
asian	0.04	0.03	0.06	0.05	0.07
hispanic	0.10	0.10	0.08	0.10	0.05
black	0.13	0.13	0.08	0.11	0.02
Geographic Variables and Zipcode Characteristics					
Northeast	0.20	0.18	0.31	0.32	0.29
North Central	0.26	0.27	0.22	0.28	0.13
South	0.35	0.35	0.30	0.25	0.39
Urban	0.25	0.23	0.45	0.48	0.40
Suburban	0.44	0.43	0.47	0.49	0.45
Distance from Cath HS	0.32	0.34	0.11	0.05	0.21
(100s of kilo+A84meters)	(0.49)	(0.50)	(0.26)	(0.07)	(0.39)
Fraction black	0.11	0.11	0.13	0.13	0.11
	(0.20)	(0.20)	(0.20)	(0.21)	(0.17)
Fraction Hispanic	0.08	0.08	0.08	0.09	0.06
	(0.17)	(0.17)	(0.14)	(0.16)	(0.09)
Median income	3.12	3.06	3.62	3.56	3.70
	(1.26)	(1.20)	(1.60)	(1.61)	(1.57)
Fraction under poverty line	0.13	0.13	0.12	0.12	0.12
	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)
Fraction over 2 times	0.67	0.66	0.71	0.71	0.70
poverty line	(0.16)	(0.16)	(0.16)	(0.16)	(0.17)
Parental Background					
Father/mother present	0.66	0.65	0.80	0.79	0.81
Father's education	13.45	13.24	15.19	14.55	16.13
	(2.88)	(2.81)	(2.90)	(2.74)	(2.89)
Mother's education	12.97	12.84	14.08	13.73	14.61
	(2.29)	(2.27)	(2.17)	(2.12)	(2.14)
Log (family inc) 1987	10.27	10.21	10.84	10.71	11.02
	(0.90)	(0.88)	(0.80)	(0.80)	(0.76)
Limited English proficiency composite	0.02	0.02	0.01	0.01	0.02
Parents Catholic	0.31	0.28	0.53	0.79	0.14
8th Test Scores and Academic Performance					
reading score	50.84	50.38	54.89	54.42	55.60
	(9.96)	(9.86)	(9.93)	(9.54)	(10.50)
math score	50.94	50.54	54.40	53.91	55.16
	(9.94)	(9.90)	(9.59)	(9.03)	(10.34)
science score	50.85	50.56	53.38	52.89	54.14
	(9.93)	(9.91)	(9.75)	(9.08)	(10.64)
history/cit/geog	50.92	50.47	54.87	54.53	55.37
	(9.90)	(9.77)	1(0.17)	(9.28)	(11.37)
delinquency index	0.63	0.65	0.53	0.52	0.53
	(1.06)	(1.08)	(0.91)	(0.93)	(0.88)

continued on next page

Table 1 (continued)

	All Schools	Public Schools	All Private	Catholic Private	Other Private
Grades composite	2.94	2.92	3.11	3.13	3.07
	(0.73)	(0.74)	(0.66)	(0.62)	(0.70)
grade trouble index	0.53	0.54	0.48	0.46	0.50
(from student, 0-4)	(0.82)	(0.83)	(0.79)	(0.83)	(0.72)
student got into fight with other student	0.25	0.25	0.22	0.22	0.22
student performs below ability	0.25	0.25	0.20	0.15	0.27
student rarely completes homework	0.19	0.20	0.13	0.09	0.18
student frequently absent	0.09	0.10	0.06	0.07	0.05
student inattentive in class	0.20	0.20	0.17	0.12	0.24
students in class seen as troublemaker	0.05	0.05	0.03	0.03	0.03
child ever had behavioral problems	0.08	0.09	0.05	0.04	0.07
Parents contacted about	1.41	1.41	1.38	1.36	1.42
school behavior 1-4	(0.75)	(0.75)	(0.70)	(0.67)	(0.73)
Repeated Grade 4-8	0.06	0.06	0.03	0.02	0.04
Risk of dropping out of	0.66	0.70	0.30	0.36	0.21
school	(0.93)	(0.95)	(0.54)	(0.59)	(0.44)
Lack of Effort index (0-21)	4.03	4.03	3.81	3.49	4.3
	(2.73)	(2.72)	(2.82)	(2.63)	(3.02)
Enrolled in gifted program	0.14	0.14	0.12	0.10	0.15
Outcomes					
High School Graduation	0.871	0.863	0.942	0.976	0.887
Attend Four Year College	0.325	0.297	0.584	0,582	0.587
Missing Value Indicators					
Family Income Missing	0.092	0.089	0.122	0.055	0.099
Tests Missing	0.032	0.033	0.022	0.023	0.02
Student Behavior Missing	0.058	0.054	0.091	0.087	0.096
Distance Missing	0.056	0.053	0.084	0.055	0.123
N	*16,483*	*14193*	*2,290*	*936*	*1354*

Notes: Means for individual variables and standard deviations (n parentheses) exclude missing cases, which were assigned the sample means when the variables are used in the school choice and outcome equations. Sample sizes refer to the school choice sample, and the number of nonmissing cases varies across variable. Sample sizes for the outcome variables are smaller. Explanatory Variables are weighted using the base year through first followup panel weights. High School graduation and Attend Four Year College are weighted using base year through third followup panel weights.

Table 2
Estimation of Peer Effects Model for Public High School Graduation (δ)
(95% Confidence Intervals in Parenthesis)

Model Specification	$X'\beta$ Index	$X'\gamma$ Index
Base Case:		
Probit	-0.0453	0.3614
	(-0.0949,0.0032)	(0.0289,0.6170)
Unobservable Peer Effects:		
Unobservables like observables (g=1)	0.0157	0.0321
	(-0.0071,0.0371)	(-0.0226,0.0983)
g=0.5	0.0169	0.0315
	(-0.0072,0.0394)	(-0.0224,0.0977)
g=2.0	0.0139	0.0328
	(-0.0067,0.0348)	(-0.0229,0.0990)
Alternative School Demand:		
Allowing for Peer Interactions ($X'\beta$ index)	-0.0412	NA
	(-0.0822,0.0003)	
Allowing for Peer Interactions ($X'\gamma$ index)	NA	0.3614
		(0.0344,0.5773)

Table 3
Effect of Voucher Program on Selected Peer Characteritics, $X'\beta$ and $X'\gamma$
Basic School Choice Model ($\varphi = 0$), No unobservable School Characteristics or Peer Characteristics
(95% Confidence Intervals in Parenthesis)

	mean pub school stayers	mean movers	mean peer stayers (before)	change in peer for stayers	change in mean for pub school
Catholic	0.2780	0.4353	0.2895	-0.0115	-0.0156
	(0.2554, 0.2990)	(0.3856, 0.4892)	(0.2668, 0.3108)	(-0.0160, -0.0072)	(-0.0210, -0.0107)
Male	0.4919	0.5068	0.4939	-0.0020	-0.0015
	(0.4812, 0.5031)	(0.4696, 0.5382)	(0.4838, 0.5042)	(-0.0058, 0.0025)	(-0.0048, 0.0026)
Hispanic	0.1056	0.0898	0.1064	-0.0008	0.0016
	(0.0828, 0.1307)	(0.0572, 0.1186)	(0.0830, 0.1317)	(-0.0032, 0.0017)	(-0.0011, 0.0049)
Black	0.1276	0.0928	0.1240	0.0035	0.0035
	(0.1057, 0.1485)	(0.0607, 0.1208)	(0.1036, 0.1439)	(0.0006, 0.0068)	(0.0002, 0.0070)
Parents Present	0.6606	0.7775	0.6718	-0.0112	-0.0116
	(0.6486, 0.6735)	(0.7488, 0.8066)	(0.6608, 0.6845)	(-0.0150, -0.0078)	(-0.0148, -0.0086)
Father's Education	13.1550	14.4130	13.2247	-0.0696	-0.1251
	(13.0189, 13.2819)	(14.1205, 14.6765)	(13.0911, 13.3413)	(-0.0948, -0.0455)	(-0.1544, -0.0955)
Mother's Education	12.7661	13.7035	12.8182	-0.0521	-0.0932
	(12.6664, 12.8566)	(13.5190, 13.8718)	(12.7140, 12.9050)	(-0.0677, -0.0345)	(-0.1098, -0.0743)
log Family Income 1987	10.1844	10.5954	10.2096	-0.0252	-0.0409
	(10.1510, 10.2180)	(10.5326, 10.6475)	(10.1785, 10.2401)	(-0.0318, -0.0179)	(-0.0469, -0.0335)
Reading Score	50.2633	53.1864	50.4609	-0.1976	-0.2907
	(49.9118, 50.6556)	(52.2368, 54.0456)	(50.1259, 50.8114)	(-0.2790, -0.1050)	(-0.3831, -0.1908)
Math Score	50.4692	53.0302	50.5914	-0.1222	-0.2547
	(50.0875, 50.8268)	(52.1919, 53.9161)	(50.2488, 50.9401)	(-0.1981, -0.0433)	(-0.3576, -0.1606)
Science Score	50.5874	52.5289	50.7145	-0.1272	-0.1931
	(50.2154, 50.9960)	(51.7007, 53.3522)	(50.3547, 51.0720)	(-0.2160, -0.0339)	(-0.2888, -0.0935)
History Score	50.3366	53.1312	50.4650	-0.1284	-0.2779
	(49.9890, 50.7166)	(52.1221, 54.0374)	(50.1281, 50.8234)	(-0.2181, -0.0375)	(-0.3760, -0.1665)
$X\beta$	1.9545	1.3358	1.9301	0.0244	0.0615
	(1.8324, 2.3478)	(1.1658, 1.5206)	(1.8070, 2.3132)	(0.0213, 0.0378)	(0.0529, 0.0991)
$X\gamma$	0.3673	0.4144	0.3710	-0.0037	-0.0047
	(0.2047, 0.5242)	(0.2498, 0.5719)	(0.2082, 0.5283)	(-0.0050, -0.0023)	(-0.0061, -0.0034)

Table 4
Estimates of Effects of Voucher on Public Schools Students
(95% Confidence Intervals in Parenthesis)

<table>
<tr><th>Model Specification</th><th>$X'\beta$ Index</th><th>$X'\gamma$ Index</th></tr>
<tr><td>1: **Base Case:**</td><td></td><td></td></tr>
<tr><td>Probit</td><td>-0.0011
(-0.0017,0.001)</td><td>-0.0013
(-0.0023,-0.0001)</td></tr>
<tr><td>2: **Peers Affect School Choice :** $\varphi \neq 0$</td><td></td><td></td></tr>
<tr><td>Peers Affect School Choice ($X'\beta$ index)</td><td>-0.0012
(-0.0023,-0.0006)</td><td>NA</td></tr>
<tr><td>Peers Affect School Choice ($X'\gamma$ index)</td><td>NA</td><td>-0.0014
(-0.0021,-0.0001)</td></tr>
<tr><td>3: **Unobservable Peer Effects:**</td><td></td><td></td></tr>
<tr><td>Unobservables like observables (g=1)</td><td>0.0006
(-0.0002,0.0016)</td><td>-0.0002
(-0.0006,0.0001)</td></tr>
<tr><td>g=0.5</td><td>0.0005
(-0.0002,0.0012)</td><td>-0.0001
(-0.0004,0.0001)</td></tr>
<tr><td>g=1.5</td><td>0.0007
(-0.0004,0.0019)</td><td>-0.0003
(-0.0007,0.0002)</td></tr>
<tr><td>4: **Alternative School Demand:**</td><td></td><td></td></tr>
<tr><td>Catholic School is only Alternative</td><td>-0.0010
(-0.0031,0.0005)</td><td>-0.0013
(-0.0022,-0.0001)</td></tr>
<tr><td>Targeted toward Low Income Families</td><td>-0.0010
(-0.0023,-0.0001)</td><td>-0.0013
(-0.0022,-0.0002)</td></tr>
<tr><td>Targeted toward Urban Districts</td><td>0.0071
(-0.0053,0.0634)</td><td>0.0026
(-0.0012,0.0070)</td></tr>
<tr><td>Targeted toward Low Income Neighborhoods</td><td>-0.0018
(-0.0091,0.0058)</td><td>-0.0013
(-0.0037,0.0038)</td></tr>
<tr><td>Nested Logit</td><td>-0.0011
(-0.0030,0.0014)</td><td>-0.0014
(-0.0023,-0.0002)</td></tr>
<tr><td>Callibrated to Milwaukee</td><td>0.0021
(-0.0016,0.0060)</td><td>0.0019
(-0.0011,0.0051)</td></tr>
<tr><td>Selection on $X'\gamma$ only</td><td>NA</td><td>-0.0059
(-0.0098,-0.0003)</td></tr>
<tr><td>5: **Alt. Assumptions. About Peer Effects on Outcomes:**</td><td></td><td></td></tr>
<tr><td>Peers same as direct effect ($\delta = 1$)</td><td>NA</td><td>-0.0037
(-0.0050,-0.0023)</td></tr>
<tr><td>Peers half of direct effect ($\delta = 0.5$)</td><td>NA</td><td>-0.0018
(-0.0025, -0.0011)</td></tr>
<tr><td>Peer effects only operate through Family Income</td><td colspan="2">-0.0011
(-0.0025,0.0006)</td></tr>
<tr><td>Peer effects only operate through African American</td><td colspan="2">-0.0002
(-0.0005,0.0001)</td></tr>
<tr><td>Peer effects only operate through test score index</td><td>-0.0005
(-0.0019,0.0005)</td><td>-0.0008
(-0.0015,0.0001)</td></tr>
<tr><td>Peer effects only operate through Father's Education</td><td colspan="2">-0.0008
(-0.0020,-0.0004)</td></tr>
<tr><td>6: **Alternative Peer Effects and demand systems:**</td><td></td><td></td></tr>
<tr><td>$\delta = 1$ and choice only on $X'\gamma$</td><td>NA</td><td>-0.0163
(-0.181,-0.0146)</td></tr>
<tr><td>$\delta = 1$ and Milwaukee Callibration</td><td>NA</td><td>0.0044
(0.0017,0.0074)</td></tr>
</table>

Appendix Table A1

Probit Model for Public School Attendance, Full Sample

	Probit Coef	95% Confidence Interval L Bound	95% Confidence Interval U Bound	Marginal Effect on Pr(Public=1)
	(1)	(2)	(3)	(4)
Constant	7.200	5.544	8.644	
Male	-0.094	-0.218	0.047	-0.017
Hispanic	0.222	-0.066	0.621	0.039
Black	0.093	-0.193	0.476	0.016
Parental Background				
Catholic	-0.517	-0.717	-0.344	-0.091
Both parents present	-0.217	-0.357	-0.076	-0.038
Father's education	-0.051	-0.086	-0.017	-0.009
Mother's education	-0.048	-0.075	-0.018	-0.008
log income 1987	-0.288	-0.393	-0.169	-0.051
Limited English proficiency (0,1)	-0.275	-0.893	0.758	-0.049
8th Grade Tests and Grades				
reading score	-0.016	-0.026	-0.006	-0.003
math score	0.002	-0.009	0.011	0.000
science score	0.009	0.000	0.020	0.002
history score	-0.009	-0.019	0.002	-0.002
Grades Composite (0-4)	0.034	-0.103	0.194	0.006
8th Grade Behavior and Performance in School Measures				
Delinquency Index	0.014	-0.071	0.108	0.003
Student got into a fight	0.019	-0.123	0.152	0.003
Student performs below ability	-0.154	-0.402	0.117	-0.027
student rarely completes homework	0.335	0.035	0.685	0.059
Student frequently absent	0.127	-0.238	0.532	0.022
Student inattentive in class	-0.149	-0.409	0.133	-0.026
Student frequently disruptive	-0.217	-0.491	0.143	-0.038
Parent believes child has a behavioral problem in school	0.180	-0.088	0.513	0.032
Repeated a grade 4-8 (0,1)	0.173	-0.143	0.565	0.031
Dropout Risk Composite (0-6)	0.037	-0.047	0.153	0.007
Lack of Effort Index	-0.024	-0.060	0.009	-0.004
Enrolled in Gifted Program	0.381	0.031	0.818	0.067
Location Measures				
North East	-0.123	-0.478	0.253	-0.022
North Central	0.103	-0.287	0.444	0.018
south	-0.165	-0.525	0.142	-0.029
urban	-0.721	-1.052	-0.419	-0.127
suburban	-0.278	-0.542	-0.018	-0.049
Distance	0.893	0.288	2.556	0.158
Distance Squared	-0.173	-0.691	-0.058	-0.031

Notes. The sample size is 16483. Column 1 reports MLE probit coefficient estimates. Columns 2 and 3 report bootstrap estimates of the lower and upper bound of the 95% confidence interval. The estimates account for correlation across students who attended the same 8th grades and/or high schools. They are based on 500 bootstrap replications. The fourth column reports marginal effects on the probability of attending public high school when X'Beta is 1.27588, which corresponds the the value at which the probability of attending public high school equals the weighted mean (.899). The model also contains missing indicators for Distance, log family income in 1987 and an indicator that is one if all test scores are missing. There is one indicator for missing data on Student performs below ability, Student rarely completes homework, Student frequently absent, student inattentive in class, and/or Student frequently disruptive. NELS:88 base-year to third year follow-up panel weights are used.

Table A2)

Effects of Students Own Characteristics on Public High School Graduation (Y)

Linear Probability Models with HS Fixed Effects

		95% Conf. interval	
	Regression Coef (1)	Lower Bound (2)	Upper Bound (3)
Male	0.0207	0.0066	0.0367
Hispanic	0.0079	-0.0243	0.0428
Black	0.0646	0.0258	0.096
Parental Background			
Catholic	0.0298	0.0131	0.0454
Both parents present	0.012	-0.0086	0.0325
Father's education	0.0044	0.001	0.0078
Mother's education	0.0008	-0.0028	0.0046
log income 1987	0.0193	0.0036	0.0353
limited English proficiency (0,1)	0.0829	0.0291	0.1352
8th Test Scores and Grades			
reading score	0.0003	-0.0008	0.0014
math score	0.0009	-0.0001	0.002
science score	0.0004	-0.0007	0.0015
history/civics/geog. score	0.0003	-0.0007	0.0015
Grades Composite (8th grade, 0-4)	0.036	0.0212	0.0513
Eighth Grade Behavioral and Performance In School Measures			
delinquency index	-0.0175	-0.029	-0.0059
student got into a fight	-0.0181	-0.0402	0.0033
student performs below ability	-0.0259	-0.0618	0.0122
student rarely completes homework	-0.0912	-0.1445	-0.0412
student frequently absent	-0.1666	-0.2154	-0.1186
student inattentive in class	-0.0319	-0.0723	0.0111
student frequently disruptive	-0.0211	-0.0664	0.0261
Parent believes child has a behavioral problem in school	-0.0429	-0.0793	-0.0035
repeated a grade 4-8	-0.1477	-0.1895	-0.1041
Dropout risk composite	-0.0241	-0.0371	-0.0113
Lack of Effort index	-0.0032	-0.0073	0.0009
Enrolled in Gifted Program	0.0131	-0.005	0.0304

Note: Column 1 of the table reports weighted least squares estimates from a regression of high school graduation with high school fixed effects included. Columns 2 and 3 report the lower and upper bounds of the 95 percent confidence interval. They are calculated from 1000 bootstrap replications. The model also includes three missing data indicators---see the note to Table A1. The sample is restricted to public high school students, and the sample size used in the calculation is 9260. Schools with only one sampled student are dropped. NELS:88 base-year to 3rd follow-up panel weights are used.