Estimating the Cream Skimming Effect of School Choice

Joseph G. Altonji

Yale University and National Bureau of Economic Research

Ching-I Huang

National Taiwan University

Christopher R. Taber

University of Wisconsin and National Bureau of Economic Research

We derive a formula to determine the degree to which a school choice program may harm public school stayers by luring the best students to other schools. The "cream skimming" effect is increasing in the degree of heterogeneity within schools, the school choice take-up rate of strong students relative to weak students, and the dependence of school outcomes on student body quality. We use the formula to investigate the cream skimming effect of hypothetical voucher programs on the high school graduation rate and other outcomes of the students who would remain in public school. We find small effects across a wide variety of model specifications and types of modest voucher programs.

I. Introduction

Dissatisfaction with the performance of the US educational system, particularly in minority urban school districts, has led to experimentation

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, Adam Kapor, Charles Manski, Derek

[*Journal of Political Economy,* 2015, vol. 123, no. 2]

with a variety of school choice programs. A full evaluation of any school choice program involves estimating its effect on all students. There are potentially four groups of students who are affected by a program: those who take advantage of the program to move to a private school, charter school, or other alternative school; those who remain in public school; those who remain in the alternative schools; and possibly private or charter school students who move to public school as a result of the program. Furthermore, to assess the overall effect of any school choice program on educational outcomes of these four groups, one must address (at least) four questions. First, by how much do children who exercise their option to switch schools benefit? Second, does increased competition lead existing schools to become more efficient? Third, to what extent do changes in financial resources associated with the program affect outcomes? Fourth, 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 four issues are central to assessing the major types of choice programs: vouchers, charter schools, and choice within the public school system. The relative importance of them will depend on the structure of the choice program. Thus in designing school choice, one needs to separately consider all four issues.

There is a large literature on the direct benefits of private schools and charter schools, a substantial literature on the response to competition, and a small literature on school finance effects.[1] However, there is very little research on isolating and understanding the cream skimming effect on public school stayers. This is the focus of our paper. We are not assuming that competition and financial aspects are unimportant for

Neal, Christopher Neilson, Jesse Rothstein, Robert Shimer, Seth Zimmerman, and three anonymous referees for helpful discussions and comments. Earlier drafts were circulated under the title "Estimating the Cream Skimming Effect of Private School Vouchers on Public School Students." Mistakes are our responsibility.

[1] Much of the US 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) suggest that students who attend Catholic schools perform substantially better than they would have 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 test scores (see Neal [2002] for an overview). Rouse and Barrow (2009) survey the evidence from a number of voucher experimental or quasi-experimental studies on the achievement effects of private school attendance and charter school attendance (see, e.g., Krueger and Zhu 2004). In both instances they conclude that the achievement effects are small. Rouse and Barrow also summarize the substantial literature on whether competition from private schools, charter schools, and/or other public schools improves public school performance. The study by Card, Dooley, and Payne (2010) is a good recent example. See also Hoxby (2003). The evidence is mixed. Given limited experience with large-scale voucher programs, the evidence concerning effects on per capita spending on public school students is limited. Ferreyra (2007) and several of the papers that use calibrated simulation models consider effects of voucher programs on public financing of schools.

public school stayers; indeed other research suggests that they may be. Rather, our goal is to develop and apply tools for estimating the cream skimming effect because it is potentially a key aspect of a voucher program and other choice programs.

What do we mean by the "cream skimming effect"? We use the term to include the full set of channels through which the composition of students in a school affects how a particular student performs. The most obvious mechanism is peer effects that occur at the classroom level or at the school level. Other mechanisms are the influence that average family background and predetermined student body characteristics have on the composition and effort level of the teachers, the performance of the principal, parental participation in school, and even how the school fares in obtaining financial resources. Some of these effects would take a few years to play out following a change in student body characteristics while others (e.g., classroom peer effects) would change quickly. We use the term "student body effects" to emphasize that we are considering more than just peer interactions.

Quantifying the cream skimming effect of a choice program is difficult. As we explain below, it is hard to conceive of a controlled experiment, let alone a natural experiment, that could directly identify the cream skimming effect of a particular school choice program without a model of school choice and school outcomes. The analysis must address individual heterogeneity, both observed and unobserved, because without heterogeneity there can be no cream skimming. But the cream skimming effect is driven by interactions among group members rather than by the behavior of an individual, which is why several papers study the effects of vouchers using general equilibrium models. At a minimum, one needs an econometric framework with which to aggregate the effects of choice programs on school choice and school outcomes that accounts for heterogeneity.

James Heckman has spent most of his professional career dedicated to econometric methods that are well identified but also allow for external validity and counterfactual policy analysis (see, e.g., Heckman 2001). Our approach is inspired by these general goals. Starting from knowledge of who currently attends public school, one must assess who will move in response to the program, estimate parameters relating student body characteristics to student outcomes, and then use these estimates to simulate the effect of the change in characteristics on those who remain in public school. We do not estimate a full structural model with all school preference parameters and school production function parameters specified. However, we put enough structure on the problem that we are able to estimate the policy counterfactuals. We view our approach as similar in spirit to Heckman and Vytlacil's (2001, 2005, 2007b) approach for estimating "policy-relevant treatment effects."

Since voucher and other types of school choice programs are very complicated objects with many different components, a key to success is finding a simplifying structure. Perhaps the most important contribution of this paper is a simple formula for the cream skimming effect. The formula shows that for a broad class of models of school choice and school composition effects on education outcomes, the cream skimming effect is determined by the covariance between a school choice term and a school composition term.[2] The school choice term is the relative probability that a student will move to an alternative school in response to the choice program. The school composition term is a weighted average of the differences between the student's characteristics and the average characteristics of his or her classmates, where characteristics with larger student body effects receive more weight. 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 student body 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 movers will be the same as the stayers, leaving the composition of a school unchanged. The second is the absence of a relationship between the probability that a student leaves and student characteristics that affect school performance, because then the voucher will not change the average values of relevant student body characteristics. Finally, the cream skimming effect will be zero if student body effects on outcomes are zero.

The cream skimming formula can be used to consider many types of choice programs under different assumptions about how selection into the program operates.[3] We focus on voucher programs for private schools because our data are much more informative about selection into private schools. 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 four stages. The first stage is to estimate the preference parameters determining who attends a public school. The second is to estimate the effects of student body characteristics on outcomes. The third stage simulates the school choice response to a specific hypothetical voucher program. The final stage is to apply our formula and compute the cream skimming effect.

[2] In our model we take the default public school as given and fixed. While we never explicitly estimate a model that allows for migration, in Sec. VII.A.3 we show that the cream skimming formula can be easily extended to incorporate it.

[3] Our use of the cream skimming formula to simplify the problem has parallels in the "sufficient statistic" approach to the welfare analysis. See Chetty (2009) for a survey. Our research strategy is also related to what Heckman and Vytlacil (2007a) call Marschak's maxim (Marschak 1953).

In the first stage of our base case model, we estimate family preferences for public school choice given the status quo of no voucher program. In our base estimation this is simply a probit for attending a public school. We also estimate several alternative specifications of the demand for public schools, including a formal utility specification that models the extent to which price effects vary with family income, 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 demand 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 student body quality influences school choice. The main difficulty stems from the fact that we do not have information on the potential public high school classmates of those who choose private schools. To handle this case, we develop a new methodology for estimating binary consumer choice models (1) when demand depends on the characteristics of the other buyers and (2) when researchers have information on the characteristics of the other buyers only if they chose the product. Our methodology could be applied in other situations in which consumer demand depends on the characteristics of other consumers, consumer characteristics are correlated with location, and sampling is choice based because of the costs of a random sample. Our approach builds on a long line of research involving choice-based availability of endogenous regressors pioneered in Heckman's (1974) study of the effects of wages on labor supply.

The second stage is to estimate the effects of student body variables on high school graduation. We start with the standard procedure of regressing estimates of high school fixed effects for the outcome on observed student body characteristics and address a number of econometric issues, including measurement error. We extend the approach to account for the effects of the voucher on unobserved student body characteristics that affect outcomes. For example, the usual school-level parental background measures such as average family income and average parental education are not the only parental attributes that may influence school quality. Indeed, much of the debate about cream skimming relates to unobservables such as the priority that parents place on education. 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 school quality through peer effects and other channels. However, we show below that student characteristics influence the cream skimming effect *only to the extent that they are related to school choice*. This fact means that in evaluating the cream skimming effect, we can restrict our attention to functions of the indices of observed and unobserved variables that determine school choice.

In the third stage we model various hypothetical voucher programs. Knowledge of the preferences and the program allows us to compute the relative probability that public school students will remain in public school given a specified level of the voucher. This permits us to obtain the distribution of observed and unobserved characteristics of students who will remain in public school. We do this by using the relative probabilities of continued public school attendance to reweight the distribution for public school students under the status quo. By comparing the 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. An advantage of our approach is that in the base case we do not need variation in tuition or voucher levels to estimate which students are likely to respond to the voucher program.

In the final stage, we weight our estimates of the shift in the observed and the unobserved characteristics of the peers of students who remain in public school with the estimate of how student body composition affects high school graduation. The weighted sum is the effect of the voucher program on the high school graduation rate of the public school stayers.

For our basic model, we find that the cream skimming effect of a universal voucher that would induce 10 percent of public school students to move to private schools would reduce the high school graduation rate of the students who remain by 0.0014 (i.e., about 0.1 of a percentage point). Since nine students remain for everyone who moves, this is large enough to offset about 20 percent of the direct benefit to the student who moves, which we take as 0.06. Nevertheless, the cream skimming effect on the students who remain is small. We also find relatively small cream skimming effects using college attendance and log earnings as the outcomes measures (−0.0007 and 0.002, respectively). We obtain similar results for a variety of different specifications of the school choice and school outcome models. We also investigate voucher programs that are targeted to low-income families, low-income schools, and urban areas and consider the voucher program in place in Milwaukee in 1991.[4] Overall, our results are robust and indicate that voucher programs of the types we consider 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 contexts such as existing charter school programs given the

[4] We analyze the 1991 program because the data are publicly available.

available evidence on selection into charter schools based on observables. However, one must be cautious here because the pattern of cream skimming associated with the entrance of charter schools with designs that attract highly motivated students and parents, such as the Knowledge Is Power Program schools, might not be well predicted by the private school choice model that we estimate. Furthermore, our results should be interpreted as the effects one would see from relatively moderate-sized voucher programs like the ones in existence today. We cannot draw conclusions about a huge voucher program in which the whole public school system is completely redesigned so that essentially everyone receives a voucher or about programs that lead to large changes in residential location and in the political economy of school finance.

Although our results are very robust to reasonable changes in the school demand and school outcome specifications, it is natural to ask whether a quasi-experimental approach could be used to provide a more direct assessment of the cream skimming effect. The answer is no, even if one sets aside the difficulty of finding exogenous variation in the availability of choice and the difficulty of generalizing from a particular school choice program.[5] The more fundamental problem is that one cannot identify the cream skimming effect on stayers even with a seemingly ideal experiment. To see why, consider an experiment in which vouchers are randomly introduced into some school districts but not others. In such an experiment, a researcher could identify only three types of students: individuals in the control districts, individuals in the treatment districts who stay in public school, and individuals in the treatment districts who enroll in voucher schools. The problem is that one cannot observe which control students would have remained in public school and which ones would have moved.[6] Consequently, one could never hope to estimate the effect of the program on public school stayers. Even if the researcher could somehow figure out the counterfactual of which school a control would have enrolled in, he could not separate the competition effect from the cream skimming effect. By comparing all the treatments with all the controls, the researcher could estimate the full effect of the particular voucher program. That is, he could estimate the combined effect

[5] See Heckman (1997, 2001) and Heckman and Vytlacil (2007a) for a more complete discussion of this issue.

[6] This is related to a point made in Heckman, Smith, and Taber (1998). The individuals who do not receive the treatment are still "partially treated" in the sense that they are affected by the change in student body composition. Consequently, one cannot use the Bloom (1984) estimator. However, experimental and quasi-experimental variation would be very helpful in estimating the school choice and school outcome parameters that enter the cream skimming effect formula. Such variation is used in some of the literature cited in the next paragraph. Experimental and quasi-experimental designs have also played a very important role in research on the effectiveness of private schools and charter schools (see fn. 1).

of all four of the mechanisms we discuss above. The overall effect is certainly of great interest. However, it is also of interest to policy makers designing future programs to understand the contribution of each of the mechanisms rather than just the overall effect of one particular program.

We do not know of another study that is directly comparable to ours. Several papers have used micro data sets such as NELS:88 to estimate models of public/private school choice. The papers find that family income, parental education, and student ability boost private school attendance. Our basic school choice model draws on this literature.[7] The literature on voucher programs in the United States 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 that the degree of positive selection in such programs is relatively small. Figlio, Hart, and Metzger (2010) study a private school tuition scholarship program in Florida open to families with an annual income below 175 percent of the poverty line. They find that selection is negative rather than positive. There are no universal voucher programs to study in the United States, but Hsieh and Urquiola (2006) find that Chile's universal voucher program induced higher-income and higher-ability children to move to private schools. Ladd (2002, 22) concludes that selection in New Zealand's choice program worked in a direction similar to that of 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 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 (see RPP International 2001; Bulkley and Fisler 2002; Lacireno-Paquet, Holyoke, and Moser 2002; Zimmer et al. 2009).

Our work is also related to a few papers on the general equilibrium effects of voucher programs with student body effects, including Manski (1992), Epple and Romano (1998, 2002, 2003), Caucutt (2002), Epple, Newlon, and Romano (2002), and Epple et al. (2004). 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 students who remain

[7] See, e.g., Figlio and Stone (2001) and Epple, Figlio, and Romano (2004) and the references they provide.

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). The author 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 school quality production function on the basis of location and schooling decisions. This is an important limitation because student body 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 parents care about outcomes and think that student body effects are important even though they may not be, and (*c*) the possibility that parents care about student body quality per se. In contrast, we are the first to directly estimate the effects of student body characteristics on high school graduation and then simulate the effects of cream skimming on this outcome for public school students.[8] Our analysis complements the general equilibrium papers.

The paper continues in Section II, where we present our school choice model, define the cream skimming parameter, and derive the key equation that determines the effect. Section III discusses the NELS:88 data and provides descriptive statistics. In Section IV, we discuss estimation and present results for our baseline model. In Section V, we allow student body quality to affect school choice, and in Section VI, we allow for unobserved school characteristics that influence choice and for unobserved student body characteristics. Section VII examines the sensitivity of our results to some additional alternative assumptions about student body effects and about school choice. Section VIII considers targeted voucher programs and the Milwaukee program. In Section IX, we consider cream skimming under the most extreme assumption about school choice and an extreme assumption about the size of student body effects. We present conclusions in Section X.

II. A Model of School Choice and Outcomes and a Formula for the Cream Skimming Effect

In this section, we present a model of school choice and a model of how classmates affect school outcomes. We then use these models to define

[8] Ferreyra (2009) does use a school quality measure—test score information—in her study of the effects of school finance reforms. Her study uses district-level data for the Detroit metropolitan area. Her model is similar to those in Nechyba (1999) and Ferreyra (2007) but focuses on school district choice rather than the choice between public and private schools. Epple, Romano, and Sieg (2006) examine sorting across universities of different quality levels.

our parameter of interest—the cream skimming effect—and to analyze the factors that determine it.

A. *School Choice*

Each student i is assigned to a particular school district. Denote the public school in that district as S_i.[9] School choice programs are indexed by τ. We do not observe voucher programs in our data. We denote this status quo as $\tau = 0$. We use a binary variable P_i^τ to denote whether the student i would choose his default public school under program τ. In general, the choice P_i^τ depends on τ both through the direct effect of the program on the student and through the effect of the program on the choices of other students. Most of our empirical results are for a universal voucher program that provides the same voucher amount to all students, but we consider other cases as well. Since in our data $\tau = 0$, we denote the individual's observed choice as P_i^0. When we use the expression $P_i^\tau = 1$ as a conditioning argument, we are conditioning on i who would choose public school when the school choice program is τ. Since alternative τ is counterfactual, P_i^τ is not observable in the data $\tau \neq 0$.

By assuming that each student is assigned to a specific public school, we have assumed away migration effects. These are prominent in several of the general equilibrium analyses of voucher programs discussed in the introduction. In Section VII.A.3, we show that our basic formula continues to apply with migration. We also argue informally that the direct effect of migration on cream skimming as well as indirect effects on cream skimming that are induced by changes in support for public schools are unlikely to change our conclusion that the cream skimming effect is small, at least for programs that are too small to lead to large changes in residence patterns.

B. *School Outcomes, School Quality, and Student Body 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. The outcome $Y_i(\tau)$ is determined by

$$Y_i(\tau) = X_i'\gamma + \theta(S_i,\ \tau) + \varepsilon_i, \tag{1}$$

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

[9] Note that we do not mean that students are randomly assigned to schools. Indeed, substantial heterogeneity in students across school districts is a key feature of our empirical results.

influence the outcome and possibly school choice, 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 student body effects that change as different students attend public school. Note that (1) rules out interactions between X_i and θ, which could be easily added. We have not done so as we do not expect them to change our basic findings.[10]

For any public school s and any program τ, student body effects are a function of the vector

$$\bar{Z}(s, \tau) \equiv E(Z_i | S_i = s, P_i^\tau = 1),$$

where Z_i is the vector of observed and unobserved student characteristics determined prior to high school that influence other students and $\bar{Z}(s, \tau)$ is the average of Z_i for students who choose to attend s 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 of underlying variables, we assume that the school effect can be expressed as

$$\theta(s, \tau) = \bar{Z}(s, \tau)'\delta + Q_s'\Theta_Q + \xi_s, \tag{2}$$

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. They include characteristics of the building and qualities of the principal and the teachers in a district that do not respond to changes in $\bar{Z}(s, \tau)$. As we noted earlier, competition effects and public finance effects induced by the voucher might lead to additional changes in these variables, but this is not the focus of this study.[11]

C. *The Cream Skimming Effect* $\pi^p(\tau)$

For individual i we define the cream skimming effect of voucher program τ conditional on staying in public school as

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

[10] Interactions between $\theta(S_i, \tau)$ and ε_i would be difficult to identify.

[11] Again, peer effects are only a part of δ. Also, we are sidestepping the reflection problem discussed by Manski (1993). It does not matter for our simulations whether peer effects operate through covariates or outcomes. The part of δ in (2) that captures peer effects is the vector of reduced-form coefficients of a model with reflection. In the case of both college enrollment and earnings, what we term student body effects may also capture a signaling-type mechanism in which colleges statistically discriminate in admissions and firms statistically discriminate in employment and pay on the basis of student body quality (see MacLeod and Urquiola 2013). The effects of statistical discrimination on post–high school options would also affect the return to graduating from high school, but not the other outcome we consider—test scores. We cannot isolate this part of δ.

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

$$\begin{aligned}\pi^p(\tau) &\equiv E(\pi_i(\tau)|P_i^\tau = 1,\ P_i^0 = 1)\\ &= [E(\overline{Z}(S_i,\ \tau)|P_i^\tau = P_i^0 = 1) - E(\overline{Z}(S_i,\ 0)|P_i^\tau = P_i^0 = 1)]'\delta.\end{aligned} \tag{3}$$

Thus we need to identify δ and the difference between $E(\overline{Z}(S_i, \tau)|P_i^\tau = P_i^0 = 1)$ and $E(\overline{Z}(S_i,\ 0)|P_i^\tau = P_i^0 = 1)$ to identify the cream skimming effect.

We now derive the formula for $\pi^p(\tau)$ that underlies our empirical investigation. The term $E(\overline{Z}(S_i,\ 0)|P_i^\tau = P_i^0 = 1)$ involves P_i^τ, which is not observed because it depends on a hypothetical voucher that has not been implemented. 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 χ_i as a vector consisting of a subset of fixed school district–specific variables (such as the income distribution in S_i) that determine $\overline{Z}(S_i,\ 0)$ and $\overline{Z}(S_i,\ \tau)$ as well as some individual covariates. The key assumptions concerning χ_i are that

$$E(\overline{Z}(S_i,\ 0)|\chi_i,\ P_i^\tau = P_i^0 = 1) = E(\overline{Z}(S_i,\ 0)|\chi_i,\ P_i^0 = 1), \tag{4}$$

$$E(\overline{Z}(S_i,\ \tau)|\chi_i,\ P_i^\tau = P_i^0 = 1) = E(\overline{Z}(S_i,\ \tau)|\chi_i,\ P_i^0 = 1). \tag{5}$$

These conditions say that the decision to remain in public school ($P_i^\tau = 1$) does not contain information about public school peers beyond what is contained in χ_i. The conditions do not restrict selection on who attends public school. That selection is already embodied in $\overline{Z}(S_i,\ 0)$. In our models these assumptions are innocuous. In our base case we have selection only on observables, so $\overline{Z}(S_i,\ 0)$ is essentially observed and is included in χ_i. We can use a similar condition for our empirical analogue of equation (5).

Let G denote a generic distribution and let $G(\chi_i|P_i^\tau = P_i^0 = 1)$ be the distribution of χ_i for public school stayers. From Bayes's theorem,

$$\begin{aligned} dG(\chi_i|P_i^\tau = P_i^0 = 1) &= \frac{\Pr(P_i^\tau = 1|P_i^0 = 1,\ \chi_i)}{\Pr(P_i^\tau = 1|P_i^0 = 1)}\, dG(\chi_i|P_i^0 = 1)\\ &= \psi(\tau,\ \chi_i)\, dG(\chi_i|P_i^0 = 1),\end{aligned} \tag{6}$$

where the weighting function

$$\psi(\tau,\ \chi_i) \equiv \frac{\Pr(P_i^\tau = 1|P_i^0 = 1,\ \chi_i)}{\Pr(P_i^\tau = 1|P_i^0 = 1)}$$

is the relative probability of remaining in public school after the voucher program τ is put into effect, conditional on χ_i.

The weight $\psi(\tau, \chi_i)$ is the link between the observable distribution of χ_i for all public school students to the distribution of those who stay after the voucher. It is central to the analysis because, as we show next, it can be used to infer average student body characteristics after the voucher program is implemented from the distribution in the absence of the program.

One may rewrite $E(\overline{Z}(S_i, 0)|P_i^\tau = P_i^0 = 1)$ as

$$\begin{aligned} E(\overline{Z}(S_i, 0)|P_i^\tau = P_i^0 = 1) &= \int E(\overline{Z}(S_i, 0)|\chi_i, P_i^\tau = P_i^0 = 1)\,dG(\chi_i|P_i^\tau = P_i^0 = 1) \\ &= \int E(\overline{Z}(S_i, 0)|\chi_i, P_i^\tau = P_i^0 = 1)\psi(\tau, \chi_i)\,dG(\chi_i|P_i^0 = 1) \\ &= \int E(\overline{Z}(S_i, 0)|\chi_i, P_i^0 = 1)\psi(\tau, \chi_i)\,dG(\chi_i|P_i^0 = 1) \\ &= E(\overline{Z}(S_i, 0)\psi(\tau, \chi_i)|P_i^0 = 1), \end{aligned} \tag{7}$$

where the first equality comes from the law of iterated expectations, the second comes from the definition of ψ, the third applies equation (4), and the fourth line comes from the law of iterated expectations.

Using the exact same argument with $\overline{Z}(S_i, \tau)$ taking the place of $\overline{Z}(S_i, 0)$ and using equation (5),

$$E(\overline{Z}(S_i, \tau)|P_i^\tau = P_i^0 = 1) = E(\overline{Z}(S_i, \tau)\psi(\tau, \chi_i)|P_i^0 = 1). \tag{8}$$

Using equations (7), (8), and (3), we can write $\pi^p(\tau)$ as

$$\pi^p(\tau) = E(\psi(\tau, \chi_i)[\overline{Z}(S_i, \tau) - \overline{Z}(S_i, 0)]'\delta|P_i^0 = 1). \tag{9}$$

The cream skimming formula simplifies further if the gain from choosing the default school is monotone in the sense that no student chooses to attend the default school under the proposed choice policy if he or she would not under the status quo. That is, by monotonicity we mean that if $P_i^\tau = 1$, then $P_i^0 = 1$. This is a very natural assumption in the case of expansion of charter schools or introduction of a voucher program, although it would not make sense for an expansion of choices within public schools.[12] Label the schools in the United States as $s_1, \ldots, s_L$. Then

$$\begin{aligned} &E(\overline{Z}(S_i, \tau)|P_i^\tau = 1, P_i^0 = 1) \\ &\quad = E(\overline{Z}(S_i, \tau)|P_i^\tau = 1) \\ &\quad = \sum_{\ell=1}^{L} \overline{Z}(s_\ell, \tau)\Pr(S_i = s_\ell|P_i^\tau = 1) \\ &\quad = \sum_{\ell=1}^{L} E(Z_i|S_i = s_\ell, P_i^\tau = 1)\Pr(S_i = s_\ell|P_i^\tau = 1) \\ &\quad = E(Z_i|P_i^\tau = 1). \end{aligned}$$

The first equality comes from monotonicity, the second explicitly states what the expectation means, the third replaces $\overline{Z}(s_\ell, \tau)$ with its formal definition, and the fourth uses the law of iterated expectations. This expression uses the fact that the average peer characteristic in a school is also the average characteristic in the school.

Using the same argument that we use for equation (7), we can show that

$$E(\overline{Z}(S_i, \tau)|P_i^\tau = P_i^0 = 1) = E(Z_i\psi(\tau, \chi_i)|P_i^0 = 1)$$

so that under monotonicity (3) reduces to[13]

$$\begin{aligned} \pi^p(\tau) &= E(\psi(\tau, \chi_i)[Z_i - \overline{Z}(S_i, 0)]'\delta|P_i^0 = 1) \\ &= \mathrm{Cov}(\psi(\tau, \chi_i), [Z_i - \overline{Z}(S_i, 0)]'\delta|P_i^0 = 1). \end{aligned} \tag{10}$$

Here we can also replace the assumption (5) with the simpler assumption

$$E(Z_i|\chi_i, P_i^\tau = P_i^0 = 1) = E(Z_i|\chi_i, P_i^0 = 1), \tag{11}$$

which holds trivially if Z_i is included in χ_i, as in our base case model below.

In practice, we do not observe all the classmates of a student, but only a random sample. Let $\overline{Z}_{S_{-i}}$ be the "i left-out means" consisting of the average value of Z for sample members who attended i's public school S_i, with i excluded. Then since

$$E(\overline{Z}_{S_{-i}}|\chi_i, P_i^\tau = P_i^0 = 1) = E(\overline{Z}(S_i, 0)|\chi_i, P_i^\tau = P_i^0 = 1),$$

[12] While we estimate some models in which monotonicity does not have to hold, in every case we present below it does hold given the model estimates. The simplified version of (9) eases interpretation of the empirical work.

[13] Going from the expectation to the covariance in the second equality uses the additional fact that $E(Z_i|P_i^0 = 1) = E(\overline{Z}(S_i, 0)|P_i^0 = 1)$.

we can also express the cream skimming effect as[14]

$$\begin{aligned}\pi^p(\tau) &= E(\psi(\tau,\ \chi_i)[Z_i - \overline{Z}_{S_{-i}}]'\delta | P_i^0 = 1)\\ &= \mathrm{Cov}(\psi(\tau,\ \chi_i),\ [Z_i - \overline{Z}_{S_{-i}}]'\delta | P_i^0 = 1),\end{aligned} \tag{12}$$

where we also replace the assumption (4) with

$$E(\overline{Z}_{S_{-i}} | \chi_i,\ P_i^\tau = P_i^0 = 1) = E(\overline{Z}_{S_{-i}} | \chi_i,\ P_i^0 = 1), \tag{13}$$

and we include $\overline{Z}_{S_{-i}}$ in χ_i so that the assumption is satisfied.

Equation (10) (or analogously [12]) shows that the cream skimming effect $\pi^p(\tau)$ is the covariance between $\psi(\tau,\ \chi_i)$ and $[Z_i - \overline{Z}(S_i,\ 0)]'\delta$. It is easy to see that $\pi^p(\tau)$ depends on three factors. The first is the extent and the nature of the variation in $\psi(\tau,\ \chi_i)$. If $\psi(\tau,\ \chi_i)$ 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 $\psi(\tau,\ \chi_i)$ does vary but is unrelated to variation within a school, there would be no cream skimming. The cream skimming effect would be zero. 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 $[Z_i - \overline{Z}(S_i,\ 0)]'\delta$). Targeting the voucher toward students with low values of $[Z_i - \overline{Z}(S_i,\ 0)]'\delta$ will lower $\psi(\tau,\ \chi_i)$ for those students and move $\pi^p(\tau)$ in a positive direction.

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

The third determinant is the magnitude of the student body coefficients δ. The value of $\pi^p(\tau)$ will be identically zero if there are no student body effects ($\delta = 0$). More generally, the more school composition matters for school outcomes, the more important the cream skimming effect could potentially be.

Thus one can see the importance of all three factors. The cream skimming effect will be zero if any of the three channels is zero, not just if all three are zero. By the same logic, to get a large value of $\pi^p(\tau)$, all the channels must be sizable. We obtain small values for $\pi^p(\tau)$ below because

[14] We leave Z_i out because random variation in Z across the sample of 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 $\overline{Z}(P_i,\ 0)$.

of the combination of factors, not a single one. We show that when one plugs reasonable estimates of the three channels into the formula, small estimates of $\pi^p(\tau)$ come out.

In Section VII.A.3, we explain how to generalize the cream skimming formula to the case of endogenous choice of district S_i, but we do not account for this in the empirical implementation.

III. Data

NELS:88 is a National Center for Education Statistics (NCES) survey that began in the spring of 1988. A total of 1,032 schools contributed as many as 26 eighth-grade students to the base year survey, resulting in 24,599 eighth graders participating.[15] Subsamples of these individuals were reinterviewed in 1990, 1992, 1994, and 2000. The NCES attempted to contact only 20,062 base year respondents in the first and second follow-ups and only 14,041 in the 1994 survey. Additional observations are lost because of attrition. A subsample of 12,144 individuals were reinterviewed in 2000, when most respondents were 26 years old. 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 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 surveys to ascertain aptitude and achievement in math, science, reading, and history. We use the behavior measures and eighth-grade test scores as person-specific control variables and student body measures. They have the advantage of being determined prior to high school.

Our main outcome measure is a high school graduation indicator, which is one if the respondent graduated from high school by the date of the 1994 survey and zero otherwise. The school choice variables are mutually exclusive indicators for whether the current or last school in which the individual was enrolled as of 1990 (2 years after the eighth-grade year) was a public high school, a Catholic high school, or a non-Catholic private high school.[16] Unless noted otherwise, the results in the paper are weighted.[17] Definitions of variables are provided in online ap-

[15] This description draws heavily on Altonji, Elder, and Taber (2002).

[16] A student who started in a private high school and transferred to a public school prior to the tenth-grade survey is coded as attending a public high school. In the case of Catholic schools, Altonji et al. (2002) present evidence that this is a minor issue.

[17] The sampling scheme in the NELS:88 is complicated. See Spencer et al. (1990) and Haggerty et al. (1996) for details. The weights depend in part on school choice and on outcomes, so it is important to weight. We use the third follow-up panel weights (f3pnlwt) for all analyses involving high school graduation or college attendance. The twelfth-grade

pendix 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 appendix tables A1 and A2. Because of the complexity of the estimator of $\pi^p(\tau)$ and its components, we use a block bootstrap method to compute standard errors and confidence intervals for most of the parameters. The blocks allow for correlation in the error terms among students who attend the same eighth grade and among students who attend the same high school. The blocks consist of students from each set of eighth grades that sent at least one student to a common high school. See appendix 1 for more detail. The distribution of N_s, the number of sample members in each high school, is concentrated between six and 18 observations in our effective sample.

Descriptive statistics.—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 schools 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 eighth-grade achievement scores. They also have a 0.63 advantage in log family income. The gap in eighth-grade math scores between private high school students and public high school students is 0.39 standard deviations.

Table 1 also shows that private high school students look stronger on a number of measures of eighth-grade behavior. For example, they score lower on an index of delinquency, fight less with other students, and have fewer behavior problems. They are more likely to complete homework, are much less likely to have repeated at least one grade between fourth and eighth grades, and score much lower on a composite measure of the risk of dropping out. The large gap between private high school students 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 school. With regard to outcomes, students who attend private high schools are much more likely to graduate from high school than public high school students (.94 vs. .86) and much more likely to be attending

math test and the labor income analyses use the second and the fourth follow-up panel weights, respectively. For the school choice models, we use the first follow-up panel weights (f1pnlwt). Weighting is particularly tricky when student body effects and/or unobserved school characteristics influence school choice, because then the appropriate weights depend on the collective behavior of the group of NELS:88 eighth graders in the attendance area of a particular public high school, not just the behavior of individual students. We discuss the issues in online apps. 3 and 4, respectively, but do not provide a full solution.

a 4-year college 2 years after the normal high school graduation year (.57 vs. .29).

IV. Results for the Basic Model

In this section we present cream skimming estimates for our baseline specification. For this case, we exclude student body characteristics and unobserved school district–specific attributes from the school choice model. We also assume that only observed student body 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 specific voucher program τ on the characteristics of those who remain in public school as well as estimates of $\pi^p(\tau)$. The steps of the estimation procedure are as follows.

1. Estimate the school choice preference parameters (Sec. IV.A).
2. Estimate the determinants of the outcome, including the coefficients on student characteristics and the school composition effects (Sec. IV.B).
3. Given a specific hypothetical voucher program, use the interaction between the program and preferences to estimate $\psi(\tau, \chi_i)$ (Sec. IV.C).
4. Calculate the treatment effect $\pi^p(\tau)$ (Sec. IV.D).

A. *Estimates of the Basic School Choice Model*

In our base case we consider a very simple version of the choice model and voucher program in which public school attendance is determined as a probit

$$P_i^\tau = 1(X_i'\beta + Q_{S_i}'\beta_Q - t(\tau) + u_i \geq 0), \tag{14}$$

where $1(\cdot)$ is the indicator function, $t(\tau)$ is the level of the universal voucher under program τ, and u_i is a standard normal and is independent of X_i as well as the observable covariates of other students who attend school S_i.[18] In the data there is no voucher program, so the voucher subsidy $t(0) = 0$. Furthermore, because we do not have data on $t(\tau)$, we normalize $\mathrm{Var}(u_i) = 1$. This implicitly defines the scale of t such that a

[18] Since the variables in Q_{S_i} are measured in eighth grade while Q_{S_i} is defined by the high school, there is a small amount of within–high school variation in these covariates. We take the average value of these variables within a school so that there is no within-school variation. This makes essentially no difference in the estimation.

TABLE 1
DESCRIPTIVE STATISTICS: FULL SAMPLE BY HIGH SCHOOL TYPE

Variable	All Schools	Public Schools	All Private	Catholic Private	Other Private
Demographics:					
Male	.50	.50	.53	.54	.52
Hispanic	.10	.10	.08	.10	.05
Black	.13	.13	.08	.11	.02
Geographic variables and zip code characteristics:					
Northeast	.20	.18	.31	.32	.29
North central	.26	.27	.22	.28	.13
South	.35	.35	.30	.25	.39
Urban	.25	.23	.45	.48	.40
Suburban	.44	.43	.47	.49	.45
Distance from Catholic high school (100s of kilometers)	.30	.32	.10	.05	.18
	(.48)	(.49)	(.24)	(.06)	(.36)
Fraction black	.11	.10	.12	.13	.10
	(.18)	(.19)	(.19)	(.20)	(.16)
Fraction Hispanic	.08	.08	.07	.09	.05
	(.16)	(.17)	(.13)	(.15)	(.09)
Median income	2.98	2.94	3.36	3.44	3.24
	(1.35)	(1.30)	(1.73)	(1.62)	(1.88)
Fraction under poverty line	.12	.13	.11	.11	.10
	(.09)	(.09)	(.10)	(.09)	(.11)
Fraction over two times poverty line	.64	.64	.66	.68	.61
	(.20)	(.20)	(.24)	(.20)	(.28)
Parental background:					
Both parents present	.66	.65	.80	.79	.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 income) 1987	10.27	10.21	10.84	10.71	11.02
	(.90)	(.88)	(.80)	(.80)	(.76)
Limited English proficiency	.02	.02	.01	.01	.02
Parents Catholic	.31	.28	.53	.79	.14
8th-grade 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)	(10.48)
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)	(10.65)
History score	50.92	50.47	54.87	54.53	55.37
	(9.90)	(9.77)	(10.17)	(9.28)	(11.37)
Delinquency index	.63	.65	.53	.52	.53
	(1.06)	(1.08)	(.91)	(.93)	(.88)
Grades composite	2.94	2.92	3.11	3.13	3.07
	(.73)	(.74)	(.66)	(.62)	(.70)
Student got into fight with other student	.25	.25	.22	.22	.22
Student performs below ability	.25	.25	.20	.15	.27
Student rarely completes homework	.19	.20	.13	.09	.18
Student frequently absent	.09	.10	.06	.07	.05
Student inattentive in class	.20	.20	.17	.12	.24

TABLE 1 (*Continued*)

Variable	All Schools	Public Schools	All Private	Catholic Private	Other Private
Student frequently disruptive	.12	.12	.12	.07	.19
Parent believes child has a behavior problem	.08	.09	.05	.04	.07
Repeated a grade	.06	.06	.03	.02	.04
Dropout risk composite	.66	.70	.30	.36	.21
	(.93)	(.95)	(.54)	(.59)	(.44)
Lack of effort index	4.01	4.03	3.81	3.49	4.30
	(2.73)	(2.72)	(2.82)	(2.63)	(3.02)
Enrolled in a gifted program	.14	.14	.12	.10	.15
Outcomes:					
High school graduation	.87	.86	.94	.98	.89
12th-grade reading test	50.89	50.45	54.77	54.21	55.68
	(9.71)	(9.70)	(8.87)	(8.92)	(8.73)
12th-grade math test	51.01	50.53	55.24	54.97	55.69
	(9.73)	(9.71)	(8.90)	(8.36)	(9.69)
Log earnings in 1999	9.96	9.94	10.07	10.18	9.90
	(.78)	(.78)	(.77)	(.67)	(.87)
Attended 4-year college	.32	.29	.57	.57	.56
Missing value indicators:					
Family income missing	.09	.09	.12	.14	.10
Tests missing	.03	.03	.02	.02	.02
Student behavior missing	.06	.05	.09	.09	.10
Distance missing	.06	.05	.08	.06	.13
Census data missing	.04	.04	.07	.03	.13
Sample size	16,483	14,193	2,290	963	1,354

NOTE.—Means for individual variables and standard deviations (in 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 variables. Sample sizes for the outcome variables are smaller. Explanatory variables are weighted using the base year through the first follow-up panel weights. High school graduation and attend 4-year college are weighted using base year through the third follow-up panel weights. The 12th-grade test scores and log earnings in 1999 are weighted using the second follow-up panel weights and the fourth follow-up panel weights, respectively.

unit change in t has the same effect on school choice as a one standard deviation change in the u_i.

The above model is a special case of more complicated school choice models we will estimate in later sections. In Section V, we allow the choice to depend not only on the student's own characteristics but on the characteristics of the public school peers. In Section VII.A.2, we work with an explicit utility-maximizing model, of which the probit model (14) is a special case. In Section VI, we allow for unobservable student body characteristics and fixed unobservable school characteristics. These extensions dramatically complicate the analysis but do not alter our basic conclusion.

Specification (14) assumes that vouchers are equivalent to an intercept shift in the latent variable that determines school choice. Our base case does not allow for interactions. Excluding them will tend to overstate the importance of cream skimming for real-world voucher programs, for two reasons. First, as we will discuss formally in Section VII.A.2, income effects should lead low-income families to be more price sensitive, and these families typically have negative student body effects. Second, the estimates of β will capture both student preferences about where to go to school and private schools' preferences for students through admission or financial aid decisions. The above specification is for a universal voucher that lowers net tuition for everyone by the same amount without changing differential treatment. The students who move to private schools in response to a voucher program that prohibits positive selection in admissions and aid would probably have weaker peer characteristics than our model would predict.

Online appendix table A1 presents maximum likelihood estimation (MLE) probit estimates of β for the basic school choice model (14) 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 student body effects are not contaminated by failure to control for the direct effects of a student's own characteristics and to ensure that the models give strong predictors of preference for public school and of high school graduation. In addition to *Catholic*, which is one 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 family income); multiple measures of aptitude and achievement, including grades and eighth-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 .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. A 2-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 from .899 to .848. Of course, these estimates hold the cognitive and behavior measures constant. Reading enters with a statistically significant 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 they are not statistically significant. Urban and suburban students are much less likely to attend public school than rural 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.

B. *Estimation of the School Outcome Parameters*

We estimate the school outcome model in two steps. First we estimate the coefficient vector γ from equation (1) by ordinary least squares (OLS) regression of $Y_i(0)$ on X_i with public school fixed effects. We then use the estimated school fixed effect in the first stage to estimate δ in our second stage.

Online appendix table A2 presents estimates of γ, the effect of student's own characteristics on high school graduation, holding constant high school characteristics common to all students. 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 interval estimates are also reported.[19]

Because we work with a very rich model, the estimates for specific family background variables, eighth-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 .15 lower for students who repeat a grade, .17 lower for students who are frequently absent, and .09 lower for students who rarely complete homework.

In this base case, we assume that all elements of Z_i are observed (this is relaxed in Sec. VI). Furthermore, we assume that ξ_{s_i} in (2) is uncorrelated with Z_i and Q_{s_i}. (We can define Θ_Q and ξ_{s_i} so that ξ_{s_i} and Q_{s_i} are uncorrelated.) In this case, estimation of δ would be straightforward if we observed the full population of students in each high school. We would just regress the estimated fixed effect on the student body characteristics ($\overline{Z}(S_i, 0)$) controlling for other aspects of the school (Q_{s_i}).

We cannot construct $\overline{Z}(S_i, 0)$ exactly because we do not observe all the students who attend the high school. However, we do observe a sample

[19] Correlation between ε_i and u_i could lead to sample selection bias given that we restrict the sample to public school students. We ignore this problem here given that relatively few students choose private school. We do correct for selection bias when we allow for unobservables in Sec. VI.

of students from each public school and thus can approach estimation of δ as an errors-in-variable problem in which we observe $\bar{Z}(S_i, 0)$ with error. With two people in each school, the solution would be a straightforward application of errors in variables with two noisy measures; we would use one person's Z_i as an instrument for the other one. That is, let i and j index two people chosen at random from school s. Then

$$Z_i \equiv E(Z_i | S_i = s) + \omega_i = \bar{Z}(s, 0) + \omega_i,$$
$$Z_j \equiv E(Z_j | S_i = s) + \omega_j = \bar{Z}(s, 0) + \omega_j.$$

The fact that i and j are chosen at random implies that $E(\omega_j | Z_i) = 0$. Thus this is a classic errors-in-variables problem in which we can use Z_i as an instrument for Z_j to estimate δ.

The estimation problem is more complicated with more than two peers as there are many possible instruments and thus many ways to implement the instrumental variable (IV) procedure. We could use only two observations from each school and use the procedure above, but this would involve throwing out most of our data. On the basis of Devereux's (2007) discussion of alternative approaches, we chose to estimate δ using the jackknife instrumental variable estimator (JIVE). Specifically, rewrite (2) as

$$\hat{\theta}(S_i, 0) = Z_i'\delta + Q_{S_i}'\Theta_Q + \varepsilon_{\theta i}, \tag{15}$$

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

$$\varepsilon_{\theta i} = [\bar{Z}(S_i, 0)' - Z_i']\delta + \xi_s + [\hat{\theta}(S_i, 0) - \theta(S_i, 0)].$$

We estimate δ and Θ_Q by IV regression. The instruments are Q_{S_i} and the i left-out means $\bar{Z}_{S_{-i}}$ consisting of the average value of Z for sample members who attended i's public school S_i, with i excluded.[20] Note that

[20] Because we must deal with sampling error in our measures of student body characteristics, the use of nonlinear alternatives to the linear probability model with fixed effects would greatly complicate the analysis. Note that JIVE does not deal with correlation between Z and ξ_s, but keep in mind that the student body effects parameter δ includes all the channels through which who attends a school influences school quality, not just peer effects in the classroom and school. This means that much of the discussion of omitted variable bias in the peer effects literature is not a concern here. Nevertheless, there are several other sources of bias in $\hat{\delta}$. First, bias in $\hat{\gamma}$ will spill over into bias $\hat{\delta}$ in the opposite direction. Second, advantaged students tend to go to better schools, suggesting a positive correlation between elements of $\bar{Z}_{S_{-i}}$ representing high socioeconomic status and the fixed unobserved component of school quality ξ_s even though the link is not causal. The fact that some high schools have more than one feeder school will bias $\hat{\delta}$ to the extent that the mean of Z_i varies across feeder schools for a given high school and the sample is not representative of the mix of students from the various feeder schools. This is likely to bias $\hat{\delta}$ downward. On the other hand, these students were schoolmates during eighth grade as

JIVE is not an alternative estimator to IV; it is just a relatively efficient way to use instruments when many are available. As it turns out, when δ is unrestricted, it is poorly determined in the sample because there is substantial collinearity among the $\overline{Z}_{S_{-i}}$.[21] Consequently, we focus on models that impose index restrictions on $\overline{Z}(s,\ \tau)'\delta$. For the base case we assume that $Z_i'\delta = \delta_{X'\gamma} X_i'\gamma$, where γ is a coefficient from the outcome equation (1) relating i's characteristics X_i to the outcome Y_i. The assumption implies that

$$\overline{Z}(s,\ \tau)'\delta = \delta_{X'\gamma}\overline{X}(s,\ \tau)'\gamma. \tag{16}$$

The equation restricts the student body effects on Y_i of school averages of father's education, family income, and so forth to be proportional to the direct effects of these on Y_i. One might expect the degree to which characteristics of student i influence the outcomes of other students is related to the degree to which they influence i directly, although the proportionality restriction is unlikely to hold exactly. We choose this specification for three reasons. First, we find it intuitively appealing that student characteristics matter for other students in proportion to the degree to which they matter for one's own outcome. Second, the outcome variable and $\overline{X}(s,\ \tau)'\gamma$ are in the same units. This makes the $\delta_{X'\gamma}$ parameter easy to interpret, as we describe below. Finally, it can be formally justified as coming from a model with endogenous peer effects in the sense of Manski (1993). We show this in online appendix 2.

Because religious preference plays a very special role in the decision to attend a Catholic high school, Catholic is set to zero when we evaluate the outcome index $X_i'\gamma$ for the purpose of imposing index restrictions on the student body effect parameters δ. Under a voucher program, the fraction of private high schools that are Catholic would probably decline.

We estimate $\delta_{X'\gamma}$ by replacing $Z_i'\delta$ with $\delta_{X'\gamma}X_i'\hat{\gamma}$ in (15) and using $\overline{X}'_{S_{-i}}\hat{\gamma}$ as the instruments for $X_i'\hat{\gamma}$ while controlling for Q_{S_i}. Panel A of table 2 reports $\hat{\delta}_{X'\gamma}$, with 95 percent confidence intervals in parentheses. These are based on 1,000 bootstrap replications. The vector Q_{S_i} consists of the region indicators, urban and suburban indicators, and a quadratic func-

well. Since δ is defined to be the effect of high school student body composition, $\hat{\delta}$ will be biased upward to the extent that eighth-grade student body effects continue to matter for high school. Since the net effect of the various biases is ambiguous, we also experiment with alternative values of δ.

[21] The collinearity among the $\overline{Z}_{S_{-i}}$ reflects that fact that student body traits such as the means of eighth-grade test scores, mother's and father's education, and family income tend to move together across schools. All the elements of Z_i have to be treated as endogenous variables. Furthermore, sampling variation in the i left-out school averages $\overline{Z}_{S_{-i}}$ due to the relatively small sample sizes for each school reduces the size of the first-stage coefficients in the JIVE procedure. The large number of endogenous regressors and collinearity among the instruments and relatively small first-stage coefficients make estimation of δ difficult.

TABLE 2
Estimation of the Effects of Student Body Composition on Public High School Graduation (δ)

	$X'\beta$ Index: $\delta_{X'\beta}$ (1)	$X'\gamma$ Index $\delta_{X'\gamma}$ (2)
A. Base case		.3680
		(.0307, .6324)
B. Base case with effects proportional to $X'\beta$	−.0450	
	(−.0985, .0041)	
C. Unobservable student body effects:		
$g = 1$	.0161	.0331
	(−.0066, .0383)	(−.0188, .0983)
$g = .5$	.0170	.0325
	(−.0078, .0404)	(−.0186, .0971)
$g = 1.5$	.0146	.0338
	(−.0055, .0349)	(−.0189, .0994)
D. Student body quality affects school choice:		
$X'\gamma$ index		.3680
		(.0581, .6002)
$X'\beta$ index	−.0389	
	(−.0826, −.0008)	

Note.—The table reports estimates of $\delta_{X'\gamma}$ and $\delta_{X'\beta}$. Panel A is the base case and assumes that the student body effect is proportional to $X'\gamma$. Panel B assumes that the effect is proportional to $X'\beta$. Panel C allows unobserved student body characteristics to affect outcomes. The effect of the unobserved index is $g\delta_{X'\beta}$ for the $X'\beta$ specification and $g\delta_{X'\gamma}$ for the $\overline{X}'\gamma$ specification, where g is assigned the values indicated in the row labels. Panel D reports $\delta_{X'\gamma}$ and $\delta_{X'\beta}$ when student body characteristics affect school choice ($\varphi \neq 0$). Numbers in parentheses are 95 percent confidence intervals.

tion of distance from the closest Catholic school (coefficients not shown). We find that $\hat{\delta}_{X'\gamma}$ is 0.368. Given that $\overline{X}(s,\ \tau)'\hat{\gamma}$ and Y are in the same units, the point estimate says that an increase in $\overline{X}(s,\ \tau)'\hat{\gamma}$ of an amount Δ raises the expected number of students who graduate from high school s by 1.368Δ. This value is the sum of Δ, the direct effect of $\overline{X}_i'\hat{\gamma}$ on i, plus the student body effect 0.368Δ. Consequently, the fraction $0.270 = 0.368/1.368$ of the effect of $\overline{X}(s,\ \tau)'\hat{\gamma}$ on the graduation rate for a given high school operates through student body effects. Thus $\hat{\delta}_{X'\gamma}$ is substantial, although it includes more than pure classroom peer effects. The estimate of $\hat{\delta}_{X'\gamma}$ is statistically distinct from zero, but the 95 percent confidence interval is fairly wide: (0.03, 0.63).

C. *Construction of the Weighting Function*

With some abuse of notation, we now use τ to refer specifically to a universal voucher generous enough to induce 10 percent of public school students to move to private school rather than as the index for a general set of school choice programs.[22] Since we are considering a universal voucher, we also use the simpler notation t rather than spell out $t(\tau)$ every time. Given estimates of the model, we next define the magnitude of t for the counterfactual and then construct an estimate of the weighting function $\psi(\tau, \chi_i)$ to estimate the cream skimming effect.

We do not need to know the price elasticity of demand for private schools or the size and effect of a reduction in nontuition costs that result from entry of additional private schools. Instead, we define the "size" t of the voucher subsidy (in units of u_i) in terms of the percentage of people induced to attend private school by the voucher. For our base case we define t to be the value that would induce 10 percent of public school students to move. Letting Φ be the cumulative density function of a standard normal, we estimate t as the value that solves[23]

$$-0.10 = \frac{\sum_i [\Phi(X_i'\hat{\beta} + Q_{S_i}'\hat{\beta}_Q - \hat{t}) - \Phi(X_i'\hat{\beta} + Q_{S_i}'\hat{\beta}_Q)]}{\sum_i \Phi(X_i'\hat{\beta} + Q_{S_i}'\hat{\beta}_Q)}. \tag{17}$$

Given the probit model assumption and the fact that Z_i is observed, (11) and (13) are satisfied when $\chi_i = \{X_i, Q_{S_i}, \overline{X}_{S_{-i}}\}$. We construct

$$\hat{\psi}(\tau, \chi_i) \equiv \frac{\dfrac{\Phi(X_i'\hat{\beta} + Q_{S_i}'\hat{\beta}_Q - \hat{t})}{\Phi(X_i'\hat{\beta} + Q_{S_i}'\hat{\beta}_Q)}}{\dfrac{1}{N_{P^0}} \sum_{j, P_j^0 = 1} \dfrac{\Phi(X_j'\hat{\beta} + Q_{S_j}'\hat{\beta}_Q - \hat{t})}{\Phi(X_j'\hat{\beta} + Q_{S_j}'\hat{\beta}_Q)}}, \tag{18}$$

where N_{p^0} is the number of sample members in public school.[24]

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

We obtain a consistent estimator $\hat{\pi}^p(\tau)$ by replacing the right-hand side of (12) with its sample analogue

$$\hat{\pi}^p(\tau) = \frac{1}{N_{P^0}} \sum_{\{i : P_i^0 = 1\}} \hat{\psi}(\tau, \chi_i)(X_i - \overline{X}_{S_{-i}})'\hat{\gamma}\hat{\delta}_{X'\gamma}. \tag{19}$$

[22] A more precise but cumbersome alternative would be to refer to the program using a specific value of τ, say $\tau = \tau_{10\%}^{\text{universal}}$.

[23] In the base case the 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. The value of $\hat{t}$ that solves (12) varies across the bootstrap replications used to construct the confidence intervals for the cream skimming effect.

[24] Note that χ_i depends on $\overline{X}_{S_{-i}}$, but $\overline{X}_{S_{-i}}$ does not enter $\hat{\psi}(\tau, \chi_i)$. This fact comes from the assumption above that $\overline{X}_{S_{-i}}$ does not enter school choice model (14) and is uncorrelated with u_i—an assumption that is relaxed in Sec. V.

Table 3 presents the base case estimates of the main parameter of interest—the cream skimming effect $\pi^p(\tau)$. The point estimate of $\pi^p(\tau)$ is -0.0014 and the lower bound to the confidence interval is -0.0023. For comparison, the public school dropout rate is 0.137.

To put these numbers in perspective, we compare the direct benefits to students who are induced to move to the harm for students who are left behind, weighting by the size of the groups. Suppose that moving from public school to private school leads to an increase in the graduation rate of 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 et al. (2005) obtain for Catholic high schools when they address the problem of selection on unobservables. The voucher program induces 10 percent of public school students to move, leaving nine students in public school for everyone who moves. The cream skimming estimate of -0.0014 implies that for each student who moves to private school, the average graduation rate for students who were in public school prior to the voucher rises by $(0.06 - 0.0014 \times 9)/10 = 0.0047$. The gain of 0.06 for each student who moves is partially offset by a decline of $0.0014 \times 9 = 0.013$ in the expected number of high school graduates among the nine students who remain. This offset is about 20 percent of the direct benefit received by the child who switches to private school. With the lower-bound estimate of -0.0023, the negative impact on the number of graduates is 0.021 among the nine stayers. Thus the average graduation rate considering both movers and stayers rises by 0.0039.

TABLE 3
ESTIMATES OF THE CREAM SKIMMING EFFECT $\pi^p(\tau)$ OF VOUCHERS ON HIGH SCHOOL GRADUATION RATES OF PUBLIC SCHOOL STAYERS

	Estimate
A. Base case	−.0014
	(−.0023, −.0001)
B. Base case with effects proportional to $X'\beta$	−.0011
	(−.0026, .0001)

NOTE.—The table reports estimates of $\pi^p(\tau)$ using the base case school choice model to estimate $\psi(\tau, \chi_i)$. These are the estimates of the cream skimming effect on the high school graduation rate of public school stayers following the introduction of a universal voucher program that induces 10 percent of public school students to move to private school. The estimates are for the base model and for the case in which composition effects are proportional to $X'\beta$. The specifications for the student body effects are (16) and (20). The corresponding formula for $\hat{\pi}_i(\tau)$ for the base case is given in (19) with an analogous expression for the $X'\beta$ case with $\hat{\gamma}\hat{\delta}_{X'\gamma}$ replaced with $\hat{\beta}\hat{\delta}_{X'\beta}$. Panels A and B of table 2 reports the estimates of $\delta_{X'\gamma}$ and $\delta_{X'\beta}$. Numbers in parentheses are 95 percent confidence intervals.

E. Estimates Assuming Student Body Effects Are Proportional to $X'\beta$

In this section we assume that student body effects are proportional to the mean of the school choice index $X_i'\beta$ rather than outcome index $X_i'\gamma$:

$$\overline{Z}(s,\ \tau)'\delta = Q_s'\delta_s + \delta_{X'\beta}\overline{X}(s,\ \tau)'\beta + \varpi_s. \tag{20}$$

We refer to this as the "$X'\beta$ index" or the "school choice index" specification.[25] While this might seem to be a less natural restriction, in appendix 2 we show that up to a first-order approximation, student characteristics matter for cream skimming only to the extent that they depend on $X_i'\beta$. The intuition follows from the cream skimming formula (10), which says the components that make up the index $Z_i'\delta$ matter for cream skimming only to the extent that are they are correlated with $\psi(\tau,\ \chi_i)$, the relative probability that a student will respond to the voucher.[26] More formally, note first that in the basic model, X_i includes all observed student characteristics that are correlated with choice and that u_i is defined to be independent of X_i and Z_i prior to conditioning on choice. Specifically, we show in appendix 2 that model (20) can be used if the relationship between an individual's $Z_i'\gamma$ and her $X_i'\beta$ can be approximated as

$$E(Z_i'\delta|X_i'\beta,\ S_i = s) \approx Q_s'\delta_s + \delta_{X'\beta}X_i'\beta + \varpi_s,$$

where ϖ_s is independent of Q_s and the distribution of X_i within a school.

We estimate $\delta_{X'\beta}$ using the same approach we use above to estimate $\delta_{X'\gamma}$. Table 2 reports that $\hat{\delta}_{X'\beta}$ is -0.045 with a confidence interval from -0.099 to 0.004. To get a sense of the magnitude, note that a decline in $\overline{X}'\hat{\beta}$ of 0.3 would lower the probability of attending a public school by roughly 0.046. The 0.3 decline in $\overline{X}'\hat{\beta}$ would increase the high school graduation probability by roughly $-0.3 \times (-0.045) = 0.014$.

Row B in table 3 reports that the point estimate of $\pi^p(\tau)$ is -0.0011 and the confidence interval is tight: -0.0026 to 0.0001. The lower-bound estimate implies a small negative effect on the graduation rate of stayers relative to the public school graduation rate.

When we use 0.06 as the gain for each student who moves and the fact that the voucher program induces 10 percent of public school students to move, the point estimate of -0.0011 for the $X'\beta$ index implies that the overall graduation rate for students who were in public school prior to the voucher rises by $(0.06 - 0.0011 \times 9)/10 = 0.005$. The lower-bound estimate is $(0.06 - 0.0026 \times 9)/10 = 0.0037$. These numbers are close to

[25] Note that γ is constructed from a fixed-effect model and so will already be purged of school fixed effects. Since β comes from the probit, it will not be. This is why we need to include the $Q_s'\delta_s$ term. The details are in the appendix.

[26] The intuition is also related to the fact that in the standard selection model, the selection bias is a function of the observable index (Heckman 1979, 199).

the values for the base case specification using the $X'\gamma$ index restriction. Our justification for the $X'\beta$ index is completely different from our justification for the $X'\gamma$ index. The fact that the basic results are very similar suggests to us that our choice of indices is not driving the results.

F. *The Effects of the Voucher Program on the Characteristics of Public School Students*

Having presented the cream skimming estimate, we now backtrack and more deeply examine a key determinant of it—the degree to which the voucher alters the characteristics of public school students. We begin by comparing the mean characteristics of public school stayers and movers for our base model.

Point estimates and 95 percent 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 4. The results show that the mean for movers is larger for two parents present, father's education, mother's education, log family 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 reason is that the stayer-mover difference for a particular variable is strongly affected by how it is correlated with other variables that influence school choice. The difference between the mean of $X_i'\gamma$ for the two groups is a good summary of how movers and stayers differ in terms of traits that matter for high school graduation. The values imply that the graduation rate for stayers would be 0.047 lower than the rate for movers, with school quality held constant (table 4, cols. 1 and 2, bottom row).[27]

Column 4 of the table also reports the change in the average value of schoolmate characteristics: the average value of Z of the peers of those who stay in public schools. This comes from the formula

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

The changes are small. For example, there is little change in racial/ethnic composition of schoolmates of public school stayers. The prevalence of two-parent households drops by only -0.01. Father's and

[27] We exclude Catholic from X_i when computing $X_i'\gamma$ and $X_i'\beta$ in table 4.

mother's education drop by 0.07 and 0.05, respectively, and the log of parental income drops by 0.025. The math test score declines by 0.12, which is only 0.012 standard deviations at the individual level.[28] Given the values of the changes in student body characteristics in column 4, all one needs to do to obtain the overall cream skimming effect is to multiply by δ. Given the small values, it would take large student body coefficients to obtain large overall cream skimming effects on outcomes. In fact, the base case estimate of the cream skimming effect in table 3 is simply the product of $-.0037$, the estimate of the average change in the school means of $\overline{X}'\gamma$ for stayers (table 4, col. 4, bottom row), and 0.3680, the estimate of $\hat{\delta}_{X'\gamma}$.[29]

Overall, the results suggest that a universal voucher program of the magnitude that we consider is unlikely to change the composition of public schools by very much.

G. *Results for College Attendance, Earnings, and Math Scores*

We focus on high school graduation because the evidence for a substantial benefit from private school is strongest for this outcome and because it is a critical degree in the United States and other developed countries. However, table 5 reports small cream skimming effects for other important outcomes. In the case of college enrollment 2 years after graduation, $\hat{\pi}^p(\tau)$ is -0.0007 with a lower bound of -0.0016 when we use the $X'\gamma$ index restriction. For comparison, the mean of college enrollment for public school students is 0.297. The point estimate implies that the cream skimming effect of a voucher that induces one in 10 students to move would reduce the number of students who enroll in college by only -0.0063. There is less evidence on the effects of private schools on college attendance to compare this value to. It is small relative to Altonji et al.'s (2005) lower-bound estimate of the effect of

[28] Column 5 reports the difference in the mean characteristics of public school stayers and the mean of the characteristics of all public school students (both movers and stayers) before the voucher. These changes are substantially larger than the change in the mean of the peers of stayers reported in col. 4. This reflects the fact that the peers of stayers prior to the voucher are less advantaged than the peers of movers. For example, the mean of the log of parental income of public school students declines by 0.041 while the change in the mean of the peers of stayers is -0.025.

[29] Similarly, when we use the $X'\beta$ index restriction (20), the estimate of $\hat{\pi}^p(\tau)$ in table 3 is calculated as

$$\hat{\pi}^p(\tau) = \left[\frac{1}{N_{P^0}} \sum_{i,P_i^0=1} \hat{\psi}(\tau,\chi_i)(X_i'\beta - \overline{X}_{S_{-i}}'\beta)\right]\hat{\delta}_{X'\beta}$$
$$= 0.0234 \times (-0.0450) = -0.0011.$$

The first number comes from the second to last row of table 4, col. 4, and the second number comes from panel B of table 2.

TABLE 4

Effect of a Universal Voucher Program on Selected Student Body Characteristics and $X_i'\beta$ and $X_i'\gamma$: Basic School Choice Model: No Unobservable School Characteristics or Peer Characteristics

	Mean Public School Stayers (1)	Mean Movers (2)	Mean Peer Stayers (before) (3)	Change in Peer for Stayers (4)	Change in Mean for Public School (5)
Catholic	.2781	.4347	.2894	−.0113	−.0156
	(.2545, .2984)	(.3857, .4899)	(.2677, .3085)	(−.0162, −.0072)	(−.0215, −.0105)
Male	.4918	.5076	.4939	−.0021	−.0016
	(.4796, .5030)	(.4709, .5420)	(.4839, .5039)	(−.0061, .0023)	(−.0053, .0024)
Hispanic	.1056	.0899	.1063	−.0007	.0016
	(.0825, .1304)	(.0549, .1187)	(.0825, .1315)	(−.0032, .0019)	(−.0015, .0050)
Black	.1276	.0929	.1240	.0035	.0034
	(.1075, .1478)	(.0555, .1218)	(.1050, .1436)	(.0008, .0069)	(.0002, .0072)
Both parents present	.6606	.7772	.6718	−.0112	−.0116
	(.6476, .6739)	(.7454, .8075)	(.6602, .6835)	(−.0156, −.0071)	(−.0150, −.0080)
Father's education	13.1552	14.4125	13.2255	−.0703	−.1249
	(13.0301, 13.2773)	(14.1131, 14.6876)	(13.1052, 13.3374)	(−.0969, −.0432)	(−.1543, −.0945)
Mother's education	12.7660	13.7058	12.8186	−.0526	−.0934
	(12.6757, 12.8699)	(13.5186, 13.8746)	(12.7256, 12.9130)	(−.0684, −.0346)	(−.1092, −.0749)
Log family income	10.1845	10.5949	10.2098	−.0254	−.0408
	(10.1493, 10.2233)	(10.5289, 10.6483)	(10.1764, 10.2474)	(−.0325, −.0179)	(−.0473, −.0337)

Reading score	50.2649	53.1744	50.4619	−.1970	−.2891
	(49.9076, 50.6598)	(52.2086, 53.9628)	(50.1021, 50.8200)	(−.2722, −.1058)	(−.3772, −.1799)
Math score	50.4698	53.0269	50.5936	−.1238	−.2541
	(50.0966, 50.8565)	(52.1854, 53.8713)	(50.2363, 50.9519)	(−.2042, −.0412)	(−.3449, −.1612)
Science score	50.5888	52.5179	50.7164	−.1276	−.1917
	(50.2001, 51.0101)	(51.6711, 53.3211)	(50.3503, 51.0917)	(−.2203, −.0315)	(−.2841, −.0903)
History score	50.3369	53.1305	50.4661	−.1292	−.2776
	(49.9555, 50.7369)	(52.0307, 54.0025)	(50.0959, 50.8694)	(−.2128, −.0325)	(−.3688, −.1653)
$X'\beta$	2.0365	1.6794	2.0131	.0234	.0355
	(1.6010, 2.5277)	(1.2107, 2.1393)	(1.5719, 2.4955)	(.0200, .0370)	(.0298, .0524)
$X'\gamma$	.3963	.4437	.4000	−.0037	−.0047
	(.2353, .5412)	(.2771, .5894)	(.2394, .5455)	(−.0051, −.0023)	(−.0060, −.0034)

NOTE.—Means for stayers (col. 1) and for movers (col. 2) are calculated using the estimated public school choice weights $\psi(\tau, \chi_i)$ to reweight the sample of students who attended public high school in NELS:88. The weights were obtained using the base case model of school choice. Column 3 is the sample mean of the characteristics of schoolmates of public school stayers prior to the voucher. Column 4 is the change in the mean of schoolmate characteristics of public school stayers following the voucher. The formula used to estimate it is in Sec. IV.F. Column 5 reports the difference between the mean characteristics of public school stayers and the mean characteristics of public school students (both movers and stayers) before the voucher. Numbers in parentheses are 95 percent confidence intervals.

TABLE 5
Estimates of the Cream Skimming Effect $\pi^{p}(\tau)$ for Alternative Outcome Variables

	$X'\beta$ Index	$X'\gamma$ Index
College attendance	−.0003	−.0007
	(−.0018, .0013)	(−.0016, .0004)
12th-grade reading	.0063	.0072
	(−.0203, .0365)	(−.0058, .0226))
12th-grade math	−.0227	−.0068
	(−.0455, .0011)	(−.0173, .0035)
Log earnings	−.0006	.0020
	(−.0033, .0019)	(.0004, .0054)

Note.—This table is analogous to table 3 except that different outcomes are considered. The outcomes are full-time enrollment in a 4-year college in 1994, twelfth-grade reading and math scores (standard deviation ≈ 10), and the log of earnings in 1999 for those with positive earnings. Numbers in parentheses are 95 percent confidence intervals.

Catholic school attendance on college enrollment using NELS:88 data of 0.03. It is very small relative to the probit estimate of 0.124 for all private schools in our sample treating private school choice as exogenous conditional on a rich set of controls (not reported). With the 0.03 estimate, the net effect of the voucher program on the average college attendance rate for students who were in public school prior to the voucher is $(0.03 - 0.0007 \times 9)/10 = 0.0024$.

The estimates of the cream skimming effect on the log of annual earnings in 1999 (when the sample is about 25 years old) are also small. The lower bound of the 95 percent confidence interval is actually positive in the case of the $X'\gamma$ index. In the case of twelfth-grade math test scores, the most negative estimate is the lower bound of −0.046 using the $X'\beta$ index, which is only −0.0046 standard deviations of the test.[30]

V. Allowing Public School Choice to Depend on Student Body Quality

In this section, we add the student body quality index $\bar{Z}(S_i, \tau)'\delta$ to the school choice model (14). The model for a universal voucher program becomes[31]

$$P_i^{\tau} = 1(X_i'\beta + Q_{S_i}'\beta_Q + \varphi\bar{Z}(S_i, \tau)'\delta - t(\tau) + u_i \geq 0). \tag{21}$$

Doing so dramatically complicates the analysis for a number of reasons. The most important is that $\bar{Z}(S_i, 0)$ and its sample counterpart for i, $\bar{Z}_{S_{-i}}$,

[30] We do not calculate the net effect of the voucher for test scores, but the evidence for a positive direct effect of private school attendance is weaker than for high school graduation rates. See Rouse and Barrow (2009) for a brief review.

[31] We continue to abstract from competition effects.

are not observed for students who choose private 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 observed only for those who choose the good. Our approach requires data on a set of observables at the district level, W_{S_i}, that shift the mean of the distribution of $X_i'\beta$ and $Z_i'\delta$ over all students in district S_i, both public and private. Our strategy is to specify the distributions of $X_i'\beta$ and $Z_i'\delta$ conditional on W_{S_i}. These distributions, together with the school choice model (21), provide a way to infer the average characteristics of the students who choose the public school S_i, which in turn determines $\overline{Z}(S_i, \tau)'\delta$.[32]

We focus first on the case in which $Z_i'\delta = \delta_{X'\gamma}X_i'\gamma$. We start by expressing the values of $X_i'\beta$ and $X_i'\gamma$ for students who are assigned to district S_i as

$$X_i'\beta + Q_{S_i}'\beta_Q = \mu_{S_i}^1 + \eta_i^1; \quad X_i'\gamma = \mu_{S_i}^2 + \eta_i^2,$$

where (μ_s^1, μ_s^2) are the mean values for district s, (η_i^1, η_i^2) is $N(0, \Sigma_\eta)$, and we define $\sigma_{\eta j\ell}$ to be the (j, ℓ) element of Σ_η.

In this case, the above model implies that

$$\begin{aligned}
&\overline{X}(s, \tau)'\gamma \\
&\quad = E(X_i'\gamma | S_i = s,\ P_i^\tau = 1) \\
&\quad = \mu_s^2 + E(\eta_i^2 | \mu_s^1 + \eta_i^1 + \varphi\delta_{X'\gamma}E(X_i'\gamma|S_i = s, P_i^\tau = 1) - t(\tau) + u_i > 0) \\
&\quad = \mu_s^2 + \frac{\sigma_{\eta 12}}{\sqrt{1+\sigma_{\eta 11}}}\lambda\left(\frac{\mu_s^1 - t(\tau) + \varphi\delta_{X'\gamma}E(X_i'\gamma|S_i = s,\ P_i^\tau = 1)}{\sqrt{1+\sigma_{\eta 11}}}\right),
\end{aligned} \tag{22}$$

where λ is the inverse Mills ratio. We solve (22) numerically for $E(X_i'\gamma| S_i = s,\ P_i^\tau = 1)$. Multiple equilibria are possible because the demand for public school depends on the choices of other students. Fortunately, this has not been a problem in practice.

[32] Bayer, Ferreira, and McMillan'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, Levinsohn, and Pakes (1995) and is very different from ours. Ferreyra (2009, 2007) estimates an equilibrium model that allows for schoolmates to affect choices using district-level data.

The biggest issue in estimating this version of the model is that we cannot simply run a probit of public school attendance on an estimate of $E(X_i'\gamma|S_i = s,\ P_i^\tau = 1)$ because for the private school students we do not have any data on the peers they would have if they attended public school. To identify the model, we use a further assumption that the district means μ_s^1 and μ_s^2 differ across s and depend on the observed district-level characteristics W_s through the equations

$$\mu_s^1 = W_s'\alpha_1 + e_s^1, \tag{23}$$

$$\mu_s^2 = W_s'\alpha_2 + e_s^2, \tag{24}$$

where $(e_s^1,\ e_s^2)$ is $N(0,\ \Sigma_e)$. 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. In this sense we are imposing some exclusion restrictions on the W_s for identification purposes. Intuitively we first estimate the relationship between W_{s_i} and $X_i'\gamma$ to learn about the relationship between W_{s_i} and school composition (correcting for selection). We then look at the relationship between W_{s_i} and public school attendance (conditional on X_i) to identify the importance of school composition on public school attendance.

The estimation procedure for the model parameters is complicated, so we relegate the details to online appendix 3. Here is a rough description. We first estimate γ from (1) using OLS with public school fixed effects as before. We iterate on the following procedure. Taking values of $(\alpha_1, \alpha_2, \Sigma_\eta, \Sigma_e)$ and the distribution of $E(X_i'\gamma|S_i = s,\ P_i^\tau = 1)$ conditional on W_{s_i} as given, we estimate β, β_Q, and the product $\varphi\delta_{X'\gamma}$ by maximum likelihood. In this step, identification of $\varphi\delta_{X'\gamma}$ comes from the variation in W_{s_i} conditional on X_i. Given the estimates of β, β_Q, γ, and $\varphi\delta_{X'\gamma}$, we use OLS and (23) and (24) 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 α_1, α_2, β, β_Q, and γ as known. Finally, taking α_1, α_2, β, β_Q, φ, Σ_η, and Σ_e as given, for each s we update $E(X_i'\gamma|S_i = s,\ P_i^\tau = 1)$ as the fixed point of (22). We iterate on these estimation steps until the model parameters converge.

In addition to Q_{S_i}, the vector W_{S_i} consists of the average of the demographic characteristics for the zip code in which a student's eighth-grade school is located. 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 percentage of the population below the poverty line, and the percentage of the population with income more than double the poverty line. The vector W_{S_i} explains an additional 11.6 percent of the cross–high school variance in $X_i'\beta + Q_{s_i}'\beta_Q$ and 14.6 percent of the cross–high school variance in $X_i'\gamma$ conditional on other area characteristics Q_{S_i} that we control for. The Q_{S_i} variables include census region, urbanicity, and distance from the nearest Catholic high school.

The value of $\widehat{\varphi\delta}_{X'\gamma}$, the estimate of the effect of $\overline{X}'\hat{\gamma}$ on the latent variable for public school choice, is 4.82 but (borderline) is not significant, with a 90 percent confidence interval of (−1.30, 12.74).[33] What does this estimate imply about the sensitivity of school choice to public school quality? The value of $\widehat{\varphi\delta}_{X'\gamma}$ implies that the student body effect of a reduction of 0.1 in $\overline{X}'\hat{\gamma}$ lowers the fraction of students who choose public school from 0.899 to 0.816. (The simulated standard deviation of $\overline{X}'\hat{\gamma}$ for public high school students is 0.050.)[34] The value of $\hat{\delta}_{X'\gamma}$ (0.368) implies that the student body effect of a reduction of 0.1 in $\overline{X}'\hat{\gamma}$ lowers the graduation rate by 0.0368. Thus, the estimates indicate that school choice is fairly sensitive to school quality, or at least to the component of school quality associated with student body composition.

Note, however, that families may have difficulty distinguishing differences in value added from differences in average school outcomes. They may choose schools on the basis of average graduation rates rather than school quality per se. If they focus on average outcomes, then the coefficient on $\overline{X}'\hat{\gamma}$ in the choice equation is an estimate of $\varphi(1 + \hat{\delta}_{X'\gamma})$ rather than $\varphi\delta_{X'\gamma}$.[35] Since the total effect on graduation rates of a reduction of 0.1 in $\overline{X}'\hat{\gamma}$ is the sum of the direct effect of 0.1 and the student body effect of 0.368×0.1, the drop from 0.899 to 0.816 in the fraction of students who choose public school should be interpreted as the demand response to a difference of $0.1 + 0.0368$ in the graduation rate for the high school. However, this alternative interpretation of the coefficient on $\overline{X}'\hat{\gamma}$ does not change the estimate of the cream skimming effect. (Similar remarks apply to the interpretation of the public school coefficient on $\overline{X}'\hat{\beta}$ when we use the $X'\beta$ index restriction on student body effects.) Of course, it is possible that school composition affects 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).

As it turns out, the point estimate of the cream skimming effect is −0.0014 (panel A of table 6). This is very close to the base case results with φ set to zero.

[33] The coefficients on the other variables in the choice equation are similar to those for the base model reported in app. table A1 and similar to those in the $X'\beta$ peer model that allows student body quality to affect the school choice.

[34] This reduction in the public school attendance probability is about the same amount as the combined effect of 4-year reductions in both mother's education and father's education on the probability that a particular student attends public school.

[35] This would be rational if colleges and firms judged students in part by the average performance in the schools they attend or by their perception of average student body quality.

TABLE 6

Estimates of the Cream Skimming Effect $\pi^p(\tau)$ When Student Body Quality Affects Choice and When Unobserved Student Body Characteristics Affect School Quality

	$X'\beta$ Index	$X'\gamma$ Index
A. Student body quality affects school choice: $\varphi \neq 0$:		
$X'\gamma$ index	NA	−.0014
		(−.0025, −.0002)
$X'\beta$ index	−.0011	NA
	(−.0027, .0000)	
B. Unobserved student body characteristics affect school quality:		
$g = 1$	.0006	−.0002
	(−.0003, .0016)	(−.0005, .0001)
$g = .5$	.0005	−.0001
	(−.0002, .0012)	(−.0004, .0001)
$g = 1.5$	.0008	−.0003
	(−.0003, .0019)	(−.0007, .0002)

Note.—Panel A of the table reports estimates of $\pi^p(\tau)$ using the school choice model with student body effects on demand ($\varphi \neq 0$) to estimate $\psi(\tau, \chi_i)$. The two columns, labeled $X'\beta$ and $X'\gamma$, use (20) and (16), respectively, as the specification for the effect of student body characteristics on school quality. The column labels also indicate which student body quality index is used in the school choice model. The cream skimming formulas are in app. 3. The estimates in panel B allow for fixed unobserved school characteristics that do not respond to the voucher program to influence choice and school outcomes. They also allow unobserved student body characteristics to influence school quality as specified in (26). The parameter g is assigned the values indicated in the row labels. See panel C of table 2. The cream skimming formulas and details of the estimation can be found in app. 4. Numbers in parentheses are 95 percent confidence intervals.

We also estimate the model with the restriction $\overline{Z}(s, \tau)\delta = Q_s'\delta_s + \delta_{X'\beta}\overline{X}(s, \tau)'\beta + \varpi_s$ imposed. This simplifies the estimation procedure because we no longer have to model the distribution of $X_i'\beta | W_s$ separately from the distribution of $X_i'\gamma | W_s$. Under the assumptions of the model in the $X'\beta$ index case, the coefficient on $\overline{X}(s, \tau)'\beta$ in the school choice model is $\varphi\delta_{X'\beta}$. The estimate is −0.52 and the 90 percent confidence interval is (−0.96, 0.05). The point estimate implies that a small increase in the student body average $\overline{X}'\beta$ lowers the probability that student i will attend public school by about 52 percent as much as an equivalent increase in $X_i'\beta$ raises the attendance probability for i. Consider an increase in $\overline{X}'\hat{\beta}$ of 0.946. The estimate of $\delta_{X'\beta}$ implies that such an increase would lower the graduation rate by 0.0368, which is the drop associated with a 0.10 decline in $\overline{X}'\hat{\gamma}$ that we considered when discussing the $X'\gamma$ index case above.[36] The estimates of $\varphi\delta_{X'\beta}$ imply that an increase in $\overline{X}'\hat{\beta}$ of 0.946 induces a decline in the public school attendance rate from 0.899 to 0.814. Thus the effect is very similar to the $X'\gamma$ model.

[36] Table 2 reports that $\hat{\delta}_{X'\beta}$ is −0.0389, which is close to the estimate of $\delta_{X'\beta}$ in the base case model with $\varphi = 0$. An increase in $\overline{X}'\hat{\beta}$ of 0.946 lowers the graduation probability by $0.946 \times 0.0389 = 0.0368 = 0.368 \times 0.1$, where from table 2 we also see that $\hat{\delta}_{X'\gamma} = 0.368$.

The point estimate of the cream skimming effect is -0.0011 (row 2 of table 6). This is also very close to the base case result for the $X'\beta$ index specification with φ set to zero.

We conclude that student body composition influences school choice but that accounting for this does not have much of an effect on the size of the cream skimming effect. To see how these two results can be reconciled, note first that students on the margin of attending are the most likely to be affected by a voucher, at least in the case of a small 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. The size of the multiplier will influence the fraction of students who move as a 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 a student's own characteristics in the school choice model, accounting for student body effects in the school choice model might make a bigger difference. For example, if the highest-performing students from the richest families are the most sensitive to student body composition, student body effects would be larger. We leave this potentially important but difficult 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 zero in what follows.

VI. Unobservable Student Body Effects and Fixed Unobservable School Characteristics

Thus far, we have estimated the student body effects that operate through covariates that are observable to the econometrician. If student body effects also involve unobservables, we might understate the true effects. Here and in online appendix 4, we present a methodology for estimating the cream skimming effect in the presence of unobservable student body effects on outcomes and unobservable school characteristics that influence school choice but do not depend on the voucher. Unfortunately, we do not have "instruments" to identify the coefficient governing the effect of the unobservable student body characteristics, which makes estimation very difficult. Consequently, we estimate our model under a wide range of assumptions about the coefficient governing the effect of the unobservables and show that the finding of a small cream skimming effect is robust.

Define

$$\bar{u}(s,\ \tau) = E(u_i | S_i = s,\ P_i^\tau = 1);$$
$$\bar{\varepsilon}(s,\ \tau) = E(\varepsilon_i | S_i = s,\ P_i^\tau = 1),$$

where, as a reminder, u_i is the error term in the selection equation (14), ε_i is the student-specific error term in the outcome equation (1), and $\bar{u}(s, \tau)$ and $\bar{\varepsilon}(s, \tau)$ 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, write the student body effect $\overline{Z}(s, \tau)'\delta$ on the school outcome as

$$\overline{Z}(s, \tau)'\delta = \overline{X}(s, \tau)'\delta_x + \delta_u\bar{u}(s, \tau) + \delta_\varepsilon\bar{\varepsilon}(s, \tau).$$

Our results up to this point imposed that $\delta_u = \delta_\varepsilon = 0$ because we assumed that only observable student characteristics influence the outcomes of other students, so that $\overline{Z}(s, \tau)'\delta = \overline{X}(s, \tau)'\delta_x$. For example, in the $X'\gamma$ index specification (16), we restricted $\delta_x = \delta_{X'\gamma}\gamma$ and $\delta_u = \delta_\varepsilon = 0$. We now generalize that case to allow the student body effect $\overline{Z}(s, \tau)'\delta$ to also depend on the school mean of the unobservable student outcome determinant ε_i. The extension is

$$\overline{Z}(s, \tau)'\delta = \delta_{X'\gamma}[\overline{X}(s, \tau)'\gamma + g\bar{\varepsilon}(s, \tau)]. \tag{25}$$

Conditional on γ, which is identified from the fixed-effect outcome equation, g is identified under our functional form assumptions. However, we do not rely on the functional form assumptions and instead produce results for the values $g = 0.5$, $g = 1$, and $g = 1.5$.

We also generalize the $X'\beta$ index version of the model to allow student body effects to depend on both observed and unobserved student characteristics that affect school choice, with

$$\overline{Z}(s, \tau)'\delta = Q_s'\delta_s + \delta_{X'\beta}[\overline{X}(s, \tau)'\beta + Q_s'\beta_Q + g\bar{u}(s, \tau)] + \varpi_s, \tag{26}$$

where, as in the base model, ϖ_s is independent of Q_s and the determinants of X_i and u_i. In appendix 2 we justify (26) by showing that up to a first-order approximation, observed and unobserved student body characteristics matter for cream skimming only to the extent that they depend on $\overline{X}(s, \tau)'\beta$ and $\bar{u}(s, \tau)$.

Estimation of the $X'\gamma$ index and $X'\beta$ index versions of the model with unobservable student body effects is very similar, so we present the estimation framework for both models. The procedure is simpler in the case of the $X'\beta$ index version (26) because one does not have to estimate any of the determinants of $X_i'\gamma$.

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

$$X_i'\beta + Q_{S_i}'\beta_Q = \mu_{S_i}^1 + \eta_i^1; \qquad X_i'\gamma = \mu_{S_i}^2 + \eta_i^2,$$

where s is the school level and i is the individual level. Analogously, decompose u_i and ε_i as

$$u_i = v^1_{S_i} + \omega^1_i; \qquad \varepsilon_i = v^2_{S_i} + \omega^2_i,$$

where $v^1_{S_i}$ and $v^2_{S_i}$ are the means for district S_i and ω^1_i and ω^2_i are the student-specific idiosyncratic components. The larger the variance of $v^1_{S_i}$ and $v^2_{S_i}$, the more important the differences across districts in unobserved characteristics that affect school choice and outcomes.

We assume that the vectors (η^1_i, η^2_i) and (ω^1_i, ω^2_i) are uncorrelated but jointly normal.[37]

Let Σ_μ, Σ_v, Σ_η, and Σ_ω be the 2×2 variance covariance matrices of (μ^1_s, μ^2_s), (v^1_s, v^2_s), (η^1_i, η^2_i), and (ω^1_i, ω^2_i), respectively. We refer to the (j, ℓ) element of Σ_η and Σ_ω as $\sigma_{\eta j\ell}$ and $\sigma_{\omega j\ell}$. We use an "observables are like the unobservables" assumption that states that there is a single scalar a such that

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

and

$$\begin{bmatrix} \sigma_{\omega 11} & \sigma_{\omega 12} \\ \sigma_{\omega 12} & \sigma_{\omega 22} \end{bmatrix} = a \begin{bmatrix} \sigma_{\eta 11} & \sigma_{\eta 12} \\ \sigma_{\eta 12} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{\omega 22} \end{bmatrix}.$$

Since $\sigma_{\omega 22}$ does not play a role in the analysis, we do not need to restrict it. The covariance restrictions impose that the relative variances and the covariance of the unobservable cross-district components affecting 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 $\sigma_{\omega 22}$, which is not relevant in our approach.

Appendix 4 discusses estimation of the model parameters. We obtain β from a standard probit model for public school choice. We obtain γ using fixed-effects regression on the public school sample, correcting for sample selection by including the inverse Mills ratio term for public school choice in the regression. Estimation of the other model parameters, including $\delta_{X'\beta}$, Σ_μ, and Σ_η, is complicated, so we leave the discussion to appendix 4.

We now provide an expression for the cream skimming effect under (26) (the $X'\beta$ index). As in Section IV, we consider the change from no

[37] The assumption of zero correlation between (ω^1_i, ω^2_i) and (η^1_i, η^2_i) is natural given that they are the student-specific components of $(X'_i\beta, X'_i\gamma)$ and (u_i, ε_i), respectively, and u_i and ε_i are orthogonal to X_i. Note that the terms (μ^1_s, μ^2_s) may depend on W_s through (23) and (24), in which case we assume further that $(W'_s\alpha_1, W'_s\alpha_2)$ is jointly normal. However, we do not need to incorporate W_s and estimation of α_1 and α_2 into the analysis in this section.

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

$$\begin{aligned}
&\pi^p(\tau;\ \mu_s^1,\ v_s^1)\\
&\quad = \delta_{X'\beta} E(X_i'\beta + Q_s'\beta_Q + g u_i | S_i = s,\ P_i^\tau = 1)\\
&\qquad - \delta_{X'\beta} E(X_i'\beta + Q_s'\beta_Q + g u_i | S_i = s,\ P_i^0 = 1)\\
&\quad = \delta_{X'\beta}[\{\mu_s^1 + g v_s^1 + E(\eta_i^1 + g\omega_i^1 | S_i = s,\ P_i^\tau = 1)\} \qquad (27)\\
&\qquad - \{\mu_s^1 + g v_s^1 + E(\eta_i^1 + g\omega_i^1 | S_i = s,\ P_i^0 = 1)\}]\\
&\quad = \delta_{X'\beta} \frac{\sigma_{\eta 11} + g\sigma_{\omega 11}}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}} \left[\lambda\left(\frac{\mu_s^1 + v_s^1 - t}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right) - \lambda\left(\frac{\mu_s^1 + v_s^1}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right)\right].
\end{aligned}$$

Note that term $\mu_s^1 + g v_s^1$ 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.[38]

When the student body effect is proportional to the expectation of the outcome index $X_i'\gamma + g\varepsilon_i$ conditional on public school attendance, the cream skimming effect for school s is

$$\begin{aligned}
\pi^p(\tau;\ \mu_s^1,\ v_s^1) = \delta_{X'\gamma} \frac{\sigma_{\eta 12} + g\sigma_{\omega 12}}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}} \Bigg[\lambda\left(\frac{\mu_s^1 + v_s^1 - t}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right)\\
- \lambda\left(\frac{\mu_s^1 + v_s^1}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right)\Bigg]. \qquad (28)
\end{aligned}$$

To use (27) or (28) to estimate cream skimming effects, we have to address the fact that we do not observe $(\mu_s^1,\ v_s^1)$, which are arguments of cream skimming effect functions for each school. We use Bayes's rule to infer the conditional distribution of $(\mu_s^1,\ v_s^1)$ given the characteristics of the other students who choose the public school, with selection taken into account. We then integrate $(\mu_s^1,\ v_s^1)$ out and identify the average of

[38] 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 as the district mean of unobserved family characteristics that influence school choice are all determinants of v_s^1, 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 that most of the elements of X_i are student level might lead one to question the "unobservables are like the observables" restriction. One could explore sensitivity of the results to allowing the proportionality factor a to differ in the two covariance matrix restrictions, but we have not done so.

$\pi^p(\tau;\ \mu_s^1,\ v_s^1)$ over the distribution of $(\mu_s^1,\ v_s^1)$ of those who stay in public school. The details are in appendix 4.

The results are as follows. In table 2, the student body effect parameter $\hat{\delta}_{X'\beta}$ is 0.0170 when $g = 0.5$, 0.0161 when $g = 1$, and 0.0146 when $g = 1.5$, which is the opposite sign of the base model, though none of these are statistically significant. The estimates of $\delta_{X'\gamma}$ are 0.0325 when $g = 0.5$, 0.0331 when $g = 1$, and 0.0338 when $g = 1.5$, which are substantially lower and more precise than the point estimate of 0.3680 for $\delta_{X'\gamma}$ in the base model with $g = 0$.[39]

The estimates of the cream skimming effect are in panel B of table 6. Regardless of the assumption about g, the estimates are slightly positive when we use the $X'\beta$ index model. The estimates for the $X'\gamma$ model are about -0.0002. All the estimates are close to zero and precisely estimated.

Given the complexity of allowing school choice to depend on the choices of other students and to allow for unobservable student body effects and given the evidence that accounting for these features makes little difference, we do not allow for them in the rest of the paper.

VII. Alternative Assumptions about School Choice and about the Effects of Student Body on Outcomes

In the next two subsections we experiment with alternative specifications of the school choice model and the outcome model.

A. *Alternative School Choice Models*

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. The first row of table 7 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.[40]

[39] Taken at face value, these results suggest that the unobserved student body characteristic $\bar{\varepsilon}(s,\ 0)$ matters less for school quality than the observed characteristics. The observed variables do contain many that one might think would be important, including family income, parental education, eighth-grade test scores, and eighth-grade measures of behavior. However, we would not want to make too much of this given that the lower bound of the 95 percent confidence interval for $\delta_{X'\gamma}$ in the base case is 0.0307. The estimation methodology in the current section is also very different.

[40] In the case of specification (20), we use the index of coefficients that determine whether public school is chosen over the two alternatives. We also tried a trinomial probit

TABLE 7
Estimates of the Cream Skimming Effect $\pi^p(\tau)$ on Public School Stayers: Alternative School Demand Models

	$X'\beta$ Index	$X'\gamma$ Index
Nested logit with Catholic and non-Catholic private schools	−.0012	−.0015
	(−.0029, .0004)	(−.0024, −.0002)
Alternative assumptions about utility function:		
Utility function parameter $\zeta = .01$	−.0010	−.0014
	(−.0024, .0002)	(−.0022, −.0001)
Utility function parameter $\zeta = .5$	−.0006	−.0008
	(−.0016, .0001)	(−.0016, −.0000)
Log utility function $\zeta = 1$	−.0001	−.0002
	(−.0007, .0001)	(−.0007, .0003)

Note.—The first row reports estimates of $\pi^p(\tau)$ based on a nested logit school choice specification that places the public school in one nest and the Catholic and non-Catholic school options in the other. The bottom rows allow for nonlinear utility of consumption using the choice model (29) for the values of ζ in the table. The columns, labeled $X'\beta$ and $X'\gamma$, use (20) and (16), respectively, as the specification for the effect of student body characteristics on school quality. Numbers in parentheses are 95 percent confidence intervals.

2. Allowing for Nonlinear Utility of Consumption

Here we modify the probit specification to allow for a more general specification of private school preferences. Assume that parents' preferences are separable in consumption and the gain from attending public school. Let $c^{1-\zeta}/(1-\zeta)$ be the contribution of consumption c to utility and let $V_i^P(\tau)$ be the utility value the parents of student i place on public school relative to the best private school alternative, holding c constant. We allow these tastes to depend on τ, potentially through school composition effects at the public school or other factors. Let t_0 denote private school tuition and $t(\tau)$ be the voucher program τ. Then the parents choose c_i and P_i^τ to maximize

$$a_c \frac{c_i^{1-\zeta}}{1-\zeta} + V_i^P(\tau)P_i^\tau$$

subject to

$$c_i + [t_0 - t(\tau)](1 - P_i^\tau) \leq I_i.$$

version of the base specification of the choice model. The point estimates $\pi^p(\tau)$ are generally consistent with the results that we obtained using a binomial choice model for the cases that we considered. However, we experienced numerical difficulties in estimating the trinomial choice model, which precluded the use of bootstrap methods to compute confidence intervals. The model with unobservable student body effects and unobserved school choice components that are specific to s and the model in which student body quality influences demand both generalize to the trinomial case in a natural way. However, we have not estimated the trinomial versions for computational reasons.

Using the budget constraint and the objective function, parents choose public school if

$$a_c\left(\frac{I_i^{1-\zeta}}{1-\zeta} - \frac{\{I_i - [t_0 - t(\tau)]\}^{1-\zeta}}{1-\zeta}\right) + V_i^P(\tau) \geq 0.$$

We parameterize the taste for public school relative to private school as

$$V_i^P(\tau) = X_i'\beta + Q_{S_i}'\beta_Q + u_i.$$

Since this is a static problem, one should not think of ζ as the intertemporal elasticity of consumption. Here it dictates the magnitude of the income effect on private schools. When $\zeta = 0$, there is no income effect; but the larger ζ is, the larger the income effect becomes.

Thus parents send their kids to public school ($P_i^\tau = 1$) if

$$a_c\left(\frac{I_i^{1-\zeta}}{1-\zeta} - \frac{\{I_i - [t_0 - t(\tau)]\}^{1-\zeta}}{1-\zeta}\right) + X_i'\beta + Q_{S_i}'\beta_Q + u_i \geq 0. \tag{29}$$

Note that when $\zeta = 0$, this reduces to

$$a_c[t_0 - t(\tau)] + X_i'\beta + Q_{S_i}'\beta_Q + u_i \geq 0.$$

Since $a_c t_0$ does not vary across individuals, it can be incorporated into the intercept. The parameter a_c in front of $t(\tau)$ makes it clear why we cannot identify the dollar level of tuition in the probit model.

We use the linear utility case ($\zeta = 0$) as the base case in Section IV for three reasons.[41] First, it is convenient because we do not need to worry about t_0. Second, we experimented with estimation of ζ using variation in I_i as the source of identification and find that a value very close to zero fits the data best.[42] Third, allowing $\zeta > 0$ would, if anything, reinforce our conclusion that the cream skimming effect is small. Why? In general, as ζ increases, the effect of the voucher on school choice is more sensitive to income. As it becomes more and more sensitive to in-

[41] We do include log income in the base model as a control variable to account for correlation between taste for private schooling and income. Epple et al. (2004) provide evidence that when one controls for private school fixed effects, higher-income families pay higher tuition net of aid. Consequently, the income coefficient may also be picking up price variation.

[42] Very strong assumptions are required. Dynarski, Gruber, and Li (2009) use tuition discounts for number of children to estimate the price elasticity of demand for Catholic primary schools. Their results also suggest that lower–socioeconomic status families are more responsive to the tuition. This supports our conclusion that the use of linear utility in the base model tends to overestimate the cream skimming effect.

come, the composition of peers leaving the school shifts toward low-income students. Since low-income students tend to be disadvantaged in many dimensions, this should tend to reduce cream skimming. Thus, by focusing on the linear case ($\zeta = 0$), we estimate the upper bounds of the effect—something we verify in this section.

The key difference between (29) and the base case probit specification is the interaction between the subsidy and income. In the base case model, log family income enters X_i as a regressor, but it does not interact with the subsidy any more than any other variable. In the more general model, the derivative of the utility of public relative to private school with respect to the subsidy (i.e., the left-hand side of eq. [29]) is

$$- a_c\{I_i - [t_0 - t(\tau)]\}^{-\zeta},$$

which is negative and larger in absolute value for lower-income individuals when $\zeta > 0$. The expanded model captures the idea that lower-income families are more price sensitive than high-income families.

We test the sensitivity of the results to alternative values of ζ. To reduce the influence of very low reports of family income, we assume that the government provides an income floor and set I_i to the maximum of reported family income (in 1988 dollars) and \$10,000. This compares to the 10th percentile value of \$8,750 for cases that provide income information. We set t_0 to \$4,000, which is very close to the average value of maximum tuition reported by the high schools attended by private school students in NELS.[43] As above, when we estimate the model, we set $t(0) = 0$ and we set $t(\tau)$ to a value that is sufficient to induce 10 percent of the public school students to switch to private schools.

Table 7 reports values of $\pi(\tau)$ for the case $\zeta = 0.01$, which is essentially linear utility, for $\zeta = 0.5$ and for $\zeta = 1$, which is log utility.[44] The most negative values are obtained when $\zeta = 0.01$, but the values are virtually identical to what we obtain in our base specification, which simply includes the log of family income as a control. In the log utility case, there

[43] Schools also report the fraction of students who receive aid and the percentage of students who receive financial aid or scholarship support. One can obtain a lower bound on net tuition by multiplying the maximum tuition figure by one minus the fraction who receive aid or scholarship support. The mean of this net tuition bound is \$2,804. We did not attempt to use the private school tuition data to estimate ζ and a_c because private school tuition is not available for those who choose public school and because the tuition level may be related to quality.

[44] Note that ζ is identified by the utility specification and variation in I_i. Income is an important determinant of school choice in the sense that the likelihood of the probit model is considerably higher at the MLE value of $\zeta = 0.001$ than when ζ is set to zero. We used $\zeta = 0.01$ as the lowest value we consider in table 7 rather than the MLE value for two reasons. The first is the computational difficulty of bootstrapping while estimating ζ. The second is that the estimates of the cream skimming effect are essentially the same whether we use $\zeta = 0.001$ or $\zeta = 0.01$, as one might expect.

is essentially no cream skimming. As mentioned above, this is not surprising. To see the intuition, note that the larger ζ is, the more rapidly the marginal utility of consumption is decreasing. The tuition subsidy will be worth less to richer families in utility terms. Consequently, higher values of ζ have the effect of lowering $\psi(\tau, \chi_i)$ for poorer families relative to richer families, leading to less cream skimming. Thus, the restriction $\zeta = 0$ in the base case would appear to overstate the negative impact of cream skimming.

3. Cream Skimming with Migration and Political Economy Effects

We have treated location choice as exogenous and have assumed away migration effects. These are prominent in several of the papers mentioned in the introduction that analyze voucher programs using calibrated general equilibrium models. Allowing for migration effects would be a very interesting and important extension to our approach. In the extended framework, parents would simultaneously choose a school district from the set in the region and decide whether to attend public school. They may respond to a voucher program by changing location and/or by switching between the private and public school sectors, as in Epple and Romano (1998), Nechyba (1999, 2000), and Ferreyra (2007). Consequently, both S_i and P_i depend on τ.

While we do not explicitly estimate a model of migration choice, we discuss how our methodology could be extended to such a setting. Denoting S_i^τ and S_i^0 as the district choice with and without the voucher program τ, the cream skimming effect is

$$\pi^p(\tau) \equiv [E(\overline{Z}(S_i^\tau, \tau)|P_i^\tau = P_i^0 = 1) - E(\overline{Z}(S_i^0, 0)|P_i^\tau = P_i^0 = 1)]'\delta. \quad (30)$$

A complication is that P_i^0 and S_i^0 are observed, but P_i^τ and S_i^τ are not. It is trivial to show that our argument in Section II.C extends to this case. One can take every equation in that section and replace $\overline{Z}(S_i, 0)$ with $\overline{Z}(S_i^0, 0)$ and $\overline{Z}(S_i, \tau)$ with $\overline{Z}(S_i^\tau, \tau)$ and the arguments go through exactly. That is, with location choice,

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

Calculating $\overline{Z}(S_i^\tau, \tau)$ and $\psi(\tau, \chi_i)$ would presumably be much harder as the model is more complicated, but the formula still holds.

Furthermore, if monotonicity holds and we observe a subsample of individuals at a school, we can continue to use

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

In this case only $\psi(\tau, \chi_i)$ is more complicated to estimate. (By monotonicity we continue to mean that $P_i^\tau = 1$ if $P_i^0 = 1$.)

Extending the empirical analysis to endogenize location would be a major undertaking. We leave it for future research.[45] Would endogenizing location choice change our conclusions about the cream skimming effect? Any consideration of this issue is necessarily speculative in the absence of the extended model. However, there is reason to believe that our finding that the cream skimming effect is small is likely to hold, at least for a voucher that induces only a modest fraction of students to leave public school. Consider the monotonicity case above. Under monotonicity, all that would be different from our current model in terms of determining the size of the cream skimming effect is $\psi(\tau, \chi_i)$. Recall that the cream skimming effect depends on the composition of students who are induced to attend private schools by the expansion of vouchers. In particular, it does not matter where those people lived prior to the reform, where they lived after the reform, or where any public school stayers move in response to the reform. We see no ex ante reason to expect that adding migration to the model would lead the take-up of the voucher program to depend much more strongly on student characteristics that have large spillovers. Of course monotonicity likely does not hold exactly, and this result depends on our linear-in-expectations assumption about peer effects, but it seems unlikely to us that relaxing these restrictions would fundamentally alter the basic findings.

To be more specific, consider a district-level voucher program. Presumably when a district introduces a voucher program, the first-order effect of migration will involve families moving into the district to take advantage of the voucher and send their children to private school. While this first-order effect could influence the finances of the district or have a competition effect on the behavior of public schools, it will not directly change the composition of the public school students in the district. However, by increasing the demand for housing in the district, it could lead some public school families to leave the district. Presumably these would be lower-income families, which would reduce cream skimming and strengthen the case that the cream skimming effect is small. On the other hand, Nechyba's (2000) analysis suggests that if the voucher is targeted to a low-income district with heterogeneous housing, the families moving into that district would choose the higher-income neighbor-

[45] The first complication is that the school choice function (S_i^τ) involves school district characteristics, and school district choice is endogenous. The distribution of district characteristics within a region would affect both district choice and the public/private decision. As a starting point for the analysis, one might assume that school choice and district choice depend only on observed characteristics of the student, the district, and the districts in each region. The second complication is that the school choice model underlying the ψ function would be quite complex. One could start with (14) as the net utility gain from attending public school rather than private school conditional on residence in S_i and combine it with a specification for the utility of choosing district S_i and attending public school. Allowing both the choice of district and the choice of public school conditional on district would substantially complicate the analysis in Sec. V.

hoods, crowding out families with incomes that are high relative to the district. This would boost the cream skimming effect.

Migration effects also involve people moving into or out of the district in response to changes in public school quality. The change in public school quality is ambiguous. Competition effects would make public schools better, cream skimming would probably make them worse, and financial considerations could work in either direction. If the schools improved, presumably the parents who move in would be of higher socioeconomic background, which would mitigate the cream skimming effect. If public schools became worse, presumably some students might leave, but with migration there are two places to go. Either families could switch school districts and send their child to a different public school or they could take advantage of the new voucher system and send their child to private school. We suspect that the second avenue would be more popular. However, movement from public school to private school is not migration: it is the source of the effect that we are trying to estimate.

Much of the same logic applies to a state-level or national voucher program that is universal. Although the voucher could be used in any district, families with students in public school prior to the voucher would presumably be more likely to stay in the district and move to a private school than to leave the district and move to a private school. Nechyba's (2000) analysis does suggest strongly that the fraction of families who respond to a voucher of a given size is larger with migration, but it is less clear that migration alters the relative response of families with favorable peer characteristics. The relative response is crucial in determining the size of the cream skimming effect.

Of course, if the national voucher program fundamentally changed the way public schooling is provided and financed, one should be very cautious about what the migration effects might be. Consequently, we cannot rule out a large cream skimming effect for a voucher program that is big enough to induce a large fraction of public school students to change sectors.

B. Alternative School Composition Models

In our base case we assume that student body characteristics influence high school graduation through either the $X'\gamma$ or the $X'\beta$ index. In this subsection we explore the sensitivity to allowing these effects to operate through more specific channels. 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. The top row of table 8 reports results based on estimation of (2) 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.

TABLE 8
Estimates of the Cream Skimming Effect $\pi^p(\tau)$ of Vouchers on Public School Stayers under Alternative Assumptions about the Effects of Student Body Composition on School Quality

Assumption about How School Composition Affects School Quality	Estimate
Composition effects operate only through family income	−.0011 (−.0027, .0007)
Composition effects operate only through African American	−.0002 (−.0005, .0001)
Composition effects operate only through $X'\beta$ test score index	−.0005 (−.0018, .0005)
Composition effects operate only through $X'\gamma$ test score index	−.0008 (−.0015, .0001)
Composition effects operate only through father's education	−.0008 (−.0021, .0000)

Note.—The table reports estimates of $\pi^p(\tau)$ using the base case school choice model to estimate $\psi(\tau, \chi_i)$ and the assumption about student body effect $Z'\delta$ indicated in the row. In the case of test scores in row 3, the test score coefficients from the school choice model are used to form the test score index. In row 4 the elements of γ corresponding to the test scores are used to form the index. Numbers in parentheses are 95 percent confidence intervals.

The next row of table 8 is based on restricting Z_i to consist of only the fraction African American. For this specification the estimate of the cream skimming effect is essentially zero. Note that from table 4, for public school stayers, the fraction of peers who are African American increases by only 0.0035.[46]

The next two rows of the table report estimates under the assumption that student body effects are a linear combination of average test scores. In the $X'\beta$ index row, we use test score coefficients from the school choice equation to form the index. In the $X'\gamma$ index row, the test score coefficients are from the outcome equation. Both of the estimates of π are closer to zero than the base case. Finally, in the last row, student body effects operate only through father's education. The estimate of the cream skimming effect is only -0.0008.

VIII. Alternative Voucher Programs

To this point we have focused on simulating the effects of universal vouchers. We did so in part because we believe that cream skimming effects would be largest in universal programs. However, most of the voucher programs in the United States are targeted to either low-income students or students from low-performing schools. They are also more common in urban areas. In this section we discuss some alternative voucher programs.

[46] This number is small in part because most white public school students have few African American schoolmates. The average number of black peers is 0.07 for whites while it is 0.49 for blacks.

A. *A Voucher That Draws 30 Percent of Public School Students*

Our base case model draws 10 percent of students from public schools. We consider a program three times as large. The cream skimming effect for this program is -0.0039, roughly triple the value for the 10 percent program (table 9). In thinking about the overall cost-benefit analysis, we get a similar overall effect. In our base case, for every one student who leaves, nine receive the cream skimming effect. Under the 30 percent program, for every student who moves, $7/3 = 2.33$ receive the cream skimming effect. So in this case, 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 $0.06 - 0.0039 \times 2.33 = 0.051$. This value is slightly higher than the estimate of 0.048 that we find in the base case. However, one should be very cautious about extrapolating our results to such a large program because changes in location decisions and large-scale entry of private schools might change the nature of demand.

B. *Programs Targeted to Urban Students or to Low-Income Students*

We consider the cream skimming effect of vouchers targeted to urban families after reestimating our base model on the urban sample.[47] The point estimates of the change in peer characteristics of stayers under an urban voucher program are somewhat higher: -0.0022 for both cases (table 9, row 2). However, the estimates are noisy for the $X'\beta$ index case, with a lower-bound estimate of -0.0057, which is substantial.

Next we consider a program that limits eligibility to families whose incomes are in the lowest 20 percent of our sample. For this case we do not reestimate the model but use our base case, and we again calculate the value of the voucher that would induce 10 percent 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 for the $X'\gamma$ index is only 0.0001.[48]

[47] We reestimate for this sample for two reasons. First, we can do it easily. Since everyone in urban schools has peers in urban schools, we can easily condition on it. Second, one might expect selection into private school to be different in urban areas than in other environments.

[48] 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 percent of the eligible population. This is only 2 percent of the full population. In this case, if one wishes to compare the private gain in the high school graduation rate 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

TABLE 9
The Cream Skimming Effect $\pi^p(\tau)$ of Alternative Voucher Progams on Public Schools Stayers

	$X'\beta$ Index	$X'\gamma$ Index
30% of public school students move	−.0032	−.0039
	(−.0077, .0006)	(−.0066, −.0004)
Vouchers targeted to urban districts	−.0022	−.0022
	(−.0057, .0003)	(−.0040, .0007)
Vouchers targeted to low-income families	.0003	.0001
	(−.0000, .0006)	(−.0001, .0005)
Vouchers targeted to low-income neighborhoods	−.0018	−.0013
	(−.0100, .0051)	(−.0045, .0039)
Calibrated to Milwaukee	.0016	.0009
	(−.0017, .0056)	(−.0005, .0026)

Note.—See the text for details. Numbers in parentheses are 95 percent confidence intervals.

Programs that target all students in low-income school districts regardless of income would have a different selection effect. We investigate this by reestimating the base models on the subsample of schools in zip codes with poverty rates above 16 percent, which is about 25 percent of the full sample. When student body effects depend on the outcome index $X'\gamma$, the point estimate is very close to the base case (-0.0013), but the confidence interval is considerably wider. When student body effects depend on the school choice index $X'\beta$, the point estimate is -0.0018 but is imprecise.[49]

C. *The Milwaukee Parental Choice Voucher Program*

We also loosely consider a voucher program modeled on the initial implementation of the publicly funded Milwaukee Parental Choice Voucher Program. That program restricted eligibility to students from households with income less than or equal to 1.75 times the poverty line who were not already in private school. The data are not rich enough to estimate all aspects of our model (in particular, δ), but we can use it as an example of how one could use evidence from a specific voucher program to form $\psi(\tau, \chi_i)$.

We do this by estimating a school choice model with data from the Milwaukee voucher experience in the early 1990s. For details, see Witte (1992), Witte and Thorn (1995), and online appendix 1. The data set

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.

[49] See table 9, row 4. The estimates are somewhat sensitive to the poverty rate cutoff we choose, which is to be expected given the sampling error.

consists of information on those who took up the voucher and a control sample. 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, gender, race, a dummy variable for married, father's and mother's education, a dummy for whether father's education is observable, the log of family income, and reading and math test scores (standardized). We should emphasize that the variables are "similar" across data sets, not identical, so this should be viewed as more of a sensitivity analysis than a formal evaluation of the Milwaukee choice program. We estimate the choice parameters (β) using the data from Milwaukee only but then estimate the outcome equation and student body effects from the urban subsample of NELS:88. The values of β will reflect both preferences and the relationship between X and voucher eligibility.

Since the sample is choice based, we use a logit model rather than a probit. This gives consistent estimates of the slope coefficients but not the intercept. The program was in early stages during the period in which we have data and involved only a small number of students but grew to roughly 20 percent of the Milwaukee student population by 2010. Furthermore, we are using all students, not just high school students. Thus we do not want to use the data to estimate the size of the program. Instead, we choose the intercept to move 10 percent of students to make this thought experiment comparable to our base case. Given the logit specification, the choice of intercept does not alter the relative probabilities that students choose private school. Let $\Lambda(X_i'\beta)$ be the estimated probability of being in a public school as opposed to a voucher school in Milwaukee. We then estimate the effect of this "voucher program" on NELS public school students by assuming that if it was implemented, a student's probability of remaining in public school would be $\Lambda(X_i'\beta)$. In practice, we apply our formula to the NELS data with a more limited set of X_i covariates and weight using

$$\hat{\psi}(\tau,\ \chi_i) = \frac{\Lambda(X_i'\hat{\beta})}{1/N_{P^0}\sum_{j,P_j^0=1}\Lambda(X_j'\hat{\beta})}.$$

We obtain a small positive estimate of the cream skimming effect for both student body effect specifications (table 9, bottom row). The difference between the small positive estimate and the small negative estimate for our base case model reflects the targeting of the program, differences in school choice preferences between Milwaukee and the rest of the country, the fact that most of the Milwaukee program refers to elementary schools rather than high schools, as well as differences in the variables used.[50]

[50] Studies of the Parental Choice Voucher Program as well as a privately funded voucher program in Milwaukee show positive selection into private schools conditional on income

IX. The Cream Skimming Effect under Extreme Assumptions about School Choice and Student Body Effects

To gain a deeper understanding of what is driving our small effects, in this section we explore what it might take to find large effects. Specifically, we make extreme assumptions about the amount of cream skimming and the size of student body effects to see how large our estimates can become. First, we simulate a universal voucher program in which sorting into new voucher schools is determined exclusively by $X_i'\gamma$. For the $X_i'\gamma$ index model (16), this is the largest amount of cream skimming that would be possible. To accomplish this, we rank public school students by the outcome index $X_i'\hat{\gamma}$ and assume that the top 10 percent will move. Under this assumption, the average value of $X_i'\hat{\gamma}$ is 0.57 for movers and 0.38 for stayers. The average change in $\overline{X}(s, \tau)'\hat{\gamma}$ for stayers is -0.0163. Using (16) as the student body effects specification, we obtain $-0.0060 = 0.3680 \times (-0.0163)$ 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 student body effects is extreme. To see this point, contrast these results with the final row of table 4. When we freely estimate the school choice model, the mean of $X_i'\gamma$ for movers is only 0.4437 and the change in $\overline{X}(s, \tau)'\gamma$ for stayers is only -0.0037. Furthermore, prior to the voucher, the difference in the mean of $X'\gamma$ for public and private school students is only 0.0743. Clearly the decision to attend public school depends on a lot more than just $X_i'\gamma$. The calculation does establish that with much more cream skimming, one could obtain larger cream skimming effects.

Second, we make an extreme assumption about student body effects on the graduation rate. As noted earlier, our point estimate of $\delta_{X'\gamma}$ is 0.3680 when we exclude unobservable student body effects (table 2, col. 2, row 1). If we set $\delta_{X'\gamma}$ to one, almost three times the point estimate, the cream skimming effect is -0.0037, with a confidence interval from -0.0051 to -0.0023 (table 10, row 3). The decline of -0.0037 in the graduation rate is relatively small in an absolute sense, but it is large enough to cancel out more than half of the direct positive effect of moving one out of 10 students to private schools, assuming a private school effect of 0.06. However, a value of one for $\delta_{X'\gamma}$ says that the external effect that $X_i'\gamma$ has on the graduation rate of a high school operating through $\overline{X}(s, \tau)'\gamma$ is as large as the direct effect of $X_i'\gamma$ on the graduation rate of person i. This seems extreme.

eligibility. Figlio et al.'s (2010) finding of negative selection into private school among households eligible for Florida's voucher program for low-income households, if anything, reinforces our conclusion of a positive cream skimming effect for a voucher program targeted to low-income households.

TABLE 10
ESTIMATES OF THE CREAM SKIMMING EFFECT $\pi^p(\tau)$ OF A VOUCHER PROGRAM ON PUBLIC SCHOOL STUDENTS UNDER EXTREME ASSUMPTIONS

	$\overline{X}'\gamma$ Index
Base case	−.0014 (−.0023, −.0001)
School choice determined entirely by $X'\gamma$	−.0060 (−.0104, −.0005)
Student body effect equals the direct effect ($\delta_{X'\gamma} = 1$)	−.0037 (−.0051, −.0023)
No sorting in public schools	−.0017 (−.0029, −.0001)
$\delta_{X'\gamma} = 1$ and choice determined entirely by $X'\gamma$	−.0163 (−.0182, −.0145)
$\delta_{X'\gamma} = 1$, choice determined entirely by $X'\gamma$, and no sorting	−.0189 (−.0213, −.0173)

NOTE.—The table reports estimates of $\pi^p(\tau)$ under extreme assumptions of the nature of cream skimming and student body effects. The first row reports the base case estimate of the cream skimming effect. The second row assumes that the top 10 percent of students ranked by $X_i'\gamma$ move to the public school and uses the base case specification (16) for $Z'\delta$. The third row uses the base case school choice model and assumes that the effect of the $\overline{X}'\gamma$ index on outcomes is the same as the direct effect of $X_i'\gamma$ on student i. The fourth row uses the base case for student body effects and school choice but assumes that there is no sorting in public schools prior to the voucher. The fifth row imposes the extreme school choice assumption from the second row and the extreme student body effects assumption from the third row. The last row imposes the extreme school choice assumption, the extreme student body effects assumption, and the no-sorting assumption from the middle row. Numbers in parentheses are 95 percent confidence intervals.

In our third case we compute $\pi^p(\tau)$ using the base case model estimates for the school choice and high school graduation equations but imposing the extreme assumption of no sorting across school districts. We use exactly the same parameters we use for the base model, but in calculating the cream skimming effect, we replace the formula

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

with

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

Here $\overline{X}_P$ is the mean value of X_i for all public school students, not just the peers of the student. This maximizes the third determinant of the size of the cream skimming effect—the degree of heterogeneity within schools. For the $X_i'\gamma$ index model, the magnitude of $\pi^p(\tau)$ rises from −0.0014 to −0.0017, with a confidence interval from −0.0029 to −0.0001. The increase is similar for the $X_i'\beta$ model: from −0.0011 to −0.0016 (not shown).

In the fifth row of table 10, we combine the assumption that sorting into voucher schools is determined exclusively by $X_i'\hat{\gamma}$ with the assumption that $\delta_{X'\gamma} = 1$. This makes two of the three factors in the cream skimming effect formula large: selection out of public into private school is chosen to be as large as possible and the student body group effect is set to a very high value. In this case the estimate of $\pi^p(\tau)$ is large: -0.0163. This value is more than 10 times the base case estimate. When we add the counterfactual assumption of no sorting across public schools to the other two extreme assumptions, the estimate of $\pi^p(\tau)$ is -0.0189 as documented in the final row of table 10.

We conclude that not one specific factor drives our small estimates, but rather the combination of all three factors.

X. 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 student body effects, the cream skimming effect is determined by the covariance of the relative probability that a student will stay in public 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 schoolmates. The weights of the index are the coefficients relating outcomes to student body 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 using alternative assumptions regarding school choice and regarding the effects of school composition on outcomes. We provide a method for allowing school choice to depend on public school student body characteristics even when these characteristics are not directly observed for those who choose a private school. We also provide a way to allow for both unobserved fixed characteristics of schools that influence school choice and for unobserved characteristics of schoolmates 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 a negative response to competitive pressure or through an effect on the financial resources available in public schools, 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 monotonicity, the only difference would come from the relative probability of remaining in the default school, $\psi(\tau, \chi_i)$. In order for the overall effect to be sizable, one would need the amount of cream skimming to be more severe than what we have simulated for the types of voucher programs we consider. 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. In his career James Heckman has pushed economists toward working on ex ante policy evaluation. He has also pushed empirical work in labor economics toward trying to make real progress on important social issues, not in just estimating abstract structural parameters. This paper responds to both of these initiatives. School choice programs are a very important social issue, and we know very little about the magnitude of the cream skimming effect. Our analysis will not be the last word on the subject. But by engaging in ex ante evaluation and by considering a variety of possible programs rather than evaluating a single program, we arrive at a provisional message for policy makers, at least for a modest program that does not lead to wholesale geographic migration and exit from public schools. When considering the costs and benefits of modest voucher programs, the potential costs associated with cream skimming effects appear to be small.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2002. “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools.” Manuscript, Northwestern Univ.

———. 2005. “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools.” *J.P.E.* 113 (1): 151–84.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. “A Unified Framework for Measuring Preferences for Schools and Neighborhoods.” *J.P.E.* 115 (4): 588–638.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. “Automobile Prices in Market Equilibrium.” *Econometrica* 63 (4): 841–90.

Bloom, Howard S. 1984. “Accounting for No-Shows in Experimental Evaluation Designs.” *Evaluation Design* 8 (2): 225–46.

Bulkley, Katrina, and Jennifer Fisler. 2002. “An Overview of the Research on Charter Schools.” Web Paper Series no. WP-01 (June), Consortium Policy Res. Educ., Philadelphia.

Card, David, Martin D. Dooley, and A. Abigail Payne. 2010. “School Competition and Efficiency with Publicly Funded Catholic Schools.” *American Econ. J.: Appl. Econ.* 2 (4): 150–76.

Caucutt, Elizabeth. 2002. “Educational Vouchers When There Are Peer Group Effects—Size Matters.” *Internat. Econ. Rev.* 43 (1): 195–222.

Chetty, Raj. 2009. “Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods.” *Ann. Rev. Econ.* 1 (June): 451–88.

Devereux, Paul. 2007. “Improved Errors-in-Variables Estimators for Grouped Data.” *J. Bus. and Econ. Statis.* 25 (3): 278–87.

Dynarski, Susan, Jonathan Gruber, and Danielle Li. 2009. “Cheaper by the Dozen: Using Sibling Discounts at Catholic Schools to Estimate the Price Elasticity of Private School Attendance.” Working Paper no. 15461 (October), NBER, Cambridge, MA.

Epple, Dennis, David Figlio, and Richard Romano. 2004. “Competition between Private and Public Schools: Testing Stratification and Pricing Predictions.” *J. Public Econ.* 88 (7): 1215–45.

Epple, Dennis, Elizabeth Newlon, and Richard Romano. 2002. “Ability Tracking, School Competition, and the Distribution of Educational Benefits.” *J. Public Econ.* 83 (1): 1–48.

Epple, Dennis, and Richard Romano. 1998. “Competition between Private and Public Schools, Vouchers, and Peer-Group Effects.” *A.E.R.* 88 (1): 33–62.

———. 2002. “Educational Vouchers and Cream Skimming.” Working Paper no. 9354 (November), NBER, Cambridge, MA.

———. 2003. “Neighborhood Schools, Choice, and the Distribution of Educational Benefits.” In *The Economics of School Choice*, edited by Caroline M. Hoxby, 227–86. Chicago: Univ. Chicago Press.

Epple, Dennis, Richard Romano, and Holger Sieg. 2006. “Admission, Tuition, and Financial Aid Policies in the Market for Higher Education.” *Econometrica* 74 (4): 885–928.

Evans, William N., and Robert M. Schwab. 1995. “Finishing High School and Starting College: Do Catholic Schools Make a Difference?” *Q.J.E.* 110 (4): 941–74.

Ferreyra, Maria Marta. 2007. “Estimating the Effects of Private School Vouchers in Multi-district Economies.” *A.E.R.* 97 (3): 789–817.

———. 2009. “An Empirical Framework for Large-Scale Policy Analysis, with an Application to School Finance Reform in Michigan.” *American Econ. J.: Econ. Policy* 1 (1): 147–80.

Figlio, David N., Cassandra M. D. Hart, and Molly Metzger. 2010. “Who Uses a Means-Tested Scholarship, and What Do They Choose?” *Econ. Educ. Rev.* 29 (2): 301–17.

Figlio, David N., and Joe A. Stone. 2001. “Can Public Policy Affect Private School Cream Skimming?” *J. Urban Econ.* 49 (2): 240–66.

Grogger, Jeff, and Derek A. Neal. 2000. “Further Evidence on the Effects of Catholic Secondary Schooling.” *Brookings-Wharton Papers Urban Affairs*, 151–201.

Haggerty, Catherine, Bernard Dugoni, Laura Reed, Ann Cederlund, and John Taylor. 1996. *National Education Longitudinal Study: 1988–1994 Methodology Report.* NCES 96-174. Washington, DC: US Dept. Education, Office Educational Res. and Improvement.

Heckman, James J. 1974. “Shadow Prices, Market Wages, and Labor Supply.” *Econometrica* 42 (4): 679–94.

———. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* 47 (1): 153–61.

———. 1990. “Varieties of Selection Bias.” *A.E.R.* 80 (2): 313–18.

———. 1997. “Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations.” *J. Human Resources* 32 (3): 441–62.

———. 2001. “Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture.” *J.P.E.* 109 (4): 673–748.

Heckman, James J., Jeffrey Smith, and Christopher Taber. 1998. "Accounting for Dropouts in Evaluations of Social Programs." *Rev. Econ. and Statis.* 80 (1): 1–14.
Heckman, James J., and Edward Vytlacil. 2001. "Policy Relevant Treatment Effects." *A.E.R. Papers and Proc.* 91 (2): 107–11.
———. 2005. "Structural Equations, Treatment Effects and Econometric Policy Evaluation." *Econometrica* 73 (3): 669–738.
———. 2007a. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*, vol. 6B, edited by James J. Heckman and Edward E. Leamer, 4779–4874. Amsterdam: North-Holland.
———. 2007b. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." In *Handbook of Econometrics*, vol. 6B, edited by James J. Heckman and Edward E. Leamer, 4875–5143. Amsterdam: North-Holland.
Howell, William, and Paul Peterson. 2002. *The Education Gap.* Washington, DC: Brookings Inst. Press.
Hoxby, Caroline M. 2003. "School Choice and School Productivity: Could School Choice Be a Tide That Lifts All Boats?" In *The Economics of School Choice*, edited by Caroline M. Hoxby, 287–342. Chicago: Univ. Chicago Press.
Hsieh, Chang-Tai, and Miguel Urquiola. 2006. "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program." *J. Public Econ.* 90 (8–9): 1477–1503.
Krueger, Alan, and Pei Zhu. 2004. "Another Look at the New York City School Voucher Experiment." *American Behavioral Scientist* 47 (5): 658–98.
Lacireno-Paquet, Natalie, Thomas T. Holyoke, and Michele Moser. 2002. "Creaming versus Cropping: Charter School Enrollment Practices in Response to Market Incentives." *Educ. Evaluation and Policy Analysis* 24 (2): 145–58.
Ladd, Helen F. 2002. "School Vouchers: A Critical View." *J. Econ. Perspectives* 16 (4): 3–24.
MacLeod, Bentley W., and Miguel Urquiola. 2013. "Anti-lemons: Reputation and Educational Quality." Manuscript, Dept. Econ., Columbia Univ.
Manski, Charles F. 1992. "Educational Choice (Vouchers) and Social Mobility." *Econ. Educ. Rev.* 1 (4): 351–69.
———. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Rev. Econ. Studies* 60 (3): 531–42.
Marschak, Jacob. 1953. "Economic Measurements for Policy and Prediction." In *Studies in Econometric Method*, edited by William C. Hood and Tjalling C. Koopmans, 1–26. New York: Wiley.
Neal, Derek. 1997. "The Effects of Catholic Secondary Schooling on Educational Attainment." *J. Labor Econ.* 15 (1): 98–123.
———. 2002. "How Vouchers Could Change the Market for Education." *J. Econ. Perspectives* 16 (4): 24–44.
Nechyba, Thomas J. 1999. "School Finance Induced Migration Patterns: The Impact of Private School Vouchers." *J. Public Econ. Theory* 1 (1): 5–50.
———. 2000. "Mobility, Targeting and Private School Vouchers." *A.E.R.* 90 (1): 130–46.
———. 2003. "Introducing School Choice into Multidistrict Public School Systems." In *The Economics of School Choice*, edited by Caroline M. Hoxby, 145–94. Chicago: Univ. Chicago Press.
Rouse, Cecilia E., and Lisa Barrow. 2009. "School Vouchers and Student Achievement: Recent Evidence and Remaining Questions." *Ann. Rev. Econ.* 1:17–42.

RPP International. 2001. *Challenge and Opportunity: The Impact of Charter Schools on School Districts.* Washington, DC: Office Educ. Res. and Improvement, US Dept. Educ.

Spencer, Bruce D., Martin R. Frankel, Steven J. Ingels, Kenneth A. Rasinski, and Roger Tourangeau. 1990. *National Education Longitudinal Study of 1988 Base Year Sample Design Report.* NCES 90-463. Washington, DC: US Dept. Education, Office Educational Res. and Improvement.

Witte, John F. 1992. "Private School versus Public School Achievement: Are There Findings That Should Affect the Educational Choice Debate?" *Econ. Educ. Rev.* 11 (4): 371–94.

Witte, John F., and Christopher A. Thorn. 1995. "The Milwaukee Parental Choice Program, 1990/1991–1994/1995." Data and Information Services Center, Madison, WI. http://www.disc.wisc.edu/archive/choice/.

Zimmer, Ron, Brian Gill, Kevin Booker, Stéphane Lavertu, and John Witte. 2009. "Do Charter Schools 'Cream Skim' Students and Increase Racial-Ethnic Segregation?" Paper presented at "School Choice and School Improvement: Research in State, District and Community Contexts," Vanderbilt Univ., October 25–27. http://www.vanderbilt.edu/schoolchoice/conference/papers/Zimmer_COMPLETE.pdf.