Appendix 1: Data

A1.1 Sample Sizes By High School and Eighth Grade

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

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

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

A1.2 Description of Variables

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

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

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

Family Background Measures:

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

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

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

Geographic Variables:

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

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

Fraction black (*p008002*/*p001001*), fraction Hispanic (p0100001/p0010001), an indicator for whether fraction black is missing, median income, the fraction of the population below the poverty line ((*p1210001*+*p1210002*+*p121000* and the fraction of the population with income more than double the poverty line (*p1210009*/*p0010001*) are from the 1990 Census for the zip code of the high school. Missing values are set to the sample mean.

Eighth Grade Test Score Measures:

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

Eighth Grade Behavioral and Performance-in-School Measures:

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

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

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

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

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

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

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

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

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

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

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

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

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

Lack of Effort index: The base year student variable *bys75* measures “How many days of school did you miss over the past four weeks? The values are 0 (none) 1 (1 to 2) 2 (3 or 4) 3 (5 to 10) 4 (more than 10). *bys76* measures “How often do you cut or skip classes?” 0 (0) 1 (< once per week), 2 (at least once per week), 3 (daily). *bys77* is the response to “how many times were you late for school over the past four weeks?”: 0

(0), 1 (1 or 2 days) 2 (3 or 4) 3 (5-10) 4 (more than 10). *bys78a, bys78b and bys78c* are responses to "How often do you come to class without pencil or paper when needed?", "How often do you come to class without books", and "How often do you come to class without homework. Each is coded as 3 (usually), 2 (often), 1 (seldom), 0 (never).The index is $[bys75 + bys76 + bys77 + (4 - bys78a) + (4 - bys78b) + (4 - bys78c)]$ and ranges from 0 to 20.

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

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

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

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

Outcome Measures:

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

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

Log earnings in 1999: The log of the response to the question "First, including all of the wages, salaries, and commissions you earning in 1999, about how much did you earn from *employment* before taxes and other deductions?"(*f4hi99*). The sample is restricted to individuals with positive earnings when considering this outcome measure.

12th grade Math Test Score and 12th grade reading test score: We use the item response theory scaled scores). Missing value treatment: Observations with missing values on an outcome are excluded from the equation for that outcome.

Milwaukee School Data

We obtain our data from an online data archive (http://www.disc.wisc.edu/archive/choice/index.html) from Witte and Thorn (1995). We used all waves of the Choice sample, but only wave 1991 of the Control sample since that is the only year that the survey data were available. We used only observations for which we had all variables (except for father's education, for which we also included an indicator for missing values). We deleted records for individuals who applied to the choice program but did not enroll because we had very few of those observations after deleting individuals with missing data.

Appendix 2: Formal Justification for Linear Index Assumptions

The $X'\gamma$ index Model

As we state in the text, we use the $X'\gamma$ index restriction because we find it intuitively appealing and because the parameter $\delta_{X'\gamma}$ is easy to interpret. However, it can be formally justified by assuming that student body effects take the form of a pure "endogenous effect model" in the sense of Manski (1993). In our notation, when $S_i = s$,

$$\begin{aligned} Y_i(\tau) =& X_i'\gamma + \delta_y \overline{Y}(S_i,\tau) + Q_{s_i}'\Theta_Q^* + \xi_{s_i}^* + \varepsilon_i \\ =& X_i'\gamma + \delta_{X'\gamma}\left(\overline{X}(S_i,\tau)'\gamma + \overline{\varepsilon}(s,\tau)\right) + Q_{s_i}'\Theta_Q + \xi_{S_i} + \varepsilon_i \end{aligned}$$

where

$$\begin{aligned} \delta_{X'\gamma} =& \frac{\delta_y}{1-\delta_y} \\ \Theta_Q =& \frac{\Theta_Q^*}{1-\delta_y} \\ \xi_s =& \frac{\xi_s^*}{1-\delta_y}. \end{aligned}$$

This model is analogous to the model with unobservables in Section 6 of the paper described by equation (25) with g=1.

In the simpler version of the model (16) in which student body effects are assumed to only involve observed student characteristics X_i, the $X'\gamma$ index restriction can be justified if in addition school choice does not depend upon unobservables that affect Y_i. That is, if u_i is independent of ε_i (and maintaining the assumptions that X_i and Q_{S_i} are independent of u_i and ϵ_i) then

$$\begin{aligned} \overline{\varepsilon}(s,\tau) =& E\left(\varepsilon_i \mid S_i = s, X_i'\beta + Q_{S_i}'\beta_Q + t(\tau) + u_i \geq 0\right) \\ =& E\left(\varepsilon_i \mid S_i = s\right). \end{aligned}$$

Since $\overline{\varepsilon}(s,\tau)$ does not depend on τ it can be incorporated into ξ_s. Thus we can treat the model as

$$Y_i(\tau) = X_i'\gamma + \delta_{x'\gamma}\overline{X}(s,\tau)'\gamma + Q_s'\Theta_Q + \widetilde{\xi}_s + \varepsilon_i$$

where $\widetilde{\xi}_s = \xi_s + \delta_{x'\gamma}E(\varepsilon_i \mid S_i = s)$.

The $X'\beta$ Index Model

For a given school s, consider a nonparametric regression of $Z_i'\delta$ on $X_i'\beta$

$$Z_i'\delta = d_s(X_i'\beta) + e_{Zi},$$

where $d_s(X_i'\beta) = E(Z_i'\delta \mid X_i'\beta, S_i = s)$ so $E(e_{Zi} \mid X_i'\beta, S_i = s) = 0$. A first order approximation to this is

$$d_s(X_i'\beta) \approx c_s + \delta_{X\beta}X_i'\beta.$$

Assume further that the coefficient $\delta_{X\beta s}$ does not vary across s.[53] Then one obtains

$$d_s(X_i'\beta) \approx c_s + \delta_{X\beta}X_i'\beta.$$

Now recall that

$$\begin{aligned}\pi^p(\tau) &= E(\psi(\tau,\chi_i)[Z_i - \overline{Z}(S_i,0)]'\delta | P_i^0 = 1) \\ &= E(\psi(\tau,\chi_i)Z_i{}'\delta | P_i^0 = 1) - E(\psi(\tau,\chi_i)\overline{Z}(S_i,0)'\delta | P_i^0 = 1) \qquad \text{(A2-1)}\end{aligned}$$

We can write

$$\begin{aligned}E(\psi(\tau,\chi_i)Z_i{}'\delta | P_i^0 = 1) &= E\left(E\left[\psi(\tau,\chi_i)Z_i{}'\delta | X_i'\beta, S_i, X_i'\beta + Q_{S_i}'\beta_Q + t(\tau) + u_i \geq 0\right] \mid P_i^0 = 1\right) \\ &= E\left(\psi(\tau,\chi_i)d_{S_i}(X_i'\beta) \mid P_i^0 = 1\right). \\ &\approx E\left(\psi(\tau,\chi_i)c_{S_i}\right) \mid P_i^0 = 1\right) + \delta_{X\beta}E\left(\psi(\tau,\chi_i)X_i'\beta \mid P_i^0 = 1\right),\end{aligned}$$

where the first equality comes from the law of iterated expectations. The second equality uses the definition of d_s and the fact that conditional on $X_i'\beta$ and S_i, neither Q_{S_i} nor u_i have predictive power. The third applies the approximation.

[53] Allowing for a random coefficient that is orthogonal to other aspects of the model would be straight forward. If it was not orthogonal, it would be more difficult. This seems feasible, but we have not worked out the details.

For the second term of (A2-1) first note that

$$\begin{aligned}
\overline{Z}(s,0)'\delta =& E\{Z_i'\delta | P_i^0 = 1, S_i = s\}\\
=& E\{E\left(Z_i'\delta \mid X_i'\beta, P_i^0 = 1, S_i = s\right) | P_i^0 = 1, S_i = s\}\\
=& E\{d_s\left(X_i'\beta\right) | P_i^0 = 1, S_i = s\}\\
\approx& E\{c_s | P_i^0 = 1, S_i = s\} + \delta_{X\beta} E\{X_i'\beta | P_i^0 = 1, S_i = s\}\\
=& c_s + \delta_{X\beta}\overline{X}(S_i,0)'\beta.
\end{aligned}$$

The first equality comes from the definition of $\overline{Z}(s,0)$. The second uses the law of iterated expectations. The third uses the definition of d_s (incorporating the fact that P_i^0 has no predictive power conditional on $X_i'\beta$ and S_i). The fourth equality uses the approximation and the fifth simplifies the expression.

Thus

$$E(\psi(\tau,\chi_i)\overline{Z}(S_i,0)'\delta | P_i^0 = 1) \approx E\left(\psi(\tau,\chi_i)c_{S_i}\right) \mid P_i^0 = 1\right) + \delta_{X\beta} E\left(\psi(\tau,\chi_i)\overline{X}(S_i,0)'\beta \mid P_i^0 = 1\right).$$

Replacing the two terms of (A2-1) with the approximations leads to

$$\begin{aligned}
\pi^p(\tau) \approx& E\left(\psi(\tau,\chi_i)c_s \mid P_i^0 = 1\right) + \delta_{X\beta} E\left(\psi(\tau,\chi_i)X_i'\beta \mid P_i^0 = 1\right)\\
&- \left[E(\psi(\tau,\chi_i)c_s | P_i^0 = 1) + \delta_{X\beta} E(\psi(\tau,\chi_i)\overline{X}(S_i,0)'\beta | P_i^0 = 1)\right]\\
=& \delta_{X'\beta} E\left(\psi(\tau,\chi_i)\left(X_i'\beta - \overline{X}(S_i,0)'\beta\right) \mid P_i^0 = 1\right),
\end{aligned}$$

which is what we wanted to show.

The argument above would be fine if we didn't need to estimate $\delta_{X'\beta}$. However, we need an additional assumption to establish consistency of the IV estimator that we use to estimate $\delta_{X'\beta}$. As mentioned in the text, we assume that

$$c_s = Q_s'\delta_s + \varpi_s,$$

with ϖ_s independent of Q_s and the process determining X_i within a school.

Extension to Incorporate Unobservables

This argument generalizes to the equation (26). Write $Z_i'\delta$ as

$$Z_i'\delta = d_s\left(X_i'\beta, u_i\right) + e_{Zi}$$

where $d_s(X_i'\beta, u_i) = E(Z_i'\delta \mid X_i'\beta, u_i, S_i = s)$. Analogous to the base case above, we use the linear approximation

$$d_s(X_i'\beta, u_i) \approx c_s + \delta_{X'\beta} X_i'\beta + g\delta_{X'\beta} u_i$$

where $g\delta_{X'\beta}$ is the coefficient on the index u_i of unobservable student characteristics that influence school choice. Then as above

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

A major difference from the base case model is that now u_i is part of χ_i. The first term of this expression is

$$\begin{aligned}E(\psi(\tau, \chi_i) Z_i{}'\delta | P_i^0 = 1) =& E\left(E\left[\psi(\tau, \chi_i) Z_i{}'\delta | X_i'\beta, u_i, S_i, X_i'\beta + Q_{S_i}'\beta_Q + u_i \geq 0\right] \mid P_i^0 = 1\right) \\ =& E\left(\psi(\tau, \chi_i) d_{S_i}(X_i'\beta, u_i) \mid P_i^0 = 1\right). \\ \approx& E\left(\psi(\tau, \chi_i) c_{S_i}\right) \mid P_i^0 = 1\right) + \delta_{X\beta} E\left(\psi(\tau, \chi_i) X_i'\beta \mid P_i^0 = 1\right) \\ &+ g\delta_{X'\beta} E\left(\psi(\tau, \chi_i) u_i \mid P_i^0 = 1\right).\end{aligned}$$

The second term of $\pi^p(\tau)$ may be approximated as

$$\begin{aligned}\overline{Z}(s, 0){}'\delta =& E\{Z_i'\delta | P_i^0 = 1, S_i = s\} \\ =& E\{E\left(Z_i'\delta \mid X_i'\beta, u_i, P_i^0 = 1, S_i = s\right) | P_i^0 = 1, S_i = s\} \\ =& E\{d_s(X_i'\beta, u_i) | P_i^0 = 1, S_i = s\} \\ \approx& E\{c_s | P_i^0 = 1, S_i = s\} + \delta_{X\beta} E\{X_i'\beta | P_i^0 = 1, S_i = s\} + g\delta_{X\beta} E\{u_i | P_i^0 = 1, S_i = s\} \\ =& c_s + \delta_{X\beta}\overline{X}(S_i, 0){}'\beta + g\delta_{X\beta}\overline{u}(S_i, 0).\end{aligned}$$

Plugging this into our expression

$$\begin{aligned}& E(\psi(\tau, \chi_i)\overline{Z}(S_i, 0){}'\delta | P_i^0 = 1) \\ \approx& E\left(\psi(\tau, \chi_i) c_{S_i}\right) \mid P_i^0 = 1\right) \\ &+ \delta_{X\beta} E\left(\psi(\tau, \chi_i)\overline{X}(S_i, 0){}'\beta \mid P_i^0 = 1\right) + g\delta_{X\beta} E\left(\psi(\tau, \chi_i)\overline{u}(S_i, 0) \mid P_i^0 = 1\right).\end{aligned}$$

Putting it all together

$$\text{(A2-2)} \qquad \begin{aligned}\pi^p(\tau) \approx & E\left(\psi(\tau,\chi_i)c_s \mid P_i^0=1\right)+\delta_{X\beta}E\left(\psi(\tau,\chi_i)\left[X_i'\beta+gu_i\right] \mid P_i^0=1\right)\\ & -\left[E(\psi(\tau,\chi_i)c_s|P_i^0=1)+\delta_{X\beta}E(\psi(\tau,\chi_i)\left[\overline{X}\left(S_i,0\right)'\beta+g\overline{u}\left(S_i,0\right)\right]|P_i^0=1)\right]\\ = & \delta_{X\beta}E\left(\psi(\tau,\chi_i)\left(X_i'\beta+gu_i-\overline{X}\left(S_i,0\right)'\beta-g\overline{u}\left(S_i,0\right)\right) \mid P_i^0=1\right)\end{aligned}$$

As in the observable case we further assume that

$$c_s = Q_s'\delta_s+\varpi_s$$

with ϖ_s independent of Q_s as well as $(\mu_s^1,\mu_s^2,v_s^1,v_s^2)$ and the distribution of $(\eta_i^1,\eta_i^2,\omega_i^1,\omega_2^2)$ for individuals in the school district.

Appendix 3: Estimation with Student Body Composition Effects on School Choice.

In this appendix we describe the estimation of the model when peers influence school choice, which is considered in Section 5. We first discuss estimation of the model when peers affect demand through the $\overline{X}\left(s,\tau\right)'\gamma$ index, and then explain how the $\overline{X}\left(s,\tau\right)'\beta$ case differs. The data comes from the "no voucher" regime, and we suppress the indicator for the voucher program regime unless it is needed for clarity.

To simplify the expressions below, define the conditional expectation $\overline{\mu}^2\left(\tau,\mu_s^1,\mu_s^2\right)\equiv E\left(X_i'\gamma \mid S_i=s,P_i^\tau=1\right)$. Keep in mind that we denote the absence of a voucher program as $\tau=0$, with $t(0)=0$. We use ϕ to denote a univariate or bivariate normal density and Φ to denote a univariate or bivariate normal cdf with specified variance. If no variance argument is given, they refer to standard normals.

The $X'\gamma$ Case

We first estimate γ using the fixed effect for the base model. We then estimate the education outcome parameters $\delta_{X'\gamma}$ and $(\delta_Q+\Theta_Q)$ using the JIVE estimator discussed in Section 4.2.

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

A: β, β_Q and $\varphi\delta_{X'\gamma}$

B: α_1 and α_2

C: Σ_η and Σ_e.

D: $\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)$ over the support of $\left(\mu_s^1,\mu_s^2\right)$ that we use in simulation

Note that we are treating $\varphi\delta_{X'\gamma}$ as one estimated parameter since this product enters as the coefficient in front of $\overline{\mu}^2\left(\tau,\mu_s^1,\mu_s^2\right)$ in the school choice equation.

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

Step 1, Given B,C,D estimate A: Given the other parameters, we estimate β, β_Q, and $\varphi\delta_{X'\gamma}$ using the likelihood for private and public schools. The log likelihood for individual i is

(A3-1)

$$
\begin{aligned}
&P_i^0 w_i \log(\\
&\left.\frac{[\int \Phi\left(X_i'\beta+Q_{S_i}'\beta_Q+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)\right)\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)d\Phi(\mu_s^1,\mu_s^2;W_{S_i},\Sigma_e)}{\int\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)d\Phi(\mu_s^1,\mu_s^2;W_{S_i},\Sigma_e)}\right) \\
&+\left(1-P_i^0\right)w_i log(1- \\
&\left.\frac{\int \Phi\left(X_i'\beta+Q_{S_i}'\beta_Q+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)\right)\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)d\Phi(\mu_s^1,\mu_s^2;W_{S_i},\Sigma_e)}{\int\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)d\Phi(\mu_s^1,\mu_s^2;W_{S_i},\Sigma_e)}\right),
\end{aligned}
$$

where w_i is the sample weight for individual i. We treat public and private school students symmetrically and in particular do not use the data on public peers to help update β in this step.[54] Instead, we fix β^* and β_Q^* in the above equation at the estimated value from the previous iteration of Step 1 rather than letting it change as we maximize the above likelihood function with respect to β, β_Q, and $\varphi\delta_{X'\gamma}$. This means that the update for $\widehat{\beta}$ and $\hat{\beta_Q}$ is chosen to maximize the likelihood of the school choice model rather than to make the $X_i'\beta$ distribution look approximately normal.

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

Step 2 Given A,C,D estimate B: This step is quite simple. We can estimate α_1 and α_2 by regressing $X_i'\beta+Q_{S_i}'\beta_Q$ and $X_i'\gamma$ on W_{S_i}. Thus we only need to know β and γ for this step.

Step 3 Given A,B,D estimate C: Taking the rest of the model as given, we estimate Σ_η and Σ_e using

[54] This is inefficient but still consistent. We cannot use the model to group private school students because the model does not say which private school students attending private school choose.

the likelihood for public students only:

$$\text{(A3-2)} \qquad \frac{\int \mathcal{L}_s\left(\Sigma_\eta \mid \mu_s^1, \mu_s^2\right) \Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right) d\Phi(\mu_s^1, \mu_s^2; W_{S_i}, \Sigma_e)}{\int \Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right) d\Phi(\mu_s^1, \mu_s^2; W_{S_i}, \Sigma_e)}$$

where

$$\mathcal{L}_s\left(\Sigma_\eta \mid \mu_s^1, \mu_s^2\right) = \prod_{\{i:S_i=s\}} \left(\frac{\Phi\left(X_i'\beta + Q_{S_i}'\beta_Q + \varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)\right) \phi\left(X_i'\beta + Q_{S_i}'\beta_Q - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right)}{\Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right)} \right)$$

To see where the above equation comes from, one must consider the NELS sampling frame. In particular schools with larger values of $\Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right)$ tend to be bigger. We make the simplifying assumption that the NELS sampling frame fixes the number of interviews at each public high school independent of the size of the school, but then oversamples bigger schools. In this case, $\mathcal{L}_s$ should be the likelihood of observing a particular realization of the vectors of values of $X_i'\beta$ and $X_i'\gamma$ for a particular sample of students from the school. Then, being loose with notation, the likelihood for a particular school takes the form

$$\int \left[\prod_{\{i:S_i=s\}} f(X_i'\beta + Q_{S_i}'\beta_Q, X_i'\gamma \mid P_i^0 = 1, \mu_s^1, \mu_s^2) \right] g_s\left(\mu_s^1, \mu_s^2 \mid W_s\right) d\mu_s^1 d\mu_s^2$$

where g_s is the probability density given the sampling scheme.

Then using Bayes' theorem

$$\begin{aligned} & f(X_i'\beta + Q_{S_i}'\beta_Q, X_i'\gamma \mid P_i^0 = 1, \mu_s^1, \mu_s^2) \\ = \; & \frac{\Pr(P_i^0 = 1 \mid X_i'\beta + Q_{S_i}'\beta_Q, X_i'\gamma, \mu_s^1, \mu_s^2) f(X_i'\beta + Q_{S_i}'\beta_Q, X_i'\gamma, \mu_s^1, \mu_s^2)}{\Pr(P_i^0 = 1 \mid \mu_s^1, \mu_s^2)} \\ = \; & \frac{\Phi\left(X_i'\beta + Q_{S_i}'\beta_Q + \varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)\right) \phi\left(X_i'\beta + Q_{S_i}'\beta_Q - \mu_s^1, X_i'\gamma - \mu_s^2; \Sigma_\eta\right)}{\Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right)} \end{aligned}$$

and

$$\begin{aligned} g_s\left(\mu_s^1, \mu_s^2 \mid W_s\right) &= g\left(\mu_s^1, \mu_s^2 \mid W_s, P_i^0 = 1\right) \\ &= \frac{Pr\left(P_i = 1 \mid W_s, \mu_s^1, \mu_s^2\right) g\left(\mu_s^1, \mu_s^2 \mid W_s\right)}{Pr\left(P_i^0 = 1 \mid W_s\right)} \\ &= \frac{\Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right) \phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)}{\int \Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right) d\Phi(\mu_s^1, \mu_s^2; W_s, \Sigma_e)} . \end{aligned}$$

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

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

$$\begin{aligned}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right) =& \mu_s^2 + E\left(\eta_i^2 \mid \mu_s^1+\eta_i^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)+\varepsilon_i>0,\mu_s^1,\mu_s^2\right)\\ =& \mu_s^2+\frac{\sigma_{\eta 12}}{\sqrt{1+\sigma_{\eta 11}}}\lambda\left(\frac{\mu_s^1+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)}{\sqrt{1+\sigma_{\eta 11}}}\right)\end{aligned}$$

We iterate on this four-step procedure until we find a fixed point estimate of β, β_Q, and $\varphi\delta_{X'\gamma}$. Each iteration is time consuming, in part because we must compute $\overline{\mu}^2\left(0,\mu_s^1,\mu_s^2\right)$ for a large number of values of $\left(\mu_s^1,\mu_s^2\right)$.[55] However, the parameter estimates converge fairly quickly.

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

$$\begin{aligned}&\Pr(P_i^\tau=1 \mid W_{S_i},X_i'\beta,Z_i'\gamma)\\ =&\int\Phi\left(X_i'\beta+Q_{S_i}'\beta_Q-t(\tau)+\varphi\delta_{X'\gamma}\overline{\mu}^2\left(\tau,\mu_s^1,\mu_s^2\right)\right)dF(\mu_s^1,\mu_s^2\mid W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q,X_i'\gamma)\end{aligned}$$

where

$$dF(\mu_s^1,\mu_s^2\mid W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q,X_i'\gamma)=\frac{\phi\left(X_i'\beta+Q_{S_i}'\beta_Q-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)d\Phi(\mu_s^1,\mu_s^2;W_{S_i},\Sigma_e)}{\int\phi\left(X_i'\beta+Q_{S_i}'\beta_Q-\mu_s^1,X_i'\gamma-\mu_s^2;\Sigma_\eta\right)d\Phi(\mu_s^1,\mu_s^2;W_{S_i},\Sigma_e)}.$$

[55] Specifically $\mu_s^1=W_s'\alpha_1+e_s^1$,$\mu_s^2=W_s'\alpha_2+e_s^2$ and we approximate the marginal distribution of each e_S^j using 6 points of Gauss-Hermite quadrature. That means for each school $\left(\mu_s^1,\mu_s^2\right)$ takes on 36 values, and these values differ over the different public schools in our sample.

We then construct

$$\psi(\tau,\chi_i)=\frac{\frac{\Pr(P_i^\tau=1|W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q,X_i'\gamma)}{\Pr(P_i^0=1|W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q,X_i'\gamma)}}{\int\frac{\Pr(P_i^\tau=1|W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q,X_i'\gamma)}{\Pr(P_i^0=1|W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q,X_i'\gamma)}dG(W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q,X_i'\gamma\mid P_i^0=1)}$$

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

Note that χ_i contains of W_{S_i} as well as X_i.

The $X'\beta$ Case

We use the same iterative procedure for the $X_i'\beta$ model, but the equations are simpler. In this case define $\overline{\mu}^1\left(\tau,\mu_s^1\right)\equiv E\left(X_i'\beta\mid P_i^\tau=1,\mu_s^1\right)$. Given our model and distribution assumptions,

$$\begin{aligned}\overline{\mu}^1\left(\tau,\mu_s^1\right)&=\mu_s^1+E\left(\eta_i^1\mid\mu_s^1+\eta_i^1-t(\tau)+\varphi\delta_{X'\beta}\overline{\mu}^1\left(\tau,\mu_s^1\right)+\varepsilon_i>0\right)\\&=\mu_s^1+\frac{\sigma_{\eta11}}{\sqrt{1+\sigma_{\eta11}}}\lambda\left(\frac{\mu_s^1-t(\tau)+\varphi\delta_{X'\beta}\overline{\mu}^1\left(\tau,\mu_s^1\right)}{\sqrt{1+\sigma_{\eta11}}}\right)\end{aligned}$$

The analogue of likelihoods (A3-1) and (A3-2) in the $X_i'\beta$ case are simpler than in the $X_i'\gamma$ case. They take on the forms

$$\begin{aligned}&P_i^0w_i\log\left(\frac{\int\Phi\left(X_i'\beta+Q_{S_i}'\beta_Q+\varphi\delta_{X'\beta}\overline{\mu}^1\left(0,\mu_s^1\right)\right)\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1;\sigma_{\eta11}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})}{\int\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1;\sigma_{\eta11}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})}\right)\\&+(1-P_i^0)w_i\times\\&\log\left(1-\frac{\int\Phi\left(X_i'\beta+Q_{S_i}'\beta_Q+\varphi\delta_{X'\beta}\overline{\mu}^1\left(0,\mu_s^1\right)\right)\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1;\sigma_{\eta11}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})}{\int\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1;\sigma_{\eta11}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})}\right),\end{aligned}$$

and the likelihood for a school is

$$\frac{\int\left[\prod_{\{i:S_i=s\}}\left(\frac{\Phi\left(X_i'\beta+Q_{S_i}'\beta_Q+\varphi\delta_{X'\beta}\overline{\mu}^1\left(0,\mu_s^1\right)\right)\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1;\sigma_{\eta11}\right)}{\Phi\left(\frac{\mu_s^1+\varphi\overline{\mu}^1\left(0,\mu_s^1\right)}{\sqrt{1+\sigma_{\eta11}}}\right)}\right)\right]\Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\beta}\overline{\mu}^1\left(0,\mu_s^1\right)}{\sqrt{1+\sigma_{\eta11}}}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})}{\int\Phi\left(\frac{\mu_s^1+\varphi\delta_{X'\beta}\overline{\mu}^1\left(0,\mu_s^1\right)}{\sqrt{1+\sigma_{\eta11}}}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})},$$

where σ_{e11} is the variance of e_s^1. The final piece is

$$\begin{aligned}&\Pr(P_i^\tau=1\mid W_i,X_i'\beta+Q_{S_i}'\beta_Q)\\&=\int\Phi\left(X_i'\beta+Q_{S_i}'\beta_Q-t(\tau)+\varphi\delta_{X'\beta}\overline{\mu}^1\left(\tau,\mu_s^1\right)\right)dF(\mu_s^1\mid W_{S_i},X_i'\beta+Q_{S_i}'\beta_Q)\\&=\frac{\int\Phi\left(X_i'\beta+Q_{S_i}'\beta_Q-t(\tau)+\varphi\delta_{X'\beta}\overline{\mu}^1\left(\tau,\mu_s^1\right)\right)\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1;\sigma_{\eta11}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})}{\int\phi\left(X_i'\beta^*+Q_{S_i}'\beta_Q^*-\mu_s^1;\sigma_{\eta11}\right)d\Phi(\mu_s^1;W_{S_i},\sigma_{e11})},\end{aligned}$$

and again

$$\psi_i(\tau,\chi_i)=\frac{\frac{\Pr(P_i^\tau=1|W_{S_i},X_i'\beta+Q'_{S_i}\beta_Q,Z_i'\gamma)}{\Pr(P_i^0=1|W_{S_i},X_i'\beta+Q'_{S_i}\beta_Q,Z_i'\gamma)}}{\int\frac{\Pr(P_i^\tau=1|W_{S_i},X_i'\beta+Q'_{S_i}\beta_Q,Z_i'\gamma)}{\Pr(P_i^0=1|W_{S_i},X_i'\beta+Q'_{S_i}\beta_Q,Z_i'\gamma)}dG(W_{S_i},X_i'\beta+Q'_{S_i}\beta_Q,Z_i'\gamma\mid P_i^0=1)}.$$

Appendix 4: Unobservable School Effects and Unobservable Peer Characteristics

In this appendix we discuss estimation of the unobservable student body effect model we describe in the text. In practice there are two issues that need to be addressed. The first is estimation of the variance/covariance components of the model. This turns out to be quite similar to the case discussed in Appendix 3.

The second issue is estimation of $\delta_{X'\beta}$ or $\delta_{X'\gamma}$ which differs from the rest of the paper. For the other cases we estimate $\delta_{X'\beta}$ using a two stage least squares approach. Here we rely on the structure of our nonlinear model so the first stage is nonlinear, but then we use OLS in the second in much the same way as in the second stage of 2SLS.

In this appendix we first discuss the general strategy and then go step by step through the details. We use ϕ to denote a normal density and Φ to denote a normal cdf with specified variance. If no variance argument is given, they refer to standard normals. Also note that all of our calculations use the NELS weights. In this appendix we ignore the weighting for expositional reasons.

To see the main complication, let X_{-i} be the matrix of values of X_j' for a random sample of individuals j who actually attend school S_i , $j\neq i$ and define
$X_i^{\beta\gamma}\equiv\left(X'_{-i}\beta,X'_{-i}\gamma,P_i^0=1,X_i'\beta,X_i'\gamma,Q_{S_i}\right)$. From equation (2)

$$E\left(\theta(S_i,\tau)\mid X_i^{\beta\gamma}\right)=E\left(\overline{Z}(S_i,\tau)'\delta\mid X_i^{\beta\gamma}\right)+Q'_{S_i}\Theta_Q.$$

We use this equation in the second stage to obtain a consistent estimate of $\delta_{X'\beta}$ by regressing the estimate of the school specific fixed effect on Q_{S_i} and a consistent estimate of $E\left(\overline{Z}(s,\tau)'\delta\mid X_i^{\beta\gamma}\right)$.

Consider the first stage. In model (26)

$$
\begin{aligned}
\text{(A4-1)}\quad & E\left(\overline{Z}(s,\tau)'\delta \mid X_i^{\beta\gamma}\right) \\
=& Q_s'\delta_Q + \delta_{X'\beta} E\left(\overline{X}(s,\tau)'\beta + Q_{S_i}'\beta_Q + g\bar{u}(s,\tau) \mid X_i^{\beta\gamma}\right) + \varpi_s \\
=& Q_s'\delta_Q + \delta_{X'\beta} E\left(X_j'\beta + Q_{s_j}'\beta_Q + gu_j \mid P_j = 1, S_j = S_i, X_i^{\beta\gamma}\right) + \varpi_s \\
=& Q_s'\delta_Q + \delta_{X'\beta} \int\int E\left(\mu_s^1 + \eta_j^1 + g\left[v_s^1 + \omega_j^1\right] \mid P_j = 1, \mu_s^1, v_s^1\right) f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 dv_s^1 + \varpi_s \\
=& Q_s'\delta_Q + \delta_{X'\beta} \int\int \left(\mu_s^1 + gv_s^1 + \frac{\sigma_{\eta 11} + g\sigma_{\omega 11}}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\lambda\left(\frac{\mu_s^1 + v_s^1}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right)\right) f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 dv_s^1 + \varpi_s
\end{aligned}
$$

where $f_{\mu^1 v^1}$ denotes the density of $\left(\mu_s^1, v_s^1\right)$ conditional on $X_i^{\beta\gamma}$. Thus to implement this we need to estimate $(\sigma_{\eta 11}, \sigma_{\omega 11})$ as well as $f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^{\beta})$.

Similarly for model (25)

$$
\begin{aligned}
\text{(A4-2)}\quad & E\left(\overline{Z}(s,\tau)'\delta \mid X_i^{\beta\gamma}\right) \\
\text{(A4-3)}\quad & =\delta_{X'\gamma} E\left(\overline{X}(s,\tau)'\gamma + g\bar{\varepsilon}(s,\tau) \mid X_i^{\beta\gamma}\right) \\
& =\delta_{X'\gamma} \int E\left(\mu_s^2 + \eta_j^2 + g\left[v_s^2 + \omega_j^2\right] \mid P_j = 1, \mu_s^1, \mu_s^2, v_s^1\right) f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 d\mu_s^2 dv_s^1 \\
& =\delta_{X'\gamma} \int \left(\mu_s^2 + \frac{g\sigma_{v12} v_s^1}{\sigma_{v11}} + \frac{\sigma_{\eta 12} + g\sigma_{\omega 12}}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\lambda\left(\frac{\mu_s^1 + v_s^1}{\sqrt{\sigma_{\eta 11} + \sigma_{\omega 11}}}\right)\right) \times \\
& \quad f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 d\mu_s^2 dv_s^1
\end{aligned}
$$

where $f_{\mu v^1}$ denotes the density of $\left(\mu_s^1, \mu_s^2, v_s^1\right)$ conditional on $X_i^{\beta\gamma}$. To implement this we need $(\sigma_{\eta 11}, \sigma_{\omega 11}, \sigma_{v11}, \sigma_{v12}, \sigma_{\eta 12}, \sigma_{\omega 12})$ as well as $f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma})$.

Our procedure is as follows.

1. We obtain β from a standard probit model for public school choice.
2. We estimate γ using fixed effects regression on the public school sample, but we correct for sample selection by including the inverse Mills-ratio term for public school choice in the regression.
3. We normalize all random variables to be mean zero, other than (μ_s^1, μ_s^2). Given β and γ we can estimate the mean and variance of $(X_i'\beta, X_i'\gamma)$. From this we know that

   $$a = \frac{1}{var(X_i'\beta)}$$

   since $Var(u_i) = 1$.

4. Next we estimate variance/covariance matrices $\Sigma_\mu, \Sigma_\eta, \Sigma_v$, and Σ_ω. Note that given information from above, knowledge of Σ_μ is sufficient for the other parameters since

$$\begin{aligned}\Sigma_\eta =& var(X_i'\beta + Q_{S_i}'\beta_Q, X_i'\gamma) - \Sigma_\mu \\ \Sigma_v =& a\Sigma_\mu \\ \sigma_{\omega 11} =& a\sigma_{\eta 11} \\ \sigma_{\omega 12}\left(\omega_i^1, \omega_i^2\right) =& a\sigma_{\eta 12}\end{aligned}$$

We estimate $\Sigma_\mu$using maximum likelihood but only using public school students. We assume that schools are sampled based on the population of students who attend (i.e. all else equal, public schools in districts with fewer students attending private schools will show up in the data more frequently).[56] This gives the likelihood function for school s:

$$\frac{\int\int\int \mathcal{L}\left(s \mid v_s^1, \mu_s^1, \mu_s^2\right)\Phi\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma_{\eta 11}(\Sigma_\mu)+\sigma_{\omega 11}(\Sigma_\mu)}}\right)d\Phi\left(v_s^1; \sigma_{v^1}^2\left(\Sigma_\mu\right)\right)d\Phi(\mu_s^1, \mu_s^2; \Sigma_\mu)}{\int\int\int \Phi\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma_{\eta 11}(\Sigma_\mu)+\sigma_{\omega 11}(\Sigma_\mu)}}\right)d\Phi\left(v_s^1; \sigma_{v11}\left(\Sigma_\mu\right)\right)d\Phi(\mu_s^1, \mu_s^2; \Sigma_\mu)}$$

where

$$\mathcal{L}\left(s \mid v_s^1, \mu_s^1, \mu_s^2\right) = \prod_{\{i:S_i=s\}}\left(\frac{\Phi\left(\frac{X_i'\beta+Q_{S_i}'\beta_Q+v_s^1}{\sqrt{\sigma_{\omega 11}(\Sigma_\mu)}}\right)\phi\left(X_i'\beta + Q_{S_i}'\beta_Q - \mu_s^1, X_i'\gamma - \mu_s^2, \Sigma_\eta\right)}{\Phi\left(\frac{\mu_s^1+v_s^1}{\sqrt{\sigma_{\eta 11}(\Sigma_\mu)+\sigma_{\omega 11}(\Sigma_\mu)}}\right)}\right)$$

where the notation $\sigma_{\eta 11}\left(\Sigma_\mu\right)$,$\sigma_{\omega 11}\left(\Sigma_\mu\right)$, and $\sigma_{v11}\left(\Sigma_\mu\right)$ denotes the fact that these objects are known functions of Σ_μ.

5. In this step we obtain our estimate of $f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma})$ for all public school observations. Using independence of v_s^1 from (μ_s^1, μ_s^2) and Bayes' theorem,

$$\begin{aligned}&f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma}) \\ &= \frac{f_{x^{\beta\gamma}}(X_i^{\beta\gamma} \mid \mu_s^1, v_s^1)\phi(\mu_s^1, \mu_s^2; \Sigma_\mu)\phi\left(v_s^1; \sigma_{v^1}^2\right)}{\int\int\int f_{x^{\beta\gamma}}(X_i^{\beta\gamma} \mid \mu_s^1, v_s^1)\phi(\mu_s^1, \mu_s^2; \Sigma_\mu)\phi\left(v_s^1; \sigma_{v^1}^2\right)d\mu_s^1 d\mu_s^2 dv_s^1},\end{aligned}$$

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

where the definition of $X'_{-i}\beta$ and $X'_{-i}\gamma$ implies that

$$f_{x^{\beta\gamma}}(X_i^{\beta\gamma} \mid \mu_s^1, \mu_s^2, v_s^1) = \frac{\prod\limits_{\{j:S_j=S_i\}} \Phi\left(\frac{X_j'\beta + v_{S_i}^1}{\sigma_{\omega^1}}\right) \prod\limits_{\{j:S_j=S_i\}} \phi\left(X_j'\beta + Q_{S_i}'\beta_Q - \mu_s^1, X_j'\gamma - \mu_s^2; \Sigma_\eta\right)}{\prod\limits_{\{j:S_j=S_i, j\neq i\}} \Phi\left(\frac{\mu_s^1 + v_s^1}{\sqrt{\sigma_{\eta^1}^2 + \sigma_{\omega^1}^2}}\right)}.$$

Given knowledge of $f_{\mu v^1}(\mu_s^1, \mu_s^2, v_s^1 \mid X_i^{\beta\gamma})$ one can obtain $f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^{\beta\gamma})$ by integrating out μ_s^2.

6. Using this density, for each g we obtain a consistent estimates of

$$E\left(\overline{X}(s,\tau)'\beta + Q_{S_i}'\beta_Q + g\bar{u}(s,\tau) \mid X_i^{\beta\gamma}\right)$$

and

$$E\left(\overline{X}(s,\tau)'\gamma + g\bar{\varepsilon}(s,\tau) \mid X_i^{\beta\gamma}\right)$$

using equations (A4-1) and (A4-3), respectively. We then regress the estimate of the school specific fixed effect on Q_s and $E\left(\overline{X}(s,\tau)'\beta + Q_{S_i}'\beta_Q + g\bar{u}(s,\tau) \mid X_i^{\beta\gamma}\right)$ to estimate $\delta_{X'\beta}$ and on $Q_s$and $E\left(\overline{X}(s,\tau)'\gamma + g\bar{\varepsilon}(s,\tau) \mid X_i^{\beta\gamma}\right)$to estimate $\delta_{X'\gamma}$.

7. Finally, we calculate the treatment effect. In the text we derive $\pi^p\left(\tau; \mu_s^1, v_s^1\right)$ for two models. (See (27) for the $X'\beta$ index model and (28) for the $X'\gamma$ model.) To obtain the average cream skimming effect, we have to integrate $\pi^p\left(\tau; \mu_s^1, v_s^1\right)$ over the distribution of the unobservables μ_s^1, v_s^1 of those who stay in public school. In the $X'\beta$ model the average value of the cream skimming effect may be written as

$$\begin{aligned}
&E\left[\pi^p\left(\tau; \mu_s^1, v_s^1\right) \mid P_i^\tau = 1\right] \\
=&\frac{E\left[\pi^p\left(\tau; \mu_s^1, v_s^1\right) P_i^\tau\right]}{Pr\left[P_i^\tau = 1\right]} \\
=&\frac{\int \pi^p\left(\tau; \mu_s^1, v_s^1\right) \frac{\Phi\left(\left(X_i'\beta + Q_{S_i}'\beta_Q - t(\tau) + v_s^1\right)/\sqrt{\sigma_{\omega 11}}\right)}{\Phi\left(\left(X_i'\beta + Q_{S_i}'\beta_Q + v_s^1\right)/\sqrt{\sigma_{\omega 11}}\right)} f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 dv_s^1 dG\left(X_i^{\beta\gamma} \mid P_i^0 = 1\right)}{\int \frac{\Phi\left(\left(X_i'\beta + Q_{S_i}'\beta_Q - t(\tau) + v_s^1\right)/\sqrt{\sigma_{\omega 11}}\right)}{\Phi\left(\left(X_i'\beta + Q_{S_i}'\beta_Q + v_s^1\right)/\sqrt{\sigma_{\omega 11}}\right)} f_{\mu^1 v^1}(\mu_s^1, v_s^1 \mid X_i^{\beta\gamma}) d\mu_s^1 dv_s^1 dG\left(X_i^{\beta\gamma} \mid P_i^0 = 1\right)},
\end{aligned}$$

where we use normality in the last line. One obtains an analogous expression for the $X'\gamma$ model. We approximate with sample analogues.

Table A1
Probit Model for Public School Attendance
Full Sample

	Probit Coefficient	Confidence Interval	Marginal Effect
Constant	6.9115	(5.2224,8.3849)	-
Male	-0.0850	(-0.2148,0.0499)	-0.0088
Hispanic	0.2217	(-0.0457,0.6413)	0.0199
Black	0.0998	(-0.2244,0.4734)	0.0097
Parental Background			
Catholic	-0.5134	(-0.7277,-0.3341)	-0.0533
Both Parents Present	-0.2170	(-0.3689,-0.0704)	-0.0225
Father's Education	-0.0511	(-0.0920,-0.0161)	-0.0053
Mother's Education	-0.0480	(-0.0757,-0.0166)	-0.0050
log Income	-0.2873	(-0.4044,-0.1598)	-0.0299
Limited English Proficiency	-0.2668	(-0.8844,0.7181)	-0.0277
8th Grade Tests and Grades			
Reading Score	-0.0160	(-0.0250,-0.0053)	-0.0017
Math Score	0.0019	(-0.0097,0.0099)	0.0002
Science Score	0.0095	(-0.0000,0.0198)	0.0010
History Score	-0.0090	(-0.0185,0.0027)	-0.00093
Grades Composite	0.0240	(-0.1135,0.1868)	0.0025
8th Grade Behavior and Performance in School Measures			
Delinquency Index	0.0149	(-0.0801,0.1063)	0.0016
Student Got into a Fight	0.0146	(-0.1302,0.1446)	0.00151
Student Performs Below Ability	-0.1561	(-0.4008,0.1181)	-0.01620
Student Rarely Completes Homework	0.3329	(0.0382,0.7047)	0.0346
Student Frequently Absent	0.0949	(-0.2655,0.4729)	0.0090
Student Inattentive in Class	-0.1446	(-0.3916,0.1069)	-0.0150
Student Frequently Disruptive	-0.2273	(-0.5195,0.1177)	-0.0236
Parent Believes Child has a Behavioral Problem in School	0.1683	(-0.0987,0.5492)	0.01747
Repeated a Grade	0.1652	(-0.1687,0.4966)	0.0172
Dropout Risk Composite	0.0358	(-0.0570,0.1503)	0.0037
Lack of Effort Index	0.0072	(-0.0226,0.0403)	0.0008
Enrolled in Gifted Program	0.3878	(0.1982,0.6438)	0.0402
Location Measures			
North East	-0.1204	(-0.4706,0.2537)	-0.0125
North Central	0.1106	(-0.3199,0.4863)	0.0115
South	-0.1605	(-0.5984,0.1804)	-0.0167
Urban	-0.7436	(-1.0679,-0.4267)	-0.0772
Suburban	-0.2776	(-0.5373,-0.0102)	-0.0288
Distance	0.9138	(0.2829,2.5733)	0.0949
Distance Squared	-0.1779	(-0.7485,-0.0593)	-0.0185

Notes: The sample size is 16483. Column 1 reports MLE probit coefficient estimates.Columns 2 and 3 report bootstrap estimates of the lower and upper bound of the 95% confidence interval. The estimates account for correlation across students who attended the same 8th grades and/or high schools. They are based on 1,000 bootstrap replications. The fourth column reports marginal effects on the probability of attending public high school when $X'\beta$ is 1.27588, which corresponds to the value at which the probability of attending public high school equals the weighted mean (.899). The mode also contains missing indicators for Distance,log family income in 1987 and an indicator that is one if all test scores are missing. There is one indicator for missing data on Student performs below ability, Student rarely completes homework, Student frequently absent, Student inattentive in class, and/or Student frequently disruptive. NELS:88 base year to third year follow up panel weights are used.

Table A2
Effects of Students Own Characteristics on Public High School Graduation
Linear Probability Models with HS Fixed Effects

	Coefficient	Confidence Interval
Male	0.0217	(0.0067,0.0371)
Hispanic	0.0076	(-0.0262,0.0461)
Black	0.0641	(0.0278,0.0940)
Parental Background		
Catholic	0.0297	(0.0121,0.0445)
Both Parents Present	0.0119	(-0.0077,0.0324)
Father's Education	0.0044	(0.0013,0.0074)
Mother's Education	0.0008	(-0.0030,0.0045)
Log Family Income	0.0196	(0.0039,0.0341)
Limited English Proficiency	0.0827	(0.0284,0.1290)
8th Grade Tests and Grades		
Reading	0.0003	(-0.0009,0.0013)
Math	0.0009	(-0.0002,0.0020)
Science	0.0004	(-0.0006,0.0015)
History	0.0003	(-0.0009,0.0014)
Grades Composite	0.0339	(0.0188,0.0490)
8th Grade Behavior and Performance in School Measures		
Delinquency Index	-0.0167	(-0.0290,-0.0044)
Student got Into a Fight	-0.0175	(-0.0403,0.0039)
Student Performs Below Ability	-0.0252	(-0.0640,0.0108)
Student Rarely Completes Homework	-0.0900	(-0.1418,-0.0402)
Student Frequently Absent	-0.1695	(-0.2213,-0.1185)
Student Inattentive in Class	-0.0313	(-0.0707,0.0125)
Student Frequently Disruptive	-0.0206	(-0.0652,0.0186)
Parent Believes Child has a Behavioral Problem in School	-0.0430	(-0.0794,-0.0044)
Repeated a Grade	-0.1483	(-0.1950,-0.1041)
Dropout Risk Composite	-0.0240	(-0.0363,-0.0116)
Lack of Effort Index	-0.0012	(-0.0046,0.0021)
Enrolled in Gifted Program	0.0135	(-0.0032,0.0307)

Notes: Column 1 of the table reports weighted least squares estimates from a regression of high school graduation with high school fixed effects included. Column 2 report the lower and upper bounds of the 95 percent confidence interval. They are calculated from 1000 bootstrap replications. The model also includes three missing data indicators see the note to Table A1. The sample is restricted to public high school students, and the sample size used in the calculation is 10795. Schools with only one sampled student are dropped. NELS:88 base year to 3rd follow up panel weights are used.