Estimating School and Neighborhood Effects Using Sorting on Observables to Control for Sorting on Unobservables

Joseph Altonji — Yale University

Richard Mansfield — Cornell University

March 16, 2014

1 Introduction

How much does the school and surrounding community that we choose for our children matter for their long run educational and labor market outcomes? In addressing this question, generations of social scientists have struggled to overcome a fundamental sorting problem: do schools differ in their average outcomes because they influence student performance, or because they have succeeded or failed in attracting the students who would have thrived regardless of the school chosen?[1]

In this paper, we use a flexible spatial equilibrium model of school choice to show that a standard method in between-school regressions, controlling for school-averages of individual-level characteristics, can yield a lower-bound estimate of the variance in school contributions to individual outcomes even when very general patterns of sorting on both observable and unobservable characteristics are allowed for. Furthermore, this lower-bound estimate can be interpreted as the component of the school contribution that was unknown or unvalued by individuals at the time the school and surrounding neighborhood was chosen.

A number of recent papers have employed experimental or quasi-experimental strategies to isolate the contribution of either schools or neighborhoods to longer run student outcomes. (Oreopoulos 2003) and (Jacob 2004) use quasi-random assignment of neighborhood in the wake of housing project closings to estimate the magnitude of neighborhood effects on student outcomes. Similarly, the Moving To Opportunity experiment, evaluated in (Kling, Liebman and Katz 2007), randomly assigned housing vouchers that required movement to a lower income neighborhood to estimate neighborhood effects. The authors report

[1] A recent example, with references to the literature, is (Altonji and Mansfield 2011a). (Altonji and Mansfield 2011b) discusses a number of econometric issues involved, not all of which we discuss here. It also presents detailed evidence on the distribution across schools of a rich set of student and family background characteristics as well as variance decompositions of school outcomes that is similar to some of the evidence below. This paper grew out of our dissatisfaction with the treatment of sorting in our prior work and other papers and can be viewed as a drastic revision of (Altonji and Mansfield 2011b).

no evidence that moving to a low-poverty neighborhood improves economic outcomes. However, (Aliprantis 2011) argues that while MTO policy effects are well-identified, if one does not impose that the poverty rate is a sufficient static for a neighborhood's contribution to outcomes, then the support of experimentally-induced changes in neighborhood quality is too narrow to draw meaningful conclusions about the importance of neighborhood quality. The experimentally-induced changes in school quality suffer from the same problem, so that little is learned about the long-run value of higher quality schools.

(Deming, Hastings, Kane and Staiger 2011) exploit randomized lottery outcomes from the school choice plan in the Charlotte-Mecklenburg district to estimate the impact of winning a lottery to attend a chosen public school on high school graduation, college enrollment, and college completion. They find that the impact of winning the lottery for students from low quality urban schools is large enough to close 75 percent of the black-white gap in graduation and 25 percent of the gap in Bachelor's Degree completion. While this is the best existing evidence about the long-run impact of attending a higher quality high school, one might nonetheless wonder how well estimates generalize beyond the Charlotte-Mecklenburg context, and beyond the impacts for students moving from the worst schools. Furthermore, their estimates represent the impacts of changing high schools for students who attended a particular elementary school and middle school, rather than the cumulative impact of choosing one neighborhood and surrounding school system relative to another. These estimates thus also reflect the time cost of attending a faraway school, the impact of moving to a radically different school environment, and the halo effect of winning a lottery and experiencing a seemingly fresh opportunity. Despite the growing popularity of open enrollment systems, most school choice is still mediated through choice of community in which to live, and most students still choose schools close to home even when given the opportunity. Thus, we aim instead to measure the importance of the combined school/neighborhood choice.

In contrast to these papers, we do not exploit any natural experiments. Instead, we show that rich observational data of the type collected by either panel surveys or administrative databases can nonetheless yield meaningful insights about the importance of school and neighborhood choices for children's later educational and labor market performance. Our key insight stems from the simple observation that average values of student characteristics differ across schools only because students with different characteristics value school or neighborhood amenities differently. If tastes or willingness to pay for particular schools were not predictable based on student characteristics, each school would have the same distribution of student characteristics. Thus, school-average values of student characteristics, if based on a sufficiently large pool of students, potentially provide considerable information about the amenities the school and neighborhood offer. In particular, we establish the following result, which is stated in Proposition 1 below. Suppose that tastes for a subset of the amenities depend on unobservable student characteristics that matter for the outcome of interest. (The subset might be the full set of amenities considered by consumers.) Assume that the rank of the matrix of coefficients relating preferences for these amenities to the observable student characteristics is equal to the dimension of the subset. One necessary condition for the rank condition to hold is that the number of observables that affect tastes for the subset greater than or equal to the dimension of the subset. A second necessary condition is that tastes for each of the elements of that subset must depend on at least one observable student characteristic. We show under these assumptions, in spatial equilibrium the average values of the observable characteristics will form basis vectors that spans the space of the subset of underlying school/neighborhood amenities, and importantly, the average values of unobservable student characteristics. Consequently, the averages of the observables can serve as controls for the unobservables.

For example, suppose that school/neighborhood combinations differ in only four dimensions that people observe and systematically care about—school quality, crime, parks, and highway access–plus a random idiosyncratic component specific to each family/location combination.[2] The price of the location will be a function of its values of the four amenities. Suppose that parental income, parental education, and number of kids influence the weight the family places on the first three amenities relative to the price of the location. They may or may not influence taste for highway access. Finally, suppose that parents differ in the importance they place on educational achievement, which influences their taste for school quality and possibly the other amenities, but not highway access. Taste for educational achievement is not observed. Our result implies that the expected value in a location of taste for educational achievement will be an exact function of the expected values of the parental income, parental education, and number of children. Consequently, the the location averages of the observables can serve as controls for taste for achievement.[3] The result also implies that only the component of school quality that was not observable when location choices are made will vary conditional on the average values of parental income, education, and children.

The locational sorting result implies that including a sufficiently large vector of school-averages of student observable characteristics can potentially control for school-average values of outcome-relevant student unobservable characteristics. However, while this control function approach solves the sorting-on-unobservables problem, it also absorbs any variation in the underlying school and neighborhood amenities that are weighted by observed and unobserved outcome-relevant student characteristics when choice of school/neighborhood is made. Consequently, any remaining between-school variation in outcomes can be attributed to two factors. The first is a component of school quality that was either unknown or unvalued at the time the school/neighborhood was chosen (including later shocks that

[2]As will be made clear below, the weights families place on the amenities may also depend on other unobserved characteristics that do not have a direct effect on the outcomes of interest. These additional characteristics are the κ_i^* variables in the analysis below.

[3](Bayer and Ross 2006) also investigate the possibility of removing bias from unobserved sorting when estimating the effects of neighborhood attributes by employing a control function approach. However, they use neighborhood house prices as a control function. This control function will not fully absorb sorting bias unless the underlying amenities that determine neighborhood desirability can combined into a single index of neighborhood quality, an assumption unlikely to hold given heterogeneity in preferences for different aspects of communities.

were common to all members of the school) The second is a component of school quality that is valued based only on student characteristics that have nothing to do with school outcomes. We treat the variance of this residual-between-school component as a lower bound on the overall contribution of schools/neighborhoods to student outcomes. We also convert this lower-bound variance estimate into lower bound estimates of the impact on outcomes of moving from a school at the 10th quantile in the distribution of school outcome contributions to a 50th or 90th quantile school (a more intuitive scale).

While we focus on the consequences of school/neighborhood choice, the insights derived from our sorting model may be applied to any setting in which individuals sort non-randomly into units and in which outcomes depend on both individual and unit contributions. For example, a number of papers have attempted to estimate the impact of competition on hospital quality by running health outcome regressions that control for both hospital characteristics as well as average patient characteristics. Our model suggests that the coefficients on hospital characteristics will only reflect the impact of the part of the variation in these characteristics that was not valued by patients when choosing hospitals. Indeed, many such contexts may offer high quality administrative or observational data, but will not feature quasi-random variation.

Implementing our approach in the school context requires rich data on student characteristics for large samples of students from a large sample of schools, as well as longer-run outcomes for these students. We use four different datasets that generally satisfy these conditions: 3 cohort-specific panel surveys (NLS72, NELS88, and ELS2002) along with administrative data from North Carolina. The advantage of the surveys is that they are representative of the full distribution of American high schools (public and private) and American high school students, they collect a wide array of student and parent characteristics, and they follow students through college and beyond. Their limitation is that they sample only a medium-sized number of students at each school. The North Carolina administrative data contains information on the universe of public school students and public schools in North Carolina, but does not include private schools, and (to this point) does not contain longer-run labor market outcomes for students.

Our North Carolina results suggest that, averaging across the student population, choosing a 90th quantile school and surrounding community instead of a 10th quantile school increases the probability of graduation by at least 8.4 percentage points. We estimate these large average impacts despite the fact that our lower bound estimate can only attribute between 1 and 4 percent of the total outcome variance to schools with certainty. However, the average impact of moving to a superior school on binary outcomes such as HS graduation or college enrollment can be quite large even if differences in school quality are small, as long as a large pool of students are near the decision margin.

While estimates derived from the panel surveys are potentially less reliable due to sampling error in school average characteristics, we nonetheless recover similar estimates of school contributions to preventing dropouts. These estimates decrease slightly over time, since a decreasing average dropout rate implies that fewer students require a shift in school environment to be induced to stay in school. Estimates of the impact of a shift in school environment on the probability of enrolling in a four-year college and on the permanent component of adult wages (only in NLS72) are similarly large.

Section II presents a simple model of long-run student outcomes. Section III presents our model of school choice, and formally derives our key control function result. Section IV presents very preliminary simulations illustrating the properties of our lower bound estimator with finite numbers of schools and finite samples of students per school. Section V describes the four datasets we use to estimate the model of outcomes. Section VI presents our results. Finally, Section VII discusses other contexts in which our control function approach might be valuable.

2 A Model of Educational Attainment and Wage Rates

2.1 The Determinants of Adult Outcomes

In this section we present the underlying econometric model of adult outcomes that provides the basis for the variance decompositions that we present below. Our formulation draws loosely on theoretical discussions in the child development literature, the educational production function literature, and the neighborhood effects literature.[4] Let $Y_{s_i i}$ denote the outcome of student i. In our application the outcomes are high school graduation, attendance at a four-year college, a measure of years of postsecondary education, and the permanent wage rate. $Y_{s_i i}$ is determined according to

$$Y_{s_i i} = X_i^* \beta^* + Z_{s_i i}^* \Gamma^* + u_{s_i, i} \tag{1}$$

The coefficients β^* and Γ^* depend implicitly upon the specific outcome under consideration as well as the time period in the case of wages. The subscript s_i denotes the neighborhood and associated school of i for the high school years, which is chosen from $s = 1...S$. The vector X_i^* is a comprehensive set of child and family characteristics that have a causal impact on student i's educational attainment and wages. Examples include race, innate ability, personality traits, values, physical attractiveness, and parental education, income, and employment. Since X_i^* may include non-linear functions of these attributes, imposing that the individual attributes enter linearly is without loss of generality.

The vector $Z_{s_i i}^*$ is an exhaustive set of school and neighborhood influences experienced by student i. $Z_{s_i i}^*$ is partly determined by the family's choice of a neighborhood and school, which is characterized by a set of features $Z_{s_i}^*$ that partially shapes the distribution

[4] A good example is (Todd and Wolpin 2003), who provide references to the literature. See also (Cunha, Heckman, Lochner and Masterov 2006).

of environment the child experiences outside the home, including neighborhood quality, school resources, and peers inside and outside of school. However, $Z^*_{s_i i}$, also varies within a school attendance area and within a school itself. Examples include the trustworthiness of immediate neighbors and distinct course tracks at a school. Some of the within-school variation is related to parent and child characteristics, X^*_i Some reflects random influences, such as random variation in the quality of teaching the child receives and random variation in peer influences. A simple way to capture the dependence of $Z^*_{s_i}$ on X^*_i, $Z^*_{s_i}$and other factors is through the equation

$$Z^*_{s_i i} = X^*_i \Pi_{Z^* X} + Z^*_{s_i} \Pi_{Z^* Z^*} + \tilde{Z}^*_{s_i i}.$$

where $\Pi_{Z^* X}$ and $\Pi_{Z^* Z^*}$ are the weights on X^*_i and $Z^*_{s_i}$ in determining $Z^*_{s_i i}$ and $\tilde{Z}^*_{s_i i}$ is purely ideosyncratic variation that is orthogonal to X^*_i and $Z^*_{s_i}$.[5] Because we are interested in the consequence of experiencing one school/neighborhood versus another, notably the consequences of variation in the neighborhood and school characteristics $Z^*_{s_i}$, we substitute for $Z^*_{s_i i}$ in (1) and re-write that equation as

$$Y_{s_i,i} = X^*_i B^* + Z^*_{s_i} G^* + \tilde{z}_{s_i,i} + u_{s_i,i} \tag{2}$$

where $B^* \equiv \beta^* + \Pi_{Z^* X^*} \Gamma^*$, $G^* \equiv \Gamma^* + \Pi_{Z^* Z^*} \Gamma^*$, and $\tilde{z}_{s_i,i} \equiv \tilde{Z}^*_{s_i i} \Gamma^*$.

The variable $u_{s_i,i}$ captures other influences on student i's outcome that are determined after secondary school that are not predictable given X_{s_i}, $Z^*_{s_i} G^*$, and $\tilde{Z}_{s_i i}$. It will prove useful to write $u_{s_i,i}$ as u_{s_i}+ u'_i, where u_{s_i} is specific to s_i and u'_i is idiosyncratic. u_{s_i} captures the effects of local shocks late in high school or during college that influence the

[5]Throughout the paper, we use the symbol Π_{DQ} to denote the vector of the partial regression coefficients relating a dependent variable or vector of dependent variables D to a vector of explanatory variables Q, holding the other variables that appear in the regression constant. In the above example, $D = Z^*_{s,i}$ and Q is X in the case of $\Pi_{Z^* X}$ and $Z^*_{s_i}$ in the case of $\Pi_{Z^*_s X}$.

decision to stay in school (such as the opening or expansion of a local college), as well as local labor market shocks after labor market entry when wages are the outcome.

Suppose we had access to data at a single point in time on each of the myriad components of X_i^* and $Z_{s_i}^*$ and were able to estimate equation (2). How would we interpret B^*? One must first realize that some components of $X_{s_i}^*$ associated with student inputs (for example, student aptitude) are determined in part by parental inputs from earlier periods (for example, parent income), as well as school and neighborhood inputs from earlier periods (for example, quality of elementary school facilities).. Likewise, parents' income may in part be determined by student aptitude and behavior, if parents work less in order to tutor their child. Such links make it difficult to interpret the coefficient associated with a given component of $X_{s_i}^*$, since once we have conditioned on the other components, we have removed many of the avenues through which the component determines Y. Furthermore, B^* captures an indirect effect of X_i^* on the environment experienced by i given choice of community. Consequently, we do not make any attempt to interpret individual components of the coefficient vector B^*, and thus do not attempt to tease apart the distinct influences of child characteristics, family characteristics, and early childhood schooling inputs, respectively. We aim instead to separate the effects of high schools and associated community influences on outcomes from student, family, and prior school/community factors (although some of our specifications using data from NELS88 examine the impact on outcomes of 8th grade schools rather than high schools). If the inequalities in outcomes are primarily attributable to differences in high school quality (as opposed to the other three classes of inputs), then policies designed to equalize high school quality have the potential to close the outcome gaps we observe.

To be more specific about what we mean by school/neighborhood effects, note that if students attended school s^1 rather than s^0, the expected difference in the outcome is $Z_{s^1}^* G^* - Z_{s^0}^* G^*$. The outcomes of a specific student i will also differ across schools because the values of $\tilde{z}_{s_i,i}$ and u_{s_i} will differ, but the former are entirely idiosyncratic and the latter common to those who attend s_i but are determined after high school is chosen. We wish to quantify the importance of differences across neighborhoods in $Z^*_{s_i}G^*$. Some of our estimates will also include differences due to u_{s_i}.

2.2 Toward an Empirical Model

In this section we discuss what parameters OLS recovers when outcomes are regressed on the subset of X^*_i and $Z^*_{s_i}$ that can be observed in a survey or administrative dataset.

If we had complete data on X^*_i and $Z^*_{s_i}$, estimation of B^* and G^* would be a straightforward exercise. With data on $\tilde{Z}_{s_i i}$, we could do even more. In practice, our measures of external influences are common to all students attending a particular high school and thus vary over schools but not within schools, so we don't observe $\tilde{Z}_{s_i i}$.[6] Also, we actually observe and make use of only a subset of the elements of X. For example, there are many characteristics of the student (e.g., physical attractiveness and temperament) and parents (e.g., parenting skill and time allocation during early childhood) that we do not measure at all. Furthermore, we only measure child and family variables at a single point in time, rather than at various stages of the child's life. Finally, we only measure and use a subset of $Z^*_{s_i}$, the external influences that are common to all students attending a particular high school and thus vary across schools but not within them.

Partition X^*_i into the observed variables X_i and unobserved variables X^u_i and partition $Z^*_{s_i}$ into Z_{s_i} and $Z^u_{s_i}$. Partition B^* into the subvectors B and B^u and partition G^* into G and G^u. Without loss of generality, redefine X^u_i so that it is orthogonal to X and redefine

[6]We do observe classmates and teacher assignments in the North Carolina data, and so it would be possible examine at least a portion of the variation in $\tilde{Z}_{s_i i}$. We leave this to future work. (Mansfield Forthcoming) examines the distribution across students in the quality of teachers that they experience. One could also examine variation in classroom peer characteristics.

Z^u to be orthogonal to Z_{s_i}. Redefine B and B^u and G and G^u accordingly. Then

$$Y_{s_i i} = X_i B + Z_{s_i} G + X_i^u B^u + Z_{s_i}^u G^u \ + v_{s_i i}^* \text{ ..} \tag{3}$$

where $v_{s_{i,i}}^* \equiv \tilde{z}_{s_i i} + u_{s_i i}$.

Project $X_i^u B^u$ onto X_i^u and Z_{s_i}:

$$X_i^u B^u = X_i \Pi_{X^u X} B^u + Z_{s_i} \Pi_{X^u Z} B^u + \tilde{X}_i^u B^u \tag{4}$$

where $\Pi_{X^u X}$ and $\Pi_{X^u Z}$ are the projection coefficients and $\tilde{X}_i^u$ is the residual vector from that regression. Because X_i^u is orthogonal to X_i, $\Pi_{X^u X}$ is a function of $cov(Z_{s_i}, X_i^u B^u)$, as we discuss further below.

One may easily show that coefficients of the regression of $Z_i^u G^u$ on X_i and Z_{s_i} are both 0.[7] Using this fact and (4), one may write (3) as

$$Y_{s_i i} = X_i[B + \Pi_{X^u X} B^u] + Z_{s_i}[G + \Pi_{X^u Z} B^u] + \tilde{X}_i^u B^u + Z_i^u G^u + \ v_{si}^*. \tag{6}$$

[7]Using notation corresponding to (4), one may write

$$Z_{s_i}^u G^u = X_i \Pi_{Z^u X} G^u + Z_{s_i} \Pi_{Z^u Z} G^u + \tilde{Z}_{s_i}^u G^u \tag{5}$$

However, $\Pi_{Z^u X}$ and $\Pi_{Z^u Z}$ are both 0, because $Z_{s_i}^u$ is orthogonal to Z_{s_i} by definition, the school mean X_{s_i} is part of Z_{s_i}, and $X_i - X_{s_i}$ is uncorrelated with $Z_{s_i}^u G^u$ by construction. To see this, partition Z_{s_i} into X_{s_i} and Z_{2_i} and rewrite (5) in terms of X_{s_i} and Z_{2s_i}. This leads to

$$\begin{aligned} Z_{s_i}^u G^u &= X_i \Pi_{Z^u X} G^u + X_{s_i} \Pi_{Z^u X_s} G^u + Z_{2s_i} \Pi_{Z^u Z_2} G^u + \tilde{Z}_{s_i}^u G^u \\ &= [X_i - X_{s_i}] \Pi_{Z^u X} G^u + X_{s_i}[\Pi_{Z^u X} G^u + \Pi_{Z^u X_s} G^u] + Z_{2s_i} \Pi_{Z^u Z_2} G^u + \tilde{Z}_{s_i}^u G^u \end{aligned}$$

Both X_{s_i} and Z_{2s_i} have a covariance of 0 with $Z_{s_i}^u$ by definition of $Z_{s_i}^u$. $[X_i - X_{s_i}]$ also has 0 covariance with $Z_{s_i}^u$. Thus, all three variables have a covariance of 0 with $Z_{s_i}^u G^u$. Consequently, all three regression coefficients are 0.

Thus $B' \equiv (B + \Pi_{X^u X} B^u)$ and $G' \equiv (G + \Pi_{X^u Z} B^u)$ are the parameters identified by an OLS regression $Y_{s_i i}$ on X_i and Z_{s_i}.

It is clear that in general, $var(Z_s G')$ differs from the variance of $var(Z^*_{s_i} G^*)$. On one hand, $var(Z_s[G + \Pi_{X^u Z} B^u])$ will tend to overstate $var(Z^*_{s_i} G^*)$ to the extent that G' are biased upward by correlation between Z_s and $X_i^u B^u$ conditional on X_i. On the other hand hand, $var(Z_{s_i}[G + \Pi_{X^u Z} B^u])$ will tend to underestimate $var(Z^*_{s_i} G^*)$ because it does not include the effect of $\tilde{Z}^u_{s_i} G^u$, which is part of the error term. Without further assumptions, one cannot use the variance across schools in the composite component $\tilde{X}_i^u B^u + \tilde{Z}_i^u G^u + v^*_{s_i i}$ to identify the contribution of schools/neighborhoods because the composite will also include the variation across schools in $\tilde{X}_i^u B^u$. The school average $\tilde{X}_i^u B^u$ is not a school/neighborhood effect. This is the essence of the problem of distinguishing school/neighborhood effects from composititon effects. To say anything more specific requires a model of sorting, to which we now turn.

3 A Multinomial Model of School Choice and Sorting

In this section we present a model of how parents/students choose schools, with the goal of placing minimal structure on parental preferences for schools. Suppose that each location s, $s \in \{1, ..., S\}$, can be characterized by a vector of K underlying amenities, $\{A^*_{1s}, \ldots, A^*_{Ks}\}$ and the price of neighborhood s, P_s.

The expected utility for the parents of student i from choosing school/neighborhood s net of the opportunity cost is denoted by $U_i(s)$, which takes the form

$$U_i(s) = \gamma_{1i} A^*_{1s} + \gamma_{2i} A^*_{2s} + \cdots + \gamma_{Ki} A^*_{Ks} - \gamma_{Pi} P_s + \epsilon^*_{s,i}. \tag{7}$$

In the above equation the $\{\gamma_{ki}\}$ are the weights that the family of i places on the various amenities, γ_{Pi} is the marginal utility of income, and $\epsilon^*_{s,i}$ is an idiosyncratic taste of the parent/student i for the particular location s. Assuming monotonicity, we can define the units of the A^*_{ks} so that utility is linear in the variable. However, we are imposing additive separability. Expected utility is taken with respect to the information available when s is chosen. The information set includes the price, the amenities, and X^*_i, but not $v^*_{s,i}$, which is an index of determinants of Y_{si} that are orthogonal to X^*_i and are determined after the start of secondary school or later, or components of neighborhood and school quality that are not observable to families when location is chosen. The set of amenities may include school/neighborhood characteristics that influence educational attainment and labor market outcomes.

The strength of tastes for particular amenities varies across parent/student combinations. In particular, each taste parameter γ_{ki} and the marginal utility of income γ_{Pi} can be written as:

$$\gamma_{ki} = \sum_{\ell=1}^{L} \delta^*_{k\ell} x^*_{\ell i} + \kappa^*_{ki};, k = 1, ..., K;, \gamma_{Pi} = \sum_{\ell=1}^{L} \delta^*_{P\ell} x^*_{\ell i} + \kappa^*_{Pi};, \tag{8}$$

where $x^*_{\ell i}$ is the ℓ^{th} element of X^*_i Thus, $\delta^*_{\ell k}$ captures the extent to which the utility weight on amenity k depends on the determinant $x^*_{\ell i}$ of student outcomes. The variable κ^*_{ki} is the component of i's taste for amenity k that is unpredictable given X^*_i. Since X^*_i is the complete set of student variables that determine $Y_{s,i}$, this means that κ^*_{ki} influences school choice but has no direct effect on student outcomes. Using (8), collecting the terms involving the κ^*_{ki}, and combining them with ϵ^*_{si}, one may rewrite (7) as

$$U_i(s) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \delta^*_{k\ell} x^*_{\ell si} A^*_{ks} + \epsilon_{si} - \gamma_{P,i} P(A^*) \tag{9}$$

where

$$\epsilon_{si} = \sum_{k=1}^{K} \kappa_{ki}^* A_{ks}^* + \epsilon_{si}^*$$

Parents i choose the school s_i if net utility $U_i(s_i)$ is the highest among the options. That is, s_i is determined by

$$s_i = \arg \max_{s=1,..,S} U_i(s) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \delta_{k\ell}^* x_{\ell si}^* A_{ks}^* + \epsilon_{si} - \gamma_{P,i} P(A^*)$$

Let L^o and L^u be the number of elements of X_i and X_i^u, respectively, where $L = L^o + L^u$. Rewrite equation (9) using matrix notation as

$$U(X_i, X_i^u, \kappa_i^*) = X_i \boldsymbol{\delta}_x A^* + X_i^u \boldsymbol{\delta}_{x^u} A^* + \kappa_i^* A^* - \gamma_{P,i} P(A^*) + \epsilon_{si}^* \tag{10}$$

where $\boldsymbol{\delta}_x$ is an $L^o \times K$ matrix with the k-th column $[\delta_{1k}, ..., \delta_{L^o k}]'$, $\boldsymbol{\delta}_{x^u}$ is an $L^u \times K$ matrix with the k-th column $[\delta_{1k}, ..., \delta_{L^u k}]'$, A^* is the $K \times 1$ matrix $[A_1^*, \ldots, A_K^*]'$, and $\kappa_i^* = [\kappa_{i1}^*, ..., \kappa_{iK}^*]$.

3.1 The Link Between Sorting on Observables and Unobservables

We now show that the school average of the unobservables X_s^u is an exact linear function of X_s. A necessary (but not sufficient) condition for the result that captures the essence is that if preferences for an amenity depend on X_i^u, they must also depend on at least one element of X_i. The result provides the justification for using X_s to control for X_s^u when assessing the importance of school/neighborhood effects on outcomes.

For analytic simplicity, assume that S is sufficiently large so that it can be well approximated by a continuum of neighborhoods that create a continuous joint distribution of amenities A^*. Thus, choosing a school is equivalent to choosing the vector of amenities that maximizes utility, given the price function $P_s = P(A^*_s)$. Assume further that in equilibrium $P(A)$ is an increasing convex function, so that prices rise at an increasing rate as amenities increase.[8]

Under these assumptions, the choice of school/neighborhood is characterized by a system of first order conditions, one for each amenity factor. The conditions are:

$$\delta_x^T X_i^T + \boldsymbol{\delta}_{x^u}^T X_i^{uT} + \kappa_i^{*T} = \gamma_{P,i} \nabla P(A^*_{s_i})$$

where the T superscript is the transpose operator and $\nabla P(A_{s_i})$ is the $K \times 1$ column vector of partial derivatives of $P(A^*)$ with respect to A^* evaluated at $A^* = A^*_{s_i}$. Since $P(A^*_{s_i})$ is strictly convex, second order conditions will be satisfied.

Note that the system of FOCs is a K-dimensional linear function of the row vectors X_i, X_i^u, and κ_i^* (since $\gamma_{P,i} \nabla P(A^*_{s_i})$ is constant across choices for i).

To see what choice of school/neighborhood s_i implies about the relationship between $X^u_{s_i}$ and X_{s_i}, it is useful to consider the projection of X_i^T and X_i^{uT} onto the index that determines location preference, $X_i^T \boldsymbol{\delta}_x + X_i^{uT} \boldsymbol{\delta}_x + \kappa_i^{*T}$. Note that since (1) X_i^u is uncorrelated with X_i by definition, and (2) κ_i^* is uncorrelated with both X_i and X_i^u, we have:

$$Cov(X_i \boldsymbol{\delta}_x + X_i^u \boldsymbol{\delta}_x + \kappa_i^*, X_i) = \boldsymbol{\delta}_x^T Var(X_i)$$
$$Cov(X_i \boldsymbol{\delta}_x + X_i^u \boldsymbol{\delta}_{x^u} + \kappa_i^*, X_i^u) = \delta_x^{uT} Var(X_i^U).$$

[8]We are making the implicit assumption that convexity of the price function holds for the tranformation of the A^*_{ks} that leads to the linear utility function above. This would hold if, for example, costs of producing amenities are linear and there is diminishing marginal utility in them.

Thus, these projections can be written as:

$$X_i = [X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*]Var(X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)^{-1}Cov(X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*, X_i) + error_i$$
$$= [X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*]Var(X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)^{-1}\boldsymbol{\delta}_x^T Var(X_i) + \ error_i$$

and

$$X_i^u = [X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*]Var(X\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)^{-1}Cov(X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*, X_i^U) + error_i$$
$$= [X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*]Var(X\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)^{-1}\boldsymbol{\delta}_{x^u}^T Var(X_i^u) + error_i$$

Using the set of first order conditions provide above, we can rewrite these equations as:

$$X_i = \gamma_{P,i}\nabla P(A_{s_i}^*)^T[Var(X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)]^{-1}\boldsymbol{\delta}_x^T Var(X_i) + \ error_i$$
$$X_i^u = \gamma_{P,i}\nabla P(A_{s_i}^*)^T[Var(X\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)]^{-1}\delta_{x^u}^T Var(X_i^u) + error_i$$

Note that since choice of s_i depends on X_i, X_i^u, and κ_i^* only through the function $[X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*]$, the error terms in the vector equations above are unrelated to s_i. Taking conditional expectations of both sides of the above equations with respect to the chosen school s_i, we obtain:

$$X_{s_i} \equiv E(X_i|s_i) = E(\gamma_{P,i}|s_i)\nabla P(A_{s_i}^*)^T[Var(X_i\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)]^{-1}\boldsymbol{\delta}_x^T Var(X_i) \quad (11)$$

$$X_{s_i}^u \equiv E(X_i^u|s_i) = E(\gamma_{P,i}|s_i)\nabla P(A_{s_i}^*)^T[Var(X\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)]^{-1}\delta_{x^u}^T Var(X_i^u) \quad (12)$$

Next, suppose that the columns of $\boldsymbol{\delta}_{x^u}$ are spanned by $\boldsymbol{\delta}_x$, which means that one can write $\boldsymbol{\delta}_{x^u}^T$ as

$$\boldsymbol{\delta}_{x^u}^T = \boldsymbol{\delta}_x^T M \tag{13}$$

for some matrix : $L^o \times L^u$ matrix M. To understand the restriction, consider the k-th row of $\boldsymbol{\delta}_{x^u}^T$. Note that the k-th row of $\boldsymbol{\delta}_{x^u}^T$ (k-th column of $\boldsymbol{\delta}_{x^u}$) is the vector of coefficients that determines how each of the L^u elements of X_i^u affect preferences for amenity A_k^*. The kth row of $\boldsymbol{\delta}_x^T$ is the corresponding set of coefficients that determine how preferences for A_k^* shift with the various elements of X_i. For each k, the restriction says that each of the weights $\delta_{x^u,\ell k}$, $\ell = 1, ..., L^u$, can be written as a linear combination of the weights $\delta_{x,\ell' k}$, $\ell' = 1, ..., L^o$. This is possible if at least one element of X affects taste for A_k^*. It will hold trivially if preferences for A_k^* do not depend on X_i^u, in which case the k-th row of $\boldsymbol{\delta}_{x^u}^T = 0$.

However, the restriction requires more than this. It would fail if the column rank of $\boldsymbol{\delta}_x$ is less than the column rank of $\boldsymbol{\delta}_{x^u}$. For example, suppose that X_{i1} influences tastes for all amenities for which γ_k depends on X^u,but none of the other elements of X_i influence tastes. In that case, all columns of $\boldsymbol{\delta}_x$ would be 0 except for the first, and the restriction will fail if X_i^u influences tastes for more than one amenity.

Substituting for $\boldsymbol{\delta}_{x^u}^T$ implies that

$$\begin{aligned} X_{s_i}^u &= E(X_i^u|s_i) = E(\gamma_{P,i}|s_i)\nabla P(A_{S_i}^*)^T[Var(X\boldsymbol{\delta}_x + X_i^u\boldsymbol{\delta}_{x^u} + \kappa_i^*)]^{-1} \\ &\quad \delta_x^T Var(X_i)[Var(X_i)]^{-1} M Var(X_i^u) \\ &= X_{s_i}[Var(X_i)]^{-1} M Var(X_i^u) \end{aligned}$$

where the third line follows from substitution using (11).

Thus, the vector $X_{s_i}^u$ will be an exact linear function of X_{s_i} if the coefficient vectors relating tastes for amenities to the X_i^u are linear combinations of the coefficient vectors relating tastes for amenities to the observables X_i. Remarkably, this is true even if X_i^u is uncorrelated with X_i.

An interesting special case is the one in which $\boldsymbol{\delta}_{x^u} = 0$, so that unobservable characteristics do not affect location preferences. When $\boldsymbol{\delta}_{x^u} = 0$, $M = 0$. One can see from the above equation or directly by substituting $\boldsymbol{\delta}_{x^u} = 0$ into (12) that in this special case $X_{s_i}^u = 0$, so that there is no variation in average unobservable characteristics across schools. That is, there is no sorting on $X_{s_i}^u$. In this case $Var(Z_s G)$ will accurately reflect the school/neighborhood contribution to outcomes.

We summarize the result in the following proposition.

Proposition 1: *Assume (i) preferences are given by (10), (ii) the price function $P(A^*)$ is increasing and strictly convex in A^*, and (iii) the columns of the coefficient matrix δ_{x^u} relating tastes for A^** to X^u *are spanned by the columns of the coefficient matrix δ_x relating tastes for A^* to X_i. Then the expectation $X_{s_i}^u$ is linearly dependent on the expectation $X_{s_i}^o$.*

The finding that $X_{s_i}^u$ is linear function of X_{s_i} suggests that school averages of student characteristics can serve as an effective control function that purges estimates of school contributions of the influence of sorting on unobservable characteristics.

However, there is a price for using X_{s_i} in this way. X_{s_i} spans the space of $X_{s_i}^u$ *because* the variation in both X_{s_i} and $X_{s_i}^u$ is driven by the same underlying variation in the desired amenity vector, $\{A_{1s}^*, A_{2s}^*, ...A_{K,s}^*\}$. Re-examining equations 11 and 12 above, we see that if δ_X (δ_X^U) is full rank, then the vector $\mathbf{A}_s = [A_{1s}^*, A_{2s}^*, ...A_{K,s}^*]'$ can also be written as a linear function of X_{s_i} ($X_{s_i}^u$).

Hence, while the inclusion of X_{s_i} in the estimated specification is effective in controlling for unobserved sorting, it also absorbs the variation in the underlying amenity factors for which X_{s_i} affects tastes. Given that parents are likely to value the contributions of schools to student outcomes, many of the characteristics that affect school quality are likely to be reflected in $\{A^*_{1s}, A^*_{2s}, ...A^*_{K,s}\}$.

To the extent that a component of school/neighborhood quality $Z^*_s G^*$ is unknown (or unvalued) by parents at the time the school/neighborhood is chosen, though, this component will not be reflected in the vector of amenities $A^*_{1s}, A^*_{2s}, ...A^*_{K,s}$ that are the basis of choice. However, it will still produce variation in average outcomes across schools. Similarly, if the outcome is measured after high school is completed, any common shocks that affect the outcomes of all those who attended a particular high school will also not be absorbed by X_{s_i}, yet will produce between-school variation in outcomes.

Proposition 1 and the above discussion has important implications concerning the common practice of controlling for average characteristics of students in a school (or class room) in addition to the characteristics of the individual students when estimating school (or teacher) value added. There are many good reasons to include such controls, including peer effects. However, our analysis implies that including these controls may lead to an understatement of the school or teacher value added if preferences for school quality (or teacher quality) in choosing a school or a teacher depends on observed student characteristics.

4 Estimating the Contribution of Schools and Neighborhoods

4.1 Variance Decomposition

In the empirical work below, we estimate models of the form

$$Y_i = X_i\beta + Z_1G_1 + Z_2G_2 + Z_s^U G^U + v^*_{s_i i} \tag{14}$$

, where $Z_{1s} = \overline{X}_{s_i}$ is a vector of school-averages of student characteristics, and Z_{2s} is a vector of observed school characteristics (such as school size or student-teacher ratio).

Consider rewriting this estimating equation as:

$$Y_i = (X_i - X_{s_i})\beta + X_s\beta + Z_{1s}G_1 + Z_{2s}G_2 + Z_s^U G^U + (v^*_{s_i i} - v^*_{s_i}) + v^*_{s_i} \tag{15}$$

Then we can decompose the variance in Y_i into observable and unobservable components of both within- and between- school variation via:

$$Var(Y_i) \tag{16}$$

$$= Var(Y_i - Y_s) + Var(Y_s) \tag{17}$$

$$= [Var((X_i - X_{s_i})\beta) + Var(v^*_{s_i i} - \overline{v^*}_{s_i})] + \tag{18}$$

$$[Var(X_s\beta) + 2Cov(X_s\beta, Z_{1s}G_1) + 2Cov(X_s\beta, Z_{2s}G_2) + Var(Z_{1s}G_1) + \tag{19}$$

$$2Cov(Z_{1s}G_1, Z_{2s}G_2) + Var(Z_{2s}G_2) + Var(Z_s^U G^U + \overline{v^*}_{s_i})] \tag{20}$$

Motivated by the model of sorting presented in Section III, we introduce two alternative lower bound estimates of the contribution of school/neighborhood choice to student outcomes.

First, due to the presence of Z_{1s}, $Var(Z_2G_2) + Var(Z_s^U G^U + v_{s_i}^*)$ will be purged of any effects of student sorting (observable or unobservable), so that it isolates only school/neighborhood factors.

However, in addition to unobserved school characteristics, $Var(Z_s^U G^U + v_{s_i}^*)$ will include common location-specific shocks (such as local employment demand shocks) that occur after high school has been completed for the chosen cohort. To the extent that these shocks should not be attributed to schools (since one could argue that they are beyond the control of school or town administrators), we also consider a second, more conservative lower bound estimate: $Var(Z_2G_2)$. This estimate only attributes to schools/neighborhooods the part of residual between-school variation that could be predicted based on observable characteristics of the schools at the time students were attending. This estimate removes true school quality variation that is orthogonal to observed characteristics, but also removes any truly idiosyncratic local shocks that occur after graduation.

Appendix Sections 1 and 2 describe the process by which the coefficients β, G_1, and G_2 are estimated, as well as the process by which the empirical variance decomposition is performed. The implementation differs depending on whether the outcome is binary or continuous.

4.2 Measuring the Effects of Shifts in School/Community Quality

The variance decompositions provide a good indication of the importance of school/community factors relative to student-specific factors. However, the effect of a shift in school/community quality from the left tail of the distribution to the right tail of the distribution might be socially significant even if most of the outcome variability is student-specific. This is particularly true in the case of binary outcomes such as high school graduation and college enrollment, where many students may be near the decision margin. Below we report estimates of the effect of a shift in the unknown/unvalued component of school/neighborhood quality from 1.28 standard deviations below the mean to 1.28 standard deviations above the mean. This would correspond to a shift from the 10th percentile to the 90th percentile if this component has a normal distribution. We interpret these as lower bound estimates of the change in outcomes from a 10th-to90th quantile shift in the full distribution of student/neighborhood quality.

The more comprehensive estimate measures the unknown component of schoo/neighborhood quality via $\widehat{Var}(Z_2G_2 + Z_s^U G^U + v_{s_i}^*)$, while the more conservative estimates that attempt to remove common shocks use $\widehat{Var}(Z_2G_2)$.

For the binary outcomes, we estimate the effect of the shift in Z_2G_2 as a weighted average over individuals i:

$$E^{noshocks}[\hat{Y}^{90} - \hat{Y}^{10}] = \tag{21}$$

$$\Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_2\hat{G}_2 + 1.28(Var(Z_2G_2))^{.5}]}{(1 + Var(Z_s^U G^U + v_{s_i}^*))^{.5}}) \tag{22}$$

$$- \Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_2\hat{G}_2 - 1.28(Var(Z_2G_2))^{.5}]}{(1 + Var(Z_s^U G^U + v_{s_i}^*))^{.5}}) \tag{23}$$

This weighted average effectively integrates over the distribution of $X_i\beta + Z_1G_1 + Z_s^U G^U + v_i^*$, but uses the empirical distributions of $X_i\beta$ and Z_1G_1 (since they are observed) instead of imposing normality.

We estimate the effect of the shift in $Z_2G_2 + Z_s^U G^U + v_{s_i}^*$ analogously via:

$$E^{w/shocks}[\hat{Y}^{90} - \hat{Y}^{10}] = \tag{24}$$

$$\Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_2\hat{G}_2 + 1.28(Var(Z_2G_2 + Z_s^U G^U + v_{s_i}^*))^{.5}]}{(1)}) \tag{25}$$

$$- \Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_2\hat{G}_2 - 1.28(Var(Z_2G_2 + Z_s^U G^U + v_{s_i}^*))^{.5}]}{(1)}) \tag{26}$$

We also report lower bound estimates of the impact of a shift from a school at the 10th percentile of quality to one at the 50th percentile. For the binary outcomes, the impact of a 10th-90th percentile shift in either Z_2G_2 or $(Z_2G_2 + Z_s^U G^U + v_{s_i}^*)$ will depend on the values of a student's observable characteristics, X_iB. Thus, we report average impacts for certain subpopulations of interest as well.

5 Monte Carlo Simulations (very preliminary)

While Proposition 1 provides a strong theoretical foundation for our control function approach to distilling school contributions to long run outcomes, it is derived from a continuous, infinite dimensional model of school choice. Furthermore, Proposition 1 refers to links between the link between the expectations of X and X^u given A_s. Given random variation associated with κ_i^*, ϵ_{si}^*, and $\gamma_{P,i}$, the realizations for a school/neighborhood a point in time might differ from the expectations.

In this section, we provide suggestive evidence that our control function approach is still viable in a context with a finite number of schools and students per school. In particular, we present the results of a series of monte carlo simulations that explore the properties of our control function across a number of key dimensions. These simulations are not intended to represent a rigorous analysis of the finite sample properties of our estimator. Instead, we focus on a stylized test case that merely serves to 1) illustrate that the control function approach has the potential to be effective in settings where a large population sorts into a fairly large number of groups and 2) highlight a few key factors that play a major role in determining the degree to which average values of observable characteristics effectively control for average values of unobservable characteristics.

5.1 Methodology

Note that while our proof of Proposition 1 does not require an analytically characterization of the equilibrium sorting of students to schools, simulating the model does require us to compute a large scale spatial equilibrium. Consequently, we alter the model presented in Section III slightly to ease the computation.

Specifically, we convert the utility function from 7 into a money-metric utility function by dividing both sides by the individual's marginal utility of income, $\gamma_{P,i}$:

$$U_i^{MM}(s) = \tilde{\gamma}_{1i} A_{1s}^* + \tilde{\gamma}_{2i} A_{2s}^* + \cdots + \tilde{\gamma}_{Ki} A_{Ks}^* + \tilde{\epsilon}_{s,i} - P_s \tag{27}$$

where

$$\begin{aligned} \tilde{\gamma}_{ki} &= \frac{\gamma_{ki}}{\gamma_{Pi}} \; \forall \; k \leq K \\ \tilde{\epsilon}_{si}^* &= \frac{\epsilon_{si}}{\gamma_{Pi}} \end{aligned} \tag{28}$$

and we recall that $\epsilon_{si} \equiv \sum_{k=1}^{K} \kappa_{ki}^* A_{ks}^* + \epsilon_{si}^*$.

We can then project these renormalized coefficients $\tilde{\gamma}_{ki}$ onto the set of observable and unobservable characteristics:

$$\tilde{\gamma}_{ki} = \sum_{\ell=1}^{L} \tilde{\delta}^*_{k\ell} x^*_{\ell i} + \tilde{\kappa}^*_{ki} \; \forall \; k < K \tag{29}$$

We can then re-express preferences via

$$U_i(s) = \sum_{k=1}^{K-1} \sum_{\ell=1}^{L} \tilde{\delta}^*_{k\ell} x^*_{\ell si} A^*_{ks} + \tilde{\epsilon}_{si} - P_s \; ,$$

where

$$\tilde{\epsilon}_{si} = \sum_{k} \kappa^*_{ki} A^*_{ks} + \tilde{\epsilon}^*_{si}.$$

This can be re-written as

$$U_i(s) = WTP_i(s) - P_s \; , \tag{30}$$

where $WTP_i(s)$ is i's willingness to pay for s.

The simulation results are presented in Table 1. The full spatial equilibrium sorting of students to schools depends on all the elements of joint distribution of $[X_i, X_i^U, \kappa_i]$ as well as the joint distribution of the amenities A^*_s and the distribution of the idiosyncratic tastes ϵ^*_{is}. Rather than attempt to provide a full characterization of how the finite sample properties depend on all of these relationships, we instead consider a stylized but conservative case in which 1) all of the elements of $[X_i, X_i^U, \kappa_i]$ are i.i.d and normally distributed (so that each characteristic is orthogonal to all the others), 2) all of the amenities A^*_s are i.i.d and normally distributed, 3) The constants $\tilde{\delta}_{k\ell}$ represent draws from a multivariate normal distribution with an identity variance matrix, and 4) $\epsilon^*_{is} = 0 \; \forall \; (i, s)$. These restrictions correspond to a scenario in which there is considerable sorting into schools along many dimensions of school amenities and along many observable and unobservable dimensions of student quality. It represents a conservative case because one might expect that in reality a few key observable (and unobservable) individual level factors (e.g. parental income, education, and wealth) and a few key school/neighborhood amenities (ethnic composition, crime, principal quality) drive most of the systematic sorting of students to schools. Given these restrictions, the model can be completed by choosing particular sets of 6 remaining parameters. The first parameter sets the total number of students choosing a school/neighborhood. It is denoted "# Stu" in Table 1. We choose among three values: 25,000, 50,000, and 100,000. The second parameter sets the total number school/neighborhood combinations available (denoted "# Sch") to be either 50 or 100. The combination of these two parameters determine the average number of students per school (either 250,500,1000, or 2000). For simplicity, we impose that each school has capacity equal to this average student/school ratio.[9]

The third parameter (denoted "# Cons."), captures the number of schools in the consideration set for each household. This captures the possibility that most parents only realistically research a limited number of possible locations. We implement this by distributing schools uniformly throughout the unit square, and drawing a random latitude/longitude combination for each household. The households then consider the preset number of schools that are closest to their location. Thus, consideration sets of different households are overlapping.

The fourth and fifth parameters (denoted "# Obs." and "# Unobs.") capture the number of observed and unobserved characteristics that affect outcomes. The sixth parameter determines the dimension of the amenity vector over which households have preferences, which is set at either 10 or 20. It is always less then or equal to the number of observed characteristics, so that $\boldsymbol{\delta}_x$ may span $\boldsymbol{\delta}_x^u$, as required by Proposition 1.

The seventh column, denoted "R-sq (All)", presents the R-squared from a regression of

[9]We believe that this is essentially without loss of generality. Without a finite elasticity of supply of land/school vacancies though, it is hard to avoid having tiny school sizes in locations with low values of amenities that tend to be highly desired. Fixed costs would prevent this.

the total contribution of school-averages of unobservable characteristics ($\sum_m X^u_{s_i m}\beta^u$) to each school's average outcome (which captures the potential bias from unobservable sorting) on the full vector of school-averages of observed characteristic, X_{s_i}. The R-squared should converge to 1 as the number of students per school gets large. However, the rate at which it does so is important for the efficacy of the control function approach.

The 8th, 9th, and 10th columns (denoted "R-sq (10)", "R-sq (20)", and "R-sq (40)", respectively) capture R-squared calculated when random samples of 10, 20, or 40 students from each school are used to estimate the regression (and calculate the school averages X_{s_i} and $X^u_{s_i}$).

We draw X_i, X^u_i, κ^*_i, and $\{\epsilon_{is}\}$ from the distributions described above to calculate the willingness-to-pay of each household for each school.[10] Since our method does not require observation of the equilibrium price function $P(A^*)$, rather than iterating on an excess demand function to find the equilibrium matching, we instead exploit the fact that a perfectly competitive market will always lead to a pareto efficient allocation. The problem of allocating students to schools to maximize total consumer surplus can be written as a linear programming problem, and solved quickly at relatively large scale using the simplex method combined with sparse matrix techniques. [11]

5.2 Simulation Results

The first takeaway from the set of simulations is that our control function approach is effective even with reasonably-sized school sizes (most of the schools in the North Carolina sample enroll between 250 and 2000 students) and a moderate number of available schools.

[10]To minimize the statistical "chatter" introduced by the particular $\tilde{\delta}$ matrix that we happened to draw, we drew six different $\tilde{\delta}$ matrices from the prescribed distribution, ran the simulations for all 36 parameter sets for each of these matrices, and then averaged the results across the six iterations within each parameter set.

[11]The problem can actually be classified as a binary assignment problem (a subset of linear programming problems), but we were unable to implement the standard binary assignment algorithms at scale.

For each of the 36 parameter sets we simulated, at least 83 percent of the variance in the school-level contribution of unobserved student characteristics can be predicted by a linear combination of school-average observable characteristics, with the R-squared exceeding .98 in several specifications.

A closer look reveals a number of insights about the factors determining the performance of the control function approach. First, the efficacy does depend somewhat critically on the number of individuals per group. Comparing rows (1) and (13), which increases school sizes from 500 to 1000 holding fixed other parameters, we see that the R-squared increases from .890 to .914. Indeed, the lowest R-squared values in the table appear in specifications where school sizes are at their smallest (250).

Second, the number of schools has very little impact on performance. Row (19) contains results from a specification that doubles the number of schools from (1), but holds school sizes and other parameters fixed. The R-squared is virtually unchanged, moving from .890 to .893 (Rows (13) and (31) show the same pattern).

Third, restricting the number of schools in each household's consideration set reduces the control function's ability to absorb unobservable sorting, but only slightly. Comparing rows (2) and (5), for example, we see a small increase in the R-squared when increasing the consideration set from 10 to the 50, while we observe no change at all when comparing rows (1) and (5). Thus, our approach does not require households to be considering large numbers of schools.

Fourth, increasing the dimension of the underlying amenity space results in a substantial decrease in the performance of the control function (compare, for example, rows (5) and (6), or (11) and (12)). Indeed, we also ran the simulation under a less conservative scenario in which we restrict $\delta_{lk} = 1 \; \forall (l, k)$, which is tantamount to restricting the amenity space to a single dimension of school/neighborhood "quality". In this setting we find that at least 97 percent of the variance in the school-level contribution of unobserved student characteristics can be predicted by a linear combination of school-average observable characteristics in all 36 parameter sets.

Fifth, increasing the number of observable (and unobservable) characteristics increases performance (compare rows (1) and (2)). This highlights the importance of collecting data on a wide variety of student/parent inputs that capture different dimensions of taste (as the panel surveys we use do).

Finally, the performance of the control function suffers when estimation is based on small subsamples of students at each school. Examining Row (1), we see that the R-squared falls from .890 to only .374 when school averages are merely approximated based on samples of 10 students (Column 8). Increasing the sample size to 20 students per school (Column 9) raises the R-squared to .494, while increasing it further to 40 students per school raises the R-squared to .616. In all of these simulations, we assumed that the strength of sorting on unobservables mirrored the strength of sorting on unobservables. In results not shown, we also experimented with weakening the degree of sorting on unobservables by making δ_x^U smaller in magnitude and increasing the variance of κ_i^* to compensate. While the control function absorbs a smaller *fraction* of the between-school variance in unobservable outcome-relevant characteristics when sorting on these characteristics is weak, this is precisely the case when the magnitude of the between-school variance in unobservables is small. Thus, there is very little potential bias to be absorbed!

While these finite-sample monte carlo simulations should be considered preliminary, they do suggest that our control function approach could work quite well in some school choice settings, particularly when either the population of students in each school is observed or the number of underlying amenity factors is small relative to the number of underlying factors represented by the observed individual characteristics.

6 Data

Our analysis uses data from four distinct sources. The first three sources consist of panel surveys conducted by the National Center for Education Statistics: the National Longitudinal Study of 1972 (NLS72), the National Educational Longitudinal Survey of 1988 (NELS88), and the Educational Longitudinal Survey of 2002 (ELS2002), and ELS02. These data sources possess a number of common properties that make them well suited for our analysis. First, each samples an entire cohort of American students. The cohorts are students who were 12th graders in 1972 in the case of NLS72, 8th graders in 1988 for NELS88, and 10th graders in 2002 for ELS02. Second, each source provides a representative sample of American high schools or 8th grades and samples of students are selected within each school. Both public and private schools are represented.[12] Enough students are sampled from each school to permit construction of estimates of the school means of a large array of student-specific variables and to provide sufficient within-school variation to support a between-/within-school variance decomposition. Third, each survey administered questionnaires to school administrators in addition to all sampled individuals at each school. This provides us with a rich set of both individual-level and school-level variables to examine, allowing a meaningful decomposition of observable versus unobservable variation at both levels of observation. Fourth, each survey collects follow-up information from each student past high school graduation, facilitating analysis of the impact of high school environment on two or more of the outcomes economists and policymakers care most about: the dropout decision, college enrollment and completion decisions, and wage profiles.

While these common properties are very helpful, each survey displays idiosyncratic features and questions that complicate efforts to compare results across time. We develop

[12]We include private schools because they are an important part of the education landscape. However, the connection between characteristics of the school and characteristics of the neighborhood may be weaker for private school students.

comparable measures for all of the variables in our baseline specification, restricting attention only to variables that are available and measured consistently across all three datasets. In addition, in the baseline specification we only use student-level characteristics that are unlikely to be affected by the high school the child attends. However, we also provide decompositions which include in X_{si} scores from standardized tests taken by students in high school as proxies for ability. These scores may be influenced directly by high school inputs, so including them could cause an underestimate of the contribution of school-level inputs. On the other hand, excluding them could instead cause an overestimate of the contribution of school-level inputs, since we run the risk of understating the extent of ability differences among students who attend different schools.

Restricting our analysis to measures that are common across datasets, however, prevents us from exploiting the full power of these rich datasets to explain the distribution of an important set of outcomes. Thus, since NELS88 and ELS02 feature considerably greater overlap in survey questions, we also constructed a larger set of common variables for these two datasets, which we labeled our "full" specification. We include in the full specification measures of student behavior and parental expectations that, like test scores, are not clearly exogenous, but may allow us to more accurately characterize differences in the backgrounds of students attending different schools. For NLS72, the specification we label "w/tests" consists of the variables from the baseline specification plus student test scores. Table 2 lists the final choices of individual-level and school-level explanatory measures used in each dataset.

The one major drawback associated with the three panel surveys is that only around 20 students per school are generally sampled. The simulation results presented in Section V suggest that samples of these size can erode the ability of sample school averages of observable characteristics to serve as an effective control function for variation in average unobservable student contributions across schools. Consequently, we also exploit administrative data from North Carolina on the universe of public schools and public school students (including charter schools) in the state.

Since the North Carolina data contains information on every student at each school, it does not suffer from the same small subsample problem as the panel surveys. On the other hand, the set of observable characteristics is not quite as diverse as in the panel surveys, though it is surprisingly rich for administrative data. In addition to test scores from each grade 3-8, the NC data contain information on each student's race and gender, the student's history of free lunch eligibility and limited english proficiency status, whether the student has been deemed "gifted", parental education, hours per week spent reading for leisure and watching TV. Table 3 provides a full list of the student- and school-level variables included in specifications using the North Carolina data.

More importantly, the data we possess does not link student records to college attendance or future wages, so that the only outcome we observe is high school graduation. Nonetheless, by comparing the graduation results between the North Carolina and panel survey datasets, we can gain some assurance that the small sample problem is distorting results too badly. We explore this issue in greater depth in the next section.

The outcome variables are defined as follows. $COLL$, the measure of college attendance, is an indicator for whether the student is enrolled in a four year college in the second year beyond the high school graduation year of his/her cohort. It is available in each dataset except the North Carolina data.[13] For NELS88 and ELS 2002 $HSGRAD$ is an indicator for whether a student has a high school diploma (not including a GED) as of two years after the high school graduation year of his/her cohort. For the North Carolina data, $HSGRAD$ indicates whether the student is classified as graduated for the official state reporting requirement. Notice, though, that since ELS02 first surveys students in 10th

[13]However, in NLS72 enrollment status is reported in January-March of the second full school year after graduation, while in NELS88 and ELS02 it is reported in October.

grade, it misses a substantial fraction of the early dropouts. Indeed, in NELS88, about one third of the 16 percent who eventually drop out do so before the first follow up survey in the middle of 10th grade. The North Carolina data considers students as eligible for official dropout statistics if they are enrolled in a North Carolina school at the beginning of 9th grade, so there is little scope for underestimating the incidence of dropout. Given that NLS72 first surveys students in 12th grade, we cannot properly examine dropout behavior in this dataset. However, because NLS72 re-surveys students in 1979 and 1986, when respondents are around 25 and 32 years old, respectively, we can use it to analyze completed years of postsecondary education and wages during adulthood. We use years of academic education as of 1979, because attrition and subsampling reduced the 1986 sample by a considerable amount relative to the 1979 follow-up survey, and most respondents have completed their education as of 1979. For the wage analysis, we include only respondents who report wages in both 1979 and 1986.

In each specification, we restrict our sample to those individuals whose school administrator filled out a school survey, and who have non-missing information on the outcome variable and the following key characteristics: race, gender, SES, test scores, region, and urban/rural status.[14] We then impute values for the other explanatory variables to preserve the sample size, since no one other variable is critical to our analysis.[15] Finally, each specification makes use of a set of panel weights. The appropriate weights depend on the analysis. Our rationale for using weights and the details of how we construct them are provided in Appendix Section 3

[14]SES and Urban/rural status are not available in the North Carolina data.

[15]This results in sample sizes for the four year college enrollment analyses of: 12,100 for NLS72, 10,990 for NELS88 using the grade 8 school, 10,710 for NELS88 using the grade 10 school, and 12,440 for ELS02. The sample sizes for the high school graduation analyses are 11,340 for NELS88 (using grade 8 school), 11,040 for NELS88 (using grade 10 school) and 12,370 for ELS02, respectively. The analysis of years of postsecondary education uses 12,070 observations from NLS72, and the wage analysis uses 4,930 individuals with 9,860 wage observations. We also create a missing indicator for mother's education, and include mother's education combined with the missing indicator when performing imputation, along with school averages of all the key characteristics above.

7 Results

7.1 High School Graduation Results from the North Carolina Administrative Data

The first column of Table 4 displays the variance decomposition associated with the latent variable determining high school graduation using the baseline specification from the North Carolina administrative data. Each entry in Panel A presents the fraction of variance contributed by the component associated with the row label.

The 4th row reveals that only 8.5% of the variance in the underlying latent variable consists of differences in school-average dropout rates across schools, with the remaining variation coming from the identities of the dropouts within schools. We also see that while within-school variation in observable characteristics, $Var(X_iB - X_sB)$ composes up a substantial 12.44 percent of the variance, its between-school counterpart $Var(X_sB)$ only accounts for 1.81 percent, for an intraclass correlation of 1.8/14.25 = .127, suggesting that observable characteristics only drive a small amount of the variance in school preferences. Once we remove the between school variation that could be potentially attributed to student sorting, $Var(X_sB) + Var(Z_{1s}G_1) + 2Cov(Z_{1s}G_1, X_sB) + 2Cov(Z_{2s}G_2, X_sB) + 2Cov(Z_{1s}G_1, Z_{2s}G_2)$, we are left with the two components $Var(Z_2G_2)$ and $Var(Z_s^UG^U + v_{s_i}^*)$. These two components combine to produce our lower bound estimate of the school contribution: 4.9 percent of the total student-level variance. Since $Var(Z_s^UG^U + v_{s_i}^*$ may partially reflect common shocks determined near the end of high school, which may or may not be appropriate to include with other school contributions, $Var(Z_2G_2)$ provides a more conservative estimate of 1.81 percent of the total variance attributable to school/neighborhood choice.

The two rows in Panel B use these two alternative lower bound variance estimates to form estimates of the average impact across the student distribution of moving from a school at the 10th percentile of the distribution of school/neighborhood contributions to graduation to a school at the 90th percentile. We can think of this as a thought experiment in which two students at each quantile in the student background distrbution are placed either in the 10th or the 90th quantile school, and the difference in the graduation status of these two pairs is summed over all such pairs.

The estimate that excludes common shocks suggests that, averaged across the student distribution, attending a 90th quantile school increases graduation rates by 10.6 percentage points relative to a school at the 10th quantile. Notice that this estimate is quite large despite the fact that the fraction of variance upon which it is based is quite small, 1.81 percent. This is driven partly by the use of the probit function and the assumption of normality. If the true distribution of latent student contributions is normal, and the graduation rate in North Carolina is not too high (around 80 percent), then there is likely to be large mass of students near the decision margin. Thus, even a small push from the surrounding school/neighborhood environment may be enough to induce a significant fraction of students to graduate. The estimate including common shocks is even higher: 17.4 percentage points.

7.2 Evaluating the Magnitude of Bias from Limited Samples of Students Per School

Before considering estimates from the three survey datasets, however, we first use the North Carolina sample to better gauge the biases produced by the student sampling schemes used by each survey. The monte carlo simulations in Section V suggested that estimation based on subsamples of 20 students per school (similar to those in the three datasets) could result in a substantial decrease in the ability of school-average observables to capture sorting on unobservables. However, these simulations are based on particular assumptions about the dimensionality of the underlying desired amenities, the joint distribution of the observable and unobervable characteristics, and the degree to which these characteristics predict tastes for schools/neighborhoods.

A potentially more direct way to determine whether the sampling schemes are biasing our bound estimates is to draw a sample students from North Carolina schools using either the NLS72, NELS88, or ELS2002 sampling schemes and re-estimate the model for high school graduation using each this sample. By comparing the results derived from such samples to the true results based on the universe of students discussed above, we can determine which if any of the survey datasets is likely to produce reliable results. Columns 2-5 of Table 4 present the results of this exercise. To remove the chatter produced by a single draw from these sampling schemes, each column displays averages over 100 samples drawn from each sampling scheme.

Comparing Column 1 with Column 2, which uses the distribution of school sample sizes observed among 10th grade schools in the NELS88, we see that small samples at each school can produce a considerable bias. Looking at the last two rows of Panel A, we that the NELS grade 10 size distribution overstates the true variance fraction for the lower bound without common shocks, $Var(Z_2G_2)$, by 1.7 percent, and the lower bound with common shocks, $Var(Z_2G_2 + Z_s^U G^U + v_{s_i}^*)$, by 1.4 percent. These translate to overestimates of the impact of a 10th-90th quantile shift in school quality of 3.9 percentage points and 2.2 percentage points, respectively.

However, grade 10 schools in the NELS88 have particularly small samples, since they were not part of the original sampling frame, and only produce large samples of students to the extent that many students from a given grade 8 school attend the same grade 10 school. If we instead compare Column 1 with Column 3, which uses the NELS grade 8 school to classify students, we observe much smaller biases: the lower bound variance fraction without common shocks is overstated by 0.5 percent in the NELS grade 8 sample, and the estimate that includes common shocks is understated by 0.25 percent. The biases in 10th-90th shift estimates are 1.1 percentage points and -1.0 percentage points, respectively. The NLS72 and ELS2002 comparisons, displayed in Columns 4 and 5, are very similar to the NELS grade 8 results. These results are comforting, and suggest that the estimates from these samples may overstate the lower bound slightly in the estimates that attempt to exclude common shocks, but may even understate appropriate lower bound estimates that include common shocks. Due the poor performance of the NELS grade 10 school sample size distribution, however, we do not report any NELS88 results that group students by their grade 10 school.

7.3 High School Graduation and College Enrollment Results from NLS72, NELS88, and ELS2002

Panel A of Table 5 displays the "w/shocks" and "no shocks" lower bound estimates of the fraction of variance attributable to school/neighborhood choices. The first column re-displays the results from the baseline specification using the North Carolina data, while the second column displays results from the full specification that includes past test scores and measures of behavior that could potentially have been altered by the school. Including these additional observable characteristics does reduce both lower bound estimates, from 1.8 percent of total variance to 1.3 percent for the estimate that excludes common shocks and from 4.9 percent to 3.6 percent for the estimate that includes common shocks.

Comparing the North Carolina results to those of NELS88 (Columns 3 and 4), however, we see that both lower bound estimates are surprisingly stable between administrative and survey datasets. Comparing the NELS88 results to the ELS2002 results (Columns 5 and 6), we see that the results also differ very little between the 1988 and 2002 cohorts. This is true for both the baseline and full specifications.

Panel B of Table 5 presents results for the decomposition of the latent index determining enrollment in a 4-year college. Comparing the baseline specifications from NLS72, NELS88, and ELS2002 (Columns 1, 3, and 5), we again observe a surprising consistency in both lower bound estimates of the school/neighborhood contribution across datasets and generations. Estimates that exclude common shocks attribute at least 2.4 to 2.7 percent of outcome variance to schools/neighborhoods, while estimates that include common shocks attribute 5.3 to 5.6 percent. Including test scores and behavioral variables (for NELS88 and ELS2002) reduces these lower bound estimates in a consistent fashion across the three panel surveys (Columns 2, 4, and 6), with the estimates that exclude common shocks dropping to 1.6 to 2.1 percent, and the estimates that include common shocks dropping to 3.6 to 4.1 percent.

Table 6 converts these variance fractions into the more easily interpreted average impacts of a 10th-to-90th quantile shift in school/neighborhood environment. Notice first that despite the striking similarity across datasets in the estimated lower bound fractions of variance attributable to schools, the 10th-90th impact estimates differ significantly across datasets, particularly for the high school graduation results in Panel A. This is due to differences in the sample average graduation rates across the datasets. The gradation rate is 80 percent[16] in the North Carolina data, 86 percent in NELS88, and 90 percent in ELS2002. As a result, a shift of the same magnitude will induce a greater increase in North Carolina than in ELS2002, because there seem to be fewer students near the decision margin. Intuitively, as the sample average converges to 100 percent graduation, the variation in the latent index determining the personal relative benefit from graduating becomes less relevant, as the entire population is far from the decision threshold at 0. Likewise, the sample average college enrollment rate is 27 percent in NLS72, 31 percent in NELS88, and 37 percent in

[16] re-check this!

ELS2002. Since more of the students are not close to the college attendance threshold in 1972, fewer of them reach the decision margin for a given shift in school/neighborhood environment, relative to the cohorts from later generations.

The next striking feature of the results is the magnitude of the estimated changes in both graduation and enrollment rates. Even when common shocks are excluded, and the full specification is considered, lower bound estimates of the impacts of the 10th-to-90th shifts in school quality are associated with an increase in the graduation rate of between 5.6 and 8.4 percentage points, depending on the dataset. This would be enough to halve the dropout rate in each sample. Including common shocks scales up the estimated impacts to between 7.6 and 15.2 percentage points, about three quarters of the dropout rate in each sample. When less dramatic 10th-to-50th quantile shifts are considered (Rows 3 and 4), estimated average impacts are still between 3 and 4 percentage points for lower bounds that exclude common shocks, and between 4 and 8 percentage points for the lower bounds that include common shocks.

Furthermore, these estimates put the full distribution of students at a 10th quantile school to begin with, when many of the students with superior background characteristics would be quite unlikely to be observed in such an environment. A more realistic estimate might put greater weight on estimates produced for the kinds of students most likely to be observed in 10th quantile schools. Note, though, that our method does not allow us to discern the quality of any given school. Average impacts for particular subpopulations, however, will be explored in the next section.

The corresponding estimates for college enrollment, displayed in Panel B, are of a similarly large magnitude. When common shocks are excluded and the full specification is considered, the lower bound on the estimated increase in the 4-year enrollment rate from moving every student (one at a time) from the 10th to the 90th quantile school/neighborhood are between 11 and 14 percentage points (Row 1 of Panel B). Including the residual between-school component boosts the range of estimates to 17 to 19 percentage points. 10th-to-50th quantile shifts still produce average estimated impacts between 5 and 9 percentage points.[17]

7.4 Heterogeneous Effects of 10th-90th Percentile Shifts in School Quality

To what extent do the 10-90 differentials reported in Table 6 conceal heterogeneity in the impact of moving schools across students with varying student backgrounds? Because of nonlinearity in the probit function that links Y_i to the binary outcome indicators for high school graduation and enrollment in a 4-year college, the sensitivity to school quality is higher for groups with values of $X_{si}\hat{B}$ that place them closer to a probability of .5. High school graduation is therefore more sensitive to school quality for disadvantaged groups and less sensitive for advantaged groups. The opposite tends to be true for college enrollment.

Table 8 reports the lower bounds (excluding and including common shocks) for the effect of a 10th to 90th percentile shift in school quality on graduation rates for two extreme cases: students whose value of the background index $X_i\hat{B}$ places them at the 10th quantile of the $X_i\hat{B}$ distribution (Rows 1 and 2), and students at the 90th quantile of the $X_i\hat{B}$ distribution (Rows 3 and 4). For the North Carolina sample and the full specification, the lower bound estimates that exclude common shocks suggest a 12.7 percentage point increase for students at the 10th quantile and a 3.6 percentage point increase for students at the 90th quantile, while the lower bound estimates that include common shocks are 22.9 and 6.3 percentage points, respectively. For NELS88 grade 8 (Column 4), the numbers

[17]Note that in the case of college enrollment, many of the students who would be induced to attend college by moving from a 10th to 90th quantile school are those from more privileged backgrounds who are already attending high quality schools. Thus, unlike the case of high school graduation, the 10-90 estimates may overstate the college enrollment gains to improving school quality for the kinds of students attending low quality schools.

are smaller, particularly for the 90th quantile: lower bound estimates that exclude common shocks are 13.0 percentage points and 0.6 percentage points, and lower bound estimates that exclude common shocks are 21.5 percentage points and 1.0 percentage points in the full specification. This reflects the fact that the average dropout rate is lower for the NELS88 than for the state of North Carolina between 2007 and 2009. ELS2002 results are generally similar to NELS88 grade 10, but slightly smaller. The results show that advantaged students tend to graduate high school regardless of the school they attend, while disadvantaged students are strongly affected by school quality.

Table 8 also reports the average impact of a 10th-90th shift on high school graduation rates for three subpopulations of interest: black students, white students with single mothers who did not attend college, and white students with both parents present, at least one of whom completed college. For the full specification in the North Carolina sample, the estimates without and with common shocks are 8.5 and 15.2 percentage points for black students. The estimates for white students with single mothers who did not attend college are 11.4 and 20.6 percentage points, while the estimates for white students with both parents, at least one of whom completed college, are 4.7 and 8.4. The estimates are consistently smaller in the NELS88 and ELS samples, but are still between 7 and 10 percentage points for black students and for white students with single mothers who did not attend college.

Table 9 reports a corresponding set of results for enrollment in a 4-year college. The college enrollment rates for students at the 10th percentile of the $X_i\hat{B}$ distribution are substantially less sensitive to school quality, reflecting the fact that most such students are nowhere near the college enrollment margin. The lower bound estimates that exclude and include common shocks are between 3 and 6 percentage points and between 4 and 8 percentage points, respectively, depending on the sample. By contrast, the lower bound estimates excluding and including common shocks for students at the 90th percentile of the $X_i\hat{B}$ distribution are between 14 and 17 percentage points and 19 and 26 percentage points, respectively. The values for blacks and for whites with non-college-educated single mothers are similar to the results for the full sample, while the values for whites with college educated parents are close to those for the 90th percentile of the $X_i\hat{B}$ distribution.

Overall, it appears that, except for the lowest stratum of student background, there are considerable pools of students that are close enough to the decision margin for a major shift in school quality to be a deciding factor in determining college enrollment.

7.5 NLS Results for Years of Postsecondary Education and Permanent Log Wages

Table 7 displays the lower bound estimates of the impact of 10th-to-90th and 10th-to-50th shifts in school quality on years of postsecondary education and permanent wages for the NLS72 sample. The baseline lower bound estimate that excludes common shocks implies that a 10-90 shift in school quality increases years of postsecondary education by .39 years, while including standardized tests among the observable characteristics reduces this estimate to .27 years. Note, though, that since the NLS72 data is collected in 12th grade, the standardized test scores are particularly likely to reflect high school quality, making the w/tests specification a likely underestimate. Adding the variance in the unobserved between-school component raises these estimates to .56 and .45 years respectively. Even 10th-to-50th quantile shifts are half as large by construction, since no non-linear transformation takes place when the outcome is continuous. Nonetheless, they suggest a substantive impact of shifts in school quality on years of college education: the corresponding four estimates are .19, .14, .28, and .20 years of college, respectively.

Columns 3-6 contain analogous estimates for the permanent component of log wages. Columns 3-4 reflect specifications in which years of postsecondary education is not included as a control, while columns 5-6 includes years of postsecondary education to focus on the effect on log wages that does not occur via postsecondary education. In practice, the two sets of estimates are quite similar. The estimates that exclude common shocks imply that a 10-90 shift in school quality increases wages by around 12 percent. The 10-50 shifts are again half as large at around 6 percent. Estimates that include common shocks imply that a 10-90 shift in school quality increases wages by around 20 percent. Thus, at least for the 1972 cohort, shifts in school quality seem to have important impacts as well for longer run outcomes of particular importance for worker welfare.

8 Discussion

The key takeaway from the results presented in this paper should be that even very conservative estimates of the contribution of schools and surrounding neighborhoods to later outcomes suggest that improving school/neighborhood environments could have a very large impact on high school graduation rates and college enrollment rates. Indeed, this is quite consistent with the lottery-based estimates of (Deming et al. 2011), and suggests that their results are likely to generalize beyond the specific high poverty Charlotte context they consider. Their results, perhaps combined with the Moving to Opportunity results, suggest that perhaps schools make a more important part of the contribution of the external environment than do neighborhoods, though the two may be complementary.

Of course, there are clearly extensions to our framework that are likely to produce a more nuanced interpretation. In particular, our model of outcomes does not explicitly allow for complementarities and other interactions between school and student quality.[18] For example, it could be the case that the types of students who attend low quality schools

[18]Recall, though, that our model does allow for school treatments to differ across students within a school. This variation is not captured, however, in our lower bound estimates, which focus only on correctly attributing across-school variation to schools/neighborhoods versus students.

are those that are least likely to profit from improvements in school quality, although the (Deming et al. 2011) results suggest otherwise. Similarly, our static model does not reflect the many school transfers and residential moves that take place during students' educational careers that are explicitly motivated by the perceived payoff to a better school.

A final point worth re-emphasizing is that the control function approach we utilize here could be employed in any setting in which individuals sort into units, and where both the individuals and the units themselves have independent impacts on individual outcomes. While we mentioned the patient/hospital case in the introduction, one could easily apply this framework to product choice in a differentiated goods market, in which individuals with different characteristics place different weights on different product characteristics (a la (Rosen 1974)). For example, one could evaluate the success of any advertising campaign where the object of interest is the degree to which the public has internalized and acted on the information provided about a particular product feature that affects a measurable individual outcome. One can simply regress average outcomes for each given product (e.g. cavity rate among users each toothpaste) on the average characteristics of the buyers, if sufficient information on buyers is collected. If the value of the particular product feature (e.g. the amount of a particular anti-cavity agent) predicts the residual difference in average outcomes across products, then buyers have not fully internalized (or do not care) about the role of this product feature in producing the measured outcome.

A final appropriate application for our model relates to government regulation. The standard textbook treatment of occupational safety regulation (e.g. (Ehrenberg and Smith 2010)) suggests that government intervention only increases worker welfare if the safety risks are unknown at the time the occupation is chosen; otherwise such regulations remove the opportunity for risk-loving workers to get paid welfare-enhancing compensating differentials for taking on risky jobs. The sorting model we presented suggests that the residual from a regression of occupation-average age at death on a large vector of occupation-average worker characteristics can potentially isolate the part of the long run occupational contribution to health that was unknown to workers when they chose the occupation. It addresses the concern that sorting on occupational sorting on unobserved characteristics that influence mortality are responsible for occupation differences in mortality rates. Thus, one can directly identify the occupations that merit government-supported information campaigns or other safety regulations. We are currently pursuing this application.

References

Aliprantis, Dionissi (2011) 'Assessing the evidence on neighborhood effects from moving to opportunity.' Technical Report

Altonji, Joseph G, and Richard K Mansfield (2011a) 'The role of family, school, and community characteristics in inequality in education and labor–market outcomes.' *Whither opportunity? Rising inequality, schools, and childrens life chances* pp. 339–58

Altonji, Joseph G, and Richard Mansfield (2011b) 'The contribution of family, school, and community characteristics to inequality in education and labor market outcomes.' *Unpublished Manuscript*

Baker, Michael, and Gary Solon (2003) 'Earnings dynamics and inequality among canadian men, 1976-1992: Evidence from longitudinal income tax records.' *Journal of Labor Economics* 21(2), 267–288

Bayer, Patrick, and Stephen L Ross (2006) 'Identifying individual and group effects in the presence of sorting: A neighborhood effects application.' Technical Report, National Bureau of Economic Research

Cunha, Flavio, James J Heckman, Lance Lochner, and Dimitriy V Masterov (2006) 'Interpreting the evidence on life cycle skill formation.' *Handbook of the Economics of Education* 1, 697–812

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger (2011) 'School choice, school quality and postsecondary attainment.' Technical Report, National Bureau of Economic Research

Ehrenberg, Ronald G, and Robert S Smith (2010) 'Modern labor economics'

Haider, Steven J (2001) 'Earnings instability and earnings inequality of males in the united states: 1967–1991.' *Journal of Labor Economics* 19(4), 799–836

Jacob, Brian A (2004) 'Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago.' *The American Economic Review* 94(1), 233–258

Jencks, Christopher S, and Marsha D Brown (1975) 'Effects of high schools on their students.' *Harvard Educational Review* 45(3), 273–324

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz (2007) 'Experimental analysis of neighborhood effects.' *Econometrica* 75(1), 83–119

Mansfield, Richard (Forthcoming) 'Teacher Quality and Student Inequality.' *Journal of Labor Economics*

Meghir, Costas, and Luigi Pistaferri (2004) 'Income variance dynamics and heterogeneity.' *Econometrica* 72(1), 1–32

Oreopoulos, Philip (2003) 'The long-run consequences of living in a poor neighborhood.' *The quarterly journal of economics* 118(4), 1533–1575

Rosen, Sherwin (1974) 'Hedonic prices and implicit markets: product differentiation in pure competition.' *The journal of political economy* pp. 34–55

Todd, Petra, and Kenneth Wolpin (2003) 'On the Specification and Estimation of the Production Function for Cognitive Achievement.' *The Economic Journal* 113, F3–F33

9 Tables and Figures

Table 1: Monte Carlo Simulation Results

Row	# Stu.	# Sch.	# Cons.	# Obs.	# Unobs.	# Amen.	R-Sq (All)	R-Sq (10)	R-sq (20)	R-sq (40)
(1)	25000	50	10	10	10	10	0.890	0.374	0.494	0.616
(2)	25000	50	10	20	20	10	0.967	0.549	0.673	0.770
(3)	25000	50	10	20	20	20	0.934	0.528	0.629	0.727
(4)	25000	50	50	10	10	10	0.890	0.400	0.528	0.644
(5)	25000	50	50	20	20	10	0.978	0.599	0.711	0.815
(6)	25000	50	50	20	20	20	0.933	0.546	0.639	0.738
(7)	25000	100	10	10	10	10	0.836	0.300	0.452	0.591
(8)	25000	100	10	20	20	10	0.927	0.392	0.549	0.701
(9)	25000	100	10	20	20	20	0.872	0.364	0.503	0.639
(10)	25000	100	50	10	10	10	0.855	0.339	0.483	0.623
(11)	25000	100	50	20	20	10	0.943	0.456	0.615	0.752
(12)	25000	100	50	20	20	20	0.872	0.410	0.536	0.663
(13)	50000	50	10	10	10	10	0.914	0.372	0.499	0.619
(14)	50000	50	10	20	20	10	0.981	0.552	0.664	0.770
(15)	50000	50	10	20	20	20	0.964	0.539	0.650	0.740
(16)	50000	50	50	10	10	10	0.942	0.391	0.523	0.652
(17)	50000	50	50	20	20	10	0.987	0.589	0.710	0.811
(18)	50000	50	50	20	20	20	0.956	0.563	0.658	0.756
(19)	50000	100	10	10	10	10	0.893	0.294	0.449	0.590
(20)	50000	100	10	20	20	10	0.960	0.403	0.548	0.696
(21)	50000	100	10	20	20	20	0.918	0.382	0.517	0.653
(22)	50000	100	50	10	10	10	0.891	0.333	0.479	0.611
(23)	50000	100	50	20	20	10	0.973	0.456	0.618	0.755
(24)	50000	100	50	20	20	20	0.909	0.409	0.544	0.664
(25)	100000	50	10	10	10	10	0.962	0.371	0.502	0.632
(26)	100000	50	10	20	20	10	0.992	0.539	0.659	0.777
(27)	100000	50	10	20	20	20	0.972	0.544	0.651	0.757
(28)	100000	50	50	10	10	10	0.962	0.398	0.530	0.650
(29)	100000	50	50	20	20	10	0.994	0.592	0.709	0.811
(30)	100000	50	50	20	20	20	0.968	0.550	0.665	0.740
(31)	100000	100	10	10	10	10	0.912	0.287	0.430	0.573
(32)	100000	100	10	20	20	10	0.980	0.394	0.548	0.701
(33)	100000	100	10	20	20	20	0.949	0.374	0.520	0.660
(34)	100000	100	50	10	10	10	0.924	0.332	0.483	0.613
(35)	100000	100	50	20	20	10	0.985	0.455	0.612	0.751
(36)	100000	100	50	20	20	20	0.944	0.416	0.551	0.676

Table 2: Variables Used in Baseline and Full (in Italics) Specifications

Student Characteristics
Female, Black, Hispanic, Asian, *Immigrant*

Student Ability
*Math Standardized Score**, *Reading Standardized Score**

Student Behavior
Hrs./Wk. Spent on Homework, *Parents Often Check Homework*, *Hrs./Wk. Spent on Leisure Reading*, *Hrs./Wk. Spent Watching TV*, *Often Arrives at Class Without a Pencil*, *Physical Fight This Year*

Family Background
Standardized SES, Number of Siblings, Both Bio. Parents Present, Mother and Male Guardian Present, Father and Female Guardian Present, Mother Only Present, Father Only Present, Father's Years of Education, Mother's Years of Education, Moth. Yrs. Ed. Missing, English Spoken at Home, Log(Family Income), *Immigrant Mother*, *Immigrant Father*, *Employed Mother*, *Employed Father*, *Parents are Married*

Parental Expectations
Mother's Desired Yrs. of Ed., *Father's Desired Yrs. of Ed.*

School Characteristics
School is Catholic, School is Private Non-Catholic, Student-Teacher Ratio, Pct. Teacher Turnover Since Last Year, Pct. on College Prep. Track, Pct. of Teachers w/ Master's Degrees or More, Average Pct. Daily Attendance, School Pct. Minority, School Teacher Pct. Minority, Total School Enrollment *Log(Min. Teacher Salary)*, *School Pct. Free/Reduced Price Lunch*, *School Pct. LEP*, *School Pct. Special Ed.*, *School Pct. Remedial Reading*, *School Pct. Remedial Math*

Neighborhood Characteristics
School in Urban Area, School in Suburban Area, School in Rural Area, School in Northeast U.S. Region, School in South U.S. Region, School in Midwest U.S. Region, School in West U.S. Region

*Standardized test scores are also included in the w/tests specifications, along with all of the baseline variables.

Table 3: Variables Included in Specifications Using North Carolina Administrative Data

Student Characteristics
Female, Black, Hispanic, Asian
Student Ability
Math Standardized Score (Grades 7 & 8), Reading Standardized Score (Grades 7 & 8) Designated Gifted Student (Math), Designated Gifted Student (Reading)
Student Behavior
Hrs./Wk. Spent on Homework (Indicator Variables), Hrs./Wk. Spent on Leisure Reading (Indicator Variables) Hrs./Wk. Spent Watching TV (Indicator Variables)
Family Background
Responding Parent Educational Attainment Category Indicator Variables Ever Eligible for Free/Reduced Price Lunch Currently Limited English Proficiency Ever Limited English Proficiency
School Characteristics
Magnet School, Charter School, Student-Teacher Ratio, Pct. Teacher Turnover Since Last Year Pct. on College Prep. Track Pct. of Teachers w/ Master's Degrees or More Average Pct. Daily Attendance, School Teacher Pct. Highly Qualified Total School Enrollment
Neighborhood Characteristics
Urbanicity Indicator Variables (12 Categories)

School averages of all individual-level variables are also included in each specification.
Classroom averages of all individual-level variables are also employed in some specifications.
See Section 10 for details.

Table 4: Bias from Observing Subsamples of Students from Each School: Comparing Results from the Full North Carolina Sample to Results from Subsamples Mirroring the Sampling Schemes of NLS72, NELS88, and ELS2002

Panel A: Fractions of Total Outcome Variance

Row	Full NC Sample	NELSg10	NELSg8	ELS2002	NLS72
Within School:					
Total $Var(Y_{is} - Y_s)$	0.9153	0.8763	0.9131	0.9120	0.9126
Observable Student-Level (Within): $Var((X_{si} - X_s)B)$	0.1244	0.1301	0.1296	0.1285	0.1296
Unobservable Student-Level (Within) $Var(V_{si})$	0.7909	0.7461	0.7834	0.7834	0.7828
Between School:					
Total $Var(Y_s)$	0.0847	0.1237	0.0869	0.088	0.0874
Observable Student-Level: $Var(X_sB)$	0.0181	0.0179	0.0183	0.0184	0.018
Student-Level/ School-Level Covariance $2 * Cov(X_sB, Z1_sG + Z2_sG)$	0.0165	0.0187	0.0170	0.175	0.0175
School-Avg. Student-Level/ School Char. Covariance $2 * Cov(Z1_sG, Z2_sG)$	-0.0166	-0.0053	0.0061	-0.0054	-0.0047
School-Avg. Student-Level $Var(Z_{1s})$	0.0178	0.029	0.0137	0.0139	0.0125
School Char. $Var(Z_{2s}G_2)$	0.0181	0.0353	0.023	0.0238	0.0269
Unobservable School-Level $Var(Z_s^U)$	0.0309	0.0283	0.0211	0.0199	0.0173

Panel B: 10th to 90th Quantile Shifts in School Quality

Row	Full NC Sample	NLS72	NELSg8	NELSg10	ELS2002
10-90 Lower Bound no unobs $Var(Z_{2s}G_2)$	0.1056	0.1435	0.1167	0.1177	0.1254
10-90 Lower Bound w/unobs $Var(Z_{2s}G_2 + Z_s^U)$	0.1742	0.1959	0.164	0.1626	0.1631

Table 5: The Fraction of Variance in the Latent Index Determining High School Graduation Attributable to School/Neighborhood Quality: Approximate Lower Bound Estimates

Panel A: High School Graduation

Upper/Lower Bound	NC		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs	0.018	0.013	0.018	0.011	0.021	0.016
$Var(Z_{2s}G_2)$	(0.024)	(0.023)	(0.024)	(0.015)	(0.021)	(0.016)
LB w/ unobs	0.049	0.036	0.049	0.031	0.044	0.029
$Var(Z_{2s}G_2 + Z_s^U)$	(0.092)	(0.078)	(0.092)	(0.078)	(0.092)	(0.076)

Panel B: Enrollment in a Four Year College

Upper/Lower Bound	NLS		NELS gr8		ELS	
	Baseline	w/Tests	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs	0.027	0.016	0.026	0.021	0.024	0.019
$Var(Z_{2s}G_2)$	(0.026)	(0.021)	(0.032)	(0.027)	(0.024)	(0.019)
LB w/ unobs	0.053	0.041	0.056	0.042	0.055	0.036
$Var(Z_{2s}G_2 + Z_s^U)$	(0.226)	(0.188)	(0.260)	(0.225)	(0.237)	(0.188)

Table 6: Effect on Outcomes of Transferring from a School at the 10th Percentile of the Distribution of School Quality to a School at the 50th or 90th Percentile: Approximate Lower Bound Estimates

Panel A: High School Graduation						
Upper/Lower Bound	NC		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.106	0.084	0.076	0.063	0.064	0.056
$Var(Z_{2s}G_2)$	(0.060)	(0.060)	(0.060)	(0.046)	(0.064)	(0.056)
LB w/ unobs: 10th-90th	0.174	0.152	0.127	0.104	0.092	0.076
$Var(Z_{2s}G_2 + Z_s^U)$	(0.092)	(0.078)	(0.092)	(0.078)	(0.092)	(0.076)
LB no unobs: 10th-50th	0.056	0.044	0.042	0.034	0.036	0.031
$Var(Z_{2s}G_2)$	(0.034)	(0.034)	(0.034)	(0.025)	(0.036)	(0.031)
LB w/ unobs: 10th-50th	0.096	0.083	0.074	0.058	0.055	0.043
$Var(Var(Z_{2s}G_2 + Z_s^U)$	(0.056)	(0.046)	(0.056)	(0.047)	(0.055)	(0.043)

Panel B: Enrollment in a Four Year College						
Upper/Lower Bound	NLS		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.142	0.108	0.153	0.132	0.155	0.136
$Var(Z_{2s}G_2)$	(0.142)	(0.108)	(0.153)	(0.158)	(0.155)	(0.136)
LB w/ unobs: 10th-90th	0.200	0.174	0.226	0.188	0.237	0.188
$Var(Z_{2s}G_2 + Z_s^U)$	(0.200)	(0.174)	(0.226)	(0.188)	(0.237)	(0.188)
LB no unobs: 10th-50th	0.066	0.051	0.073	0.063	0.075	0.066
$Var(Z_{2s}G_2)$	(0.073)	(0.063)	(0.084)	(0.076)	(0.075)	(0.066)
LB w/ unobs: 10th-50th	0.090	0.080	0.105	0.088	0.112	0.090
$Var(Z_{2s}G_2 + Z_s^U)$	(0.105)	(0.088)	(0.121)	(0.106)	(0.112)	(0.090)

Table 7: Effect on Outcomes of Transferring from a School at the 10th Percentile of the Distribution of School Quality to a School at the 50th or 90th Percentile: Approximate Lower Bound Estimates

Panel C: Years of Postsecondary Education and Permanent Wages (NLS72 data)						
Upper/Lower Bound	Yrs. Postsec. Ed.		Perm. Wages No Post-sec Ed.		Perm. Wages w/ Post-sec Ed.	
	Baseline	w/Tests	Baseline	w/Tests	Baseline	w/Tests
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.389	0.270	0.123	0.133	0.124	0.131
$Var(Z_{2s}G_2)$	(0.049)	(0.044)	(0.064)	(0.065)	(0.066)	(0.066)
LB w/unobs: 10th-90th	0.559	0.409	0.207	0.205	0.269	0.263
$Var(Z_{2s}G_2 + Z_s^U)$	(0.048)	(0.039)	(0.016)	(0.016)	(0.015)	(0.015)
LB no unobs: 10th-50th	0.194	0.135	0.062	0.067	0.062	0.066
$Var(Z_{2s}G_2)$	(0.025)	(0.022)	(0.032)	(0.032)	(0.033)	(0.033)
LB w/unobs: 10th-50th	0.280	0.204	0.103	0.102	0.134	0.131
$Var(Z_{2s}G_2 + Z_s^U)$	(0.024)	(0.019)	(0.008)	(0.008)	(0.008)	(0.008)

w/Tests specification includes student ability measures.
No Post-sec Ed. refers to specifications in which we do not include years of completed post-secondary education as an element of X_{si}.
w/ Post-sec Ed. refers to specifications in which we include years of completed post-secondary education as an element of X_{si}.

Table 8: The Impact of 10th-90th Percentile Shifts in School Quality on High School Graduation Rates for Selected Subpopulations

	NC		NELS gr8		ELS	
Subpopulation	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
XB: 10th Quantile						
LB no unobs	0.146	0.127	0.132	0.130	0.106	0.114
$Var(Z_{2s}G_2)$	(0.106)	(0.126)	(0.132)	(0.130)	(0.106)	(0.114)
LB w/ unobs	0.242	0.229	0.221	0.215	0.153	0.154
$Var(Z_{2s}G_2 + Z_s^U)$	(0.164)	(0.166)	(0.221)	(0.215)	(0.153)	(0.154)
XB: 90th Quantile						
LB no unobs	0.060	0.036	0.025	0.006	0.019	0.009
$Var(Z_{2s}G_2)$	(0.020)	(0.008)	(0.025)	(0.006)	(0.019)	(0.009)
LB w/ unobs	0.098	0.063	0.040	0.010	0.028	0.012
$Var(Z_{2s}G_2 + Z_s^U)$	(0.031)	(0.011)	(0.040)	(0.010)	(0.028)	(0.012)
Black						
LB no unobs	0.107	0.085	0.076	0.070	0.074	0.070
$Var(Z_{2s}G_2)$	(0.067)	(0.069)	(0.076)	(0.070)	(0.074)	(0.070)
LB w/ unobs	0.176	0.152	0.126	0.114	0.107	0.094
$Var(Z_{2s}G_2 + Z_s^U)$	(0.102)	(0.091)	(0.126)	(0.114)	(0.107)	(0.094)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.142	0.114	0.121	0.104	0.092	0.079
$Var(Z_{2s}G_2)$	(0.095)	(0.101)	(0.121)	(0.104)	(0.092)	(0.079)
LB w/ unobs	0.235	0.206	0.202	0.172	0.133	0.106
$Var(Z_{2s}G_2 + Z_s^U)$	(0.147)	(0.134)	(0.202)	(0.172)	(0.133)	(0.106)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.062	0.047	0.034	0.022	0.028	0.020
$Var(Z_{2s}G_2)$	(0.026)	(0.022)	(0.034)	(0.022)	(0.028)	(0.020)
LB w/ unobs	0.102	0.084	0.055	0.036	0.039	0.027
$Var(Z_{2s}G_2 + Z_s^U)$	(0.040)	(0.028)	(0.055)	(0.036)	(0.039)	(0.027)

NELS gr8 refers to a decomposition that uses the 8th grade school as the class variable, and uses 8th grade measures of student behavior and parental expectations, and 8th grade test scores in the full specification.

"Lower Bound w/unobs" and "Lower Bound no unobs" refer to lower bound estimates of the increase in the probability of graduation associated with a move from the 10th percentile school to the 90th percentile school, independent of differences in student composition, that exclude and include common shocks to all members of a school that take place after high school begins, respectively.

XB: 10th (90th) Quantile reports results for students whose values of $X_{si}B$ equal the estimated 10th (90th) quantile value of the $X_{si}B$ distribution. See Section 6.

Table 9: The Impact of 10th-90th Percentile Shifts in School Quality on Four-Year College Enrollment Rates for Selected Subpopulations

	NLS		NELS gr8		ELS	
Subpopulation	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
XB: 10th Quantile						
LB no unobs	0.088	0.025	0.082	0.044	0.110	0.056
$Var(Z_{2s}G_2)$	(0.082)	(0.044)	(0.107)	(0.051)	(0.110)	(0.056)
LB w/ unobs	0.124	0.039	0.119	0.061	0.167	0.076
$Var(Z_{2s}G_2 + Z_s^U)$	(0.119)	(0.061)	(0.158)	(0.071)	(0.167)	(0.076)
XB: 90th Quantile						
LB no unobs	0.190	0.165	0.190	0.162	0.168	0.139
$Var(Z_{2s}G_2)$	(0.190)	(0.162)	(0.201)	(0.175)	(0.168)	(0.139)
LB w/ unobs	0.269	0.268	0.282	0.230	0.258	0.193
$Var(Z_{2s}G_2 + Z_s^U)$	(0.283)	(0.231)	(0.299)	(0.249)	(0.258)	(0.193)
Black						
LB no unobs	0.132	0.098	0.148	0.130	0.145	0.125
$Var(Z_{2s}G_2)$	(0.148)	(0.130)	(0.171)	(0.153)	(0.145)	(0.125)
LB w/ unobs	0.186	0.158	0.220	0.185	0.221	0.173
$Var(Z_{2s}G_2 + Z_s^U)$	(0.220)	(0.185)	(0.254)	(0.218)	(0.221)	(0.173)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.110	0.090	0.111	0.090	0.089	0.129
$Var(Z_{2s}G_2)$	(0.111)	(0.089)	(0.133)	(0.115)	(0.142)	(0.129)
LB w/ unobs	0.154	0.144	0.164	0.126	0.217	0.179
$Var(Z_{2s}G_2 + Z_s^U)$	(0.164)	(0.126)	(0.197)	(0.164)	(0.217)	(0.179)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.184	0.146	0.187	0.163	0.173	0.154
$Var(Z_{2s}G_2)$	(0.187)	(0.163)	(0.203)	(0.184)	(0.173)	(0.154)
LB w/ unobs	0.260	0.237	0.279	0.233	0.265	0.213
$Var(Z_{2s}G_2 + Z_s^U)$	(0.279)	(0.233)	(0.301)	(0.263)	(0.265)	(0.213)

Notes: See Table 8

Appendix

1 Estimation of Model Parameters

In this section we discuss estimation of the coefficients B and G. The estimation strategy depends on the outcome, so we consider the outcomes in turn.

1.1 Years of Postsecondary Academic Education

Parameter estimation is most straightforward in the case of years of postsecondary academic education. We estimate B using ordinary least squares regression with high school fixed effects, which controls for all observed and unobserved school and neighborhood influences.

Recall that Z_s is comprised of two components: $Z_s = [Z_s^1; Z_s^2]$. Z_s^2 consists of school and neighborhood characteristics for which direct measures are available, such as student/teacher ratio, city size, and school type. Z_s^1 consists of school wide averages for each variable in X_{si}, such as parental education or income, which we do not observe directly but must estimate from sample members at each school. Consequently, the makeup of Z_s^1 differs across specifications that use different X vectors. G_1 and G_2 are the corresponding subsets of the coefficients in G.

We replace Z_s^1 with $\bar{Z}_s^1$, where $\bar{Z}_s^1$ is the average of X_{si} computed over all available students from the school.[19] We estimate G by applying least squares regression to

$$Y_{si} - X_{si}\hat{B} = [\overline{Z}_s^1 G_1 + Z_s^2 G_2] \equiv \overline{Z}_s G + e_{si}$$

using the appropriate panel weights from the surveys.

1.2 Permanent Wage Rates

Abstracting from the effects of labor market experience and a time trend, let the log wage W_{sit} of individual i, from school s, at time t be governed by

$$W_{sit} = W_{si} + e_{sit} + \xi_{sit}.$$

In the above equation W_{si} is i's "permanent" log wage (given that he/she attended high school s) as of the time by which most students have completed education and spent at least a couple of years in the labor market, which we take to be 1979 in the case of NLS72. e_{sit} is a random walk component that evolves as a result of luck in the job search process or within a company, and changes in motivation or productivity due to health and other factors. We normalize e_{sit} to be 0 in 1979.[20] ξ_{sit} includes measurement error and relatively short term factors that have little influence on the lifetime earnings of an individual. The determination of the permanent wage is given by 1 with Y_{si} defined to be W_{si}. After substituting for W_{si}, the wage equation is

$$W_{sit} = X_{si}B + Z_s G + Z_s^U G^U + v_{s_i i}^* + e_{sit} + \xi_{sit}.$$

[19] A substantial number of persons who appear in the base year of the surveys can be used to construct $\bar{Z}_s^1$ but cannot be used to estimate (1.1) because some variables, such as test scores, are missing, or because the students are not included in the follow-up surveys that provide the measure of Y_{si}. As we discuss in the data section, we impute missing values for most of our explanatory variables prior to estimating B and G, but we do not use the imputed values when constructing the school averages.

[20] We include e_{sit} as well as ξ_{it} because the earnings dynamics literature typically finds evidence of a highly persistent wage component. Several studies cannot reject the hypothesis that e_{sit} is a random walk. Recent examples include (Baker and Solon 2003), (Haider 2001), and (Meghir and Pistaferri 2004).

We estimate B by OLS with school fixed effects included.[21]

Let $\tilde{W}_{sit} \equiv W_{sit} - X_i \hat{B}$. We estimate G by applying OLS to

$$\tilde{W}_{sit} = \bar{Z}_s G + Z_s^U G^U + v^*_{s_i i} + e_{sit} + \xi_{sit} \tag{31}$$

The presence of ξ_{sit} complicates the variance decompositions, as we discuss below.

1.3 High School Graduation and College Enrollment

The methods outlined in Sections 3.1 need to be adapted for binary measures such as high school graduation and college attendance. Consequently, for high school graduation we reinterpret Y_{si} to be the latent variable that determines the indicator for whether a student graduates, $HSGRAD_{si}$. That is,

$$HSGRAD_{si} = 1(Y_{si} > 0).$$

Or, after substituting for Y_{si},

$$HSGRAD_{si} = 1(X_{si}B + Z_s G + Z_s^U G^U + v^*_{s_i i} > 0) \tag{32}$$

We replace Z_s with $\bar{Z}_s$ and estimate the equation

$$HSGRAD_{si} = 1(X_{si}B + \bar{Z}_s G + (Z_s - \bar{Z}_s)G + Z_s^U G^U + v^*_{s_i i} > 0) \tag{33}$$

using maximum likelihood probit. The procedure for enrollment in a four-year college is

[21] In reality, we also include a vector T_{it} consisting of a dummy indicator for the year 1979 (relative to 1986), years of work experience of i at time t, and experience squared. Let Ψ be the corresponding vector of wage coefficients. We adjust wages for differences in labor market experience and for whether the data are from 1979 or 1986 by subtracting $T_{it}\hat{\Psi}$ from the wage prior to performing the variance decompositions. The estimate of $\hat{\Psi}$ depends on whether tests, postsecondary education, or both are in X_{si}. We report results with and without these variables. In our main specification, we exclude postsecondary education from X_{si}.

analogous to that of high school graduation.

2 Decomposing the Variance in Educational Attainment and Wages

In this section we discuss an analysis of variance based on equation that can be used to place a lower bound on the importance of factors that are common to students from the same school.[22] As with parameter estimation, the details of our procedure depend upon the outcome. We begin with years of postsecondary education.

2.1 Years of Postsecondary Education

One may decompose $Var(Y_{si})$ into its within and between school components

$$Var(Y_{si}) = Var(Y_{si} - Y_s) + Var(Y_s)$$

where $(Y_{si} - Y_s)$ is the part of Y_{si} that varies across students in school s and Y_s is the average outcome for students from s. We estimate $Var(Y_{si} - Y_s)$ by using the sample variances of $Var(Y_{si} - \overline{Y}_s)$ with an appropriate correction for degrees of freedom lost in using the sample mean $\overline{Y}_s$ in place of Y_s. Then $Var(Y_s)$ can be estimated as

$$\widehat{Var}(Y_s) = \widehat{Var}(Y_{si}) - \widehat{Var}(Y_{si} - Y_s).$$

Then, from (2),

$$(Y_{si} - Y_s) = (X_{si} - X_s)B + (v_i^* - v_{s_i}^*)$$

[22](Jencks and Brown 1975) propose and implement a similar decomposition.

and

$$Y_s = X_s B + Z_s G + Z_s^U G^U + v_{s_i}^*.$$

Thus, one may express the outcome variance as[23]

$$Var(Y_i) = [Var((X_i - X_{s_i})\beta) + Var(v_{s_i i}^* - \overline{v^*}_{s_i})] + \tag{34}$$

$$[Var(X_s\beta) + 2Cov(X_s\beta, Z_{1s}G_1) + 2Cov(X_s\beta, Z_{2s}G_2) + Var(Z_{1s}G_1) + \tag{35}$$

$$2Cov(Z_{1s}G_1, Z_{2s}G_2) + Var(Z_{2s}G_2) + Var(Z_s^U G^U + v_{s_i}^*)] \tag{36}$$

Given an estimate of B, $Var((X_i - X_s)B)$ can be estimated using its corresponding sample variance, $Var((X_{si} - \overline{X}_s)B)$. $Var(v_i^* - v_s^*)$ can then be estimated as $\widehat{Var}(Y_{si} - Y_s) - \widehat{Var}((X_{si} - X_s)B)$, and $Var(X_s B)$ can be calculated as $\widehat{Var}(X_{si}B) - \widehat{Var}((X_{si} - X_s)B)$. One can also estimate the components $Var(Z_{1s}G_1)$,$Var(Z_{2s}G_2)$ of the school/community contribution and the common terms $2Cov(X_s B, Z_{1s}G_1)$ and $2Cov(X_s B, Z_{2s}G_2)$ using the estimates of B, G_1, G_2 and the data $\bar{X}_s \equiv \bar{Z}_{1s}$ and Z_{2s}. $Var(Z_s^U G^U + v_{s_i}^*)$ can be calculated as

$$\begin{aligned}
&\widehat{Var}(Z_s^U G^U + v_{s_i}^*) = \\
&\widehat{Var}(Y_s) - \widehat{Var}(X_s B) - \widehat{Var}(Z_{1s}G_1) - \widehat{Var}(Z_{2s}G_2) \\
&- 2\widehat{Cov}(X_s B, Z_{1s}G_1) - 2\widehat{Cov}(X_s B, Z_{2s}G_2) - 2\widehat{Cov}(Z_{1s}G_1, Z_{2s}G_2)
\end{aligned}$$

However, $Var(Z_s^U G^U)$ is not identified separately from the common shock component $Var(v_{s_i}^*)$ and $Cov(v_{s_i}^*, Z_s^U G^U)$ without further assumptions.

[23]The equation below imposes $Cov(X_{si}B, v_i^*) = 0$, which is implied by our definition of B and v_i^*. The equation also assumes $Cov(Z_s, Z_s^U G^U) = 0$, which is implied by our definition of G and $Z_s^U G^U$ We do not need to separately consider $Cov(X_s B, Z_s^U G^U)$ because the elements of X_s are included in Z_s, and so $Cov(X_s B, Z_s^U G^U)$ is also 0.

2.2 Permanent Wage Rates

We focus on decomposing the permanent wage component W_{si}. We take advantage of the existence of panel data on wages in NLS72 and work with a balanced sample of individuals who report wages in both 1979 and 1986 (the fourth and fifth follow-ups, respectively). We estimate the variance in the permanent component of the wage, $Var(W_{si})$, using the covariance between wage observations from the same individual in different years

$$\begin{aligned} Cov(W_{sit}, W_{sit'}) &= Cov(W_{si} + e_{sit} + \xi_{sit}, W_{si} + e_{sit'} + \xi_{sit'}) \\ &= Var(W_{si}), \end{aligned}$$

where $Cov(\xi_{sit}, \xi_{sit'})$ is assumed to be 0 given that the observations are seven years apart and $Cov(e_{sit}, e_{sit'}) = 0$ from normalizing e_{sit} to be 0 in 1979. We use the sample estimate of $Cov(W_{sit}, W_{sit'})$ as our estimate of $Var(W_{si})$. We estimate this covariance by subtracting out the global mean for W_{sit}, calculating the wage product $(W_{sit})(W_{sit'})$ for each individual, and taking a weighted average across all the individuals in the sample using the weights discussed in Section 2 of the Appendix, adjusting for degrees of freedom. Similarly, we estimate the between-school component of the permanent wage, $Var(W_s)$, by estimating the covariance between wage observations for different years (1979 and 1986) from different individuals from the same school. Specifically, we use the moment condition

$$\begin{aligned} Cov(W_{sit}, W_{sjt'}) &= Cov(W_{si} + e_{sit} + \xi_{sit}, W_{sj} + e_{sjt'} + \xi_{sjt'}), i \neq j, t \neq t' \\ &= Var(W_s), \end{aligned}$$

where $Cov(e_{sit}, e_{sjt'})$ is defined to be 0, and $Cov(\xi_{sit}, \xi_{sjt'})$ is assumed to be 0. We estimate this covariance by first calculating $((W_{sit}W_{sjt'}) + (W_{sit'}W_{sjt}))/2$ for each pair of individuals i and j at school s and then computing the weighted mean for each school s. We then average across schools, weighting each school by the sum of the weights of the individuals who contributed to the school-specific estimate.

We estimate the corresponding within school component using

$$\widehat{Var}(W_{si} - W_s) = \widehat{Var}(W_{si}) - \widehat{Var}(W_s).$$

Given $\widehat{Var}(W_{si})$, $\widehat{Var}(W_{si} - W_s)$, $\widehat{Var}(W_s)$, $\hat{G}_1$, $\hat{G}_2$, and $\hat{B}$, estimation of the contributions of $X_{si}B$, $Z_{1s}G_1$, $Z_{2s}G_2$, v_i^*, and $Z_s^U G^U$ to $Var(W_{si})$ proceeds as in previous subsection.

2.3 High School Graduation and College Enrollment

For both of our binary outcomes, high school graduation and enrollment in a four-year college, we decompose the latent variable that determines the outcome. Given that there is no natural scale to the variance of the latent variable, we normalize $Var(v_i^* - v_s^*)$ to one, and define the total variance of the latent variable to be

$$Var(Y_{si}) = \widehat{Var}(X_s B) + \widehat{Var}(Z_s G) + \widehat{Cov}(X_s B, Z_s G) + \widehat{Var}(Z_s^U G^U + v_{s_i}^*) + 1$$

Given that the raw variance component estimates are subject to the choice of normalization, we instead report fractions of the variance contributed by the various components.

3 Construction and Use of Weights

In the NLS72 analyses of four-year college enrollment and postsecondary years of education, we use a set of panel weights (w22) designed to make nationally representative a sample of respondents who completed the base-year and fourth-follow up (1979) questionnaires. For the NLS72 wage analysis, we chose a set of panel weights (comvrwt) designed for all 1986 survey respondents for whom information exists on 5 of 6 key characteristics: high school grades, high school program, educational attainment as of 1986, gender, race, and socioeconomic status. Since there are very few 1986 respondents who did not also respond in 1979, this weight matches the wage sample fairly well. For the NELS88 sample, we use a set of weights (f3pnlwt) designed to make nationally representative the sample of respondents who completed the first four rounds of questionnaires (through 1994, when our outcomes are measured). For the ELS02 sample, we use a set of weights (f2bywt) designed to make nationally representative a sample of respondents who completed the second follow up questionnaire (2006) and for whom information was available on certain key baseline characteristics (gathered either in the base year questionnaire or the first follow-up). This seemed most appropriate given that our outcomes are measured in the 2006 questionnaire and we require non-missing observations on key characteristics for inclusion in the sample.

We use panel weights in the estimation for a number of reasons. The first is to reduce the influence of choice-based sampling, which is an issue in NELS88 and in the wage analysis based on NLS72. The second is to correct for non-random attrition from follow-up surveys. The third is a pragmatic adjustment to account for the possibility that the link between the observables and outcomes involves interaction terms or nonlinearities that we do not include. The weighted estimates may provide a better indication of average effects in such a setting. Finally, various populations and school types were oversampled in the three datasets, so that applying weights makes our sample more representative of the universe of American 8th graders, 10th graders, and 12th graders, respectively. Note, though, that we do not adjust weights for item non-response associated with the key variables required for inclusion in our sample. Thus, even after weighting, our estimates do not represent estimates of population parameters for the populations of American high school students of which the surveys were designed to be representative.