Group-Average Observables as Controls for Sorting on Unobservables When Estimating Group Treatment Effects: the Case of School and Neighborhood Effects

Joseph Altonji
Yale University and NBER

Richard Mansfield
Cornell University

September 12, 2014

Abstract

We consider the classic problem of estimating group treatment effects, such as school effects, on an outcome when people sort into groups based on observed and unobserved individual characteristics that affect the outcome. We show that under some circumstances standard choice models imply that group averages of observed individual-level characteristics can serve as controls for *all* of the across-group variation in the unobservable individual characteristics. This permits estimation of a lower bound for the variance of group treatment effects across groups. We use the idea to provide lower bound estimates of the effect of school systems and associated neighborhoods on high school graduation, college attendance, and earnings. Across three data sets, our most conservative estimates indicate that choosing a 90th quantile school and surrounding community instead of a 10th quantile school increases the probability of high school graduation by between 0.047 and 0.085 and increases the college attendance probability by between 0.11 and 0.13. We also find large effects on adult earnings. We discuss a number of applications of our methodology, including the measurement of teacher value added.

This research uses data from the National Center for Education Statistics as well as from North Carolina Education Research Data Center at Duke University. We acknowledge both the U.S. Department of Education and the North Carolina Department of Public Instruction for collecting and providing this information.

1 Introduction

Society is replete with contexts in which a) individual outcomes of interest depend on both individual and group-level inputs, and b) the group is endogenously chosen either by the individuals themselves or by administrators, partly based on the individual's own inputs. Examples of such combinations of outcomes and group inputs include health outcomes or costs and hospitals, earnings and workplace characteristics, and earnings and the university one attended.[1] Generations of social scientists have been interested in evaluating the productivity of such group-level inputs, but have struggled to overcome the fundamental sorting problem: do groups differ in their average outcomes because the group-level characteristics influence individual outcomes, or because the groups have succeeded or failed in attracting the individuals who would have thrived regardless of the group chosen? In some cases, sources of exogenous variation are available that may be used to assess the consequences of a particular group treatment. However, assessing the overall distribution of hospital or university quality, for example, is much more difficult.

In this paper, we develop a novel control function approach that allows us to place a lower bound on the variance in the contributions of group-level treatments to individual outcomes. In particular, we show that under some circumstances the standard method of controlling for group averages of observed individual-level characteristics in between group regressions, can absorb *all* of the between-group variation in both observable and unobservable individual inputs. This is true even when very general patterns of sorting on both observable and unobservable individual characteristics are involved. We further show that the lower-bound estimate we recover can be interpreted as the component of the group contribution that was unknown or unvalued by the individuals or administrators at the time they chose groups.

We apply our methodological insight and demonstrate its empirical value using the context of school choice, where we address a classic question in social science: How much does the school and surrounding community that we choose for our children matter for their long run educational and labor market outcomes?[2]

[1] See for example, Spetz et al. (2011) on hospital outcomes and Black and Smith (2004) on universities.

[2] See Duncan and Murnane (2011) for recent papers on school and neighborhood effects, with extensive references to

Our key insight follows directly from the parent's school/neighborhood choice decision: average values of student characteristics differ across schools only because students/families with different characteristics value school or neighborhood amenities differently. If tastes or willingness to pay for particular schools and their surrounding neighborhoods were not predictable based on student and parent characteristics, each school would feature the same joint distribution of these characteristics as the population.

Because the variation in school-average individual characteristics is fundamentally driven by the variation in school and neighborhood amenities, each school-average of an individual characteristic, whether observed or unobserved by the econometrician, will simply represent a different function of the same (full) vector of amenity factors (denoted $\mathbf{A}$) that parents consider when making their school choices. Thus, the sets of school averages of observed and unobserved characteristics, denoted $\mathbf{X}$ and $\mathbf{X}^U$ respectively, simply represent different vector-valued functions of the same common set of amenities: $\mathbf{X} = f(\mathbf{A})$ and $\mathbf{X}^U = f^U(\mathbf{A})$. If the dimension of the underlying amenity space is smaller than the number of observed characteristics, then under certain conditions one can potentially invert this vector-valued function to express the amenities in terms of school-averages of observed characteristics: $\mathbf{A} = f^{-1}(\mathbf{X})$. But this implies that the vector of school averages of unobserved characteristics can also be written as a function of observed characteristics: $\mathbf{X}^U = f^U(f^{-1}(\mathbf{X}))$.

We formalize this control function intuition by introducing and analytically solving a multi-dimensional spatial equilibrium model of school choice, and provide conditions under which this mapping is exact. We show that when parent's utility functions have a standard linear hedonic form and the joint distributions of student characteristics and amenities feature linear conditional expectations, the sorting-driven variation in average (observed or unobserved) student contributions across schools can be fully absorbed by a *linear* control function consisting of school-averages of observed characteristics. As we make precise in Proposition 1 below, X and X^u need not affect preferences for all of the amenities A. Roughly speaking, the keys are that (1) X affects preferences for all amenities that the uncorrelated elements of X^U shift preferences for, and (2) that there are

the literature.

enough elements of X to span this amenity space.

To take a simple example, suppose that school system/neighborhood combinations could be ordered into a single index of perceived quality–plus a random idiosyncratic taste component specific to each family/location combination. Suppose further that parental education (observed) and the uncorrelated part of parental wealth (unobserved) both predict willingness-to-pay for each neighborhood and directly contribute to student outcomes. In equilibrium the expected values of both parent's education and residual parent wealth will be increasing in perceived neighborhood quality, so that neighborhood-average parents' education will be a perfect proxy for neighborhood-average residual parent wealth.[3] If instead, however, school quality and, say, restaurant quality were not perfectly correlated, and only residual parent wealth predicted willingness-to-pay for restaurant quality, then differences in restaurant quality conditional on school quality would drive between-neighborhood differences in average residual parent wealth that would create between-neighborhood outcome variation that average parent's education could not predict. In this case we would need to control for the neighborhood average of another observable characteristic (e.g. parents' income) that also predicted willingness-to-pay for restaurant quality.

However, while this control function approach solves the sorting-on-unobservables problem, it also absorbs any variation in the underlying school and neighborhood amenities for which observed and unobserved outcome-relevant student characteristics affect tastes when the 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 were common to all members of the school). The second is a component of school quality whose valuation by parents is affected by only by student characteristics that do not influence the outcome (e.g., college attendance) under study. The variance of this residual between-school component therefore serves as a lower bound on the overall contribution of schools/neighborhoods to student outcomes. In the above example, only the component of school quality that was not observable when location

[3]To offset the between-neighborhood correlation and preserve the zero unconditional correlation, students at high quality schools with low residual wealth are even more likely to have high parent's education, so that a negative within-school correlation between the observed and unobserved characteristics will exist.

choices are made will vary conditional on the average values of parental income and education.

We also adapt the school choice model to the context of student assignment to classrooms within schools in order to demonstrate that our control function approach can be used in contexts where an administrator is endogenously choosing the group assignments for all individuals based on both observable and unobservable student traits.[4] We show that introducing classroom averages of observed student characteristics can potentially remove the bias from estimates of teacher value-added that stems from non-random assignment of students to teachers as well as part of the bias from non-random assignment of teachers to other classroom amenities (such as difficulty levels, periods of the day, classroom technology). As long as students are sorted based on a noisy signal of the teacher's quality (e.g. the principal's perception of teacher quality), identifying variation in true teacher quality necessary to recover point estimates of teacher value-added will still exist even conditional on average classroom characteristics (provided that the teacher is observed teaching multiple classrooms of students). Our analysis may help explain the very good performance of teacher value added models in predicting not only contemporaneous test scores but also future test scores and earnings. (See Chetty et al. (2014a) and Chetty et al. (2014b)).

The empirical part of the paper applies the control function approach in the school choice context. Implementation 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: three cohort-specific panel surveys (NLS72, NELS88, and ELS2002) along with administrative data from North Carolina. The advantage of the panel 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

[4]Non-random assignment of students to teachers is a central concern of the teacher value added literature. Rothstein (2010) considers the implications of classroom assignment based on student observables and unobservables for teacher value added estimation and concludes that it is a serious problem. Chetty et al. (2014a) question his interpretation of the evidence.

market outcomes for students.

For each dataset, we provide lower bound estimates of the overall contribution of differences in the school system and neighborhood quality to the variance in student outcomes: high school graduation, enrollment in a four-year college, and adult wages (NLS72 only).[5] In addition, we also convert each lower-bound variance estimate into a lower bound estimate of the impact on the chosen outcome of moving from a school system and surrounding neighborhood at the 10th quantile in the distribution of school contributions to a 50th or 90th quantile system (a more intuitive scale).

Even our most conservative 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. In the NELS88 and ELS2002 the corresponding are estimates of 4.7 and 6.8 percentage points, respectively, although these may be less reliable due to sampling error in school average characteristics. We estimate large average impacts despite the fact that our lower bound estimate only attributes between 1 and 4 percent of the total variance in the latent index determining graduation to schools. 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.

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: choosing a 90th quantile school and surrounding community instead of a 10th quantile school increases the probability of four-year college enrollment by at least 11-13 percentage points (across all three survey datasets), and increases adult wages by 19 percent (in NLS72).

The methodological part of the paper draws on and contributes to a number of literatures. First, the basic idea that observed choices reveal information about choice-relevant factors unobserved by the econometrician has been utilized in a number of settings, including the estimation of firm pro-

[5]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.

duction functions[6], labor supply functions [7], distinguishing between uncertainty and heterogeneity in earnings[8], and even neighborhood effects [9].

Second, this paper contributes to the theoretical sorting literature by presenting an analytical solution to a multidimensional sorting problem.[10] Third, this paper also overlaps with the literature on identification in multinomial choice models in which preferences for observed and (in some papers) unobserved product characteristics depend on both observed and unobserved consumer attributes. Much of this literature is focused on estimating preferences and the sensitivity of choice to relative prices and product characteristics in world in which product prices will depend on product characteristics.[11] To the best of our knowledge, we are the first to point out that the relationship between sorting on observables and unobservables implied by multinomial choice models and hedonic demand models implies that group averages of observables can serve as a control for group averages of unobservables.[12]

The empirical part of the paper adds to a large literature on school and neighborhood effects.[13] A number of recent papers in this literature have employed experimental or quasi-experimental strategies to isolate the contribution of either schools or neighborhoods to longer run student out-

[6]See Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2006), among others.

[7]See Metcalf (1974), Altonji (1982) and many subsequent studies of labor supply that use consumption or savings decisions as an (endogenous) proxy for the marginal utility of income or permanent wage rates.

[8]See Cunha et al. (2005), who exploit the fact that school choice will respond only to information about earnings returns that is known to the individual when the choice is made.

[9]Bayer and Ross (2006) consider using neighborhood house prices as a control function in estimating the impact of neighborhood characteristics. This approach, however, requires that the underlying amenities that determine neighborhood desirability can be combined into a single index of neighborhood quality.

[10]We consider an arbitrarily large number of amenities but solve the model only for the case in which amenities are exogenous and in particular do not depend on group characteristics. Many of the papers in the equilibrium sorting literature consider the case in which group characteristics affect choice but typically do not provide analytical solutions, although they may be computed. See Epple and Sieg (1999), Bayer et al. (2007), Bayer and Timmins (2005). Much of this literature is centrally concerned with the question of how individuals will sort in equilibrium as a function characteristics of a location, school, product, insurance contract, etc. The model we explicitly solve has a continuum of choices, and so relates closely to the hedonic demand literature, building on Rosen (1974), Heckman et al. (2010) among others.

[11]See for example, McFadden et al. (1978), McFadden (1984), Berry (1994), Berry and Pakes (2007) and Bayer and Timmins (2005).

[12]Despite the similarity in titles, our analysis is completely distinct from that of Altonji et al. (2005). That paper examines the econometric implications of how observed variables are drawn from the full set of variables that determine the outcome and the treatment variable of interest.

[13]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)

comes. 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 et al. (2007), randomly assigned housing vouchers that required movement to a lower income neighborhood to estimate neighborhood effects. None of these studies finds evidence that moving to a low-poverty neighborhood improves economic outcomes.

However, using a very different approach that exploits high quality data from tax records, Chetty and Hendren (In Progress) identify neighborhood effects on long-run outcomes that are more consistent with our results. Furthermore, Aliprantis (2011) argues that if one relaxes the assumption 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 from MTO is too narrow to draw meaningful conclusions about the importance of the full distribution of neighborhood quality. The experimentally-induced changes in school quality suffer from the same problem, so that the amount one can learn about the long-run value of higher quality schools is limited.

Deming et al. (2014), by contrast, 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 report large effects: 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, their estimates are specific to the Charlotte-Mecklenburg context, and 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.

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.

The rest of the paper proceeds as follows. Section 2 presents a simple model of long-run student outcomes. Section 3 presents our model of school choice, while Section 4 formally derives our key control function result. Section 5 adapts our model to the classroom assignment context, and derives its implications for estimating teacher value-added. Section 6 describes and presents results from a monte carlo analysis of the finite sample properties of our estimator. Section 7 describes our empirical methodology for placing lower bounds on school and neighborhood contributions to long run student outcomes. Section 8 describes the four datasets we use to estimate the model of outcomes. Section 9 presents our results. Finally, Section 10 concludes with a summary of our empirical results as well as a further discussion of 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 lower bound estimates of school/neighborhood contributions 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.[14] 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

$$Y_{s(i),i} = X_i\beta + X_i^U\beta^U + Z_{s(i)}\Gamma + Z^U_{s(i),i}\Gamma^U + u_{s(i),i}\,. \tag{1}$$

[14] A good example is Todd and Wolpin (2003), who provide references to the literature. See also Cunha et al. (2006).

The subscript $s(i)$ denotes the neighborhood and associated school of i for the high school years, which is chosen from $s = 1...S$. For the rest of the section we will usually suppress the dependence of s on i unless necessary for clarity. The row vector $[X_i, X_i^U]$ is an exhaustive set of child and family characteristics that have a causal impact on student i's educational attainment and wages; the subvector X_i is observed by the econometrician and the subvector X_i^U is unobserved. Since X_i and X_i^U may include non-linear functions of these attributes, imposing that the individual attributes enter linearly is without loss of generality.

The row vectors Z_s and $Z_{s,i}^U$ combine to form an exhaustive set of school and neighborhood influences experienced by student i. Z_s captures the influence of observed school/neighborhood-level characteristics (which in our empirical work do not vary among students within a school), and $Z_{s,i}^U$ represents the remaining unobserved school/neighborhood influences, which will vary between school attendance areas (e.g. quality of the school principal or the local crime rate) but also within a school attendance area and within a school itself (e.g. trustworthiness of immediate neighbors and distinct course tracks at the school). Indeed, some elements of $Z_{s,i}^U$ may represent the within-school components of Z_s, so that such elements will contain no between-school variation. The coefficients β, β^U, Γ, and Γ^U depend implicitly upon the specific outcome under consideration as well as the time period in the case of wages.

The component $u_{s,i}$ captures other influences on student i's outcome that are determined after secondary school but are not predictable given X_i, X_i^U, Z_s, and $Z_{s,i}^U$; these might include the opening or expansion of a local college or local labor market shocks that occur after high school is completed. It will prove useful to write $u_{s,i}$ as $u_s + u_i$, where u_s is common to all students at school s and u_i is idiosyncratic.

Note that while we allow the levels of school/neighborhood contributions Z_s or $Z_{s,i}^U$ to co-vary with individual-level contributions X_i and X_i^U, we rule out explicit structural interactions (such as products) between school and student characteristics. We discuss the issues involved in footnote 32 while describing our empirical methodology.

In practice we only have data on observed student and school inputs X_i and Z_s at a single

point in time. Thus, some components of X_i associated with student inputs (for example, student aptitude) will have been determined in part by parental inputs from earlier periods (for example, parent income). Such links make it difficult to interpret the coefficient associated with a given component of X_i, since once we have conditioned on the other components, we have removed many of the avenues through which the component determines Y. Consequently, we do not make any attempt to interpret individual components of the coefficient vectors B or B^U, and thus do not attempt to tease apart the distinct influences of child characteristics, family characteristics, and early childhood schooling inputs, respectively. Similarly, we do not attempt to remove bias in G stemming from correlations between Z_s and the omitted school/neighborhood factors $Z^U_{s,i}$. We aim instead to separate the effects of schools and associated community influences on outcomes from student, family, and prior school/community factors.

To be more specific about what we mean by school/neighborhood effects, first decompose Z^U_{si} into between- and within-school components: $Z^U_{si} = Z^U_s + (Z^U_{si} - Z^U_s)$. Then, note that if a randomly selected student attended school s^1 rather than s^0, the expected difference in his/her outcome would be $(Z_{s^1}\Gamma + Z^U_{s^1}\Gamma^U) - (Z_{s^0}\Gamma + Z^U_{s^0}\Gamma^U)$. The outcomes of a specific student i will also differ across schools because the values of $(Z^U_{si} - Z^U_s)$ and u_s will differ, but the former are entirely idiosyncratic and the latter are common to those who attend s^1 or s^0 but are determined after high school is completed. We wish to quantify the importance of differences across neighborhoods in $Z_s\Gamma + Z^U_s\Gamma^U$. Some of our estimates will also include differences due to u_s.

2.2 Toward an Empirical Model

In this section we discuss the parameters that OLS recovers when outcomes are regressed on only the student-level and school-level variables that can be observed in a survey or administrative dataset: X_i and Z_s.

To do this, consider projecting the full unobserved school contribution $Z^U_{s,i}\Gamma^U$ onto the vectors

of both the observed student level (X_i) and school-level (Z_s) variables[15]:

$$Z^U_{s,i} = \overbrace{(Z_s \Pi_{Z^U,Z} + \widetilde{Z^U_s})}^{Z^U_s} + \overbrace{(X_i \Pi_{Z^U,X} + \widetilde{(Z^U_{s,i} - Z^U_s)})}^{(Z^U_{s,i} - Z^U_s)}$$

Because we include the school-averages of each student-level variable in X_i in Z_s, standard partitioned regression results show that X_i will not predict any of the between-school components of $Z^U_{s,i}$ (denoted Z^U_s) that Z_s cannot predict. Thus, the matrix $\Pi_{Z^U,X}$ only captures the extent to which student-level observable characteristics predict the within-school variation in unobserved school characteristics. For example, parent's education may predict the course track the student is assigned to within the school. Likewise, Z_s, as a school-level variable, cannot possibly predict the student-specific deviations from the vector of school averages $(Z^U_{s,i} - Z^U_s)$. Thus, the matrix $\Pi_{Z^U,Z}$ only captures the extent to which unobserved influences common to all students at a school can be predicted by the vector of school-level observables.

Next, in order to more clearly demonstrate the impact of student sorting as separate from simple omitted variables bias, we project the vector of unobserved student-level characteristics X^U_i onto the space of observable variables in two steps instead of one. In the first step, we regress X^U_i on the student-level observable vector X_i only:

$$X^U_i = X_i \Pi_{X^U X} + \widetilde{X^u_i} \tag{2}$$

The matrix $\Pi_{X^U X}$ captures the extent to which the unobserved student-level contribution can be predicted by the observed student-level characteristics in the full population, and contributes to standard omitted variables bias in the coefficient on X_i even in the absence of non-random student sorting to schools. In the second step, we project the uncorrelated residual row vector from the first-step, $\widetilde{X^u_i}$, onto both the student-level and school-level vectors of observables (X_i and Z_s):

$$\widetilde{X^u_i} = X_i \Pi_{\tilde{X^u} X} + Z_{s_i} \Pi_{\tilde{X^u} Z} + \epsilon^{\tilde{X}}_{si} \,, \tag{3}$$

[15]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 case of $\Pi_{Z^U X}$, In this equation $D = Z^U_s$ and $Q = X_i$.

where $\epsilon_{si}^{\tilde{X}}$ is row vector. If students with greater unobservable contributions to their long run outcomes are more likely to sort into schools with particular observed characteristics Z_s, then the matrix $\Pi_{\tilde{X^u}Z} \neq \mathbf{0}$. Furthermore, even though each component of the vector $\widetilde{X_i^u}$ is uncorrelated with X_i (by construction from step 1), $\Pi_{\tilde{X^u}X}$ need not equal zero once school characteristics have been conditioned on. For example, parents with low income (included in X_i) who nonetheless choose an expensive school/neighborhood for their kids may be revealing high residual parental value for student's education outcomes (this unobserved characteristic might also improve their kids' outcomes regardless of school, thus belonging in X_i^U).

Substituting the results from these three projections for $Z_{s,i}^U$ and X_i^U in equation (1), we obtain:

$$Y_{s,i} = X_i B + Z_s G + v_s + (v_{si} - v_s), \quad where \tag{4}$$

$$B \equiv [\beta + \Pi_{Z^U X}\Gamma^U + (\Pi_{X^U X} + \Pi_{\tilde{X^U}X})\beta^U] \tag{5}$$

$$G \equiv [\Gamma + \Pi_{Z^U Z}\Gamma^U + \Pi_{\tilde{X^U}Z}\beta^U] \tag{6}$$

$$v_s \equiv \widetilde{Z_s^U}\Gamma^U + \epsilon_s^{\tilde{X}}\beta^U + u_s \tag{7}$$

$$v_{si} - v_s \equiv u_{si} + (\epsilon_{si}^{\tilde{X}} - \epsilon_s^{\tilde{X}})\beta^U + (\widetilde{Z_{si}^U - Z_s^U})\Gamma^U \tag{8}$$

The expressions for G and v_s in (21) and (22) make it clear that in the presence of general student sorting into schools/neighborhoods, both the observable school component $Z_s G$ and the unobservable component v_s reflect a mixture of school effects and student composition biases. This fact undermines any attempt to use basic variance decomposition techniques to identify or even bound the contribution of schools/neighborhoods to student outcomes. Specifically, the component $Z_s G$ will reflect $Z_s \Pi_{\tilde{X^u}Z}$, which captures differences in average unobservable student characteristics that are predictable by Z_s after conditioning on average observable student characteristics X_s. The unpredicted between-school component v_s will reflect $\epsilon_s^{\tilde{X}}\beta^U$, which captures differences in average unobservable student characteristics that are not predictable based on observed school-level characteristics or average student-level characteristics. Neither $Z_s\Pi_{\tilde{X^u}Z}$ nor $\epsilon_s^{\tilde{X}}\beta^U$ is a school/neighborhood effect, since any child who was reallocated to a school with a higher value of these components could

not expect an increase in test scores[16]. Some schools have higher average student outcomes simply because they attract more able students who would have performed better regardless of where they went to school. This is the essence of the problem of distinguishing school/neighborhood effects from composition effects. To say anything more specific requires a model of how students sort into neighborhoods and accompanying school systems, 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 school systems and associated neighborhoods, with the goal of placing minimal structure on parental preferences. In Section 4, we show that the model places restrictions on the relationship between X_s and X_s^U, the school level averages of X_i and X_i^U. Consequently, X_s can serve as a control for X_s^U.

Suppose that each location s, $s \in \{1, ..., S\}$, can be characterized by a vector of K underlying latent amenities, $\{A_{1s}, \ldots, A_{Ks}\}$ and the price of neighborhood s, P_s. We adopt a money-metric representation of the expected utility for the parents of student i from choosing school/neighborhood s, so that $U_i^*(s)$ can be interpreted as the family's consumer surplus from their choice. We assume the utility function takes the following linear form:[17]

$$U_i(s) = \boldsymbol{\gamma}_i \mathbf{A}_s - P_s + \epsilon_{s,i} \tag{9}$$

$$\text{where} \quad \boldsymbol{\gamma}_i \equiv (X_i \boldsymbol{\delta} + X_i^U \boldsymbol{\delta}^U + \kappa_i)' \tag{10}$$

$\boldsymbol{\gamma}_i$ is a $1 \times K$ vector of weights that captures the increases in family i's willingness to pay for a school per unit increase in each of its amenity factors $A_{1s}, \ldots, A_{ks}$, respectively. P_s is the price of living in the neighborhood surrounding school s, and $\epsilon_{s,i}$ is an idiosyncratic taste of the parent/student i for the particular location s.

The second equation projects the heterogeneity in strength of willingness-to-pay for particu-

[16] Note that peer effects stemming from concentration of particular types of students at a school are captured by either $Z_s\Gamma$ or $Z_s^U\Gamma^U$.

[17] The "prime" symbol denotes denotes matrix or vector transposes throughout paper.

lar amenities across parent/student combinations onto these families' observable and unobservable characteristics. In particular, suppose that X_i has L^O elements, while X_i^U has L^U elements. Then $\boldsymbol{\delta}$ ($\boldsymbol{\delta}^U$) is an $L^O \times K$ ($L^U \times K$) matrix whose $\delta_{\ell k}$-th entry ($\delta_{\ell k}^U$-th entry) captures the extent to which the willingness to pay for amenity A_k can be predicted given the determinant $x_{\ell i}$ ($x_{\ell i}^U$) of student outcomes. The $K \times 1$ vector κ_{ki} captures the components of i's taste for the amenities $A_1, \ldots, A_K$ that is unpredictable given $[X_i, X_i^U]$. Since $[X_i, X_i^U]$ is the complete set of student variables that determine $Y_{s,i}$, the elements of κ_i influence school choice but have no direct effect on student outcomes.

Expected utility is taken with respect to the information available when s is chosen. The information set includes the price, the amenities, and student/parent characteristics $[X_i, X_i^U]$. The information set does not include $u_{s,i}$, which is determined after the start of secondary school or later, nor does it include components of neighborhood and school quality that are not observable to families when a location is chosen. The set of amenities may include school/neighborhood characteristics that influence educational attainment and labor market outcomes.

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

$$s_i = \arg \max_{s=1,..,S} U_i(s)$$

4 Linking Group Observables to Group Unobservables

In this section we characterize the equilibrium allocation of students to schools from the model above under some simplifying assumptions. We approach the problem in two complementary ways. In Section 4.1, we exploit the first order condition of the family's location choice problem, taking equilibrium prices as given. In Section 4.2, we explicitly solve for the equilibrium allocation of families across schools under somewhat stronger assumptions.

4.1 Using the First Order Conditions of the Family's Choice Problem to Determine the Link Between X_s and X_s^U

For analytic simplicity, we consider a version of the school choice model in which (a) we ignore the idiosyncratic school-family taste match by setting $\epsilon_{is} = 0 \; \forall \; i, \; s$, and (b) we 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. In addition, we assume that parents behave competitively in the sense that prices are taken as given and choice is unrestricted.

Under assumptions (a) and (b), 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.[18] The optimal choice is characterized by a system of first order conditions, one for each amenity factor. The conditions are

$$\boldsymbol{\gamma}_i' \equiv \delta^{'} X_i^{'} + \delta^{U^{'}} X_i^{U\prime} + \kappa_i^{'} = \nabla P(A_{s(i)}) \, , \tag{11}$$

where $\nabla P(A_{s(i)})$ is the $K \times 1$ column vector of partial derivatives of $P(A)$ with respect to A, evaluated at $A = A_{s(i)}$. Since $P(A)$ is strictly convex, second order conditions will be satisfied.

We now use (11) to study the relationship between X_s^U and X_s. First, using the projection equation (2), we can rewrite the utility function in 9 in terms of X_i and $\tilde{X}_i^U$, where $\tilde{X}_i^U$ is the component of X_i^U that is orthogonal to X_i:

$$U_i(s) = [X_i \tilde{\boldsymbol{\delta}} + \tilde{X}_i^U \boldsymbol{\delta}^U + \kappa_i] A - P(A), \text{where} \tag{12}$$

$$\tilde{\boldsymbol{\delta}} = [\boldsymbol{\delta} + \Pi_{X^U X} \boldsymbol{\delta}^U] \tag{13}$$

Next, consider the projection of X_i and $\tilde{X}_i^U$ onto the index that determines location preference, $\boldsymbol{\gamma}_i \equiv X_i \boldsymbol{\delta}_x + X_i^u \boldsymbol{\delta}_{x^u} + \kappa_i = X_i \tilde{\boldsymbol{\delta}} + \tilde{X}_i^U \boldsymbol{\delta}^U + \kappa_i$, where we have rewritten $\boldsymbol{\gamma}_i$ to reflect equation (12). Note that since (1) $\tilde{X}_i^U$ is uncorrelated with X_i by construction, and (2) κ_i is uncorrelated

[18] In Section 4.2 we explicitly solve for $P(A_s^*)$ under some stronger assumptions and it is increasing and strictly convex.

with both X_i and $\tilde{X}_i^U$,

$$Cov(\boldsymbol{\gamma}_i, X_i) = \tilde{\boldsymbol{\delta}}' Var(X_i)$$
$$Cov(\boldsymbol{\gamma}_i, \tilde{X}_i^U) = \boldsymbol{\delta}^{U\prime} Var(\tilde{X}_i^U).$$

Thus, these projections can be written as

$$X_i = \boldsymbol{\gamma}_i Var(\boldsymbol{\gamma}_i)^{-1} \tilde{\boldsymbol{\delta}}' Var(X_i) + error_i^X$$
$$\tilde{X}_i^U = \boldsymbol{\gamma}_i Var(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{\delta}^{U\prime} Var(\tilde{X}_i^U) + error_i^{\tilde{X}^U} \ .$$

Using the first order conditions (11),we can rewrite these equations as:

$$X_i = \nabla P(A_{s_i})' Var(\boldsymbol{\gamma}_i)^{-1} \tilde{\boldsymbol{\delta}}' Var(X_i) + \ error_i$$
$$\tilde{X}_i^u = \nabla P(A_{s_i})' Var(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{\delta}^{U\prime} Var(\tilde{X}_i^U) + error_i \ .$$

Note that since choice of s_i depends on X_i, $\tilde{X}_i^U$, and κ_i only through the $K \times 1$ index vector $\boldsymbol{\gamma}_i$, 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:

$$X_{s_i} \equiv E(X_i | s_i) = \nabla P(A_{s_i})' Var(\boldsymbol{\gamma}_i)^{-1} \tilde{\boldsymbol{\delta}}' Var(X_i) \tag{14}$$

$$\tilde{X}_{s_i}^U \equiv \nabla P(A_{s_i})' Var(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{\delta}^{U\prime} Var(\tilde{X}_i^U). \tag{15}$$

Next, suppose that the columns of $\boldsymbol{\delta}^U$ are spanned by $\tilde{\boldsymbol{\delta}}$, which means that one can write $\boldsymbol{\delta}^U$ as

$$\boldsymbol{\delta}^U = R\tilde{\boldsymbol{\delta}} \tag{16}$$

for some matrix : $L^U \times L^O$ matrix R. To understand the restriction, recall that the matrix $\tilde{\boldsymbol{\delta}}$ that relates observables X_i to WTP for amenities is the sum of two components, $\boldsymbol{\delta}$ and $\Pi_{X^U X} \boldsymbol{\delta}^U$. The spanning condition requires that if WTP for any amenity is affected by a characteristic in X^U, then either WTP for this amenity must also be affected by a characteristic in X, or the characteristic in

X^U must be predictable based on a characteristic in X.

However, the restriction requires more than this. For example, suppose that X_{i1} influences tastes for all amenities for which WTP for the amenity depends on X^u, but none of the other elements of X_i influence tastes nor predicts the elements of X_i^U. In that case, the rank of $\tilde{\boldsymbol{\delta}}$ would be 1, and the row space of $\boldsymbol{\delta}^U$ would not be a linear subspace of the row space of $\boldsymbol{\delta}$, as spanning requires. Appendix Section 1 goes through several simple examples that illustrate when the spanning condition will and will not be satisfied.

Substituting the condition (16) for $\boldsymbol{\delta}^U$ in 15 implies that

$$\begin{aligned}
\tilde{X}_{s_i}^u &= E(\tilde{X}_i^u|s_i) = \nabla P(A_{S_i}^*)' Var(\boldsymbol{\gamma}_i)^{-1} \tilde{\delta}' R' Var(X_i^U) \\
&= \nabla P(A_{S_i}^*)' Var(\boldsymbol{\gamma}_i)^{-1} \tilde{\delta}' Var(X_i) Var(X_i)^{-1} R' Var(X_i^U) \\
&= X_{s_i} Var(X_i)^{-1} R' Var(X_i^U)
\end{aligned}$$

where the third line follows from substitution using (14). Note that from 2, we have:

$$X_{s(i)}^U = X_{s(i)}[\Pi_{X^U X} + Var(X_i)^{-1} R' Var(X_i^U)] \tag{17}$$

Thus, (17) implies that the between school component of the vector of student-level unobservables $X_{s_i}^U$ will be an exact linear function of X_{s_i} as long as 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.[19] We summarize the result in the following proposition.

Proposition 1: *Assume (i) preferences are given by (9), (ii)* $\varepsilon_{s(i)}$ is $0 \; \forall \; (i, s)$, *(iii) the price function* $P(A^*)$ *is increasing and strictly convex in* A^*, *and (i4) the rows of the coefficient matrix* δ^U *relating tastes for* A^* to X^u *are spanned by the rows of the coefficient matrix* $\tilde{\delta}$. *Then the*

[19] An interesting special case is the one in which $\boldsymbol{\delta}^U = 0$, so that unobservable characteristics do not affect location preferences. When $\boldsymbol{\delta}^U = 0$, $R = 0$. One can see from the above equation or directly by substituting $\boldsymbol{\delta}^U = 0$ into (15) 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.

expectation $X^U_{s_i}$ is linearly dependent on the expectation X_{s_i}.

Proposition 1 plays a central role in the empirical analysis of school effects. The finding that $X^U_{s_i}$ is a linear function of X_{s_i} means that school averages of student characteristics can serve as a 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 $X^U_{s_i}$ *because* the variation in both X_{s_i} and $X^u_{s_i}$ is driven by the same underlying variation in the desired amenity vector, $\{A_{1s}, A_{2s}, ...A_{K,s}\}$. Re-examining equations 14 and 15 above, we see that if δ_X (δ^U_X) is of full rank, then the vector $\mathbf{A}_s = [A_{1s}, A_{2s}, ...A_{Ks}]'$ can also be written as a linear function of X_{s_i} ($X^U_{s_i}$). 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 $\{A_{1s}, A_{2s}, ...A_{Ks}\}$.

On the other hand, components of school/neighborhood quality $Z_sG + Z^U_sG^U$ that are unknown (or unvalued) by parents at the time the school/neighborhood is chosen will not be reflected in the vector of amenities $A_{1s}, A_{2s}, ...A_{Ks}$ that are the basis of choice. Such components 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.

To see this more clearly, partition Z_s into two components $[Z_{1s}, Z_{2s}]$. $Z_{1s} \equiv X_{s(i)}$ is the vector of school-averages of observable student characteristics, while Z_{2s} is a vector of other observed school level characteristics not mechanically related to student composition (e.g. teacher turnover rate or student-teacher ratio). Partition the corresponding coefficient vectors $G \equiv [G_1, G_2]'$ and $\Gamma \equiv [\Gamma_1, \Gamma_2]'$ analogously. Then reconsider the equations in 8, substituting in the partitions of Z_s

and Γ:

$$Y_{s,i} = X_i B + Z_{1s} G_1 + Z_{2s} G_2 + v_s + v_{si}, \quad where \tag{18}$$

$$B \equiv [\beta + \Pi_{Z^U X} \Gamma^U + (\Pi_{X^U X} + \Pi_{\tilde{X^U} X}) \beta^U] \tag{19}$$

$$G_1 \equiv [\Gamma_1 + \Pi_{Z^U Z_1} \Gamma^U + \Pi_{\tilde{X^U} Z_1} \beta^U] \tag{20}$$

$$G_2 \equiv [\Gamma_2 + \Pi_{Z^U Z_2} \Gamma^U + \Pi_{\tilde{X^U} Z_2} \beta^U] \tag{21}$$

$$v_s \equiv \widetilde{Z_s^U} \Gamma^U + \epsilon_s^{\tilde{X}} \beta^U + u_s \tag{22}$$

$$v_{si} - v_s \equiv u_{si} + (\epsilon_{si}^{\tilde{X}} - \epsilon_s^{\tilde{X}}) \beta^U + \widetilde{(Z_{si}^U - Z_s^U)} \Gamma^U \tag{23}$$

Plugging equation 17 into equation 2 reveals that $\Pi_{\tilde{X^U} Z_1} = Var(X_i)^{-1} R' Var(X_i^U)$, $\Pi_{\tilde{X^U} Z_2} = 0$, and $\epsilon_s^{\tilde{X}} = 0$. Thus,

$$G_2 \equiv \Gamma_2 + \Pi_{Z^U Z_2} \Gamma^U \tag{24}$$

$$v_s \equiv \widetilde{Z_s^U} \Gamma^U + u_s \tag{25}$$

We see that the inclusion of $X_{s(i)}$ in Z_1 purges G_2 and v_s of biases from student sorting, so that $Var(Z_{2s} G_2)$ and $Var(v_s)$ only reflect true school/neighborhood contributions and, in the case of v_s, later common shocks. Note, though, that estimating a causal effect of a particular school characteristic in Z_2 still requires an instrument, since G_2 still reflects omitted variables bias driven by correlation between the chosen characteristic and other unobserved school characteristics in Z^U. We make no attempt to causal effects of particular school characteristics in this paper.

4.2 Solving for the Equilibrium Allocation

In this section we study the relationship between X_s and X_s^U by explicitly solving for the equilibrium allocation of families across schools under somewhat stronger assumptions. First, we assume that the joint distribution of $[X, X^U, \kappa]$ and the joint distribution of A_s are both multivariate normal. Second, we assume that A_s does not depend on which families choose s (i.e. the elements of A_s do not depend on directly on X_s or on X_s^u). This is a restrictive assumption in the school

choice setting, but may be quite plausible in other settings (e.g. differentiated product choice). Neither of these assumptions are required for Proposition 1 and so are not necessary to justify the use of school average observable characteristics as a control function for school average unobservable characteristics. We continue to assume that families freely choose locations subject to $P(A_s)$. This enables us to exploit the second welfare theorem result that any proposed allocation must be an equilibrium allocation if it satisfies the following two conditions: (a) the allocation is feasible, and (b) the allocation is pareto efficient.

An allocation in our model may be defined as a vector-valued function from $\mathcal{R}^{L^O} \times \mathcal{R}^{L^U} \times \mathcal{R}^K \rightarrow \mathcal{R}^K$ that maps a vector of student characteristics $[X, X^U, \kappa]$ into a vector of chosen school characteristics A. However, as noted earlier, when ϵ_{is} is excluded from the model, any vectors $[X, X^U, \kappa]$ sharing the same value of the index vector $\boldsymbol{\gamma}$ will feature the same willingness to pay for all possible neighborhoods. Consequently, the exact school assignment for a particular vector of characteristics $[X_i, X_i^U, \kappa_i]$ conditional on $\boldsymbol{\gamma}_i$ carries no welfare implications, which creates an infinite number of equilibrium allocations. Consequently, we first focus attention on how values of $\boldsymbol{\gamma}_i$ get mapped to amenity vectors A in equilibrium, and then examine implications for the distributions of X, X_U and κ_i within and across schools.

Define an "γ-allocation" to be a vector-valued function $\ominus : \mathcal{R}^K \rightarrow \mathcal{R}^K$ that maps a vector of K summary willingness-to-pay parameters $(M_{1i}, \ldots, M_{Ki})'$ for each of the amenities into a K-vector of chosen amenity levels $(A_1, \ldots, A_K)'$.

We restrict attention to linear γ-allocations of the form $A = \Psi M'$, where Ψ is a $K \times K$ matrix, and proceed to show that there is a unique linear γ-allocation that satisfies both feasibility and pareto efficiency. We have not yet proven but strongly suspect that the linear allocation represents the unique equilibrium γ-allocations when non-linear γ-allocations are also considered[20]

In Appendix Section 2, we show that feasibility of an allocation in the continuous context is equivalent to requiring that requires that the covariance matrices of the distributions of the amenities and the preference indices (denoted Σ_A and $\Sigma_{\boldsymbol{\gamma}'}$, respectively) be made congruent by the allocation

[20]Gretsky et al. (1992) proved the existence of an equilibrium for a class of games that includes our model.

matrix Ψ. We also show that pareto optimality of an allocation is equivalent to the requirement that the matrix Ψ be positive definite. However, there exists a unique positive-definite matrix that makes congruent to other positive definite matrices:

$$\Psi = \Sigma_{\boldsymbol{\gamma}'}^{-1/2}(\Sigma_{\boldsymbol{\gamma}'}^{1/2}\Sigma_A\Sigma_{\boldsymbol{\gamma}'}^{1/2})\Sigma_{\boldsymbol{\gamma}'}^{-1/2}. \tag{26}$$

Since the matrix Ψ is the unique matrix satisfying both the pareto efficiency and feasibility requirements, it is the unique (linear) equilibrium γ-allocation.

Furthermore, since every positive definite matrix is invertible, we can also express the vector $\boldsymbol{\gamma}_i$ for any individual as a linear function of the amenity vector of their chosen school:

$$\boldsymbol{\gamma}_i = (\Psi^{-1}A_{s(i)})'.$$

To characterize the equilibrium price function, $P(A)$, note that from (16) $\nabla P(A_{s_i}) = \boldsymbol{\gamma}_i'$. Substituting for $\boldsymbol{\gamma}_i'$ using the above equation for the equilibrium relationship between $\boldsymbol{\gamma}_i$ and $A_{s(i)}$, we obtain:

$$\nabla P(A_{s_i}) = \Psi^{-1}A_{s(i)} \tag{27}$$

The general solution to this linear first order partial differential equation is the following quadratic form:

$$P(A) = 0.5A'\Psi^{-1}A + c, \text{ for any } c \in \mathcal{R} \tag{28}$$

Furthermore, since Ψ^{-1} is the inverse of a positive definite function, it is also positive definite. And since the positive definite Ψ^{-1} represents the Hessian of the equilibrium price function, $P(A)$ must be strictly convex, as previously supposed.

4.3 The Link Between X_s and X_s^U with Finite School Choices and Idiosyncratic Tastes (Preliminary)

In formally deriving the link between X_s and X_s^U above, we focused on a simplified version of our original discrete choice model in which consumers choose from a continuum of options, and the idiosyncratic taste match component $\epsilon_{is} = 0 \; \forall \; (i, s)$. However, once we re-consider the original model, two related questions naturally arise. First, is the mean X_s^U still closely approximated by a linear function of the mean X_s? Second, is there still a perfect (possibly non-linear) link between the distribution of X_i and X_i^U across schools?

We are investigating these issues in two ways. First, in Section 6, we simulate the equilibrium of the school choice model (9) under alternative assumptions about the number of school choices, the number of elements of A_s, and other parameter choices. In many of the cases considered, X_s explains nearly all of the variation in X^u, despite the fact that a non-trivial variance in ϵ_{is} has been included. Note that in the additively separable specification for utility in (9), ϵ_{is} enters the equation symmetrically with respect to X_i and X_i^U, so that there is no a priori reason to believe that its addition should break the link between X_s and X_s^U. Indeed, in preliminary simulations not shown, we observe that increasing the variance of ϵ_{is} reduces the between school variance in both X_i and X_i^U symmetrically, but does not erode the effectiveness of X_s as a control for X_s^U if school sizes are sufficiently large (though the finite sample performance of the control function deteriorates slightly). Intuitively, as $Var(\epsilon_{is}) \to \infty$, idiosyncratic tastes fully drive choice, and the between school variation in X_i and X_i^U disappears, so that there is no more sorting problem to address.

To complement our simulations, we are also in the midst of a theoretical analysis of the joint distribution X_i and X_i^U conditional on $s\,(i) = s$ under the more general, finite choice model. We conjecture that under a spanning condition analogous to (16) the distribution $f(X_i^U | A_{s(i)}, s(i) = s)$ varies smoothly with the distribution of $f(X_i | A_{s(i)}, s(i) = s)$. More specifically, we conjecture that the quantiles of $f(X_i^U | A_{s(i)}, s(i) = s)$ are a function of the quantiles of $f(X_i | A_{s(i)}, s(i) = s)$ across s. The intuition is clearest in the case when X_i, X_i^u and A_s are all scalars. If $\delta > 0$ and

$\delta^U > 0$, then increases in A_s increase the willingness to pay to live in s of persons with high values of X_i relative to persons with low values. Consequently, the quantiles of X should be higher when A_s is higher. The same should be true for X_i^U, suggesting a link between the quantiles of the two distributions.

5 Application #2: Estimating Teacher Value-Added

In this section we show that the idea of using sorting across groups on observed agent characteristics to control for outcome-relevant unobserved characteristics can be applied to situations in which group assignments are the result of a centralized decision process rather than being competitively determined based on willingness to pay. The particular application we demonstrate is the measurement of teacher value added when a principal or other school administrator is charged with allocating students to high school classrooms. In this context, we argue that including observable peer characteristics along with teacher fixed effects will control for unobserved student sorting across classrooms (as well as partially controlling for differences in physical classroom quality) *without absorbing ALL OF the true variation in teacher quality*. This facilitates plausibly unbiased estimation of teacher value-added.

Assume for now that the administrator has already determined which teachers to allocate to which courses for which periods of the day, so that a classroom c can be effectively captured by a vector of amenity values $\mathbf{A}_c = [A_{1c}, \ldots, A_{Kc})]$. These amenities may include the principal's perceptions of various teacher attributes or skills, but could also include classroom amenities unrelated to teacher quality that might reflect whether the heating works, the quality of acoustics/lighting/classroom technology in the room, the time in the day that the class is held, or the difficulty level of the class. One could imagine that some components of A_c reflect the demographic make up of the class and are endogenous to the principal's assignment decisions. Including anticipated peer effects complicates the specification of principal preferences, since now the utility from assigning a given student to a classroom would depend on the other students assigned to the classroom.

In keeping with the basic structure of the sorting model laid out in Section 3, the value to the principal from allocating student i to classroom c is assumed to be represented by the following utility function

$$U_i^*(c) = \boldsymbol{\gamma}_i' \mathbf{A}_c + \epsilon_{c,i} \tag{29}$$

$$\text{where} \quad \boldsymbol{\gamma}_i \equiv (X_i \boldsymbol{\delta} + X_i^U \boldsymbol{\delta}^U + \kappa_i)' \tag{30}$$

We use the notation U_i^* rather than U_i to reflect the fact that U_i^* is utility while U_i is utility net of the shadow cost of assigning i to c given capacity constraints. As before, $\mathbf{X}_i$ represent a vector of student characteristics that are observed by the econometrician and are relevant for the outcome Y_i, the student's end-of-year standardized test score. Similarly, $\mathbf{X}_i^U$ is a vector of student characteristics that are unobserved by the econometrician but are observed by the principal and are relevant for test score performance, and κ_i represents a vector of student characteristics that are unobserved by the econometrician and observed by the principal, but do not affect test score performance. Finally, ϵ_{ic} captures the idiosyncratic value to the principal of allocating student i to classroom c that is unrelated to $\mathbf{A}_c$.

The δ parameter matrix might capture a principal's belief about which types of students are most likely to benefit from a better teacher or difficulty level, or it might simply reflect a desire to placate parents or students, where students/parents with certain values of $\mathbf{X}_i$ or $\mathbf{X}_i^U$ are more likely to advocate for particular classroom assignments (some parental or student characteristics may predict a stronger preference for a particular difficulty level or time of day, while others predict a stronger preference for teacher quality). Likewise, ϵ_{ci} might capture, for example, the desire to fulfill a particular parent's request that their child be assigned to the same teacher that his brother had.[21] Thus, we model parent and student preferences as affecting choice through their impact on principal preferences.

Suppose, as in Section 4.1, that we approximate the principal's allocation decision by 1) considering continuous joint distributions of both classroom amenities and student characteristics, and

[21] ϵ_{ci} might also, indirectly, capture the principal's desire to keep two troublesome students apart, or to allow two productive friends to stay together. As mentioned above, modeling such preferences over groups more directly adds considerable complexity to the model.

2) ignoring ϵ_{ci}. Implicitly, the principal places a shadow price on classrooms with characteristics A because of classroom capacity constraints, captured by the function $P(A)$. The principal's allocation problem is now isomorphic to the family's school choice problem from Section 3, and our analysis there applies here as well. In particular, a necessary and sufficient condition for the vector of classroom-average observables $E[X_i|i \in c]$ to serve as a control function for the vector of classroom-average unobservables $E[X_i^U|i \in c]$ is still the spanning condition $\boldsymbol{\delta}^U = R\tilde{\boldsymbol{\delta}}$, for some L^UxL^O matrix R. If this spanning condition holds, we can write $X^U_{c(i)} = X_{c(i)}\Pi_{X^U X} + X_{c(i)}Var(X_i)^{-1}R'Var(X_i^U)$.[22]

5.1 Distinguishing Classroom and Teacher Effects

Suppose that the true classroom contribution to a given student i's test scores stems from a vector of classroom factors $[D_{c(i)}\Gamma^{Tch}, Z^*_{c(i),i}\Gamma]$. Γ^{Tch} is a column vector that captures the true abilities of the population of teachers to increase test scores (the teachers' "value-added"), while $D_{c(i)}$ is a row vector of dummies indicating which classrooms were taught by which teachers.[23] $Z^*_{c(i),i}$ is a vector of other classroom factors that may include but is not limited to the factors $\mathbf{A}_c$ considered by the

[22] In the case in which $A_{c(i)}$ is fixed, and in addition X_i, X_i^u and A^* are jointly normal, we can solve the model if we restrict restrict our attention to linear allocations of the form $A_{c(i)} = \Psi\boldsymbol{\gamma}_i$. Let $f(\boldsymbol{\gamma}_i)$ denote the PDF of $\boldsymbol{\gamma}_i$ across the student population, and let $U(\boldsymbol{\gamma}, A) = \boldsymbol{\gamma}'A$ denote the rewritten utility function. Then we can write the principal's problem as:

$$\max_{\Psi} \int_{\underline{\gamma}}^{\overline{\gamma}} U(\boldsymbol{\gamma}, \Psi\boldsymbol{\gamma}) f(\boldsymbol{\gamma}) d\boldsymbol{\gamma}$$
$$s.t.\ E[\Psi\boldsymbol{\gamma}] = E[A]$$
$$s.t.\ Var(\Psi\boldsymbol{\gamma}) = Var(A)$$

Note that the constraints represent the same feasibility conditions as those from the school choice problem solved in Section 4.2, and pareto optimality remains a necessary condition for the objective function $\int_{\underline{\gamma}}^{\overline{\gamma}} U(\boldsymbol{\gamma}, \Psi\boldsymbol{\gamma}) f(\boldsymbol{\gamma}) d\boldsymbol{\gamma}$ to be maximized. Indeed, the equivalence of the two problems is essentially a manifestation of the first and second welfare theorems. Consequently, the solution $\Psi = \Sigma_{\boldsymbol{\gamma}'}^{-1/2}(\Sigma_{\boldsymbol{\gamma}'}^{1/2}\Sigma_A\Sigma_{\boldsymbol{\gamma}'}^{1/2})\Sigma_{\boldsymbol{\gamma}'}^{-1/2}$ will also be the same. In the case in which some of the A^* functions represent student demographics such as average parental education, the solution will be different, and the school choice problem and the class room assignment problems differ in an important way. In the school choice case parents make choices treating the A_s as fixed and do not internalize how their choices affect others. In contrast, the principal internalizes how assigning students of a given type to a classroom c affects A_c. In both cases, solving for the equilibrium is much more difficult, and multiple solutions are likely. However, at the equilibrium values of A the first order condition underlying the link between the group averages of X and X^U will still be satisfied. This is the key to our use of $X_{s(i)}$ as a control for $X^U_{s(i)}$.

[23] We assume here that teacher quality is not classroom-specific.

principal when allocating students to classrooms.[24] Including the additional i subscript allows for the possibility that these classroom factors may not be experienced symmetrically by all students in the classroom.

We can project the true value-added of the teachers in each classroom onto the vector of teacher attributes and classroom amenities perceived by the principal, denoted A_c above, as follows:

$$D_c\Gamma^{Tch} = A_c\lambda + \widetilde{D_c\Gamma^{Tch}}_c \tag{31}$$

where $\widetilde{D_c\Gamma^{Tch}}_c$ is a mean-zero error component capturing the part of the teacher's value added that is orthogonal to any of the perceived teacher skills or other classroom amenities that the principal acted upon when allocating students to classrooms. Such an error component might exist because the principal only has access to a noisy signal of teacher quality at a point in time.[25]Suppose that the vector of classroom-level characteristics available to the econometrician consists of classroom-level averages of observable student characteristics, denoted Z_{1c} along with the teacher assignment design vector, D_c. Letting $D_c \equiv Z_{2c}$ in the reformulated production function equation presented at the end of Section 3 (equation), we obtain:

$$Y_{c,i} = X_iB + Z_{1c}G_1 + D_cG_2 + v_c + v_{ci}\ . \tag{32}$$

If we assume that the spanning condition presented above holds, then as long as the principal's perception of teacher quality is noisy, so that D_c is not perfectly collinear with Z_{1c}, the vector of teacher value-added estimates recovered by OLS are given by:

$$G_2 = \Gamma^{Tch} + \Pi_{Z^UD_c}\Gamma^U. \tag{33}$$

Under what conditions is $\Pi_{Z^UD_c}\Gamma^U = 0$, so that estimates of teacher value added are unbiased? Note first that $\Pi_{Z^UD_c}\Gamma^U$ is not a bias from student sorting, but rather a potential bias stemming from

[24]Note that the experience of the current teacher might be included in $Z^*_{c(i),i}$ to capture the typical gains from learning by doing.

[25]An error component could also exist because maximizing student test scores is not the principal's objective, so that some component of a teacher's value added is not considered in allocation (even if known to the principal).

correlation between the assignment of the chosen teacher to the classroom and other aspects of the classroom environment. For example, some teachers may be more likely to teach at advantageous times of day or in better-equipped classrooms. Such classroom features might boost students' test scores irrespective of the teacher, and their contributions to performance should not be attributed to the teacher.

However, suppose that the other contributing classroom factors are considered when allocating students to classrooms, so that each element of Z_c^U is an element of (or function of) A_c. This could occur either because these environmental factors are known by the principal or because they are known by students/parents who then submit preference forms or requests that principals consider. If the vector of classroom factors is not too large, so that δ has full column rank, then equation (14) can be inverted, and $\nabla P(A_c)$ can be written as a linear function of X_c:

$$[\nabla P(A_c)]' = X_c Var(X_i)^{-1} \tilde{\boldsymbol{\delta}}' Var(\boldsymbol{\gamma}_i) \, . \tag{34}$$

If $\nabla P(A_c)$ is invertible, then this implies that classroom averages of observed student characteristics perfectly determine Z_c^U. An interesting special case is the case considered above in which the A_c do not depend on student classroom composition. In that case, $\Psi^{-1} A_{s(i)} = \nabla P(A_{s_i})$,

$$A_c = X_c Var(X_i)^{-1} \tilde{\boldsymbol{\delta}}' Var(\boldsymbol{\gamma}_i) \Psi'. \tag{35}$$

One can see that classroom averages of observed student characteristics will also perfectly predict Z_c^U, which further implies that $\Pi_{Z^U Z^1} = Var(X_i)^{-1} \tilde{\boldsymbol{\delta}}' Var(\boldsymbol{\gamma}_i) \Psi'$, and $\Pi_{Z^U Z^2} = 0$. Put another way, if all classroom factors that are inequitably distributed across teachers are also being used as a basis for student allocation to classrooms, then including classroom averages of observed student characteristics will also purge teacher value-added estimates G_2 of any bias driven by inequitable access to advantageous classroom environments.

Of course, in addition to needing large samples of students per teacher (so that the continuous model is a reasonable approximation), this argument also requires that teachers are observed teaching in multiple classroom environments (and perhaps at multiple times of day), so that the columns

of the design matrix corresponding to particular teachers is not collinear with the classroom environmental factors that schools are considering. In practice there may be classroom factors that students and principals do not consider that do not even out across teachers. The above analysis suggests, however, that the common practice of including classroom averages of student characteristics (such as in Chetty et al. (2014a)) plays a potentially powerful role in purging value-added estimates of biases stemming from non-random student sorting on unobservables and observables as well as non-random assignment of teachers to other unobserved outcome-relevant classroom environmental factors. Indeed, our analysis may partially explain the otherwise surprising finding that non-experimental OLS estimators of teacher quality produce nearly unbiased estimates of true teacher quality as ascertained by quasi-experimental and experimental estimates (Chetty et al. (2014a),Kane and Staiger (2008)).

6 Monte Carlo Simulations

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 characterizes the link between the expectations of X and X^u given A_s. However, with a finite number of students per school, random variation associated with κ_i^* and ϵ_{si}^*, will cause sample averages to deviate from their expectations, which could weaken the link between school averages of observable and unobservable characteristics.

To address these concerns, in this section 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. Note that the full spatial equilibrium sorting of students to schools depends on a huge number of parameters: all the parameters characterizing the joint distribution of $[X_i, X_i^U, \kappa_i]$, the joint distribution of the amenities A_s, and the distribution of the idiosyncratic tastes ϵ_{is}, as well as all the elements of the δ and δ^U matrices that capture how observed and unobserved characteristics affect willingness-to-pay. Thus, providing a full characterization of how the finite sample properties of

our estimator depend on all of these relationships is a daunting task and well beyond the scope of this paper. 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, 2) demonstrate that re-introducing the idiosyncratic student-school match components $\{\epsilon_{is}\}$ back into the model does not undermine the basic control function result established in Proposition 1, 3) 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, and 4) show that the control function approach is relatively robust to small departures from the conditions laid out in Proposition 1.

6.1 Methodology

The stylized test case we consider is one in which:

1. The elements of $[X_i, X_i^U, \kappa_i]$ are jointly normally distributed; the elements of κ are independent of each other and $[X_i, X_i^U]$, while each pair of characteristics in $[X_i, X_i^U]$ features a .25 correlation[26].

2. The latent amenity vectors A_s are normally distributed with a .25 correlation between any pair of amenities across schools.

3. The matrices of taste parameters $\delta_{k\ell}$ and δ^U represent draws from a multivariate normal distribution in which $corr(\delta_{k\ell}, \delta_{jm}) = \rho$ if $j = k$ or $\ell = m$, and 0 otherwise.

4. The variances of the elements of A_s, $[X_i, X_i^U, \kappa_i]$, $\epsilon_{i,s}$ are chosen to create interclass correlations for X_i and X_i^U of between .1 and .25 across specifications (in line with the range observed across the datasets we use in our empirical analysis).

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

[26]This is the average correlation between observed continuous student-level characteristics in ELS2002.

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, we complete the model by choosing particular sets of 7 remaining parameters. The first parameter sets the total number of students per school. For simplicity, we impose that each school has capacity equal to this student/school ratio.[27] The student/school ratio is denoted "# Stu" in Table 6. The second parameter sets the total number school/neighborhood combinations available (denoted "# Sch").

The third parameter (denoted "# Con."), indicates 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 "# Ob." and "# Un.") capture the number of observed and unobserved student characteristics that affect outcomes. The sixth parameter determines the dimension of the amenity vector over which households have preferences. It is always less then or equal to the number of observed characteristics, so that $\boldsymbol{\delta}$ may span $\boldsymbol{\delta}^U$, as required by Proposition 1.

The seventh parameter determines ρ, the correlation between any pair of random variables $(\Delta_{k\ell}, \Delta_{jm})$ from which each $(\delta_{k\ell}, \delta_{jm}$ is a draw. If ρ is high, then student characteristics that predict a high willingness to pay for one amenity factor will also predict a high willingness to pay for other amenity factors, and amenity factors that are strongly weighted by one characteristic are likely to be strongly weighted by other characteristics (i.e. WTP for some amenity factors may generally be sensitive to student characteristics).

[27] 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.

In addition, we also consider four additional specifications that illustrate the degree to which our control function approach is robust to various failures of the spanning condition from Proposition 1 (i.e. cases in which $\boldsymbol{\delta} \neq \Psi \boldsymbol{\delta}^U$ for any Ψ).

Our measure of control function effectiveness, denoted "R-sq (All)", is the R-squared from a regression of the total contribution of school-averages of unobservable characteristics $(X^U_{s_i} B^U)$ 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.

We also present the R-squared values calculated when random samples of 10, 20, or 40 students from each school are used to calculate the school averages $\overline{X}_{s_i}$ that compose the control function (these values are denoted "R-sq (10)", "R-sq (20)", and "R-sq (40)", respectively, in our tables).

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.[28] 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. [29]

6.2 Simulation Results

The simulation results are presented in Table 6. Row (1) presents the base parameter set to which other parameter sets will be compared. It features 1000 students per school and 50 schools in the area, all of which are considered by each family when the school choice is made. It also features 10

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

[29] 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.

amenities, 10 observable student characteristics, and 10 unobservable student characteristics. The variances of these characteristics are all identical, so that sorting on unobservables is as strong as sorting on observables, which likely to be a conservative choice). Finally, the correlation ρ between the random variables of which the taste weight matrices $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^U$ are multivariate draws is assumed to be .25.

The first takeaway from Row (1) is that our control function approach is effective even with reasonably-sized schools of 1000 students each (most of the schools in the North Carolina sample enroll between 250 and 2000 students) and a moderate number of available schools: 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 (Column 8).

The second insight revealed by Row (1) is that the performance of the control function may suffer when estimation is based on small subsamples of students at each school. We see that the R-squared falls from .972 to only .368 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 .501, while increasing it further to 40 students per school raises the R-squared to .640.

Rows (2) and (3) illustrate the impact of adapting the specification in Row (1) by decreasing or increasing the number of individuals per group. We see that the efficacy of the control function depends relatively strongly on the number of individuals per group. Decreasing school sizes from 1000 to 500 decreases the R-squared from .972 to .944, while increasing from 1000 to 2000 increases the R-squared to .986. Interestingly, increasing the number of individuals sorting into each group has almost no impact on the effectiveness of the control function if the larger number of individuals is not actually being used to construct the group averages of individual characteristics, $\overline{X}_s$; Columns (9) - (11) are nearly identical across Rows (1) - (3).

Comparing Row (4) to Row (1), we see that increasing the number of schools from 50 to 100 has very little impact on the performance of the control function when the full population of students is used to construct school averages. Interestingly, reducing the number of schools slightly ameliorates

the problems posed by using small samples of students from each school to construct $\overline{X}_s$. Row (5) shows that restricting the number of schools in each household's consideration set from 50 to 10 reduces the control function's ability to absorb unobservable sorting, but only slightly. The R-squared drops modestly from Row (1) to Row (5) across all columns (8) - (11). Nonetheless, the high R-squared in Row (5) reveals that our approach does not require households to be considering large numbers of schools.

Row (6) illustrates the impact of doubling both the number of observable and unobservable outcome relevant characteristics. By increasing the numbers of both observable and unobservable characteristics symmetrically, we can show the impact of utilizing a richer control set while holding fixed the strength of sorting on observables relative to unobservables.[30] Doubling the number of elements of X_i and X_i^U increases the R-squared from .972 in Row (1) to .982. This somewhat small increase understates the importance of the richness of the control set, since the control function was already nearly perfectly effective for the baseline parameter set. Row (10) shows that when only 20 students are used to construct sample school averages, doubling the control set from 10 to 20 characteristics increases the R-squared from .501 to .604. 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).

Row (7) shows that doubling the number of amenity factors from 5 to 10 dramatically reduces the effectiveness of the control function, dropping the R-squared from .972 in Row (1) to .914. Note, though, that increasing the dimension of the amenity space has a negligible impact when small samples of students are used to construct school averages (Columns (9) - (11)). However, Row (8), when compared to Row (6), reveals that the performance of the control function really depends on the dimension of the amenity space *relative* to the dimension of X_s, rather than the absolute number of amenities: when X_s has 20 elements, the fraction of absorbed sorting bias barely falls when the number of amenities rises from 5 to 10.

[30] 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!

Finally, Row (9) displays the results of a specification in which all of the $\delta_{k\ell}$ and $\delta k\ell^U$ elements are drawn independently ($\rho = 0$). The R-squared falls moderately across columns (8) - (11) relative to Row (1), which demonstrating the difficulty of fully characterizing the situations in which the control function will perform adequately: almost every parameter of the model affects the efficacy of the control function.

Overall, the results in Table 6 indicate that our control function approach could work quite well even in settings where 1) individuals have idiosyncratic tastes for particular groups, 2) there are only moderate number of total groups to join, and 3) only a subset of these are considered by any given individual. The simulations do suggest, however, that the control function may be less effective when 1) the dimension of underlying amenity factors is large relative to the number of independent factors represented by the observed individual characteristics, or 2) only a sample of individuals is observed in each group. We revisit the latter concern for our school effects application in Appendix Section 5.

Note that all the specifications in Table 6 consider cases in which the conditions presented in Proposition 1 are satisfied, so that we should expect the control function to perfectly absorb sorting on observables as the number of students per school gets sufficiently large. However, there may also be many contexts in which the set of observables is not sufficiently rich to make our spanning condition plausible. Thus, we are also interested in the extent which the addition of group-averages of individual characteristics can reduce bias from sorting on unobservables, even if it cannot eliminate the bias.

Table 7 presents the results from four specifications representing distinct scenarios in which our spanning condition fails. The worst-case scenario is one in which the unobservable characteristics only predict willingness-to-pay for a set of amenities that the observable characteristics do not affect taste for. We implement this scenario by setting all the parameters at their values from the baseline specification from Row (1) of Table 6, but allowing the elements of $\mathbf{X}_i^U$ to only predict WTP for the last amenity, A_5, while the elements of $\mathbf{X}_i$ only predict WTP for amenities A_1 to A_4. Since the group-averages of the observables and unobservables are functions of disjoint sets of amenities, we have no reason to expect that our control function approach would be effective, and it isn't: only

37% of the variance in X_s^U is predictable given X_s, even when the universe of students at each school is observed.

The second scenario alters the first scenario by allowing the unobservable characteristics X_i^U to predict willingness-to-pay for amenities A_1 to A_4 in addition to A_5. The control function performs somewhat better: 56% of the variance in X_s^U is absorbed by the coefficients on X_s.

These two scenarios are quite pessimistic, however. If willingness-to-pay for an amenity is unaffected by the entire vector $\mathbf{X}_i$, then it seems likely that a subset of the unobservables may not predict WTP for this amenity either. Thus, we consider two additional scenarios in which willingness-to-pay for the last amenity (A_5) is only predicted by one of the ten components of the unobserved vector $\mathbf{X}_i^U$. In the third scenario, $\mathbf{X}_{i,10}^U$ affects willingness-to-pay for A_5 only, while in the fourth scenario, $\mathbf{X}_{i,10}^U$ predicts willingness to pay for all amenities A_1 to A_5. Rows (4) and (5) reveal that our control function performs quite well in these scenarios: it absorbs 96% of the variation in X_s^U when $\mathbf{X}_{i,10}^U$ predicts willingness-to-pay for A_5 only, and 97% when $\mathbf{X}_{i,10}^U$ predicts willingness-to-pay for all five amenities (compared to the the baseline of 97.2% from Row (1)).

We conclude that our control function approach is likely to be quite robust to the violations of the spanning condition that are arguably the most plausible: those that involve just a few components of the individual's unobservable outcome contribution affecting willingness-to-pay for just a few additional amenities unweighted by X_i.

7 Estimating the Contribution of Schools and Neighborhoods

7.1 Variance Decomposition

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

$$Y_i = X_i\beta + Z_1G_1 + Z_2G_2 + v_{si}, \tag{36}$$

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:

$$Y_i = (X_i - X_{s_i})\beta + X_s\beta + Z_{1s}G_1 + Z_{2s}G_2 + Z_s^U G^U + (v_{si} - v_s) + v_s \tag{37}$$

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

$$Var(Y_i) \tag{38}$$

$$= Var(Y_i - Y_s) + Var(Y_s) \tag{39}$$

$$= [Var((X_i - X_{s_i})\beta) + Var(v_{si} - v_s)] + \tag{40}$$

$$[Var(X_s\beta) + 2Cov(X_s\beta, Z_{1s}G_1) + 2Cov(X_s\beta, Z_{2s}G_2) + Var(Z_{1s}G_1) + \tag{41}$$

$$2Cov(Z_{1s}G_1, Z_{2s}G_2) + Var(Z_{2s}G_2) + Var(v_s)] \tag{42}$$

Motivated by the model of sorting presented in Section 3, 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}, $Var(Z_{2s}G_2)+Var(v_s)$ will be purged of any effects of student sorting (observable or unobservable), so that it isolates only school/neighborhood factors.

However, in addition to the unpredicted component school/neighborhood contributions ($\widetilde{Z_s^U}\Gamma^U$), $Var(v_s)$ will include u_s, 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.

The static sorting model presented in Section 3 is silent about the point in time in a student's childhood the school/neighborhood decision is made. To illustrate how different assumptions about timing affect the interpretation of our bounds, consider first the case in which changing schools/communities is costless, so that each family decides each year where to live and send their children to school. In this case, if the data are collected in 10th grade (as in ELS2002), then any impact of prior schools/neighborhoods can be thought of as entering the outcome equation by altering the observable or unobservable student contributions X_i and X_i^U. Thus, if prior schooling inputs affect willingness-to-pay for school/neighborhood amenities, our control function argument suggests that 10th grade school averages of X_i and X_i^U will absorb all between-school variation in prior school contributions. In this case, the residual variance contributions $Var(Z_{2s}G)$ or $Var(Z_{2s}G + v_s)$ that we identify will represent a lower bound on the contributions of only the high schools and their surrounding neighborhoods to our outcomes.

Now consider the opposite extreme: moving costs are prohibitive, and each family makes a one time choice about where to settle down when they begin to have children. Suppose that the observed characteristics $\mathbf{X}_i$ are unaffected by early schooling.[31] Then the average values of $\mathbf{X}_i$ will span the subspace of the school/neighborhood amenities and of the average values of X_i^U as they existed when the family made its choice. In this scenario, the residual variance contributions $Var(\mathbf{Z}_{2s}G_2)$ or $Var(\mathbf{Z}_{2s}G_2 + v_s)$ that we identify will represent a lower bound on the variation in contributions to our later outcomes of entire sequences of schools (elementary, middle, and high) and entire childhoods of neighborhood exposure. In reality, of course, moving costs are substantial but not prohibitive, so that are estimates probably reflect a mix of elementary school and high school contributions, with a stronger weight on high school contributions. However, note that as long as high school quality in a neighborhood is positively correlated with elementary and middle school quality, a lower bound estimate of the variance in high school contributions is itself a (very conservative) lower bound estimate of the variance of contributions of entire school systems. Thus,

[31] As outlined in Section 8 below, we choose a set of variables in $\mathbf{X}_i$ that satisfies this property in our baseline specification for each dataset.

since our goal is to create an inviolable lower bound, the safest interpretation is that our estimates represent lower bounds on the variance in the cumulative effects of growing up in different school systems/neighborhoods.

Appendix Sections 3 and 4 describe the process by which the coefficients B, 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.

7.2 Measuring the Effects of Shifts in School/Community Quality

The fraction of outcome variance unambiguously attributable to school/neighborhood factors provides 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 average change in outcomes from a 10th-to-90th quantile shift in the full distribution of school/neighborhood quality, where the average is taken over the distribution of student contributions.

The more comprehensive estimate measures the unknown component of schoo/neighborhood quality via $\widehat{Var}(Z_2G_2+v_s)$, 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:

$$E^{noshocks}[\hat{Y}^{90} - \hat{Y}^{10}] = \tag{43}$$

$$\Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_{2s}\hat{G}_2 + 1.28(Var(Z_{2s}G_2))^{.5}]}{(1 + \widehat{Var}(v_s))^{.5}}) \tag{44}$$

$$- \Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_{2s}\hat{G}_2 - 1.28(Var(Z_{2s}G_2))^{.5}]}{(1 + \widehat{Var}(v_s))^{.5}}) \tag{45}$$

This weighted average effectively integrates over the distribution of $X_i\beta + Z_1G_1 + v_{s,i}$, but uses the empirical distributions of $X_i\beta$ and $Z_{1s}G_1$ (since they are observed) instead of imposing normality. Note that the scale of the latent index Y_i is unobserved, so we have normalized $Var(v_{si} - v_s)$ to 1.

We estimate the effect of the shift in $Z_{2s}G_2 + v_s$ analogously via:

$$E^{w/shocks}[\hat{Y}^{90} - \hat{Y}^{10}] = \tag{46}$$

$$\Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_{2s}\hat{G}_2 + 1.28(Var(Z_{2s}G_2 + v_s))^{.5}]}{(1)}) \tag{47}$$

$$- \Phi(\frac{[X_i\hat{B} + Z_{1s}\hat{G}_1 + \overline{Z}_{2s}\hat{G}_2 - 1.28(Var(Z_{2s}G_2 + v_s))^{.5}]}{(1)}) \tag{48}$$

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 $Z_{2s}G_2$ or $(Z_{2s}G_2 + v_s)$ 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.[32]

[32]Recall that we have ruled out interactions between X_i or X_i^U and Z_s or Z_s^U in the production of Y_i. To see how such nonseparabilities might be addressed, consider the simple case in which the interaction involves observed student and school characteristics. Suppose, for example, that low income students benefit disproportionately from a low student teacher ratio, one of the elements of Z_{2s}. One could add the interaction between family income and the student/teacher ratio to the outcome equation. If Proposition 1 holds, then the interaction between family income and the student teacher ratio will be unrelated to the error term conditional on $\bar{X}_s$, which includes the mean of family income. One can estimate the coefficient on the interaction term.

However, more generally, with nonseparable models schools may no longer be ordered. The best school for a low income student may not be the best school for a high income student. When the nonseparability involves observed variables, one may rank schools based on their average performance over the distribution of observed characteristics, and define the 10th and 90th percentile schools accordingly. Alternatively, one could identify the 10th percentile school and the 90th percentile school for each student, evaluate the difference in outcomes that the two schools, and then average over all students.

8 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). 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.[33] 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. In our previous work (Altonji and Mansfield (2011a)) we restricted attention only to variables that are available and measured consistently across all three datasets. However, because the efficacy of the control function approach introduced in this paper depends on the richness and diversity of our student-level measures, for each dataset we include in X student-level measures that may not appear in the other datasets.

[33] 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.

In our "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 results from a "full" specification which includes in X_{si} measures of student behavior, parental expectations, and student academic ability (standardized test scores). Such measures may be influenced directly by school inputs, so including them could cause an underestimate of the contribution of school-level inputs (our lower bound estimates will be too conservative). On the other hand, excluding such measures could instead cause an overestimate of the contribution of school-level inputs if this sparser set of student observables no longer satisfies the spanning condition stated in Proposition 1. In this case there would exist differences in average unobservable student contributions to outcomes across schools that are not predicted by the vector of school averages of observable characteristics. Tables 9 - 12 list 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 6 suggest that samples of this 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. Furthermore, we can use the North Carolina we assess the potential for bias in our survey-based estimates more directly by drawing samples of students from North Carolina schools using either the NLS72, NELS88, or ELS2002 sampling schemes and re-estimate the model for high school graduation using these samples. By comparing the results derived from such samples to the true results based on the universe of students in North Carolina, we can determine which if any of the survey datasets is likely to produce reliable results. Appendix ? reports the results of this exercise. It shows that using school sample sizes whose distributions match the NLS72, NELS88, or ELS2002 distributions generates only relatively minor biases, generally increasing $\hat{Var}(Z_{2s}G_2)$ and decreasing $\hat{Var}(Z_{2s}G_2 + v_s)$ by less than ten percent of their full sample values.

The North Carolina data are also the most recent: data are collected for all 2004-2006 public school 9th graders. On the other hand, 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. The set of observable characteristics is also not 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 12 provides a full list of the student- and school-level variables included in specifications using the North Carolina data.

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.[34] The sample college enrollment rate is 27 percent in NLS72, 31 percent in NELS88, and 37 percent in ELS2002. 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.10th 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

[34]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.

as of 1979. For the wage analysis, we include only respondents who report wages in both 1979 and 1986. The full variance decompositions described in Section 7 are provided for each of our outcomes in Appendix Tables 13, 2, and 15

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.[35] We then impute values for the other explanatory variables to preserve the sample size, since no one other variable is critical to our analysis.[36] 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 6

9 Results

9.1 High School Graduation

Panel A of Table 1 displays our lower bound estimates of the fraction of variance in the latent index that determines high school graduation that can be directly attributed to school/neighborhood choices for each dataset. The first row presents estimates that exclude $Var(v_s)$ (labeled "no unobs"), while the second row presents estimates that includes $Var(v_s)$ (labeled " w/ unobs"). However, recall that the motivation for excluding v_s is that it may reflect common shocks that occur after high school that may not be responsive to any changes in school or neighborhood policies. Since graduation is not really a post-secondary outcome, v_s is likely to contain only school and neighborhood contributions that are orthogonal to the observed school-level measures Z_2 (or sorting bias if the spanning condition from Proposition 1 fails). Thus, for high school graduation we focus on the re-

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

[36] 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.

sults that contain v_s. The first column displays the results from the baseline specification using the North Carolina data: our lower bound estimate is that at least 4.9 percent of the total student-level variance can be attributed exclusively to school system and neighborhood contributions. Since the set of observed individual-level measures X_i is somewhat sparse in the North Carolina data, it is possible that our control function of school-averages X_s does not span the full amenity space, so that unobservable sorting bias may contribute to this estimate. Thus, the second column displays results from the full specification that augments X_i by adding past test scores and measures of behavior. Since these measures could potentially have been altered by the school, including them removes some true school system contributions, but also makes the spanning condition in Proposition 1 more plausible. The estimated lower bound from 4.9 percent to 3.6 percent of the latent index variance.

Comparing the North Carolina results to those of NELS88 (Columns 3 and 4) and ELS2002 (Columns 5 and 6), a couple of noteworthy patterns emerge. First, across both specifications and both lower bound estimates, NELS88 features smaller fractions of outcome variance unambiguously attributable to schools/neighborhoods than either NC or ELS2002 ($\sim$ 1% relative to $\sim$ 2-3%). One possible explanation for this finding is that NELS88 school-level observables ($\mathbf{Z}_2$) reflect the 8th grade school environment while the corresponding measures in the other two datasets reflect the high school environment. It could be that the nature of the high school environment is particularly critical to dropout prevention. Second, comparing Row 2 across columns, we see that the North Carolina administrative data features the largest gap between the lower bound estimates that include versus exclude the school level residual, v_s, while the gap is negligible for ELS2002. This is not surprising; the North Carolina data has the sparsest set of school-level observables, which leads to a small $Var(Z_2G_2)$ relative to v_s, since less true variation in school quality is captured by observables. North Carolina also has the sparsest set of student-level observables (even in the full specification), which may cause v_s to contain some between school variation in student unobservables $X_s^U B^U$ that is unabsorbed by the control function (the spanning condition in Proposition 1 fails). By contrast, ELS2002 has the richest set of both student-level and school-level observables, so that there is very little residual school-level variation that cannot be captured by either the control function $\mathbf{Z}_1$ or the school-level observables $\mathbf{Z}_2$.

The small fractions of variance attributed to schools in Panel A are consistent with the considerable literature emphasizing the importance of student talent, parental inputs, and even luck relative to school and neighborhood inputs in determining who completes high school. However, to get a more intuitive sense of the difference that an effective school system and neighborhood can make, in Panel B we use these two alternative lower bound variance estimates to form estimates of the average impact on the probability of graduation across the distribution of student contributions of moving from a school at the 10th percentile of the distribution of school/neighborhood contributions 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 contribution distribution are placed either in the 10th or the 90th quantile school system, and the difference in the graduation status of these two pairs is summed over all such pairs.

The most striking feature of the results is the large magnitude of the estimated changes in graduation rates. For North Carolina, the estimate from the baseline specification suggests that, averaged across the student distribution, attending a 90th quantile school increases graduation rates by a whopping 17.4 percentage points relative to a school at the 10th quantile (from 67.6% to 85.0%) The corresponding estimates for 9.8 percentage points for NELS88 (80.7% to 90.5%) and 8.3 percentage points for ELS2002 (86.0% to 94.3%). Even the more conservative estimates from the full specification, which likely removes mostly true school/neighborhood contributions, suggest increases in graduation rates from a 10th-to-90th quantile shift of 15.2, 7.5 and 7.0 percentage points in NC, NELS88, and ELS2002, respectively. Notice further that these estimates are quite large despite the fact that the fractions of variance upon which they are based is quite small: 3.6, 1.6, and 2.5 percent for NC, NELS88 and ELS2002. One reason for this seeming disconnect is that squaring of deviations to produce variances will naturally mute moderate differences in school contributions relative to the standard deviations on which the 10-90 shifts are based. A second reason may be related to our reliance on the probit function and the assumption of normality. If the true distribution of latent student contributions is normal, and the graduation rate is not too high, 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.

Second, notice that even though the estimated lower bound fractions of variance were smaller for NELS88 than for ELS2002 in Panel A, the 10th-90th impact estimates displayed in Panel B are larger for NC. This is due to differences in the sample average graduation rates across the datasets. The graduation rate is 76 percent 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 NELS88 than in ELS2002 (and an even larger shift in NC), 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.

Given that our control function removes all variation in the amenities that families knew about or valued at the time of choice (assuming the conditions of Proposition 1 are satisfied or nearly satisfied), the large lower bound estimates suggest that school systems and neighborhoods have a considerable role to play in determining which students graduate high school.

9.2 Enrollment in a Four Year College

Panel A of Table 2 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 observe a surprising consistency in both lower bound estimates of the school/neighborhood contribution across datasets and generations. Estimates that exclude the between-school residual v_s attribute at least 2.2 to 2.6 percent of outcome variance to schools/neighborhoods, while estimates that include v_s attribute 3.8 to 4.6 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 1.5 to 1.9 percent, and the estimates that include common shocks dropping to 2.9 to 3.2 percent.

Panel B of Table 2 converts these variance fractions into the more easily interpreted average impacts of a 10th-to-90th quantile shift in school/neighborhood environment. Recall that the sample

average college enrollment rate is 27 percent in NLS72, 31 percent in NELS88, and 37 percent in 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. Despite these differences in baseline enrollment rates, the estimated lower bounds on the 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 fairly consistent across generations. When the residual component v_s is excluded and the full specification is considered, the estimates for each dataset are between 11 and 13 percentage points (Row 1, Columns 2, 4, and 6 of Panel B). Specifically, a 10th to 90th quantile shift in the school/neighborhood component $Z_{2s}G_2$ increases enrollment rates from 21.0% to 32.9% in NLS72, from 26.1% to 37.3% in NELS88, and from 30.2% to 43.4% in ELS2002. Including the residual between-school component boosts the range of estimates to 15 to 17 percentage points. Even 10th-to-50th quantile shifts still produce average estimated impacts between 5 and 8 percentage points.

As with the estimates for high school graduation, the estimates in Table 2 suggest that schools and neighborhoods also play an important role in determining who enrolls in a four-year college.

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

The estimates reported in Panel B of Tables 1 and 2 are based on starting the full distribution of students at a 10th quantile school and moving them to a 90th quantile school. However, many of the students with superior background characteristics would be quite unlikely to ever be observed in a 10th quantile 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. While our method does not allow us to discern the quality of any given school, we can nonetheless explore the extent to which the estimates in Tables 1 and 2 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 $X_{s,i}\hat{B}$ 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 enrollment in a four-year college.

Table 3 reports the lower bounds (excluding and including the school-level residual v_s) 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 (Column 2), the lower bound estimates that include the between-school residual component v_s suggest a 22.9 percentage point increase for students at the 10th quantile (43.2% to 66.1%) and a 6.3 percentage point increase for students at the 90th quantile (90.8% to 96.2%), respectively. For NELS88 grade 8 (Column 4), the numbers are smaller, particularly for the 90th quantile: lower bound estimates that include v_s are 15.9 percentage points (55.5 to 71.4) and 0.6 percentage points (99.0% to 99.7%). This partly 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 quite similar to those from NELS88. The results suggest that advantaged students tend to graduate high school regardless of the school they attend, while disadvantaged students are strongly affected by school quality.

Table 3 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 shift increases the predicted graduation rate among black students from 68.4% to 83.6% (a net gain of 15.2 percentage points). The corresponding increase for white students with single mothers who did not attend college is 20.6 percentage points (69.2% to 84.3%), while the increase for white students with both parents, at least one of whom completed college, is 8.4 percentage points (86.3% to 94.6%). The estimated increases in graduation rates are consistently smaller in the NELS88 and ELS samples, but are still between 5 and 12 percentage points for black students and for white students with single mothers who did not attend college.

Table 4 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 four year college enrollment margin. For example, the ELS2002 estimate from the full specification suggests that a 10th-90th shift in the school system/neighborhood component $Var(Z_{2s}G_2 + v_s$ would increase the four year enrollment rates of students at the 10th percentile of $X_i\hat{B}$ by 6.4 percentage points (from 2.1% to 8.6%). More generally, the lower bound estimates that exclude and include the residual v_s are between 2.7 and 5.0 percentage points and between 3.4 and 6.4 percentage points, respectively, depending on the dataset and specification. By contrast, for students at the 90th percentile of $X_i\hat{B}$ the ELS2002 estimate from the full specification suggests that a 10th-90th shift in the school system/neighborhood component $Var(Z_{2s}G_2 + v_s$ would increase enrollment rates at four year colleges by 16.7 percentage points (from 72.8% to 89.6%). The lower bound estimates excluding and including common shocks for students at the 90th percentile of the $X_i\hat{B}$ distribution are between 13 and 18 percentage points and 17 and 23 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 enrollment in a four year college.

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

Table 5 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 the between-school residual v_s implies that a 10-90 shift in school quality increases years of postsecondary education by .58 years, while including standardized tests among the observable characteristics reduces this estimate to .44 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 full specification a likely underestimate.

Adding the variance in the unexplained between-school component raises these estimates to .66 and .52 years respectively. 10th-to-50th quantile shifts are half as large by construction, since no non-linear transformation takes place when the outcome is continuous (the "latent" index is perfectly revealed). Collectively, the estimates suggest a substantive impact of shifts in school quality on years of college education.

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 the residual v_s imply that a 10-90 shift in school quality increases wages by around 17 percent. The 10-50 shifts are again half as large at around 8.5 percent. Estimates that include v_s imply that a 10-90 shift in school quality increases wages by around 19 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.

10 Discussion

The key takeaway from the results presented in this paper should be that even 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. (2014), 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 schools may constitute 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 complemen-

tarities and other interactions between school and student quality.[37] For example, it could be the case that the types of students who attend low quality schools are those that are most likely to profit from improvements in school quality, as the Deming et al. (2014) results suggest.

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 of 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 another 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 occupational sorting on unobserved characteristics that influence

[37] 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 quality versus student quality.

mortality is responsible for differences in mortality rates across occupations. Thus, one can directly identify the occupations that merit government-supported information campaigns or other safety regulations. We are currently pursuing this application.

References

Ackerberg, Daniel, Kevin Caves, and Garth Frazer (2006) 'Structural identification of production functions'

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

Altonji, Joseph G (1982) 'The intertemporal substitution model of labour market fluctuations: An empirical analysis.' *The Review of Economic Studies* 49(5), 783–824

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*

Altonji, Joseph G, Todd E Elder, and Christopher R Taber (2005) 'Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools.' *Journal of Political Economy* 113(1), 151

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 Christopher Timmins (2005) 'On the equilibrium properties of locational sorting models.' *Journal of Urban Economics* 57(3), 462–477

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

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007) 'A unified framework for measuring preferences for schools and neighborhoods.' *Journal of Political Economy* 115(4), 588–638

Berry, Steven, and Ariel Pakes (2007) 'The pure characteristics demand model.' *International Economic Review* 48(4), 1193–1225

Berry, Steven T (1994) 'Estimating discrete-choice models of product differentiation.' *The RAND Journal of Economics* pp. 242–262

Black, Dan A, and Jeffrey A Smith (2004) 'How robust is the evidence on the effects of college quality? evidence from matching.' *Journal of Econometrics* 121(1), 99–124

Chetty, Raj, and Nathaniel Hendren (In Progress) 'Quasi-experimental estimates of neighborhood effects on children's long term outcomes'

Chetty, Raj, John Friedman, and Jonah Rockoff (2014a) 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.' *American Economic Review* 104(9), 2593–2632

_ (2014b) 'Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood.' *American Economic Review* 104(9), 2633–2679

Cunha, Flavio, James Heckman, and Salvador Navarro (2005) 'Separating uncertainty from heterogeneity in life cycle earnings.' *Oxford Economic Papers* 57(2), 191–261

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, and Thomas J Kane (2014) 'School choice, school quality, and postsecondary attainment.' *American Economic Review* 104(3), 991–1013

Duncan, Greg J, and Richard J Murnane (2011) *Whither opportunity?: Rising inequality, schools, and children's life chances* (Russell Sage Foundation)

Ehrenberg, Ronald G, and Robert S Smith (2010) *Modern labor economics* (Pearson Addison Wesley)

Epple, Dennis, and Holger Sieg (1999) 'Estimating equilibrium models of local jurisdictions.' *Journal of Political Economy* 107(4), 645–681

Gretsky, Neil E, Joseph M Ostroy, and William R Zame (1992) 'The nonatomic assignment model.' *Economic Theory* 2(1), 103–127

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

Heckman, James J, Rosa L Matzkin, and Lars Nesheim (2010) 'Nonparametric identification and estimation of nonadditive hedonic models.' *Econometrica* 78(5), 1569–1591

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

Kane, Thomas J, and Douglas O Staiger (2008) 'Estimating teacher impacts on student achievement: An experimental evaluation.' Technical Report, National Bureau of Economic Research

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

Lawson, Jimmie, and Yongdo Lim (2006) 'Solving symmetric matrix word equations via symmetric space machinery.' *Linear algebra and its applications* 414(2), 560–569

Levinsohn, James, and Amil Petrin (2003) 'Estimating production functions using inputs to control for unobservables.' *The Review of Economic Studies* 70(2), 317–341

McFadden, Daniel et al. (1978) *Modelling the choice of residential location* (Institute of Transportation Studies, University of California)

McFadden, Daniel L (1984) 'Econometric analysis of qualitative response models'

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

Metcalf, Charles E (1974) 'Predicting the effects of permanent programs from a limited duration experiment.' *Journal of Human Resources* pp. 530–555

Olley, G Steven, and Ariel Pakes (1996) 'The dynamics of productivity in the telecommunications equipment industry.' *Econometrica* 64(6), 1263–1297

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

Rothstein, Jesse (2010) 'Teacher Quality in Education Production: Tracking, Decay, and Student Achievement.' *Quarterly Journal of Economics* 125(1), 175–214

Spetz, Joanne, Michael Ash, Charalampos Konstantinidis, and Carolina Herrera (2011) 'The effect of unions on the distribution of wages of hospital-employed registered nurses in the united states.' *Journal of clinical nursing* 20(1-2), 60–67

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

11 Tables and Figures

Table 1: Lower Bound Estimates of the Contribution of School Systems and Neighborhoods to High School Graduation Decisions

Panel A: Fraction of Latent Index Variance Determining Graduation Attributable to School/Neighborhood Quality

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	0.011	0.006	0.025	0.024
$Var(Z_{2s}G_2)$	(0.006)	(0.003)	(0.019)	(0.015)	(0.012)	(0.011)
LB w/ unobs	0.049	0.036	0.028	0.016	0.036	0.025
$Var(Z_{2s}G_2 + Z_s^U)$	(0.012)	(0.007)	(0.021)	(0.016)	(0.016)	(0.015)

Panel B: Effect on Enrollment Probability of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile

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	0.061	0.047	0.070	0.068
$Var(Z_{2s}G_2)$	(0.019)	(0.012)	(0.023)	(0.019)	(0.013)	(0.012)
LB w/ unobs: 10th-90th	0.174	0.152	0.098	0.075	0.083	0.070
$Var(Z_{2s}G_2 + Z_s^U)$	(0.023)	(0.016)	(0.021)	(0.019)	(0.013)	(0.013)
LB no unobs: 10th-50th	0.056	0.044	0.033	0.025	0.040	0.038
$Var(Z_{2s}G_2)$	(0.010)	(0.007)	(0.016)	(0.013)	(0.009)	(0.008)
LB w/ unobs: 10th-50th	0.096	0.083	0.055	0.041	0.048	0.039
$Var(Var(Z_{2s}G_2 + Z_s^U)$	(0.014)	(0.009)	(0.016)	(0.013)	(0.010)	(0.009)

Table 2: Lower Bound Estimates of the Contribution of School Systems and Neighborhoods to Four Year College Enrollment Decisions

Panel A: Fraction of Latent Index Variance Determining Enrollment Attributable to School/Neighborhood Quality

Lower Bound	NLS		NELS gr8		ELS	
	Baseline	w/Tests	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs	0.026	0.019	0.022	0.015	0.022	0.018
$Var(Z_{2s}G_2)$	(0.006)	(0.004)	(0.007)	(0.006)	(0.006)	(0.005)
LB w/ unobs	0.038	0.032	0.041	0.029	0.046	0.031
$Var(Z_{2s}G_2 + Z_s^U)$	(0.012)	(0.006)	(0.010)	(0.009)	(0.008)	(0.007)

Panel B: Effect on Enrollment Probability of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile

Upper/Lower Bound	NLS		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.139	0.118	0.140	0.112	0.155	0.132
$Var(Z_{2s}G_2)$	(0.013)	(0.010)	(0.019)	(0.016)	(0.017)	(0.016)
LB w/ unobs: 10th-90th	0.170	0.152	0.192	0.155	0.216	0.172
$Var(Z_{2s}G_2 + Z_s^U)$	(0.014)	(0.011)	(0.017)	(0.015)	(0.015)	(0.015)
LB no unobs: 10th-50th	0.065	0.056	0.067	0.054	0.075	0.064
$Var(Z_{2s}G_2)$	(0.006)	(0.005)	(0.008)	(0.007)	(0.008)	(0.007)
LB w/ unobs: 10th-50th	0.078	0.071	0.090	0.073	0.103	0.083
$Var(Z_{2s}G_2 + Z_s^U)$	(0.005)	(0.004)	(0.007)	(0.007)	(0.007)	(0.007)

Table 3: 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	0.110	0.099	0.123	0.140
$Var(Z_{2s}G_2)$	(0.025)	(0.018)	(0.040)	(0.040)	(0.023)	(0.024)
LB w/ unobs	0.242	0.229	0.176	0.159	0.146	0.144
$Var(Z_{2s}G_2 + Z_s^U)$	(0.031)	(0.023)	(0.038)	(0.040)	(0.023)	(0.026)
XB: 90th Quantile						
LB no unobs	0.060	0.036	0.016	0.004	0.019	0.010
$Var(Z_{2s}G_2)$	(0.011)	(0.006)	(0.007)	(0.003)	(0.004)	(0.002)
LB w/ unobs	0.098	0.063	0.026	0.006	0.022	0.010
$Var(Z_{2s}G_2 + Z_s^U)$	(0.014)	(0.007)	(0.007)	(0.003)	(0.005)	(0.003)
Black						
LB no unobs	0.107	0.085	0.061	0.053	0.079	0.082
$Var(Z_{2s}G_2)$	(0.019)	(0.012)	(0.018)	(0.024)	(0.015)	(0.015)
LB w/ unobs	0.176	0.152	0.098	0.084	0.094	0.084
$Var(Z_{2s}G_2 + Z_s^U)$	(0.023)	(0.016)	(0.018)	(0.027)	(0.015)	(0.016)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.142	0.114	0.099	0.078	0.101	0.096
$Var(Z_{2s}G_2)$	(0.024)	(0.016)	(0.037)	(0.033)	(0.021)	(0.020)
LB w/ unobs	0.235	0.206	0.159	0.125	0.120	0.099
$Var(Z_{2s}G_2 + Z_s^U)$	(0.030)	(0.021)	(0.035)	(0.033)	(0.024)	(0.023)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.062	0.047	0.025	0.016	0.032	0.016
$Var(Z_{2s}G_2)$	(0.011)	(0.007)	(0.010)	(0.006)	(0.007)	(0.005)
LB w/ unobs	0.102	0.084	0.040	0.025	0.037	0.016
$Var(Z_{2s}G_2 + Z_s^U)$	(0.014)	(0.009)	(0.009)	(0.006)	(0.007)	(0.006)

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 4: 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	0.078	0.027	0.067	0.032	0.100	0.050
$Var(Z_{2s}G_2)$	(0.008)	(0.003)	(0.011)	(0.006)	(0.011)	(0.007)
LB w/ unobs	0.094	0.034	0.091	0.046	0.138	0.064
$Var(Z_{2s}G_2 + Z_s^U)$	(0.010)	(0.005)	(0.012)	(0.007)	(0.012)	(0.008)
XB: 90th Quantile						
LB no unobs	0.191	0.182	0.176	0.128	0.166	0.128
$Var(Z_{2s}G_2)$	(0.019)	(0.017)	(0.024)	(0.023)	(0.018)	(0.017)
LB w/ unobs	0.234	0.234	0.243	0.187	0.231	0.167
$Var(Z_{2s}G_2 + Z_s^U)$	(0.009)	(0.016)	(0.024)	(0.025)	(0.018)	(0.017)
Black						
LB no unobs	0.132	0.109	0.133	0.111	0.145	0.121
$Var(Z_{2s}G_2)$	(0.015)	(0.012)	(0.019)	(0.018)	(0.016)	(0.014)
LB w/ unobs	0.161	0.140	0.182	0.152	0.201	0.158
$Var(Z_{2s}G_2 + Z_s^U)$	(0.018)	(0.015)	(0.020)	(0.020)	(0.015)	(0.014)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.110	0.099	0.098	0.074	0.140	0.124
$Var(Z_{2s}G_2)$	(0.012)	(0.010)	(0.015)	(0.013)	(0.016)	(0.015)
LB w/ unobs	0.134	0.127	0.134	0.102	0.195	0.162
$Var(Z_{2s}G_2 + Z_s^U)$	(0.015)	(0.013)	(0.017)	(0.015)	(0.017)	(0.016)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.180	0.158	0.173	0.139	0.173	0.148
$Var(Z_{2s}G_2)$	(0.018)	(0.015)	(0.023)	(0.020)	(0.019)	(0.018)
LB w/ unobs	0.220	0.204	0.238	0.192	0.242	0.193
$Var(Z_{2s}G_2 + Z_s^U)$	(0.017)	(0.015)	(0.022)	(0.020)	(0.018)	(0.018)

Notes: See Table 8

Table 5: 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)						
Lower Bound	Yrs. Postsec. Ed.		Perm. Wages No Post-sec Ed.		Perm. Wages w/ Post-sec Ed.	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th $Var(Z_{2s}G_2)$	0.578 (0.044)	0.445 (0.037)	0.152 (0.020)	0.157 (0.020)	0.155 (0.019)	0.157 (0.019)
LB w/unobs: 10th-90th $Var(Z_{2s}G_2 + Z_s^U)$	0.661 (0.046)	0.520 (0.050)	0.177 (0.067)	0.177 (0.066)	0.175 (0.067)	0.173 (0.066)
LB no unobs: 10th-50th $Var(Z_{2s}G_2)$	0.283 (0.022)	0.222 (0.018)	0.076 (0.010)	0.079 (0.010)	0.077 (0.010)	0.078 (0.010)
LB w/unobs: 10th-50th $Var(Z_{2s}G_2 + Z_s^U)$	0.331 (0.023)	0.260 (0.025)	0.088 (0.034)	0.088 (0.033)	0.087 (0.034)	0.087 (0.033)

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}.

Appendix

1 Spanning Condition Examples

Consider first a scenario in which there are two observed student characteristics $\mathbf{X} \equiv [X_1, X_2]$, two outcome-relevant unobserved student characteristics $\mathbf{X}^U = [X_1^U, X_2^U]$, and two school/neighborhood amenity factors, $\mathbf{A} = [A_1, A_2]$.

Case 1: $rank(\boldsymbol{\delta}^U) \leq rank(\boldsymbol{\delta}) = dim(\mathbf{A})$

Suppose that the matrices capturing the impact of observed unobserved student characteristics on parent willingness-to-pay for amenities $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^U$ are each full rank. For example:

$$\boldsymbol{\delta} = \left\{ \begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right\} \quad \boldsymbol{\delta}^U = \left\{ \begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array} \right\}$$

Then we can write $\boldsymbol{\delta}^U = R\boldsymbol{\delta}$, where

$$R = \left\{ \begin{array}{cc} 1 & 1 \\ 2 & -1 \end{array} \right\}$$

Thus, the spanning condition is satisfied in this case. If $\boldsymbol{\delta}^U$ were rank-deficient, then the spanning condition would still be satisfied, but R would be rank-deficient.

Now suppose that there are instead three outcome-relevant unobserved characteristics: $\mathbf{X}^U = [X_1^U, X_2^U, X_3^U]$, each of which affects willingness-to-pay for the two amenities differentially. Sup-

pose that X and $\boldsymbol{\delta}$ are unchanged from Case 1:

$$\boldsymbol{\delta} = \begin{Bmatrix} 1 & 1 \\ 0 & 1 \end{Bmatrix} \quad \boldsymbol{\delta}^U = \begin{Bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{Bmatrix}$$

Then we can write $\boldsymbol{\delta}^U = R\boldsymbol{\delta}$, where

$$R = \begin{Bmatrix} 1 & 1 \\ 2 & -1 \\ 1 & 0 \end{Bmatrix}$$

Thus, the spanning condition is satisfied in this case. We see that $dim(X)$ can be less than $dim(X^U)$ without violating the spanning condition, as long as the row rank of $\boldsymbol{\delta}$ is at least as large as the row rank of $\boldsymbol{\delta}^U$. Any scenario satisfying $rank(\boldsymbol{\delta}^U) \leq rank(\boldsymbol{\delta}) = dim(A)$ will satisfy the spanning condition in Proposition 1.

Case 2: $rank(\boldsymbol{\delta}) < rank(\boldsymbol{\delta}^U) \leq dim(\mathbf{A})$

Suppose instead that neither X_1 nor X_2 predicts willingness to pay for A_2:

$$\boldsymbol{\delta} = \begin{Bmatrix} 1 & 0 \\ 2 & 0 \end{Bmatrix} \quad \boldsymbol{\delta}^U = \begin{Bmatrix} 1 & 2 \\ 2 & 1 \end{Bmatrix}$$

Since $\boldsymbol{\delta}$ is now rank-deficient, there is no matrix R such that $R\boldsymbol{\delta} = \boldsymbol{\delta}^U$. In particular, for any matrix R, each entry in column 2 will always be zero, but the second column of $\boldsymbol{\delta}^U$ contains non-zero entries. Similarly, if both X_1 and X_2 affect willingness-to-pay for A_1 and A_2 in the same proportion, a rank-deficiency will also occur:

$$\boldsymbol{\delta} = \begin{Bmatrix} 1 & 2 \\ 2 & 4 \end{Bmatrix}$$

Here, an incremental unit of X_1 or X_2 will affect willingness-to-pay for A_2 by twice as much as it will affect willingness-to-pay for A_1. As in the previous example, there is no matrix R such that $R\boldsymbol{\delta} = \boldsymbol{\delta}^U$. For any choice of R, in each row of $R\boldsymbol{\delta}$ the second column will always be twice as large as the first column, but the second row of $\boldsymbol{\delta}^U$ has a first column entry that is only half as large as its second column entry. Both these examples violate the spanning condition. If the row rank of $\boldsymbol{\delta}$ is less than the row rank of $\boldsymbol{\delta}^U$, then the row space of $\boldsymbol{\delta}^U$ cannot possibly be a subspace of the row space of $\boldsymbol{\delta}$.

Case 3: $rank(\boldsymbol{\delta}^U) \leq rank(\boldsymbol{\delta}) < dim(\mathbf{A})$

Suppose now that both $\mathbf{X}$ and $\mathbf{X}^U$ are scalars: $\mathbf{X} \equiv X_1$, $\mathbf{X}^U \equiv X_1^U$. Consider first the case where X_1 only predicts willingness-to-pay for A_1, while X_1^U only predicts willingness-to-pay for A_2:

$$\boldsymbol{\delta} = \begin{Bmatrix} 1 & 0 \end{Bmatrix} \quad \boldsymbol{\delta}^U = \begin{Bmatrix} 0 & 1 \end{Bmatrix}$$

Regardless of the $1x1$ scalar R, the product $R\boldsymbol{\delta}$ will have a zero in the second column, which does not match $\boldsymbol{\delta}^U$. Despite the fact that $rank(\boldsymbol{\delta}) = rank(\boldsymbol{\delta}^U) = 1$, the spanning condition fails because the row space of $\boldsymbol{\delta}^U$ is not a subspace of the row space of $\boldsymbol{\delta}$.

Indeed, suppose that we alter $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^U$ so that both X_1 and X_1^U affect willingness to pay for both amenities (but in different proportions):

$$\boldsymbol{\delta} = \begin{Bmatrix} 1 & 1 \end{Bmatrix} \quad \boldsymbol{\delta}^U = \begin{Bmatrix} 2 & 4 \end{Bmatrix}$$

There is no scalar R such that $R\boldsymbol{\delta} = \boldsymbol{\delta}^U$, since any value of R will preserve the one-to-one ratio between the first and second entries in $\boldsymbol{\delta}$, while $\boldsymbol{\delta}^U$ has a one-to-two ratio between its first and second entries. The spanning condition also fails in this case because the row space of $\boldsymbol{\delta}^U$ is not a subspace of the row space of $\boldsymbol{\delta}$. This example demonstrates that if the set of factors that individuals

consider when choosing groups is large, one will generally need an equally large set of observable characteristics in order to satisfy the spanning condition in Proposition 1.

Finally, suppose that both X_1 and X_1^U only affect willingness to pay for A_1 (κ may affect taste for A_2, so that A_2 is still relevant for school choice):

$$\boldsymbol{\delta} = \left\{ \begin{array}{cc} 1 & 0 \end{array} \right\} \quad \boldsymbol{\delta}^U = \left\{ \begin{array}{cc} 2 & 0 \end{array} \right\}$$

Then for R = 2, $R\boldsymbol{\delta} = \boldsymbol{\delta}^U$, and the spanning condition is satisfied. Note that the row space of $\boldsymbol{\delta}$ is a subspace of the row space of $\boldsymbol{\delta}^U$, despite the fact that both $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^U$ are rank deficient. This last example illustrates that the observed characteristics need not predict willingness-to-pay for all choice-relevant amenities as long as the rows of $\boldsymbol{\delta}$ span the same (or a superspace) of the amenity subspace spanned by the rows of $\boldsymbol{\delta}^U$.

2 Proof of the Closed Form for the Equilibrium γ-allocation

Consider first the feasibility requirement: demand for each value of amenity vector $A_1, \ldots, A_K$ generated by the proposed γ-allocation must equal supply. In our continuous context, feasibility means that the distributions of amenity vectors supplied and demanded coincide. Let $\Sigma_{\boldsymbol{\gamma}_i'}$ denote the covariance matrix of the joint distribution of the K-vector $\boldsymbol{\gamma}_i'$ in the population, and let Σ_A denote the covariance matrix of the joint distribution of the K-vector $A_{s(i)}$ in the population, where the dependence on i indicates that we are considering the distribution that weights amenity combinations by the number of student slots. The overall capacity across all schools/neighborhoods is assumed to match the number of students, which in the continuous case means that the densities of $[X, X^U, \kappa]$ and A integrate to the same value, which we normalize to 1. Since the amenity vector demanded by a student with index $\boldsymbol{\gamma}_i$ in our proposed equilibrium is $\Psi\boldsymbol{\gamma}_i'$, our feasibility requirement can be written as:

$$\Sigma_A = Var(\Psi\boldsymbol{\gamma}') = \Psi\Sigma_{\boldsymbol{\gamma}'}\Psi' \tag{49}$$

Thus, feasibility requires that Σ_A and $\Sigma_{\gamma'}$ be made congruent by the allocation matrix Ψ. Since both ΣA and $\Sigma_{\gamma'}$ are full rank variance matrices, both are positive definite.[38] Any pair of positive definite matrices is known to be congruent to infinitely many matrices (there are infinitely many market-clearing ways to assign students to schools).

Now consider the pareto efficiency requirement. Since $P(A)$ is a transfer, and we assume that schools/landowners do not have preferences over families, requiring that the allocation S be pareto efficient is equivalent to requiring that there are no possible exchanges of neighborhoods between any pair of families that would make both families at least as well off and at least one family strictly better off.

Given a common price function $P(A)$ across families, such mutually beneficial exchanges will only exist if each family places a relatively higher valuation on the other family's chosen school than their own. Given that our money-metric utility function can be written as $M'A$ and the proposed M-allocation $A = \Psi M'$, a mutually beneficial exchange between families with the vectors M_1 and M_2 will exist iff $M_1\Psi M_1' - M_1\Psi M_2' < M_2\Psi M_1' - M_2\Psi M_2'$. Thus, ruling out such exchanges requires:

$$M_1\Psi M_1' - M_1\Psi M_2' \geq M_2\Psi M_1' - M_2\Psi M_2' \,\forall\, (M_1, M_2) \in \mathcal{R}^K x \mathcal{R}^K \tag{50}$$

But this equation can be reformulated as follows:

$$\begin{aligned}
&M_1\Psi M_1' - M_1\Psi M_2' \geq M_2\Psi M_1' - M_2\Psi M_2' \,\forall\, (M_1, M_2) \in \mathcal{R}^K x \mathcal{R}^K \\
&\Rightarrow (M_1 - M_2)\Psi M_1' - (M_1 - M_2)\Psi M_2' \geq 0 \,\forall\, (M_1, M_2) \in \mathcal{R}^K x \mathcal{R}^K \\
&\Rightarrow (M_1 - M_2)\Psi (M_1 - M_2)' \geq 0 \,\forall\, (M_1, M_2) \in \mathcal{R}^K x \mathcal{R}^K \\
&\Rightarrow \tilde{M}\Psi\tilde{M}' \geq 0 \,\forall \tilde{M} \in \mathcal{R}^K
\end{aligned} \tag{51}$$

where $\tilde{M}$ is defined to be the difference between two arbitrarily chosen row vectors M_1 and M_2. Equation 51 makes reveals that pareto efficiency requires Ψ to be a positive semi-definite or positive definite matrix. But a matrix Ψ that was merely positive semi-definite would be rank deficient, and

[38] If either Σ_A or $\Sigma_{\gamma'}$ were rank deficient, we could redefine the amenity factors to be a smaller number of linear combinations of the original K amenity factors.

would violate feasibility, since $rank(\Psi\Sigma_{M'}\Psi') \leq rank(\Psi) < rank(\Sigma_A)$. Thus, Ψ must be positive definite. But two positive definite matrices (in our case $\Sigma_{M'}$ and Σ_A) are made congruent by a unique positive definite matrix Lawson and Lim (2006):

$$\Psi = \Sigma_{M'}^{-1/2}(\Sigma_{M'}^{1/2}\Sigma_A\Sigma_{M'}^{1/2})\Sigma_{M'}^{-1/2} \tag{52}$$

Thus, since the matrix Ψ is the unique matrix satisfying both the pareto efficiency and feasibility requirements, it is the unique (linear) equilibrium M-allocation.

3 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.

3.1 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: $Z_s = [Z_{1s}; Z_{2s}]$. Z_{2s} consists of school and neighborhood characteristics for which direct measures are available, such as student/teacher ratio, city size, and school type. Z_{1s} 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_{1s} 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_{1s} with $\bar{Z}_{1s}$, where $\bar{Z}_{1s}$ is the average of X_{si} computed over all available students

from the school.[39] We estimate G by applying least squares regression to

$$Y_{si} - X_{si}\hat{B} = [\overline{Z}_{1s}G_1 + Z_{2s}G_2] \equiv \overline{Z}_s G + v_{s,i}$$

using the appropriate panel weights from the surveys.

3.2 Permanent Wage Rates

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

$$W_{s,i,t} = W_{s,i} + e_{s,i,t} + \xi_{s,i,t}.$$

In the above equation $W_{s,i}$ 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. $e_{s,i,t}$ is a random walk component that evolves as a result of luck in the job search process or within a company, or perhaps changes in motivation or productivity due to health and other factors. We normalize $e_{s,i,t}$ to be 0 in 1979.[40] $\xi_{s,i,t}$ 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 $W_{s,i}$. After substituting for W_{si}, the wage equation is

$$W_{s,i,t} = X_{s,i}B + Z_s G + v_{s,i} + e_{s,i,t} + \xi_{s,i,t}.$$

We estimate B by OLS with school fixed effects included.[41]

[39] A substantial number of students who appear in the base year of the surveys can be used to construct $\bar{Z}_{1s}$ but cannot be used to estimate (3.1) 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 $Y_{s,i}$. As we discuss in Section 8, 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.

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

[41] In reality, we also include a vector $T_{i,t}$ 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

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

$$\tilde{W}_{s,i,t} = \bar{Z}_s G + v_{s,i} + e_{s,i,t} + \xi_{s,i,t} \tag{53}$$

The presence of $\xi_{s,i,t}$ complicates the variance decompositions, as we discuss below.

3.3 High School Graduation and College Enrollment

The methods outlined in Appendix Sections 3.1 and 3.2 need to be adapted for binary measures such as high school graduation and college attendance. Consequently, for high school graduation we reinterpret $Y_{s,i}$ to be the latent variable that determines the indicator for whether a student graduates, $HSGRAD_{s,i}$. That is,

$$HSGRAD_{s,i} = 1(Y_{s,i} > 0).$$

Or, after substituting for Y_{si},

$$HSGRAD_{s,i} = 1(X_i B + Z_s G + + v_{s,i} > 0) \tag{54}$$

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

$$HSGRAD_{s,i} = 1(X_i B + \bar{Z}_s G + (Z_s - \bar{Z}_s)G + v_{s,i} > 0) \tag{55}$$

using maximum likelihood probit. The procedure for enrollment in a four-year college is analogous to that of high school graduation.

4 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

adjust wages for differences in labor market experience and for whether the data are from 1979 or 1986 by subtracting $T_{i,t}\hat{\Psi}$ 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}.

lower bound on the importance of factors that are common to students from the same school.[42] As with parameter estimation, the details of our procedure depend upon the outcome. We begin with years of postsecondary education.

4.1 Years of Postsecondary Education

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

$$Var(Y_{s,i}) = Var(Y_{s,i} - Y_s) + Var(Y_s)$$

where $(Y_{s,i} - Y_s)$ is the part of $Y_{s,i}$ that varies across students in school s and Y_s is the average outcome for students from s. We estimate $Var(Y_{s,i} - Y_s)$ by using the sample variances of $Var(Y_{s,i} - \overline{Y}_s)$ 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

$$\widehat{Var}(Y_s) = \widehat{Var}(Y_{s,i}) - \widehat{Var}(Y_{s,i} - Y_s).$$

Then, from (4),

$$(Y_{s,i} - Y_s) = (X_i - X_s)B + (v_{s,i} - v_s)$$

and

$$Y_s = X_s B + Z_s G + v_s.$$

Thus, one may express the outcome variance as[43]

$$Var(Y_i) = [Var((X_i - X_s)B) + Var(v_{s,i} - v_s)] + \tag{56}$$

$$[Var(X_s B) + 2Cov(X_s B, Z_{1s} G_1) + 2Cov(X_s B, Z_{2s} G_2) + Var(Z_{1s} G_1) + \tag{57}$$

$$2Cov(Z_{1s} G_1, Z_{2s} G_2) + Var(Z_{2s} G_2) + Var(v_s)] \tag{58}$$

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

[43] The equation below imposes $Cov(X_{s,i} B, v_{s,i} - v_s) = 0$, which is implied by our definition of B and $v_{s,i} - v_s$. The equation also assumes $Cov(Z_s, v_s) = 0$, which is implied by our definition of G and v_s (see Section 2). We do not need to separately consider $Cov(X_s B, v_s)$ because the elements of X_s are included in Z_s, and so $Cov(X_s B, v_s)$ is also 0.

Given an estimate of B, $Var((X_i - X_s)B)$ can be estimated using its corresponding sample variance, $Var((X_i - \overline{X}_s)B)$. $Var(v_{s,i} - v_s)$ can then be estimated as $\widehat{Var}(Y_{s,i} - Y_s) - \widehat{Var}((X_i - X_s)B)$, and $Var(X_sB)$ can be calculated as $\widehat{Var}(X_iB) - \widehat{Var}((X_i - X_s)B)$. 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 $2Cov(X_sB, Z_{1s}G_1)$, $2Cov(X_sB, Z_{2s}G_2)$ and $2Cov(Z_{1s}G_1, Z_{2s}G_2)$ using the estimates of B, G_1, G_2 and the data $\bar{X}_s \equiv \bar{Z}_{1s}$ and Z_{2s}. $Var(v_s)$ can be calculated as

$$\begin{aligned}
&\widehat{Var}(v_s) = \\
&\widehat{Var}(Y_s) - \widehat{Var}(X_sB) - \widehat{Var}(Z_{1s}G_1) - \widehat{Var}(Z_{2s}G_2) \\
&- 2\widehat{Cov}(X_sB, Z_{1s}G_1) - 2\widehat{Cov}(X_sB, Z_{2s}G_2) - 2\widehat{Cov}(Z_{1s}G_1, Z_{2s}G_2)
\end{aligned}$$

4.2 Permanent Wage Rates

We focus on decomposing the permanent wage component $W_{s,i}$. 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, $Var(W_{s,i})$, using the covariance between wage observations from the same individual in different years

$$\begin{aligned}
Cov(W_{s,i,t}, W_{s,i,t'}) &= Cov(W_{s,i} + e_{s,i,t} + \xi_{s,i,t}, W_{s,i} + e_{s,i,t'} + \xi_{s,i,t'}) \\
&= Var(W_{s,i}),
\end{aligned}$$

where $Cov(\xi_{s,i,t}, \xi_{s,i,t'})$ is assumed to be 0 given that the observations are seven years apart and $Cov(e_{s,i,t}, e_{s,i,t'}) = 0$ from normalizing $e_{s,i,t}$ to be 0 in 1979. We use the sample estimate of $Cov(W_{s,i,t}, W_{s,i,t'})$ as our estimate of $Var(W_{s,i})$. We estimate this covariance by subtracting out the global mean for $W_{s,i,t}$, calculating the wage product $(W_{s,i,t})(W_{s,i,t'})$ for each individual, and taking a weighted average across all the individuals in the sample using the weights discussed in Appendix Section 6, 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

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

where $Cov(e_{s,i,t}, e_{s,j,t'})$ is defined to be 0, and $Cov(\xi_{s,i,t}, \xi_{s,j,t'})$ is assumed to be 0. We estimate this covariance by first calculating $((W_{s,i,t}W_{s,j,t'})$ + $(W_{s,i,t'}W_{s,j,t}))$/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

$$\widehat{Var}(W_{s,i} - W_s) = \widehat{Var}(W_{s,i}) - \widehat{Var}(W_s).$$

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

4.3 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 $Var(v_{s,i} - v_s)$ to one, and define the total variance of the latent variable to be

$$Var(Y_{s,i}) = \widehat{Var}(X_sB) + \widehat{Var}(Z_sG) + 2\widehat{Cov}(X_sB, Z_sG) + \widehat{Var}(v_s) + 1$$

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

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

Before considering estimates from the three survey datasets, 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 6 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.

In this appendix, we assess the potential for bias in our survey-based estimates more directly by drawing samples of students from North Carolina schools using either the NLS72, NELS88, or ELS2002 sampling schemes and re-estimate the model for high school graduation using these samples. By comparing the results derived from such samples to the true results based on the universe of students in North Carolina, we can determine which if any of the survey datasets is likely to produce reliable results. To remove the chatter produced by a single draw from these sampling schemes, we computed estimate averages over 100 samples drawn from each sampling scheme.

Table 8 presents the results of this exercise. For comparison, the first column of Panel A presents the variance decomposition described in Section 7 for the full North Carolina sample, while the first column of Panel B converts these variance components isolating school/neighborhood effects into our lower bound estimates of the average impact of moving from the 10th to the 90th quantile

of the distribution of school/neighborhood contributions. Columns 2 through 5 display the results from recompute these estimates subsamples of the North Carolina population featuring with the same distributions of school-specific sample sizes as in NLS72, ELS, grade 8 schools in NELS88 and grade 10 schools in NELS88.[44] Column 2 displays the results from recomputing these estimates using subsamples of the North Carolina student population featuring the distribution school-specific sample sizes observed among 12th grade schools in NLS72. We see that the use of small student samples at each school produces may actually produce a relatively small amount of bias. Most of the rows of Panel A match quite closely across Columns 1 and 2. Of particular interest are the last two rows of Panel A: we see that the NLS72 sample size distribution overstates the true variance fraction for the lower bound without common shocks, $Var(Z_2G_2)$, by 0.88%, and understates true variance fraction for the lower bound that may include common shocks, $Var(Z_2G_2+v_s)$, by 0.38%. These translate to over/under estimates of the impact of a 10th-90th quantile shift in school quality on the probability of graduation of .0188 and .0111, respectively. Comparing the full NC sample with the NELS88 grade 8 and ELS2002 results (Columns 3 and 5), we see a similar pattern. 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.

Column 4 reports results from NELS88 in which students are grouped by their 10th grade school rather than their 8th grade school. Since grade 10 schools were not part of the original NELS88 sampling frame, they feature particularly small samples of students, 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. These results reveal that considerable bias may be produced if student samples are sufficiently small. Looking at the last two rows of Panel A, we that the NELS grade 10 sample size distribution overstates the true variance fraction for the lower bound without common shocks by 1.7 percent, and the lower bound with common shocks 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. Due the poor performance of the NELS grade 10 school sample size

[44] 10th grade schools are based on the schools in which the origina 8th grade NELS sample are observed in the first follow-up survey.

distribution in our simulation test, we do not report any NELS88 results that group students by their grade 10 school.

6 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.

7 Appendix Tables

Table 6: Monte Carlo Simulation Results: Cases in which the Spanning Condition in Proposition 1 is Satisfied ($\boldsymbol{\delta}^U = \Psi\boldsymbol{\delta}$ For Some Ψ)

Row	# Stu.	# Sch.	# Con.	# Ob.	# Un.	# Am.	δ Corr	R-Sq (All)	R-Sq (10)	R-Sq (20)	R-Sq (40)
(1)	1000	50	50	10	10	5	0.25	0.972	0.368	0.501	0.640
(2)	500	50	50	10	10	5	0.25	0.944	0.374	0.497	0.641
(3)	2000	50	50	10	10	5	0.25	0.986	0.376	0.497	0.644
(4)	1000	100	50	10	10	5	0.25	0.969	0.293	0.443	0.595
(5)	1000	50	10	10	10	5	0.25	0.968	0.354	0.479	0.619
(6)	1000	50	50	20	20	5	0.25	0.982	0.495	0.604	0.715
(7)	1000	50	50	10	10	10	0.25	0.914	0.363	0.482	0.608
(8)	1000	50	50	20	20	10	0.25	0.976	0.492	0.580	0.682
(9)	1000	50	50	10	10	5	0	0.958	0.320	0.422	0.566

Stu.: Number of students per school

Sch.: Total number of schools

Con.: Number of schools in each family's consideration set

Ob: Number of observable student characteristics

Un: Number of unobservable student characteristics

Am.: Number of latent amenity factors valued by families

δ Corr: Correlation in δ_{lk} taste parameters across student characteristics for a given amenity and across amenities for a given student characteristic

R-sq(all): Fraction of between-school variance in unobservable student characteristics $X_s^U B^U$ explained by the control function $\overline{X}_s$ (sample averages $\overline{X}_s$ computed using all students)

R-sq(10/20/40): Fraction of between-school variance in unobservable student characteristics $X_s^U B^U$ explained by the control function $\overline{X}_s$ (sample school averages $\overline{X}_s$ computed using 10/20/40 students)

Table 7: Monte Carlo Simulation Results: Cases in which the Spanning Condition in Proposition 1 Fails

Row	WTP for $A_1 - A_4$	WTP for A_5	R-Sq (All)	R-Sq (10)	R-Sq (20)	R-Sq (40)
(1)	Depends on all elements of $\mathbf{X}_i$ and $\mathbf{X}_i^U$	Depends on all elements of $\mathbf{X}_i$ and $\mathbf{X}_i^U$	0.972	0.368	0.501	0.640
(2)	Depends on each element of $\mathbf{X}_i$, independent of all $\mathbf{X}_i^U$	Depends on each element of $\mathbf{X}_i^U$, independent of $\mathbf{X}_i$	0.369	0.260	0.282	0.313
(3)	Depends on all elements of $\mathbf{X}_i$ and $\mathbf{X}_i^U$	Depends on each element of $\mathbf{X}_i^U$, independent of $\mathbf{X}_i$	0.561	0.296	0.361	0.440
(4)	Depends on all elements of $\mathbf{X}_i$ and $\mathbf{X}_i^U$	Depends on only on $\mathbf{X}_{i,10}^U$	0.961	0.367	0.490	0.635
(5)	Depends on all elements of $\mathbf{X}_i$ and $\mathbf{X}_i^U$ except $\mathbf{X}_{i,10}^U$	Depends only on $\mathbf{X}_{i,10}^U$	0.970	0.406	0.553	0.693

All specifications in Panel B share the following parameter values: # Stu. = 1000, # Sch. = 50, # Con. = 50, # Ob = 10, # Un = 10, # Am. = 5, δ Corr = 0.25 (See Table 7 for definitions of parameters).

Table 8: 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	NLS72	NELSg8	NELSg10	ELS2002
Within School:					
Total $Var(Y_{is} - Y_s)$	0.9153	0.9126	0.9131	0.8763	0.9120
Observable Student-Level (Within): $Var((X_{si} - X_s)B)$	0.1244	0.1296	0.1296	0.1301	0.1285
Unobservable Student-Level (Within) $Var(V_{si})$	0.7909	0.7828	0.7834	0.7461	0.7834
Between School:					
Total $Var(Y_s)$	0.0847	0.0874	0.0869	0.1237	0.088
Observable Student-Level: $Var(X_sB)$	0.0181	0.018	0.0183	0.0179	0.0184
Student-Level/ School-Level Covariance $2 * Cov(X_sB, Z1_sG_1 + Z2_sG_2)$	0.0165	0.0175	0.0170	0.0187	0.175
School-Avg. Student-Level/ School Char. Covariance $2 * Cov(Z1_sG_1, Z2_sG_2)$	-0.0166	-0.0047	0.0061	-0.0053	-0.0054
School-Avg. Student-Level $Var(Z_{1s}G_1)$	0.0178	0.0125	0.0137	0.0290	0.0139
School Char. $Var(Z_{2s}G_2)$	0.0181	0.0269	0.023	0.0353	0.0238
Unobservable School-Level $Var(v_s)$	0.0309	0.0173	0.0211	0.0283	0.0199

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

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

Table 9: NLS72: Variables Used in Baseline and Full (in Italics) Specifications

Student Characteristics
Race Indicators, Gender Indicator
Student Ability
*Math Standardized Score**, *Reading Standardized Score**
Student Behavior
Family Background
Standardized SES, Number of Siblings, Indicators for Presence Biological Parents at Home, Father's Years of Education, Mother's Years of Education, Moth. Yrs. Ed. Missing, Log(Family Income),1(English Spoken at Home), Indicators for Parental Religion, Indicators for Father's Occupation Group, Indicators for Mother's Occupation Group Home Environment Indicators (1st Principal Component)
Parental Expectations
School Characteristics (Treated as elements Z_1*)
School Pct. Minority
School Characteristics (Treated as elements Z_2)
1(Catholic School), 1(Private Non-Catholic School), Total School Enrollment, Student-Teacher Ratio, Pct. Teacher Turnover Since Last Year, Pct. of Teachers w/ Master's Degrees or More, School Teacher Pct. Minority, 1(Tracking System), Age of School Building, Distance to 4-year College Distance to Community College
Neighborhood Characteristics
Urbanicity Indicators (Detailed), Indicators for U.S. Census Region

*School characteristics treated as elements of Z_1 are included to reduce measurement error in school sample averages of student characteristics. They do not contribute to the estimated lower bound on the contribution of schools/neighbhorhoods to outcomes

Table 10: NELS88: Variables Used in Baseline and Full (in Italics) Specifications

Student Characteristics
Race Indicators, 1(Female), 1(Immigrant), Self-Reported Athleticism Index
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*, *Physical Fight This Year*
Family Background
Standardized SES, Number of Siblings, Indicators for Presence Biological Parents at Home, Father's Years of Education, Mother's Years of Education, Moth. Yrs. Ed. Missing, Log(Family Income),1(English Spoken at Home), Indicators for Parental Religion, 1(Parents are Married), 1(Immigrant Father), 1(Immigrant Mother), Indicators for Father's Occupation Group, Indicators for Mother's Occupation Group, Home Environment Indicators (1st Principal Component), Parental School Involvement Indicators (1st Principal Component)
Parental Expectations
Mother's Desired Yrs. of Ed., *Father's Desired Yrs. of Ed.*
School Characteristics (Treated as elements Z_1*)
School Pct. Minority, School Pct. Free/Reduced Price Lunch, School Pct. LEP, *School Pct. Special Ed.*, *School Pct. Remedial Reading*, *School Pct. Remedial Math*
School Characteristics (Treated as elements Z_2)
1(Catholic School), 1(Private Non-Catholic School), Total School Enrollment, Student-Teacher Ratio, Pct. Teacher Turnover Since Last Year, Log(Min. Teacher Salary) Pct. of Teachers w/ Master's Degrees or More, School Teacher Pct. Minority, 1(Gifted Program Exists), 1(Collectively Bargained Contract), School Security Policy Indicators (1st and 2nd Principal Components)
Neighborhood Characteristics
Urbanicity Indicators (Urban/Suburban/Rural), Indicators for U.S. Census Region

*School characteristics treated as elements of Z_1 are included to reduce measurement error in school sample averages of student characteristics. They do not contribute to the estimated lower bound on the contribution of schools/neighbhorhoods to outcomes

Table 11: ELS2002: Variables Used in Baseline and Full (in Italics) Specifications

Student Characteristics
Race Indicators, 1(Female), 1(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*, *Hrs./Wk. Spent on Computer*, *Physical Fight This Year*
Family Background
Standardized SES, Number of Siblings, Indicators for Presence Biological Parents at Home, Father's Years of Education, Mother's Years of Education, Moth. Yrs. Ed. Missing, Average of Grandparents' Education, Log(Family Income),1(English Spoken at Home), Indicators for Parental Religion, 1(Parents are Married), 1(Immigrant Father), 1(Immigrant Mother), Indicators for Father's Occupation Group, Indicators for Mother's Occupation Group, Home Environment Indicators (1st Principal Component), Parental School Involvement Indicators (1st Principal Component)
Parental Expectations
Mother's Desired Yrs. of Ed., *Father's Desired Yrs. of Ed.*
School Characteristics (Treated as elements Z_1*)
School Pct. Minority, School Pct. Free/Reduced Price Lunch, School Pct. LEP, *School Pct. Special Ed.*, *School Pct. Remedial Reading*, *School Pct. Remedial Math*, *Frequency of Fights (Administrator's Impression)*
School Characteristics (Treated as elements Z_2)
1(Catholic School), 1(Private Non-Catholic School), Total School Enrollment, Student-Teacher Ratio, Pct. Teacher Turnover Since Last Year, Log(Min. Teacher Salary) Pct. of Teachers w/ Master's Degrees or More, Pct. of Teachers w/Certification, School Teacher Pct. Minority, 1(Minimum Competency Test Exists), Teacher Evaluation Mechanism Indicators (1st Principal Component), Teacher Incentives Indicators (1st Principal Component) School Security Policy Indicators (1st and 2nd Principal Components) School Security Implementation (ELS Facility Inspection, 1st and 2nd Principal Components) School Environment Indicators (ELS Facility Inspection, 1st and 2nd Principal Components), School Facilities Indicators (Administrator Survey, 1st and 2nd Principal Components), Teacher Access to Technology Indicators (Administrator Survey, 1st Principal Component),
Neighborhood Characteristics
Urbanicity Indicators (Detailed), Indicators for U.S. Census Region Neighborhood Crime Level Category (Sch. Administrator Survey)

*School characteristics treated as elements of Z_1 are included to reduce measurement error in school sample averages of student characteristics. They do not contribute to the estimated lower bound on the contribution of schools/neighbhorhoods to outcomes

Table 12: 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 13: Decomposition of Variance in Latent Index Determining High School Graduation for Baseline and Full Specifications from Each Dataset: NC, NLS72, NELS88, and ELS2002

	NC		NELS gr8		ELS	
Fraction of Variance	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.915	0.919	0.830	0.836	0.874	0.881
$Var(Y_{is} - Y_s)$	(0.014)	(0.014)	(0.026)	(0.022)	(0.019)	(0.019)
Observable Student-Level (Within):	0.124	0.213	0.162	0.292	0.134	0.221
$Var((X_{si} - X_s)B)$	(0.005)	(0.006)	(0.022)	(0.018)	(0.027)	(0.036)
Unobservable Student-Level (Within)	0.791	0.706	0.668	0.543	0.740	0.660
$Var(V_{si})$	(0.013)	(0.012)	(0.031)	(0.023)	(0.028)	(0.032)
Between School:						
Total	0.085	0.081	0.170	0.164	0.126	0.119
$Var(Y_s)$	(0.014)	(0.014)	(0.026)	(0.022)	(0.019)	(0.019)
Observable Student-Level:	0.018	0.033	0.073	0.109	0.037	0.060
$Var(X_s B)$	(0.002)	(0.003)	(0.015)	(0.016)	(0.005)	(0.008)
Student-Level/						
School-Level Covariance	0.016	0.010	0.025	0.007	0.025	0.006
$2 * Cov(X_s B, Z1_s G_1 + Z2_s G_2)$	(0.003)	(0.005)	(0.012)	(0.025)	(0.009)	(0.011)
School-Avg. Student-Level/						
School Char. Covariance	-0.017	-0.008	0.007	0.004	0.001	-0.002
$2 * Cov(Z1_s G_1, Z2_s G_2)$	(0.006)	(0.004)	(0.011)	(0.010)	(0.014)	(0.012)
School-Avg. Student-Level	0.018	0.009	0.037	0.029	0.028	0.029
$Var(Z_{1s} G_1)$	(0.004)	(0.003)	(0.013)	(0.014)	(0.015)	(0.012)
School Char.	0.018	0.012	0.011	0.006	0.025	0.024
$Var(Z_{2s} G_2)$	(0.006)	(0.003)	(0.019)	(0.015)	(0.012)	(0.011)
Unobservable School-Level	0.031	0.026	0.017	0.010	0.010	0.001
$Var(v_s)$	(0.007)	(0.005)	(0.010)	(0.008)	(0.011)	(0.010)

Table 14: Decomposition of Variance in Latent Index Determining Enrollment in a Four-Year College from Each Dataset: NC, NLS72, NELS88, and ELS2002 (Baseline and Full Specifications)

	NLS		NELS gr8		ELS	
Fraction of Variance	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.857	0.857	0.781	0.774	0.785	0.791
$Var(Y_{is} - Y_s)$	(0.020)	(0.011)	(0.016)	(0.017)	(0.014)	(0.014)
Observable Student-Level (Within):	0.176	0.354	0.203	0.316	0.184	0.330
$Var((X_{si} - X_s)B)$	(0.091)	(0.017)	(0.012)	(0.014)	(0.015)	(0.013)
Unobservable Student-Level (Within)	0.681	0.503	0.578	0.458	0.600	0.461
$Var(V_{si})$	(0.075)	(0.016)	(0.014)	(0.011)	(0.015)	(0.012)
Between School:						
Total	0.143	0.143	0.219	0.226	0.215	0.209
$Var(Y_s)$	(0.020)	(0.011)	(0.016)	(0.017)	(0.014)	(0.014)
Observable Student-Level:	0.042	0.062	0.093	0.143	0.079	0.127
$Var(X_s B)$	(0.006)	(0.006)	(0.010)	(0.012)	(0.007)	(0.009)
Student-Level/ School-Level Covariance	0.037	0.032	0.056	0.027	0.071	0.039
$2 * Cov(X_s B, Z1_s G_1 + Z2_s G_2)$	(0.008)	(0.008)	(0.012)	(0.014)	(0.008)	(0.010)
School-Avg. Student-Level/ School Char. Covariance	0.000	-0.002	-0.001	0.005	-.003	-0.002
$2 * Cov(Z1_s G_1, Z2_s G_2)$	(0.005)	(0.004)	(0.008)	(0.006)	(0.007)	(0.006)
School-Avg. Student-Level	0.026	0.020	0.029	0.021	0.022	0.015
$Var(Z_{1s} G_1)$	(0.006)	(0.004)	(0.006)	(0.005)	(0.006)	(0.004)
School Char.	0.026	0.019	0.022	0.015	0.024	0.018
$Var(Z_{2s} G_2)$	(0.006)	(0.004)	(0.007)	(0.006)	(0.006)	(0.005)
Unobservable School-Level	0.012	0.013	0.019	0.014	0.022	0.013
$Var(v_s)$	(0.008)	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)

Table 15: Decomposition of Variance in Years of Post-Secondary Education and Adult Log Wages using NLS72 (Baseline and Full Specifications)

	Yrs. Postsec. Ed.		Perm. Wages No Post-sec Ed.		Perm. Wages w/ Post-sec Ed.	
Fraction of Variance	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.904	0.904	0.837	0.834	0.829	0.829
$Var(Y_{is} - Y_s)$	(0.008)	(0.011)	(0.016)	(0.017)	(0.014)	(0.014)
Observable Student-Level (Within):	0.154	0.280	0.140	0.174	0.212	0.224
$Var((X_{si} - X_s)B)$	(0.006)	(0.017)	(0.012)	(0.014)	(0.015)	(0.013)
Unobservable Student-Level (Within)	0.749	0.624	0.697	0.660	0.617	0.605
$Var(V_{si})$	(0.008)	(0.016)	(0.014)	(0.011)	(0.015)	(0.012)
Between School:						
Total	0.096	0.096	0.163	0.166	0.171	0.171
$Var(Y_s)$	(0.008)	(0.011)	(0.016)	(0.017)	(0.014)	(0.014)
Observable Student-Level:	0.041	0.058	0.045	0.055	0.061	0.065
$Var(X_s B)$	(0.004)	(0.006)	(0.010)	(0.012)	(0.007)	(0.009)
Student-Level/ School-Level Covariance	0.031	0.023	0.033	0.028	0.033	0.029
$2 * Cov(X_s B, Z1_s G_1 + Z2_s G_2)$	(0.005)	(0.008)	(0.012)	(0.014)	(0.008)	(0.010)
School-Avg. Student-Level/ School Char. Covariance	0.024	0.016	-0.002	0.001	-.003	0.000
$2 * Cov(Z1_s G_1, Z2_s G_2)$	(0.005)	(0.004)	(0.008)	(0.006)	(0.007)	(0.006)
School-Avg. Student-Level	0.012	0.008	0.033	0.029	0.029	0.028
$Var(Z_{1s} G_1)$	(0.006)	(0.004)	(0.006)	(0.005)	(0.006)	(0.004)
School Char.	0.017	0.010	0.039	0.041	0.039	0.040
$Var(Z_{2s} G_2)$	(0.006)	(0.004)	(0.007)	(0.006)	(0.006)	(0.005)
Unobservable School-Level	0.005	0.004	0.014	0.011	0.011	0.009
$Var(v_s)$	(0.008)	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)