12-16-2014

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

Joseph G. Altonji
Yale University

Richard K. Mansfield
Cornell University, rm743@cornell.edu

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

Abstract

We consider the classic problem of estimating group treatment effects when individuals sort based on observed and unobserved characteristics. Using a standard choice model, we show that controlling for group averages of observed individual characteristics potentially absorbs *all* the across-group variation in *unobservable* individual characteristics. We use this insight to bound the treatment effect variance of school systems and associated neighborhoods for various outcomes. Across four datasets, our conservative estimates indicate that a 90th versus 10th percentile school system increases high school graduation and college enrollment probabilities by at least 0.047 and 0.11. Other applications include measurement of teacher value-added.

Keywords

observables, unobservables, treatment effects

Disciplines

Economic Theory | Education Economics | Labor Relations | Other Economics

Comments

Suggested Citation

Altonji, J. G., & Mansfield, R. K. (2014). *Group-average observables as controls for sorting on unobservables when estimating group treatment effects: The case of school and neighborhood effects* [Electronic version]. Retrieved [insert date], from Cornell University, ILR School site: http://digitalcommons.ilr.cornell.edu/workingpapers/178/

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

Joseph G. Altonji
Yale University and NBER

Richard K. Mansfield
Cornell University

December 16, 2014

Abstract

We consider the classic problem of estimating group treatment effects when individuals sort based on observed and unobserved characteristics. Using a standard choice model, we show that controlling for group averages of observed individual characteristics potentially absorbs *all* the across-group variation in *unobservable* individual characteristics. We use this insight to bound the treatment effect variance of school systems and associated neighborhoods for various outcomes. Across four datasets, our conservative estimates indicate that a 90th versus 10th percentile school system increases high school graduation and college enrollment probabilities by at least 0.047 and 0.11. Other applications include measurement of teacher value-added.

Altonji: Department of Economics, Yale University, PO Box 208264, New Haven CT 06520-8264. joseph.altonji@yale.edu. Mansfield: Cornell University, ILR School Department of Economics 266 Ives Faculty Building Ithaca, NY 14853-3901. rkmansfield@gmail.com. We thank Steven Berry, Greg Duncan, Phil Haile, Hidehiko Ichimura, Amanda Kowalski, Costas Meghir, Richard Murnane, Douglas Staiger, Jonathan Skinner, Shintaro Yamiguchi and as well as seminar participants at Cornell University, Dartmouth College, Duke University, the Federal Reserve Bank of Cleveland, McMaster University, the NBER Economics of Education Conference, Yale University, Paris School of Economics, University of Colorado-Denver, and the University of Western Ontario for helpful comments and discussions. 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. We would also thank the Russell Sage Foundation and theYale Economics Growth Center for financial support. A portion of this research was conducted while Altonji was a visitor at the LEAP Center and the Department of Economics, Harvard University.

1 Introduction

Society is replete with contexts in which (1) a person's outcome depends on both individual and group-level inputs, and (2) the group is endogenously chosen either by the individuals themselves or by administrators, partly based on the individual's own inputs. Examples include health outcomes and hospitals, earnings and workplace characteristics, and test scores and teacher value-added.[1] Generations of social scientists have been interested in determining whether group outcomes differ because the groups 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, assessment of the overall distribution of group treatments is much more difficult, and researchers and governments frequently rely on non-experimental estimators of group treatment effects (e.g. school report cards and teacher value-added).

In this paper we show that in certain circumstances the tactic of controlling for group averages of observed individual-level characteristics, generally thought to control for "sorting on observables" only, will absorb *all* of the between-group variation in both observable and *unobservable* individual inputs. We then show how this insight can be used to estimate a lower bound on the variance in the contributions of group-level treatments to individual outcomes. We also examine the conditions under which causal effects of particular observed group characteristics can be estimated.

We apply our methodological insight and demonstrate its empirical value by addressing 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] To illustrate the sorting problem consider the following simplified production function relating education outcomes to individuals' characteristics and the inputs of the schools/neighborhoods they choose. Let Y_{si} denote the outcome (e.g. attendance at a four-year college) of student i who attends and lives near school s.[3] Y_{si} is determined according to[4]

$$Y_{s,i} = [X_i\beta + X_i^U\beta^U] + [Z_s\Gamma + Z_s^U\Gamma^U] \; . \tag{1}$$

Let the vectors X_i and X_i^U be the *complete* set of child and family characteristics that have a causal impact on student i's educational attainment. X_i is observed by the econometrician and X_i^U is

[1] Ash et al. (2012) provide an overview of the issues involved in assessing hospitals. Doyle Jr et al. (2012) also discuss the issues and provide a short literature survey. They are among a small set of studies that use a quasi-experimental design to assess effects of particular hospital characteristics on outcomes. See Chetty et al. (2014) and Rothstein (2014) for discussions and references related to the estimation of teacher value-added.

[2] See Duncan and Murnane (2011) for recent papers on school and neighborhood effects, with references to the literature. Meghir and Rivkin (2010) discuss alternative approaches to estimating school fixed effects and the effects of particular school inputs, and highlight the problem of endogenous selection of schools and neighborhoods, among other econometric issues.

[3] 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.

[4] Later we will introduce additional components to the outcome model.

unobserved. Analogously, the row vectors Z_s (observed) and Z_s^U (unobserved) capture the complete set of school and neighborhood level influences common to students live in s, so that the school/neighborhood treatment effect is given by $[Z_s\Gamma + Z_s^U\Gamma^U]$.

Unfortunately, sorting will lead the school average of X_i^U, denoted X_s^U, to vary across s. This will contaminate estimates of Γ and fixed effect estimates of the school treatment effect $Z_s\Gamma + Z_s^U\Gamma^U$. While various studies have included controls for group-level averages of individual observables (denoted X_s), the role played by such controls in mitigating sorting bias has generally been underappreciated.

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. This means that school-averages of individual characteristics such as parental education, family income, and athletic ability will be functions of the vector of amenity factors (denoted A_s) that parents consider when making their school choices. Thus, the school averages X_s and X_s^U will be different vector-valued functions of the same common set of amenities: $X_s = f(A_s)$ and $X_s^U = f^U(A_s)$. The functions f and f^U, are determined by the sorting equilibrium and reflect the equilibrium prices of the amenities. If the dimension of the underlying amenity space is smaller than the number of observed characteristics, then under certain conditions one can invert this vector-valued function to express the amenities in terms of school-averages of observed characteristics: $A_s = f^{-1}(X_s)$. But this implies that the vector of school averages of unobserved characteristics can also be written as a function of observed characteristics: $X_s^U = f^U(f^{-1}(X_s))$. This function of X_s can serve as a control function for X_s^U when estimating group effects.

We formalize this intuition by introducing a multidimensional spatial equilibrium model of school choice and providing conditions under which this mapping is exact. Given an additively separable specification of utility, we show that the average unobserved student variables X_s^U are a *linear* function of X_s. As we make precise in Proposition 1 below, X and X^U need not affect preferences for all of the amenities A. Partition X_s^U into a subset X_{1s}^U that is correlated X_s and a subset X_{2s}^U that is not correlated with X_s. Roughly speaking, the keys are that (1) X_s and/or X_{1s}^U affects preferences for all amenities that any elements of X_{2s}^U shifts preferences for, and (2) that there are enough elements of X to span this amenity space.

To take a simple example, suppose that school/neighborhood combinations differ in only one dimension that people observe and systematically care about—perceived school quality—plus a random idiosyncratic component specific to each family/location combination.[5] Suppose further that two uncorrelated characteristics, parental education (observed) and student athleticism (unobserved), both increase families' willingness-to-pay for school quality, and that both affect the outcome Y_{it} (e.g. graduation from high school). In equilibrium the expected values of both par-

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

ent's education and student athleticism will be increasing in perceived school quality, so that the neighborhood average of parents' education will be a perfect proxy for the neighborhood average of student athleticism. Now suppose that the quality of athletic facilities also varies across neighborhoods and that student athleticism influences willingness to pay for better athletic facilities but parental education does not. Then variation in the quality of athletic facilities would lead to between-neighborhood variation in average athleticism that average parental education could not predict. In this case we would need to control for the neighborhood average of another observable characteristic (e.g. parental income) that either directly affects willingness to pay for athletic facility quality or is correlated with student athleticism.

However, while this control function approach potentially solves the sorting-on-unobservables problem, the group averages X_s control for too much. They will absorb peer effects that depend on X_s and X_s^U. They will also absorb a part of the unobserved school/neighborhood quality component that is both orthogonal to the observed school characteristics and is correlated with the amenities that families consider when choosing where to live. As a result, our estimator will provide a lower bound on the variance of the overall contribution of schools/neighborhoods to student outcomes.

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) and administrative data from North Carolina.

For each dataset, we provide lower bound estimates of the overall contribution of differences between school systems and associated neighborhoods to the variance in student outcomes: high school graduation, enrollment in a four-year college, and adult wages (NLS72 only). 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 associated 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 estimates are 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 instead of a 10th quantile school and surrounding community 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 production 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 a 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 in the estimation of group treatment effects.[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 outcomes. Oreopoulos (2003) and Jacob (2004) use quasi-random assignment of neighborhood in the wake of housing project closings to estimate the magnitude of neighborhood effects on stu-

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

[7]See, for example, Metcalf (1974) or Altonji (1982)

[8]See Cunha et al. (2005).

[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]In the main text, we derive our key result using only the first order conditions of the consumer's choice problem. However, in the Appendix we present a closed form solution to the model for the special case in which amenities are exogenous and in particular do not depend on group-averages of individual 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), and Bayer and Timmins (2005). 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. Browning et al. (2014) survey the literature on sorting in marriage markets. A number of recent papers analyse labor market sorting based on firm and worker quality (e.g, Lise et al. (2013)). In parallel work Lindenlaub (2013) presents a closed form solution to the equilibrium of a labor market in which jobs differ in the skill vectors they require and workers vary in the skill vectors they supply.

[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) and Altonji et al. (2013). These papers examine 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 (2011).

dent 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.[14]

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 find large effects. Specifically, for students from low quality urban schools, the treatment effects from winning the lottery are large enough to close 75 percent of the black-white gap in graduation and 25 percent of the gap in bachelor's degree completion. On the other hand, Cullen et al. (2006) use a similar identification strategy with lotteries in Chicago Public Schools and find little effect on the high school graduation probability.

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 our model of school choice, while Section 3 formally derives our key control function result. Section 4 describes and presents results from a monte carlo analysis of the finite sample properties of our control function approach. Section 5 presents a simple production function for long-run student outcomes. Section 6 describes our empirical methodology for placing lower bounds on school and neighborhood contributions to long run student outcomes. Section 7 describes the four datasets we use to estimate the model of outcomes. Section 8 presents our results. Section 9 discusses other applications of our methodology. We give special emphasis to the problem of assessing teacher value added. Section 10 closes the paper with a brief summary of our empirical results and a discussion of potential theoretical extensions.

2 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.

Assume that each location $s \in \{1,...,S\}$ can be characterized by a vector of K underlying latent amenities $A_s \equiv [A_{1s},\ldots,A_{Ks}]'$.[15] 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

[14] Aliprantis (2011) also stresses the limitations of the MTO study for uncovering the full distribution of school system and neighborhood effects.

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

linear form:

$$U_i(s) = (\sum_{k=1}^{K} \lambda_{ki} A_{ks}) - P_s + \varepsilon_{s,i} \equiv \lambda_i A_s - P_s + \varepsilon_{s,i}. \tag{2}$$

$\lambda_i \equiv [\lambda_{1i}, \ldots, \lambda_{Ki}]$ 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 K amenity factors $A_{1s}, \ldots, A_{Ks}$, respectively. P_s is the price of living in the neighborhood surrounding school s, and $\varepsilon_{s,i}$ is an idiosyncratic taste of the parent/student i for the particular location s.

Consider projecting the willingness to pay (hereafter denoted WTP) for particular amenities across parent/student combinations onto these families' observable (X_i) and unobservable (X_i^U) characteristics. In particular, suppose that X_i has L^O elements, while X_i^U has L^U elements. Then we obtain:

$$\lambda_{ki} \equiv \sum_{\ell=1}^{L^O} X_{i\ell} \Theta_{\ell k} + \sum_{\ell=1}^{L^U} X_{i\ell}^U \Theta_{\ell k}^U + \kappa_{ki} \quad \forall\, k \in \{1, \ldots, K\}$$

By converting equation (**??**) into matrix notation and substituting it into equation (2), we obtain:

$$U_i(s) = (X_i \Theta + X_i^U \Theta^U + \kappa_i) A_s + \varepsilon_{s,i} - P_s \tag{3}$$

Θ (Θ^U) is an $L^O \times K$ ($L^U \times K$) matrix whose ℓk-th entry $\Theta_{\ell k}$ ($\Theta_{\ell k}^U$) 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 $1 \times K$ 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 attributes that determine $Y_{s,i}$, the elements of κ_i influence school choice but have no direct effect on student outcomes. Note that in the absence of restrictions on the elements of Θ and Θ^U, this formulation of utility allows for a fully general pattern of relationships between different student characteristics (observable or unobservable) and tastes for different school/neighborhood amenities (given the additive separability assumed in equation (2)).

Expected utility is taken with respect to the information available when s is chosen. The information set includes the price and the amenity vector in each school/neighborhood as well as student/parent characteristics $[X_i, X_i^U]$. The information set excludes any local shocks that are determined after the start of secondary school. It also excludes 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. Some of the amenities may include aspects of the demographic composition of the school/neighborhood and thus are outcomes of the sorting equilibrium.

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

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

3 The Link Between Group Observables and Group Unobservables

Next we characterize the equilibrium allocation of students to schools from the above choice model. For analytic simplicity, in Section 3.1 we first consider a version of the school choice model in which (a) we ignore the idiosyncratic school-family taste match by setting $\varepsilon_{si} = 0 \ \forall \ (s,i)$, 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. We revisit the finite choice case in Section 3.1.2 and Section 4.

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

Under assumptions (a) and (b), choosing a school is equivalent to choosing the vector of amenities that maximizes utility, given the price function $P_s = P(A_s)$. In addition, we assume that parents behave competitively in the sense that prices are taken as given and choice is unrestricted. We also assume that the equilibrium price function $P(A)$ is increasing and convex, so that prices rise at an increasing rate as amenities increase.[16]

The optimal choice is characterized by a system of K first order conditions, one for each amenity factor. The conditions are

$$\lambda_i' \equiv \Theta' X_i' + \Theta^{U'} X_i^{U\prime} + \kappa_i' = \nabla P(A_{s(i)}) \ , \tag{4}$$

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 $A = A_{s(i)}$. The conditions say that the family chooses a community with a level of A such that the family's willingness to pay for additional units of A is equal to their marginal cost. Since $P(A)$ is strictly convex, second order conditions will be satisfied.

We now use (4) to study the relationship between X_s^U and X_s. First, decompose X_i^U into its projection on X_i and the orthogonal component $\tilde{X}_i^U$:[17]

$$X_i^U = X_i \Pi_{X^U X} + \tilde{X}_i^U \tag{5}$$

Use the decomposition (5) to rewrite $\lambda_i = X_i\Theta + X_i^U\Theta^U + \kappa_i$ as $\lambda_i = X_i\tilde{\Theta} + \tilde{X}_i^U\Theta^U + \kappa_i$, where $\tilde{\Theta} = [\Theta + \Pi_{X^U X}\Theta^U]$. In its rewritten form, all three components of λ_i are mutually orthogonal. We are now prepared to present the main proposition of the paper:

Proposition 1: *Assume (i) preferences are given by (3), (ii) $\varepsilon_{si} = 0 \ \forall \ (s,i)$, (iii) the price function $P(A)$ is increasing and strictly convex in A, (iv) the rows of the coefficient matrix Θ^U relating tastes*

[16] In Appendix A3 we explicitly solve for $P(A_s^*)$ under stronger assumptions and show that it is increasing and strictly convex.

[17] 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 $\Pi_{X^U X}$, $D = X_i^U$ and $Q = X_i$.

for A to X^U *are spanned by the rows of the coefficient matrix* $\tilde{\Theta}$ *and (v)* $E(X_i|\lambda_i)$ and $E(X_i^U|\lambda_i)$ are linear in λ_i. *Then the expectation* X_s^U *is linearly dependent on the expectation* X_s. *Specifically*,

$$X_s^U = X_s[\Pi_{X^U X} + Var(X_i)^{-1} R' Var(X_i^U)] \quad forsome\ L^U \times L^O\ matrix\ R \tag{6}$$

3.1.1 Proof of Proposition 1:

Consider the projection of X_i and $\tilde{X}_i^U$ onto λ_i. Note that since (1) $\tilde{X}_i^U$ is uncorrelated with X_i by construction, and (2) κ_i is uncorrelated with both X_i and $\tilde{X}_i^U$,

$$Cov(\lambda_i, X_i) = \tilde{\Theta}' Var(X_i) \tag{7}$$

$$Cov(\lambda_i, \tilde{X}_i^U) = \Theta^{U'} Var(\tilde{X}_i^U). \tag{8}$$

Thus, these projections can be written as

$$X_i = \lambda_i Var(\lambda_i)^{-1} \tilde{\Theta}' Var(X_i) + error_i^X \tag{9}$$

$$\tilde{X}_i^U = \lambda_i Var(\lambda_i)^{-1} \Theta^{U'} Var(\tilde{X}_i^U) + error_i^{\widetilde{X^U}}\ . \tag{10}$$

where $E(error_i^X|\lambda_i) = E(error_i^{\widetilde{X^U}}|\lambda_i) = 0$ from basic regression theory and the linear in expectations assumption (v). Using the first order conditions (4), we can rewrite these equations as:

$$X_i = \nabla P(A_{s(i)})' Var(\lambda_i)^{-1} \tilde{\Theta}' Var(X_i) + error_i^X$$

$$\tilde{X}_i^U = \nabla P(A_{s(i)})' Var(\lambda_i)^{-1} \Theta^{U'} Var(\tilde{X}_i^U) + error_i^{\widetilde{X^U}}\ .$$

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 λ_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:

$$X_{s(i)} \equiv E(X_i|s(i)) = \nabla P(A_{s(i)})' Var(\lambda_i)^{-1} \tilde{\Theta}' Var(X_i) \tag{11}$$

$$\tilde{X}_{s(i)}^U \equiv E(\tilde{X}_i^U|s(i)) = \nabla P(A_{s(i)})' Var(\lambda_i)^{-1} \Theta^{U'} Var(\tilde{X}_i^U). \tag{12}$$

By the spanning assumption (iv),

$$\Theta^U = R\tilde{\Theta} \tag{13}$$

for some $L^U \times L^O$ matrix R. Substituting the condition (13) for Θ^U in (12) implies that

$$\begin{aligned}
\tilde{X}_{s(i)}^U &= E(\tilde{X}_i^U|s_i) = \nabla P(A_{S_i})' Var(\lambda_i)^{-1} \tilde{\Theta}' R' Var(X_i^U) \\
&= \nabla P(A_{S_i})' Var(\lambda_i)^{-1} \tilde{\Theta}' Var(X_i) Var(X_i)^{-1} R' Var(X_i^U) \\
&= X_{s(i)} Var(X_i)^{-1} R' Var(X_i^U) \qquad (14)
\end{aligned}$$

where the third line follows from substitution using (11). Combining (14) with equation (5), we have

$$X^U_{s(i)} = X_{s(i)}[\Pi_{X^U X} + Var(X_i)^{-1} R' Var(X^U_i)] \; . \tag{15}$$

This completes the proof.

3.1.2 Interpreting Proposition 1: When Will the Spanning Assumption Hold?

Proposition 1 lays out the conditions under which X^U_s, the between school component of the vector of student-level unobservables, will be an exact linear function of its observable counterpart X_s. Remarkably, the dependence between the school averages X^U_s and X_s exists even when the vector X^U_i is uncorrelated with the vector X_i at the student level.[18]

The key restriction on preferences in Proposition 1 is the spanning condition (iv), which requires that $\Theta^U = R\tilde{\Theta}$ for some $L^U \times L^O$ matrix R. In other words, it requires that the coefficient vectors relating tastes for amenities to the elements of X^U_i are linear combinations of the coefficient vectors relating tastes for amenities to the observables X_i and/or elements of X^U_i that are correlated with X_i. Given the importance and subtlety of this spanning condition, however, this subsection further develops the intuition underlying the condition and highlights cases in which it fails to hold.

Reconsider the more general function formulation used in the introduction. Let $A^X \subseteq A$ represent the subset of amenities that affect the distribution of observable school averages X_s. An amenity will be included in A^X if WTP for the amenity is affected by either X_i or elements of X^U_i correlated with X_i. Likewise, $A^{X^U} \subseteq A$ represents the subset of amenities that affect the distribution of unobservable school averages X^U_s. The between-school variation in X_s will only be driven by A^X, so that $X_s = f(A^X)$ for some vector-valued function f. Similarly, $X^U_s = f^U(A^{X^U})$. In order to be able to write $X^U_s = f^U(f^{-1}(X_s)) \equiv g(X_s)$, two conditions must be satisfied: (1) f must be invertible, so that we can write $A^X = f^{-1}(X_s)$, and (2) $A^{X^U} \subseteq A^X$, so that the amenity space that X_s spans is the relevant amenity space that drives the variation in X^U_s (i.e. the range of f^{-1} must encompass the domain of f^U).

This intuition suggests that there are two fundamental ways the spanning condition $\Theta^U = R\tilde{\Theta}$ can fail. First, X_i may affect tastes for more amenities than its own number of elements: $|A^X| > L^O$. In this case, the function $f(*)$ is not invertible.[19] In the case of the additively separable utility function from (3), $|A^X|$ is captured by the row rank of $\tilde{\Theta}$. In the context of the simple example from the introduction, this condition might fail if the only observable characteristic were parental

[18] Note that if unobservable characteristics do not affect location preferences (i.e. students do not sort based on unobservables), so that $\Theta^U = 0$, then $R = 0$. By substituting $\Theta^U = 0$ into (12), we see that in this special case $X^U_s = 0$, so that there is no across-school variation in average unobservable characteristics. In this case $Var(Z_s G)$ will accurately reflect the school/neighborhood contribution.

[19] More specifically, what is relevant for invertibility is not the number of elements of X_i (denoted L^O) per se but the number of independent taste factors that these L^O observables represent. Suppose for example, that mother's education and father's education were both observed, but affected willingness to pay for each amenity in the same relative proportions. Then adding father's education to X_i would not make $f(*)$ invertible if it were not already when only mother's education was included in X_i.

income, and the amenity space consisted of two imperfectly correlated factors: schools' quality of teachers and quality of athletic facilities (which are not perfectly correlated with one another). Even if parental income affected WTP for both amenities, one would not be able to disentangle the quality of athletic facilities from the quality of teachers based on only neighborhood averages of parental income. We would need to observe a second individual characteristic, such as parental education, in order to satisfy the spanning condition.

Second, the spanning condition can also fail if the unobservable vector X_i^U affects WTP for certain amenities that no element in X_i predicts WTP for, so that $A^{X^U} \not\subseteq A^X$. In the case of the additively separable utility function from equation (3), $A^{X^U} \subseteq A^X$ if and only if the row space of Θ^U is a linear subspace of the row space of $\tilde{\Theta}$. Note, though, that a given element of X_i, say X_{il}, can help predict WTP for a particular amenity A_k either directly by affecting taste for the amenity (so that $\Theta_{lk} \neq 0$), or indirectly by merely being correlated with an element of X_i^U that predicts taste for the amenity (so that the (l,k)-th element of $\Pi_{X^U X}\Theta^U \neq 0$). Either will yield a non-zero value of $\tilde{\Theta}_{lk}$.

In the context of our simple example from the introduction, failure of this second requirement would occur if the only observable were parental education, and parental education only directly affected WTP for average teacher quality, and was uncorrelated with the unobservables (e.g. student athleticism) that affected WTP for athletic facility quality. In this case the quality of athletic facilities in the neighborhood would be an element of A^{X^U} but not A^X, so that $A^{X^U} \not\subseteq A^X$. Consequently, variation in athletic facility quality would drive between-neighborhood variation in average student athleticism that average parental education would not capture. Appendix Section A1 goes through further examples that illustrate when the spanning condition will and will not be satisfied.

Are the conditions underlying the spanning assumption testable, and are they plausible? The plausibility of condition (1) depends on the number and breadth of coverage of variables in X_i. Condition 1 is testable. The model implies a factor structure for the vector X_s, where the number of factors is determined by the row rank of $\tilde{\Theta}$. A finding that the number of factors that determine X_s is smaller than the dimension of X_i is consistent with the assumption that $|A^X| \leq L_O$. A finding that the number of factors is at least as large as the dimension of X is also technically consistent with the assumption, but would strongly suggest that $|A^X| > L^O$.[20]

We investigate the factor structure of X_s in Section A2 of the Online Appendix. We find that for each of our three survey datasets the estimated covariance matrix of X_s is rank deficient. This means that each element of X_s can be written as a linear combination of a smaller number of latent factors (generally between 25 and 30 factors, depending on the specification and dataset). Since the rank of $Cov(X_s)$ should reflect the number of amenity factors $|A^X|$, this validates our assumption that $|A^X| \leq L_O$. Indeed, we further show that in each dataset an even smaller number of latent factors (generally around 10) can explain 90% of the variance in the expected values X_s, suggesting that the variation in student composition across schools is driven primarily by a handful of amenity factors.

Condition (2) is a statement about unobservables and is not testable without more structure than

[20]It is important to point out that Proposition 1 refers to the expected values $E[X_i|s(i) = s]$ and $E[X_i^U|s(i)] = s$. The observed values of X_s and X_s^u will deviate from the expectations when neighborhood/school sizes are not extremely large.

we impose. However, we think that it is plausible in situations in which X contains a rich set of variables that are likely to matter for student outcomes. For example, we use parental income but not parental wealth. Both are likely to directly affect the education outcomes such as college attendance, so sorting on wealth might be expected to lead to an overstatement of school effects. However, the two variables are strongly positively correlated across families, and are both are likely to influence WTP for a similar set of amenities. For both reasons the school average of parental income is likely to serve as an effective control for the school average of parental wealth.

Note that the derivation of (6) does not require one to solve the model. If some of the neighborhood amenities are functions of resident characteristics, the distribution of amenities will be endogenous, and may result in multiple equilibria. However, the derivation is based only on a set of necessary first order conditions, so the linear dependence between X_s and X_s^U will hold in any equilibrium of the model. In Appendix Section A3 we derive an analytical solution for the equilibrium mapping from student/family characteristics $[X_i, X_i^U, \kappa_i]$ to the school/neighborhood amenity vectors they choose under somewhat stronger assumptions. The most important one is that A_s does not depend on which families choose s, so that the joint distribution of amenities across neighborhoods is exogenous. This is a restrictive assumption in the school choice setting, but may be plausible in other settings (e.g. differentiated product choice). Since analytical solutions to multidimensional sorting problems are quite rare, the solution we provide is likely to be useful in other applications.[21]

Finally, recall that 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 $\varepsilon_{si} = 0 \; \forall \; (s,i)$. Note, though, that in the additively separable specification for utility in (3), 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.[22] Nonetheless, in the next section we describe results from a series of monte carlo simulations designed to address this issue along with a number of other issues related to the finite-sample performance of our control function approach.

4 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

[21] Much of the literature has either relied on models featuring a single, vertical dimension of quality or has resorted to numerical solutions. An analytical solution to the multidimensional sorting problem allows researchers to make clearer predictions about how the distribution of individuals across groups will change in response to changes in the distribution of group amenities. Such a solution may also aid structural estimation of the joint distribution of amenities and of preferences, facilitating full-scale welfare analysis of proposed policy changes.

[22] We conjecture that under a spanning condition analogous to (13) the quantiles of the conditional distribution of unobservables $f(X_i^U|A_{s(i)}, s(i) = s)$ are a function of the quantiles of the conditional distribution of observables $f(X_i|A_{s(i)}, s(i) = s)$ across s. The intuition is clearest in the case when X_i, X_i^u and A_s are all scalars. If $\Theta > 0$ and $\Theta^U > 0$, 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.

dimensional model of school choice. Furthermore, Proposition 1 characterizes the link between the expectations of X_i and X_i^U given A_s. With a finite number of students per school, random variation associated with κ_i and ε_{si} will cause school averages at point in time to deviate from their expectations. This could weaken the link between school averages of observable and unobservable characteristics.

To investigate these concerns, in this section we summarize the results of a series of monte carlo simulations that explore the properties of our control function approach across a number of key dimensions. A full description of our simulation methodology and results is contained in Appendix Section A4, while the results themselves are displayed in Appendix Tables A2 and A3.

The simulations are not intended to provide a full characterization of the finite sample properties of our estimator. Such a characterization is a daunting task given the large number of parameters that determine the full spatial equilibrium sorting of students to schools.[23] Instead, our simulations center around a stylized test case that is calibrated to represent a plausible description of the school/neighborhood choice context. These simulations serve 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 $\{\varepsilon_{si}\}$ 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.

All of our simulations consider combinations of model parameters which imply considerable sorting across schools on a vector of unobservable characteristics. Our metric for evaluating the effectiveness of our control function approach is the fraction of the between-group variance in the outcome contribution of unobservable individual-level characteristics ($Var(X_s^U\beta^U)$ in our production function below) that can be predicted using group-averages of observable characteristics. This is the R^2 from a regression of the potential bias from unobservable sorting, $X_s^U\beta^U$, on the vector X_s.

The first key result is that the control function can work 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. In many of our simulations in such settings over 97% of the variance in group-average values of unobservables $X_s^U\beta^U$ is absorbed by controlling for a sufficiently large vector of group-average observables X_s.

The second result is that the control function approach is quite robust to the violations of the spanning condition in which just a few outcome-relevant unobservables affect WTP for just a few additional amenities that are not weighted by any component of the observable vector X_i. These are arguably the most plausible cases when rich data on students and parents are available. Not

[23]The parameters include those characterizing the joint distribution of the individual characteristics affecting choice $[X_i, X_i^U, \kappa_i]$, the joint distribution of the amenities A_s, and the distribution of the idiosyncratic tastes ε_{si}. The parameters also include the Θ and Θ^U matrices that capture how observed and unobserved characteristics affect WTP.

surprisingly, the control function approach fails when the X_i and X_i^U are orthogonal to each other and X_i and X_i^U affect WTP for disjoint sets of amenities.

Finally, the third key result is that the effectiveness of our control function suffers when the group-averages of the observables X_s are constructed using small samples of group members rather than the full school population. Due to our reliance on such small samples in our three panel survey datasets in the empirical analysis below, we investigate this issue further for our particular school effects application in Section 7 and Appendix Section A7.

5 A Model of Educational Attainment and Wage Rates

5.1 The Determinants of Adult Outcomes

In this section we further develop and analyze the underlying econometric model of adult outcomes presented in the introduction. This production function 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.[24] Let $Y_{s(i)i}$ denote the outcome of student i whose family has chosen the school and surrounding neighborhood $s(i) \in \{1,\ldots,S\}$. For the rest of the section we will usually suppress the dependence of s on i unless necessary for clarity. In our application the outcomes will be comprised of 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

$$Y_{s(i),i} = X_i\beta + X_i^U\beta^U + Z_{s(i)}\Gamma + Z_{s(i),i}^U\Gamma^U + u_{s(i),i}\,. \tag{16}$$

As discussed above, the student's outcome contribution can be summarized by the index $(X_i\beta + X_i^U\beta^U)$, where 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 outcome; 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.

Analogously, the school/neighborhood outcome contribution is captured by the index $(Z_s\Gamma + Z_{s,i}^U\Gamma^U$, where 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), while $Z_{s,i}^U$ 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 or distinct course tracks at the school). Indeed, some elements of $Z_{s,i}^U$ may represent the

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

within-school components of Z_s, so that such elements will contain no between-school variation. The productivity parameters 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 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.

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 estimate the productivity parameters β or β^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 estimates of Γ stemming from correlations between Z_s and the omitted school/neighborhood factors $Z_{s,i}^U$. 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_{si}^U into between- and within-school components: $Z_{si}^U = Z_s^U + (Z_{si}^U - Z_s^U)$. 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 $(Z_{s^1}\Gamma + Z_{s^1}^U\Gamma^U) - (Z_{s^0}\Gamma + Z_{s^0}^U\Gamma^U)$. The outcomes of a specific student i will also differ across schools because the values of $(Z_{si}^U - Z_s^U)$ 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 $Z_s\Gamma + Z_s^U\Gamma^U$. Some of our estimates will also include differences due to u_s.

Note that while we allow the levels of school/neighborhood inputs Z_s or $Z_{s,i}^U$ to co-vary with individual attributes X_i and X_i^U, we rule out explicit structural interactions (such as products) between school and student characteristics. Allowing for such non-separability does not break the linear relationship between X_s and X_s^U, but would complicate the interpretation of the distribution of the school treatment effects. We discuss the issues involved in footnote 29 while describing our empirical methodology.

5.2 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. We show that Proposition 1 implies restrictions on these parameters that allow the recovery of a lower bound estimate of the contribution of schools (and groups more generally) to individual outcomes. We also present the more demanding conditions under which unbiased estimates of the causal effects of particular group-level characteristics can be recovered.

To facilitate the analysis, first partition Z_s into two subvectors $[X_s, Z_{2s}]$. X_s 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 coefficient vector $\Gamma \equiv [\Gamma_1, \Gamma_2]$ analogously.

Next, consider projecting the full vector of unobserved school inputs $Z_{s,i}^U \Gamma^U$ onto the vectors of both the observed student level (X_i) and school-level (Z_s) variables:

$$Z_{s,i}^U = \overbrace{(X_s \Pi_{Z_s^U, X_s} + Z_{2s} \Pi_{Z_s^U, Z_{2s}} + \widetilde{Z_s^U})}^{Z_s^U} + \overbrace{(X_i \Pi_{Z_{si}^U, X_i} + (\widetilde{Z_{s,i}^U - Z_s^U}))}^{(Z_{s,i}^U - Z_s^U)} . \tag{17}$$

Because we include in Z_s the school-averages of each student-level variable in X_i, standard partitioned regression results show that X_i will not predict any of the between-school components of $Z_{s,i}^U$ (denoted Z_s^U) that Z_s cannot predict. Thus, the matrix $\Pi_{Z_{si}^U, X_i}$ only captures the extent to which student-level observable characteristics predict the within-school variation in unobserved school characteristics. For example, parental education may predict the course track the student is assigned to within the school. Likewise, $Z_s \equiv [X_s, Z_{2s}]$, as a vector of school-level variables, cannot possibly predict the student-specific deviations from the vector of school averages $(Z_{s,i}^U - Z_s^U)$. Thus, the matrices $\Pi_{Z_s^U, X_s}$ and $\Pi_{Z_s^U, Z_{2s}}$ only capture 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_i^U onto the space of observable variables in two steps instead of one. In the first step, we regress X_i^U on the student-level observable vector X_i only, as in equation (5):

$$X_i^U = X_i \Pi_{X_i^U X_i} + \tilde{X}_i^U . \tag{18}$$

The matrix $\Pi_{X_i^U X_i}$ captures the extent to which the unobserved student-level contribution can be predicted by the observed student-level characteristics in the full population. It 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, $\tilde{X}_i^U$, onto both the student-level and school-level vectors of observables (X_i and Z_s):

$$\tilde{X}_i^U = X_i \Pi_{\tilde{X}_i^U X_i} + X_s \Pi_{\tilde{X}_i^U X_s} + Z_{2s} \Pi_{\tilde{X}_i^U Z_{2s}} + \varepsilon_{si}^{\tilde{X}} , \tag{19}$$

where $\varepsilon_{si}^{\tilde{X}}$ 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 matrices $\Pi_{\tilde{X_i^U} X_s}$ and $\Pi_{\tilde{X_i^U} Z_{2s}}$ need not equal 0. Furthermore, even though each component of the vector $\tilde{X}_i^U$ is uncorrelated with X_i (by construction from step 1), $\Pi_{\tilde{X_i^U} X_i}$ 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 projections (17), (18), and (19) for $Z_{s,i}^U$ and X_i^U in equation (16), we obtain:

$$Y_{s,i} = X_i B + X_s G_1 + Z_{2s} G_2 + v_s + (v_{si} - v_s), \quad where \tag{20}$$

$$B \equiv [\beta + \Pi_{Z_{si}^U X_i} \Gamma^U + (\Pi_{X_i^U X_i} + \Pi_{\tilde{X_i^U} X_i}) \beta^U] \tag{21}$$

$$G_1 \equiv [\Gamma_1 + \Pi_{Z_s^U X_s} \Gamma^U + \Pi_{\tilde{X_s^U} X_s} \beta^U] \tag{22}$$

$$G_2 \equiv [\Gamma_2 + \Pi_{Z_s^U Z_{2s}} \Gamma^U + \Pi_{\tilde{X_i^U} Z_{2s}} \beta^U] \tag{23}$$

$$v_s \equiv \widetilde{Z_s^U} \Gamma^U + \varepsilon_s^{\tilde{X}} \beta^U + u_s \tag{24}$$

$$v_{si} - v_s \equiv u_{si} + (\varepsilon_{si}^{\tilde{X}} - \varepsilon_s^{\tilde{X}}) \beta^U + (\widetilde{Z_{si}^U - Z_s^U}) \Gamma^U \tag{25}$$

The expressions for G_1, G_2 and v_s in (22), (23) and (24) reveal that the observable school components $X_s G_1$ and $Z_{2s} G_2$ and the unobservable residual component v_s all reflect a mixture of school effects and student composition biases. Specifically, the components $X_s G_1$ and $Z_{2s} G_2$ will reflect $X_s \Pi_{\tilde{X_i^u} X_s}$ and $Z_{2s} \Pi_{\tilde{X_i^u} Z_{2s}}$, respectively, which capture 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 $\varepsilon_s^{\tilde{X}} \beta^U$, which captures the part of the average unobservable student contribution that is not predictable based on observed school-level characteristics or average student-level characteristics. $X_s \Pi_{\tilde{X_i^u} X_s}$, $Z_{2s} \Pi_{\tilde{X_i^u} Z_{2s}}$ and $\varepsilon_s^{\tilde{X}} \beta^U$ are not school/neighborhood effects, since any child who was reallocated to a school with a higher value of these components could not expect an increase in test scores[25]. This analysis suggests that without further assumptions about how students sort into schools, basic regression and variance decomposition techniques cannot be used to identify or even bound the contribution of schools/neighborhoods to student outcomes. However, the next subsection shows that the assumptions laid out in Proposition 1 are sufficient to place a lower bound on the variance in school and neighborhood effects given the production function (16) presented at the beginning of the section.

5.3 Using Proposition 1 to Bound the Importance of School/Neighborhood Effects

Section 2 provides conditions under which the school-average values of student observables X_s and unobservables X_s^U are linearly dependent, as summarized in Proposition 1. To illustrate the value to identification of this result, substitute the linear mapping of X_s into $\tilde{X_s^U}$ from equation (14)

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

into the left hand side of the projection equation for $\tilde{X_s^U}$ given by (19). One can immediately see that $\Pi_{\tilde{X_i^U}X_s} = Var(X_i)^{-1}R^{'}Var(X_i^U)$, $\Pi_{\tilde{X_i^U}Z_{2s}} = 0$, and $\varepsilon_s^{\tilde{X}} = 0$. Thus,

$$G_2 \equiv \Gamma_2 + \Pi_{Z_s^U Z_{2s}}\Gamma^U \tag{26}$$

$$v_s \equiv \widetilde{Z_s^U}\Gamma^U + u_s \; . \tag{27}$$

We see that when the conditions of Proposition 1 are satisfied, the inclusion of X_s in Z_s purges both 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. However, the components $\hat{Var}(Z_{2s}G_2)$ and $\hat{Var}(v_s)$ only permit a lower bound estimate of the importance of school and neighborhood effects, for three reasons. The first and obvious one is that the causal effect of X_s on outcomes, $X_s\Gamma_1$, will be excluded from estimates of school/neighborhood effects. If peer effects are important, this could lead to a substantial underestimation of the importance of school/neighborhood effects. Second, if the school mean X_s^U has external effects, it is part of Z_s^U and therefore enters the outcome equation separately from the individual level variable $X_i^U\beta^U$. Since this component will be absorbed by $X_s\hat{G}_1$, school/neighborhood peer effects associated with X_s^U will be excluded from the estimate of school/neighborhood effects.

Finally, equation (22) reveals that X_s will also absorb part of the unobserved school contribution Z_s^U via $\Pi_{Z_s^U X_s}\Gamma^U$. To see why, note that X_s spans the space of X_s^U *because* the variation in both X_s and X_s^U is driven by the same underlying variation in the desired amenity vector, $\{A_{1s}, A_{2s}, ...A_{Ks}\}$. Re-examining equation (11) above, we see that if Θ is of full column rank and the vector-valued function $\nabla P(A_s)$ is invertible, then the system of equations can be inverted and the vector $A_s = [A_{1s}, A_{2s}, ...A_{Ks}]'$ can also be written as a linear combination of some K-length subvector of X_s:

$$A_s' = \nabla P^{-1}(X_s Var(X_i)^{-1}\tilde{\Theta}^{'-1}Var(\lambda_i)) \; . \tag{28}$$

Given that parents are likely to value the contributions of schools to student outcomes, many of the characteristics Z_s^U that affect school quality are likely to be reflected in $\{A_{1s}, A_{2s}, ...A_{Ks}\}$. Hence, while the inclusion of X_s in the estimated specification effectively removes sorting bias, it also absorbs some of the variation in the underlying amenity factors for which X_s affects taste. Furthermore, if some elements of the school-level observables Z_{2s} also serve directly as amenities in A_s, then these elements will be collinear with X_s, undermining our ability to estimate the vector G_2.

On the other hand, components of school/neighborhood quality $Z_{2s}\Gamma_2 + Z_s^U\Gamma^U$ that are either unvalued or not fully known (or knowable) 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, and will break the collinearity between X_s and Z_{2s}. 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, yet will produce between-school variation in outcomes.

5.3.1 Identification of Γ_2

The existence of $\Pi_{Z_s^U Z_{2s}} \Gamma^U$ in the expression for G_2 in (26) reveals that even when the conditions of Proposition 1 are satisfied, G_2 still reflects omitted variables bias driven by correlations between Z_{2s} and the unobserved school characteristics in Z^U. Thus, estimating the vector of causal effects Γ_2 associated with the school characteristics in Z_2 will in general still require a vector of instruments.

However, the sorting model in Section 2 also sheds light on the circumstances in which $\Pi_{Z_s^U Z_{2s}} = 0$, so that $\hat{G}_2$ represents an unbiased estimator of the vector of causal effects Γ_2. In particular, suppose that each element of Z_s^U is either an amenity considered by individuals at the time of choice or is perfectly predicted by the vector of amenities, so that $Z_s^U = \Pi_{Z_s^U A_s} A_s$, for some matrix $\Pi_{Z_s^U A_s}$. Furthermore, suppose the matrix $\tilde{\Theta}$ has full column rank and $\nabla P(A_s)$ is invertible, so that equation (28) holds. This implies that school averages of observed student characteristics X_s also perfectly determine Z_s^U. In this case, there will be no residual variation in Z_s^U for Z_{2s} to predict in equation (17), so that $\Pi_{Z_s^U Z_{2s}} = 0$, and $\hat{G}_2$ will be an unbiased estimator of Γ_2.[26]

Because we suspect that there are a large array of outcome-relevant school inputs, not all of which are directly and accurately valued by parents when choosing schools, we do not assume that $\Pi_{Z_s^U Z_{2s}} = 0$ in our empirical work. Thus, we do not attempt to interpret the individual coefficients estimated by $\hat{G}_2$.[27] However, this analysis does suggest that controlling for group-averages of individual characteristics can potentially remove part of the omitted variable bias from estimated coefficients on group-level characteristics. This is particularly true in contexts where those choosing groups are thought to consider and at least noisily observe most of the group-level characteristics expected to have substantial causal effects. We return to this point when considering the estimation of teacher value-added in Section 9.

6 Estimating the Contribution of Schools and Neighborhoods

6.1 Variance Decomposition

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

$$Y_i = X_i\beta + X_s G_1 + Z_{2s} G_2 + v_{si}, \tag{29}$$

where X_s 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:

$$Y_i = (X_i - X_s)\beta + X_s\beta + X_s G_1 + Z_{2s} G_2 + (v_{si} - v_s) + v_s \tag{30}$$

[26] $Var(Z_s^U) = 0$ will also lead to unbiased estimates of Γ_2, but requires the unrealistic assumption that all outcome-relevant school inputs are observed.

[27] See Meghir and Rivkin (2011) for a recent discussion of some of the issues in estimating the effects of particular school characteristics. A key source of bias they highlight is the vector of omitted school characteristics Z_s^U.

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

$$Var(Y_i) = Var(Y_i - Y_s) + Var(Y_s) \tag{31}$$

$$= [Var((X_i - X_s)\beta) + Var(v_{si} - v_s)] + \tag{32}$$

$$[Var(X_s\beta) + 2Cov(X_s\beta, X_sG_1) + 2Cov(X_s\beta, Z_{2s}G_2) + Var(X_sG_1) +$$

$$2Cov(X_sG_1, Z_{2s}G_2) + Var(Z_{2s}G_2) + Var(v_s)] \tag{33}$$

Motivated by the model of sorting presented in Section 2, we introduce two alternative lower bound estimates of the contribution of school/neighborhood choice to student outcomes.

The first is $Var(Z_{2s}G_2) + Var(v_s)$. Due to the presence of X_s, it will be purged of any effects of student sorting (observable or unobservable). Thus, it isolates only school/neighborhood factors. However, in addition to the unpredicted component of school/neighborhood contributions ($\widetilde{Z_s^U}\Gamma^U$), $Var(v_s)$ will include u_s, common location-specific shocks (such as local labor demand shocks) that occur after the chosen cohort has completed high school. One can argue that such shocks should not be attributed to schools because they are beyond the control of school or town administrators. Consequently, we also consider a second, more conservative lower bound estimate: $Var(Z_{2s}G_2)$. 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. $Var(Z_{2s}G_2)$ excludes true school quality variation that is orthogonal to observed characteristics, but also removes any truly idiosyncratic local shocks that occur after graduation.

Appendix Sections A5 and A6 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.

6.2 Interpreting Our Lower Bound Estimates

The static sorting model presented in Section 2 is silent about when 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 WTP 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 $Var(Z_{2s}G_2)$ or $Var(Z_{2s}G_2 + v_s)$ 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 X_i are unaffected by early schooling.[28] Then X_s will span the subspace of the school/neighborhood amenities A_s as well as X_s^U as they existed when the family made its choice. In this scenario, the residual variance contributions $Var(Z_{2s}G_2)$ or $Var(Z_{2s}G_2 + v_s)$ 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 our 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 of high school contributions is itself a (very conservative) lower bound estimate of the variance of contributions of entire school systems. Thus, since our goal is to create an inviolable lower bound, the safest interpretation is that our estimates represent lower bounds on the variance of the cumulative effects of growing up in different school systems/neighborhoods.

6.3 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 lower bounds on the effect of a shift in 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 estimates use $\widehat{Var}(Z_{2s}G_2 + v_s)$ to calculate the 10th-90th shifts, while the more conservative estimates that seek to remove common shocks use $\widehat{Var}(Z_{2s}G_2)$. For binary outcomes, we estimate the effect of the shift in $Z_{2s}G_2$ via:

$$\begin{aligned} E[\hat{Y}^{90} - \hat{Y}^{10}] = \quad & \frac{1}{I}\sum_i \Phi\Big(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} + 1.28(Var(Z_{2s}G_2))^{.5}]}{(1+\widehat{Var}(v_s))^{.5}}\Big) \\ & - \frac{1}{I}\sum_i \Phi\Big(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} - 1.28(Var(Z_{2s}G_2))^{.5}]}{(1+\widehat{Var}(v_s))^{.5}}\Big) \end{aligned} \tag{34}$$

[28] As outlined in Section 7 below, we choose a set of variables in X_i that satisfies this property in our baseline specification for each dataset.

This average effectively integrates over the distribution of $X_iB + X_sG_1 + v_{si}$, but uses the empirical distributions of X_iB and X_sG_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:

$$\begin{aligned} E[\hat{Y}^{90} - \hat{Y}^{10}] = \quad & \frac{1}{I}\sum_i \Phi(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} + 1.28(Var(Z_{2s}G_2 + v_s))^{.5}]}{(1)}) \\ & - \frac{1}{I}\sum_i \Phi(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} - 1.28(Var(Z_{2s}G_2 + v_s))^{.5}]}{(1)}) \end{aligned} \tag{35}$$

We also report lower bound estimates of the impact of a 10th-to-50th percentile shift in school/neighborhood quality. For the binary outcomes, the impact of a 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.[29]

7 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.[30] 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 the variance decomposition described above. Third, each survey administered questionnaires to school admin-

[29]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 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.

[30]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.

istrators in addition to 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 (2011)) 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.

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_i 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. Appendix Tables A5 - A8 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 discussed in Section 4 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 data to assess the potential for bias in our survey-based estimates more directly. Specifically, we draw 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 Table A4 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 $\widehat{Var}(Z_{2s}G_2)$ and decreasing $\widehat{Var}(Z_{2s}G_2 + v_s)$ 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. Table A8 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. 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.[31] The sample college enrollment rate is 27 percent in NLS72, 31 percent in NELS88, and 37 percent in ELS2002. For NELS88 and ELS2002 the measure of high school graduation 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, the measure is an indicator for whether the student is classified as graduated for the official state reporting requirement. Notice, though, that since ELS2002 first surveys students in 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 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 6 are provided for each of our outcomes in Online Appendix Tables A9, A10, and A11.

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.[32] We then impute values for the other explanatory variables to preserve the sample size, since no one other variable is critical to our analysis.[33] Finally, each specification makes use of a set of panel

[31]However, in NLS72 enrollment status is reported in January-March of the second full school year after graduation, while in NELS88 and ELS2002 it is reported in October.

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

[33]This results in sample sizes for the four year college enrollment analyses of: 12,100 for NLS72, 10,990 for NELS88,

weights. The appropriate weights depend on the analysis. Our rationale for using weights and the details of how we construct them are provided in Online Appendix Section A8.

8 Results

We now turn to results. Along with the point estimates, we report bootstrap standard errors estimates based on re-sampling schools with replacement, with 150 replications. To preserve the size distribution of the samples of students from particular schools, we divide the sample into five school sample size classes and resample schools within class.

8.1 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 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_{2s} (or sorting bias if the spanning condition from Proposition 1 fails). Thus, for high school graduation we focus on the results 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 falls 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 unambigu-

and 12,440 for ELS2002. The sample sizes for the high school graduation analyses are 11,340 for NELS88, and 12,370 for ELS2002, and 297,518 for North Carolina 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.

ously 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 (Z_{2s}) 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 $Var(v_s)$, 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 $X_s^U \beta^U$ 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 X_s or the school-level observables Z_{2s}.

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 are 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 Row 2 of Panel A, the 10th-90th impact estimates displayed in Row 2 of Panel B are larger for NELS88. 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 .

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.

8.2 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 1.8 to 2.6 percent of the outcome variance to schools/neighborhoods, while estimates that include v_s attribute 3.8 to 4.6 percent. Including test scores and behavioral variables reduces these lower bound estimates in a consistent fashion across the three panel surveys (Columns 2, 4, and 6), with the estimates that exclude the residual v_s dropping to 1.5 to 1.9 percent, and the estimates that include the residual v_s 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.

8.3 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 place greater weight on the individual-specific estimates associated with 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. Due to the 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.

8.4 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 log 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 include 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 also seem to have important impacts for longer run outcomes of prime importance for worker welfare.

9 Other Applications

The control function approach can also be applied to other situations in which selective sorting into units makes identification of the independent effect of the units difficult. Measurement of teacher quality is a particularly important application given the widespread use of teacher value added models to aid in the evaluation of teachers. It is also an example of a set of problems in which sorting into groups (classrooms in this case) is mediated by an administrator rather than the result of individual choices.

Most of the analysis in Section 2 can be adapted easily to the administrator choice context. For example, suppose that the school principal has already decided which teachers to allocate to which courses for which periods of the day. A classroom c can also be characterized by a vector of amenity values A_c. The amenities might include the principal's perceptions of various teacher attributes or skills as well as other amenities such as whether the heating system works and the difficulty level of the class. The Θ and Θ^U matrices that relate preferences for different elements of A_c to X_i and X_i^U will now reflect a principal's belief about which types of students are most likely to benefit from a better teacher, the difficulty level, etc. They might also reflect a desire to placate parents or students, where students/parents with certain values of X_i or X_i^U are more likely to advocate for particular classroom assignments.

When the amenity vector A_c is taken to be exogenous to the principal's choice (i.e. independent of classroom composition), the solution to the classroom allocation problem aligns with that of a competitive equilibrium. In Appendix Section A9, we show that in this case the unobserved classroom averages X_c^U will be a linear function of the observed averages X_c under assumptions analogous to those in Proposition 1. Exogeneity of the amenity vector may be reasonable in some high school and college contexts in which students submit course preferences and a schedule-making algorithm assigns students to classrooms.

However, in the elementary and middle school contexts, it seems likely that some elements of A_c could reflect the student makeup of the class. In such contexts the classroom sorting problem diverges from the school/neighborhood sorting problem in two important respects. First, the principal may care directly about inequality across classes in average student characteristics. Second, the principal would internalize the effect that allocating a student to a classroom c has on the classroom's composition-dependent amenities A_c, whereas parents take the school amenities A_s as given. We have not yet solved a planning problem featuring endogenous amenities.

Nevertheless, our analysis of the exogenous amenities case does suggest that the common practice of including classroom averages of student characteristics (such as in Chetty et al. (2014)) may play a potentially powerful role in purging value-added estimates of biases stemming from non-random student sorting on unobservables and observables. Furthermore, as we note in the Appendix, it may also reduce omitted variables bias from non-random assignment of teachers to other unobserved outcome-relevant classroom environmental factors such as track or time of day. While there are many caveats to our analysis, it 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. (2014), Kane and Staiger (2008)).

We also mentioned the evaluation of hospitals and hospital inputs in the introduction. Recent work by Fletcher et al. (2014) uses patient data matched to physicians to estimate the effects of physicians on health outcomes. It controls for very detailed patient characteristics but not for the physician-specific averages of patient characteristics. Our analysis suggests that adding these would allay concerns about sorting on patient unobservables.

Finally, a very different type of application of our approach 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 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.

10 Concluding Remarks

The key takeaway from the empirical analysis is that even conservative estimates of the contribution of schools and surrounding neighborhoods to later outcomes suggest that improving school/neighborhood environments could have a large impact on high school graduation rates and college enrollment rates. As we noted in the introduction, prior evidence on this topic is mixed, in part because prior research showing substantial across-school and across-neighborhood variation in outcomes is subject to concerns about sorting on unobservables that we address in this paper. Our results are quite consistent with the lottery-based estimates of Deming et al. (2014). They suggest that their results might generalize beyond the specific high poverty Charlotte context they consider. Much more speculatively, 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.

There is a long research agenda. On the empirical side, we have discussed a number of other applications for our approach to distinguishing true group effects from sorting. On the theoretical side, there are clearly extensions to our framework that are likely to produce a more nuanced interpretation. In particular, our model of outcomes does not explicitly allow for complementarities and other interactions between school and student quality.[34] For example, it could be the case that the types of students who attend low quality schools are those who are most likely to profit from improvements in school quality. Additional monte carlo analysis is needed to better understand the conditions under which the approach works well in delivering a lower bound. We are currently extending the theoretical analysis around Proposition 1 to the case when the number of choices is discrete rather than continuous, although the simulations reported in the paper make clear that the control function approach works well in that case. We view the central message of our paper to be that in many circumstances aspects of the distribution of the group averages of observed individual characteristics are likely to be a useful control for group averages of unobserved individual characteristics, not that the relationship between the observed and unobserved group averages is necessarily linear.

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

[34] Recall, though, that our model does allow for school treatments to differ across students within a school. Furthermore, the preference weights on amenities that represent school characteristics are permitted to be heterogeneous, as would be the case if parents choose locations the match to their child's needs in mind. 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 composition.

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 (2011) '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, Timothy Conley, Todd E. Elder, and Christopher R. Taber (2013) 'Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables.' Working Paper

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

Ash, Arlene S, Stephen F Fienberg, Thomas A Louis, Sharon-Lise T Normand, Therese A Stukel, and Jessica Utts (2012) 'Statistical issues in assessing hospital performance'

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

Browning, Martin, Pierre-André Chiappori, and Yoram Weiss (2014) *Economics of the Family* (Cambridge University Press)

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 (2014) 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.' *American Economic Review* 104(9), 2593–2632

Cullen, Julie Berry, Brian A Jacob, and Steven Levitt (2006) 'The effect of school choice on participants: Evidence from randomized lotteries.' *Econometrica* 74(5), 1191–1230

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

Doyle Jr, Joseph J, John A Graves, Jonathan Gruber, and Samuel Kleiner (2012) 'Do high-cost hospitals deliver better care? evidence from ambulance referral patterns.' Technical Report, National Bureau of Economic Research

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

Fletcher, Jason M, Leora I Horwitz, and Elizabeth Bradley (2014) 'Estimating the value added of attending physicians on patient outcomes.' Technical Report, National Bureau of Economic Research

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

Lindenlaub, Ilse (2013) 'Sorting multidimensional types.' Technical Report

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin (2013) 'Mismatch, sorting and wage dynamics.' Technical Report, National Bureau of Economic Research

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

Meghir, Costas, and Steven Rivkin (2010) 'Econometric Methods for Research in Education.' NBER Working Paper 16003. Prepared for the Handbook of Education., National Bureau of Economic Research, Inc

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 (2009) 'Student sorting and bias in value-added estimation: Selection on observables and unobservables.' *Education* 4(4), 537–571

_ (2014) 'Revisiting the impacts of teachers.' *Unpublished working paper. http://eml. berkeley. edu/~jrothst/workingpapers/rothstein_cfr. pdf*

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

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		NELS88 gr8		ELS2002	
	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
$Var(Z_{2s}G_2)$	(0.008)	(0.004)	(0.006)	(0.004)	(0.012)	(0.011)
LB w/ unobs	0.049	0.036	0.028	0.016	0.036	0.025
$Var(Z_{2s}G_2+v_s)$	(0.014)	(0.008)	(0.009)	(0.005)	(0.012)	(0.011)

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

Lower Bound	NC		NELS88 gr8		ELS2002	
	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
Based on $Var(Z_{2s}G_2)$	(0.022)	(0.014)	(0.014)	(0.012)	(0.013)	(0.012)
LB w/ unobs: 10th-90th	0.174	0.152	0.098	0.075	0.083	0.070
Based on $Var(Z_{2s}G_2+v_s)$	(0.026)	(0.017)	(0.017)	(0.013)	(0.013)	(0.013)
LB no unobs: 10th-50th	0.056	0.044	0.033	0.025	0.040	0.038
Based on $Var(Z_{2s}G_2)$	(0.013)	(0.008)	(0.008)	(0.007)	(0.009)	(0.008)
LB w/ unobs: 10th-50th	0.096	0.083	0.055	0.041	0.048	0.039
Based on $Var(Z_{2s}G_2+v_s)$	(0.016)	(0.010)	(0.010)	(0.008)	(0.009)	(0.008)

Bootstrap standard errors based on resampling at the school level are in parentheses.

Panel A reports 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 row labelled "LB no unobs" reports $Var(Z_{2s}G_2)$ and excludes the unobservable v_s while the row labeled "LB w/ unobs" reports $Var(Z_{2s}G_2+v_s)$.

Panel B reports estimates of the average effect of moving students from a school/neighborhood at the 10th quantile of the quality distribution to one at the 50th or 90th quantile.

The columns headed "NC" are based on the North Carolina data and refer to a decomposition that uses the 9th grade school as the group variable. The columns headed "NELS88 gr8" are based on the NELS88 sample and refer to a decomposition that uses the 8th grade school as the group variable. The columns headed "ELS2002" are based on the ELS2002 sample and refer to a decomposition that uses the 10th grade school as the group variable.

For each data set the variables in the baseline model and the full model are specified in Web Appendix Tables A5 - A8.

The full variance decompositions underlying these estimates are presented in Web Appendix Table A9.

Appendix Sections 5 and 6 discuss estimation of model parameters and the variance decompositions. Section 6.3 discusses estimation of the 10-50 and 10-90 differentials.

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	NLS72		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs	0.026	0.019	0.018	0.015	0.022	0.018
$Var(Z_{2s}G_2)$	(0.005)	(0.004)	(0.006)	(0.005)	(0.007)	(0.006)
LB w/ unobs	0.038	0.032	0.040	0.029	0.046	0.031
$Var(Z_{2s}G_2+v_s)$	(0.007)	(0.006)	(0.008)	(0.006)	(0.009)	(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

Lower Bound	NLS72		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.139	0.118	0.127	0.112	0.155	0.132
Based on $Var(Z_{2s}G_2)$	(0.013)	(0.012)	(0.018)	(0.017)	(0.019)	(0.017)
LB w/ unobs: 10th-90th	0.170	0.152	0.188	0.155	0.216	0.172
Based on $Var(Z_{2s}G_2+v_s)$	(0.017)	(0.016)	(0.020)	(0.018)	(0.021)	(0.020)
LB no unobs: 10th-50th	0.065	0.056	0.061	0.054	0.075	0.064
Based on $Var(Z_{2s}G_2)$	(0.006)	(0.005)	(0.008)	(0.008)	(0.008)	(0.008)
LB w/ unobs: 10th-50th	0.078	0.071	0.088	0.073	0.103	0.083
Based on $Var(Z_{2s}G_2+v_s)$	(0.007)	(0.007)	(0.009)	(0.008)	(0.009)	(0.009)

Bootstrap standard errors based on resampling at the school level are in parentheses.

The notes to Table 1 apply, except that Table 2 reports results for enrollment in a 4-year college two years after graduation.

The column headed NLS72 refers to a variance decomposition that uses the 12th grade school as the group variable.

Table 3: The Impact of 10th-90th Percentile Shifts in School Quality on High School Graduation Rates for Selected Subpopulations

	NC		NELS88 gr8		ELS2002	
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
Based on $Var(Z_{2s}G_2)$	(0.030)	(0.020)	(0.026)	(0.024)	(0.023)	(0.025)
LB w/ unobs	0.242	0.229	0.176	0.159	0.146	0.144
Based on $Var(Z_{2s}G_2+v_s)$	(0.035)	(0.024)	(0.031)	(0.027)	(0.024)	(0.025)
XB: 90th Quantile						
LB no unobs	0.060	0.036	0.016	0.004	0.019	0.010
Based on $Var(Z_{2s}G_2)$	(0.013)	(0.006)	(0.004)	(0.001)	(0.004)	(0.002)
LB w/ unobs	0.098	0.063	0.026	0.006	0.022	0.010
Based on $Var(Z_{2s}G_2+v_s)$	(0.016)	(0.008)	(0.005)	(0.001)	(0.004)	(0.002)
Black						
LB no unobs	0.107	0.085	0.061	0.053	0.079	0.082
Based on $Var(Z_{2s}G_2)$	(0.022)	(0.014)	(0.015)	(0.015)	(0.015)	(0.014)
LB w/ unobs	0.176	0.152	0.098	0.084	0.094	0.084
Based on $Var(Z_{2s}G_2+v_s)$	(0.026)	(0.017)	(0.018)	(0.017)	(0.015)	(0.014)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.142	0.114	0.099	0.078	0.101	0.096
Based on $Var(Z_{2s}G_2)$	(0.029)	(0.019)	(0.023)	(0.019)	(0.021)	(0.020)
LB w/ unobs	0.235	0.206	0.159	0.125	0.120	0.099
Based on $Var(Z_{2s}G_2+v_s)$	(0.034)	(0.022)	(0.028)	(0.021)	(0.022)	(0.020)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.062	0.047	0.025	0.016	0.032	0.016
Based on $Var(Z_{2s}G_2)$	(0.014)	(0.008)	(0.007)	(0.004)	(0.007)	(0.005)
LB w/ unobs	0.102	0.084	0.040	0.025	0.037	0.016
Based on $Var(Z_{2s}G_2+v_s)$	(0.016)	(0.010)	(0.007)	(0.005)	(0.007)	(0.005)

Bootstrap standard errors based on re-sampling at the school level are in parentheses.

The table reports the average effect for the subpopulation indicated by the row heading of moving students from a school/neighborhood at the 10th quantile of the quality distribution to one at the 90th quantile.

"XB: 10th Quantile" and "XB: 90th Quantile" refer to students whose values of $X_{si}B$ is equal the estimated 10th (90th) quantile value of the $X_{si}B$ distribution. See Section 8.3.

See the notes to Table 1 for row and column definitions

Table 4: The Impact of 10th-90th Percentile Shifts in School Quality on Four-Year College Enrollment Rates for Selected Subpopulations

	NLS72		NELS88 gr8		ELS2002	
Subpopulation	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
XB: 10th Quantile						
LB no unobs	0.078	0.027	0.064	0.032	0.100	0.050
Based on $Var(Z_{2s}G_2)$	(0.008)	(0.004)	(0.010)	(0.005)	(0.013)	(0.007)
LB w/ unobs	0.094	0.034	0.093	0.046	0.138	0.064
Based on $Var(Z_{2s}G_2+v_s)$	(0.010)	(0.005)	(0.011)	(0.006)	(0.015)	(0.008)
XB: 90th Quantile						
LB no unobs	0.191	0.182	0.160	0.128	0.166	0.128
Based on $Var(Z_{2s}G_2)$	(0.017)	(0.017)	(0.023)	(0.022)	(0.021)	(0.017)
LB w/ unobs	0.234	0.234	0.236	0.187	0.231	0.167
Based on $Var(Z_{2s}G_2+v_s)$	(0.023)	(0.024)	(0.025)	(0.022)	(0.023)	(0.017)
Black						
LB no unobs	0.132	0.109	0.125	0.111	0.145	0.121
Based on $Var(Z_{2s}G_2)$	(0.013)	(0.012)	(0.017)	(0.016)	(0.018)	(0.016)
LB w/ unobs	0.161	0.140	0.184	0.152	0.201	0.158
Based on $Var(Z_{2s}G_2+v_s)$	(0.017)	(0.016)	(0.019)	(0.017)	(0.020)	(0.018)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.110	0.099	0.091	0.074	0.140	0.124
Based on $Var(Z_{2s}G_2)$	(0.012)	(0.011)	(0.014)	(0.012)	(0.018)	(0.016)
LB w/ unobs	0.134	0.127	0.132	0.102	0.195	0.162
Based on $Var(Z_{2s}G_2+v_s)$	(0.014)	(0.013)	(0.016)	(0.013)	(0.020)	(0.019)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.180	0.158	0.157	0.139	0.173	0.148
Based on $Var(Z_{2s}G_2)$	(0.016)	(0.015)	(0.022)	(0.021)	(0.021)	(0.020)
LB w/ unobs	0.220	0.204	0.232	0.192	0.242	0.193
Based on $Var(Z_{2s}G_2+v_s)$	(0.022)	(0.021)	(0.024)	(0.022)	(0.023)	(0.023)

Bootstrap standard errors based on resampling at the school level are in parentheses.

The notes to Table 3 apply, except that Table 4 reports results for enrollment in a 4 year college 2 years after graduation, and the NLS72 is one of the data sets.

Table 5: Lower Bound Estimates of the Contribution of School Systems and Neighborhoods to Completed Years of Postsecondary Education and Permanent Wages (NLS72 data)

Panel A: Fraction of Variance Attributable to School/Neighborhood Quality

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	0.026	0.019	0.018	0.015	0.022	0.018
$Var(Z_{2s}G_2)$	(0.002)	(0.002)	(0.010)	(0.010)	(0.011)	(0.011)
LB w/ unobs	0.038	0.032	0.040	0.029	0.046	0.031
$Var(Z_{2s}G_2+v_s)$	(0.004)	(0.002)	(0.013)	(0.016)	(0.021)	(0.021)

Panel B: Effects on Years of Postsecondary Education and Permanent Wages of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile

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	0.578	0.445	0.152	0.157	0.155	0.157
Based on $Var(Z_{2s}G_2)$	(0.054)	(0.039)	(0.019)	(0.019)	(0.020)	(0.020)
LB w/unobs: 10th-90th	0.661	0.520	0.177	0.177	0.175	0.173
Based on $Var(Z_{2s}G_2+v_s)$	(0.069)	(0.047)	(0.028)	(0.023)	(0.031)	(0.031)
LB no unobs: 10th-50th	0.283	0.222	0.076	0.079	0.077	0.078
Based on $Var(Z_{2s}G_2)$	(0.027)	(0.019)	(0.010)	(0.010)	(0.010)	(0.010)
LB w/unobs: 10th-50th	0.331	0.260	0.088	0.088	0.087	0.087
Based on $Var(Z_{2s}G_2+v_s)$	(0.035)	(0.024)	(0.014)	(0.012)	(0.016)	(0.016)

Bootstrap standard errors based on resampling at the school level are in parentheses.

Panel A of Table 5 reports lower bound estimates of the fraction of variance of years of postsecondary education and permanent wage rates (with and without controls for postsecondary education) that can be directly attributed to school/neighborhood choices for each dataset. The sample is NLS72.

The row labelled "LB no unobs" reports $Var(Z_{2s}G_2)$ and excludes the unobservable v_s while the row labeled "LB w/ unobs" reports $Var(Z_{2s}G_2+v_s)$.

Panel B reports estimates of the average effect of moving students from a school/neighborhood at the 10th quantile of the quality distribution to one at the 50th or 90th quantile. It is equal to $2*1.28$ times the value of $[\widehat{Var}(Z_{2s}G_2+v_s)]^{0.5}$ or $[\widehat{Var}(Z_{2s}G_2)]^{0.5}$ in the corresponding column of the table.

See Web Appendix Table A5 for the variables in the baseline model and the full model. The full variance decompositions are in Appendix Table A10. Web Appendix Sections 5 and 6 discuss estimation of model parameters and the variance decompositions.

Appendix: For Online Publication Only

A1 Spanning Condition Examples

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

Case 1: $rank(\Theta^U) \leq rank(\Theta) = dim(A)$

Suppose that the matrices capturing the impact of observed unobserved student characteristics on parent WTP for amenities Θ and Θ^U are each full rank. For example:

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

Then we can write $\Theta^U = R\Theta$, where

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

Thus, the spanning condition is satisfied in this case. If Θ^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: $X^U = [X_1^U, X_2^U, X_3^U]$, each of which affects WTP for the two amenities differentially. Suppose that X and Θ are unchanged from Case 1:

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

Then we can write $\Theta^U = R\Theta$, where

$$R = \left\{ \begin{array}{cc} 1 & 1 \\ 2 & -1 \\ 1 & 0 \end{array} \right\}$$

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 Θ is at least as large as the row rank of Θ^U. Any scenario satisfying $rank(\Theta^U) \leq rank(\Theta) = dim(A)$ will satisfy the spanning condition in Proposition 1.

Case 2: $rank(\Theta) < rank(\Theta^U) \leq dim(A)$

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

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

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

$$\Theta = \left\{ \begin{array}{cc} 1 & 2 \\ 2 & 4 \end{array} \right\}.$$

Here, an incremental unit of X_1 or X_2 will affect WTP for A_2 by twice as much as it will affect WTP for A_1. As in the previous example, there is no matrix R such that $R\Theta = \Theta^U$. For any choice of R, in each row of $R\Theta$ the second column will always be twice as large as the first column, but the second row of Θ^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 Θ is less than the row rank of Θ^U, then the row space of Θ^U cannot possibly be a subspace of the row space of Θ.

Case 3: $rank(\Theta^U) \leq rank(\Theta) < dim(A)$

Suppose now that both X and X^U are scalars: $X \equiv X_1$, $X^U \equiv X_1^U$. Consider first the case where X_1 only predicts WTP for A_1, while X_1^U only predicts WTP for A_2:

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

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

Indeed, suppose that we alter Θ and Θ^U so that both X_1 and X_1^U affect willingness to pay for both amenities (but in different proportions):

$$\Theta = \left\{ \begin{array}{cc} 1 & 1 \end{array} \right\} \quad \Theta^U = \left\{ \begin{array}{cc} 2 & 4 \end{array} \right\}$$

There is no scalar R such that $R\Theta = \Theta^U$, since any value of R will preserve the one-to-one ratio between the first and second entries in Θ, while Θ^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 Θ^U is not a subspace of the row space of Θ. 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):

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

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

A2 Testing Proposition 1: Does X_s Span the Amenity Space?

As discussed in Section 3.1.2, one of the key necessary conditions for Proposition 1 to hold is that the vector of observables X_i captures enough independent factors determining families' preferences over group amenities that it can span the space of amenities for which these observables affect tastes (denoted A_s^X). Under the particular linear specification of utility featured in 2, this condition is tantamount to requiring that $rank(\delta) \geq \dim(A_s^X)$.

Note that equation 11 implies that X_s can be written as a linear combination of the gradient of equilibrium neighborhood prices with respect to amenities: $X_s = \nabla P(A_{s(i)}) * \Omega$, where $\Omega = Var(\gamma_i)^{-1}\tilde{\delta}^{'} Var(X_i)$. If the condition $rank(\delta) \geq \dim(A_s^X)$ is strictly satisfied, then the L^O elements of X_s are all linear combinations of a smaller number g of price gradient components. But this implies that $Cov(X_s)$ will be rank deficient, with $rank(Cov(X_s)) = \dim(A_s^X)$. This is a testable prediction.

More generally, suppose Proposition 1 is nearly satisfied, so that small number of amenity factors drive the vast majority of the variation in X_s, but there are several other amenities for which elements of X_i slightly influence taste. Our simulations in section 4 suggest that such minor departures from the necessary conditions laid out in Proposition 1 have very little impact on the ability of X_s effectively control for the unobservable between-school variation X_s^U. In such contexts, a small number of amenity factors should account for a very large fraction of the variation in X_s, with only a very small amount of unexplained residual variation.

We test these predictions by performing principal components analysis (PCA) on X_s. Because the sample school averages of observable characteristics $\overline{X}_s$ are noisy measures of the expected values $X_s \equiv E[X_i|s(i) = s$, we do not fit the PCA model to $\overline{X}_s$ directly. Instead, we estimate the underlying true covariance matrix $Cov(X_s)$[35], and then directly perform the principal components

[35]Specifically, we estimate $\hat{Cov}(X_i)$ and $\hat{Cov}(X_i - X_s)$ by taking the sample (weighted) covariances of X_i and $X_i - \hat{\overline{X}}_s$, performing the requisite degrees-of-freedom adjustment, and then obtaining $\hat{Cov}(X_s)$ via $\hat{Cov}(X_s) = \hat{Cov}(X_i) - \hat{Cov}(X_i - X_s)$.

analysis on the estimated covariance matrix.

In Appendix Table A1 we present the results of this exercise. Panel A provides, for each dataset we use, the number of principal components necessary to explain 75%, 90%, 95%, 99%, and 100% of the sum of the variances of the standardized values of the L^O characteristics in X_s ($\sum_l^{L^O} Var(X_{sl})$, respectively. This is the standard output from a factor analysis. In Panel B, we also provide the number of principal components necessary to explain 75%, 90%, 95%, 99%, and 100% of the variance in $X_s\hat{G}_1$, the regression index formed by using the estimated coefficients on school-level averages from our empirical analysis.

Both Panel A and Panel B provide strong evidence that $rank(\delta) \geq \dim(A_s^X)$, implying that the first necessary condition for the spanning condition $\tilde{\delta} = R\delta^U$ in Proposition 1 is satisfied in the datasets we use. Specifically, in each dataset, $Cov(X_s)$ is found to be rank deficient. For example, in the full specification using ELS2002, only 33 latent factors are needed to explain all of the variance in X_s (Panel A, Row 6, Column 6), compared to $L^O = 51$ elements of X_s. Similarly, in the NELS88 full specification, only 32 factors fully explain the variance in the 49 factors of X_s.

Furthermore, the PCA analysis also suggests that a much smaller number of factors can account for the vast majority of the variation in either $\sum_l^{L^O} Var(X_{sl}$ or $Var(X_s\hat{G}_1)$. In the ELS2002 full specification, only 19 and 11 factors are needed to explain 95% of the variation in $\sum_l^{L^O} Var(X_{sl}$ and $Var(X_s\hat{G}_1)$, respectively (Panels A and B, Row 4, Column 6). For NELS88, only 20 and 13 factors are needed to explain 95% of the variation in the corresponding two measures (Panels A and B, Row 4, Column 4).

Note, though, that our principal components analysis does not inform us about the second necessary condition for X_s to effectively control for X_s^U: for each component of X_i^U, either there must exist an element of the observed vector X_i that is correlated with this unobservable, or the set of amenities for which it shifts preferences must also be a subset of the amenities for which elements of X_i shift preferences ($A_s^{X^U} \subset A_s^X$). As mentioned in Section 3.1.2, we believe that the richness and size of the set of observables used in our datasets make this second necessary condition plausible.

A3 Solving for the Equilibrium Allocation

In this section we further explore the relationship student characteristics, school choices, and school and neighborhood amenities by explicitly solving the model from Section 2 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).[36] We continue to assume that families freely choose locations subject to $P(A_s)$. This

[36]Neither of these assumptions are required for Proposition 1 and so are not necessary to justify the use of school

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.

Recall that when ε_{is} is excluded from the model, any vectors $[X, X^U, \kappa]$ sharing the same value of the index vector λ will feature the same willingness to pay for all possible neighborhoods. As a result, the exact school assignment for a particular vector of characteristics $[X_i, X_i^U, \kappa_i]$ conditional on λ_i carries no welfare implications, which creates an infinite number of equilibrium allocations. Consequently, we focus on how values of λ_i get mapped to amenity vectors A in equilibrium (which we call "λ-allocations").

We restrict attention to linear λ-allocations of the form $A = \Psi\lambda'$, where Ψ is a $K \times K$ matrix, and we proceed to show that there is a unique matrix Ψ that satisfies both feasibility and pareto efficiency.[37]

Consider first the feasibility requirement: demand for each value of amenity vector $A_1, \ldots, A_K$ generated by the proposed λ-allocation must equal supply. In our continuous context, feasibility means that the distributions of amenity vectors supplied and demanded coincide. Let $\Sigma_{\lambda_i'}$ denote the covariance matrix of the joint distribution of the K-vector λ_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 λ_i in our proposed equilibrium is $\Psi\lambda_i'$, our feasibility requirement can be written as:

$$\Sigma_A = Var(\Psi\lambda') = \Psi\Sigma_{\lambda'}\Psi' \tag{36}$$

Thus, feasibility requires that Σ_A and $\Sigma_{\lambda'}$ be made congruent by the allocation matrix Ψ. Since both ΣA and $\Sigma_{\lambda'}$ 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.

average observable characteristics as a control function for school average unobservable characteristics.

[37]We have not yet proven but strongly suspect that the linear allocation represents the unique equilibrium λ-allocations when non-linear λ-allocations are also considered. Gretsky et al. (1992) has already proven the existence of an equilibrium for a class of games that includes our model.

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

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 $\lambda' A$ and the proposed λ-allocation $A = \Psi\lambda'$, a mutually beneficial exchange between families with the vectors λ_1 and λ_2 will exist iff $\lambda_1\Psi\lambda_1' - \lambda_1\Psi\lambda_2' < \lambda_2\Psi\lambda_1' - \lambda_2\Psi\lambda_2'$. Thus, ruling out such exchanges requires:

$$\lambda_1\Psi\lambda_1' - \lambda_1\Psi\lambda_2' \geq \lambda_2\Psi\lambda_1' - \lambda_2\Psi\lambda_2' \; \forall \; (\lambda_1, \lambda_2) \in \mathscr{R}^K x \mathscr{R}^K \tag{37}$$

But this equation can be reformulated as follows:

$$\begin{aligned}
&\lambda_1\Psi\lambda_1' - \lambda_1\Psi\lambda_2' \geq \lambda_2\Psi\lambda_1' - \lambda_2\Psi\lambda_2' \; \forall \; (\lambda_1, \lambda_2) \in \mathscr{R}^K x \mathscr{R}^K \\
&\Rightarrow (\lambda_1 - \lambda_2)\Psi\lambda_1' - (\lambda_1 - \lambda_2)\Psi\lambda_2' \geq 0 \; \forall \; (\lambda_1, \lambda_2) \in \mathscr{R}^K x \mathscr{R}^K \\
&\Rightarrow (\lambda_1 - \lambda_2)\Psi(\lambda_1 - \lambda_2)' \geq 0 \; \forall \; (\lambda_1, \lambda_2) \in \mathscr{R}^K x \mathscr{R}^K \\
&\Rightarrow \tilde{\lambda}\Psi\tilde{\lambda}' \geq 0 \; \forall \tilde{\lambda} \in \mathscr{R}^K
\end{aligned} \tag{38}$$

where $\tilde{\lambda}$ is defined to be the difference between two arbitrarily chosen row vectors λ_1 and λ_2. Equation 38 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 would violate feasibility, since $rank(\Psi\Sigma_{\lambda'}\Psi') \leq rank(\Psi) < rank(\Sigma_A)$. Thus, Ψ must be positive definite. But two positive definite matrices (in our case $\Sigma_{\lambda'}$ and Σ_A) are made congruent by a unique positive definite matrix Lawson and Lim (2006):

$$\Psi = \Sigma_{\lambda'}^{-1/2}(\Sigma_{\lambda'}^{1/2}\Sigma_A\Sigma_{\lambda'}^{1/2})\Sigma_{\lambda'}^{-1/2} \tag{39}$$

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

Furthermore, since every positive definite matrix is invertible, we can also express the vector λ_i for any individual as a linear function of the amenity vector of their chosen school:

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

To characterize the equilibrium price function, $P(A)$, note that from (4) $\nabla P(A_{s_i}) = \lambda_i'$. Substituting for λ_i' using the above equation for the equilibrium relationship between λ_i and $A_{s(i)}$, we obtain:

$$\nabla P(A_{s_i}) = \Psi^{-1}A_{s(i)} \tag{40}$$

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

$$P(A) = 0.5A'\Psi^{-1}A + c, \text{ for any } c \in \mathscr{R} \tag{41}$$

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.

A4 Monte Carlo Simulations Exploring Finite-Sample Properties

This section describes a set of monte carlo simulations designed to explore the finite-sample properties of our control function estimator across a number of key dimensions. As discussed in section 4, a full characterization of these finite-sample properties is not feasible, so we focus on a stylized test case that is rich enough to reveal the strengths and weaknesses of our approach. Section A4.1 lays out the simulation methodology, while Section A4.2 presents and interprets the results.

A4.1 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[39].
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 $\Theta_{k\ell}$ and Θ^U represent draws from a multivariate normal distribution in which $corr(\Theta_{k\ell}, \Theta_{jm}) = \rho$ if $j = k$ or $\ell = m$, and 0 otherwise.
4. The variances of the elements of A_s, $[X_i, X_i^U, \kappa_i]$, $\varepsilon_{i,s}$ are chosen to create interclass correlations for X_i and X_i^U of between .1 and .25 across specifications. These values are in line with the range observed across the datasets used in the empirical analysis.

Our test case implies considerable sorting into schools along many dimensions of school amenities and along many observable and unobservable dimensions of student quality. It represents a conservative case because one might expect that in reality a few key observable (and unobservable) individual level factors (e.g. parental income, education, and wealth) and a few key school/neighborhood amenities (ethnic composition, crime, principal quality) drive most of the systematic sorting of students to schools. Given restrictions 1-4, we complete the model by choosing particular sets of 7 remaining parameters. The first parameter is students per school. For simplicity, we impose that each school has capacity equal to this student/school ratio.[40] The student/school ratio is denoted “# Stu” in Appendix Table A2. The second parameter is the total number school/neighborhood combinations available (denoted “# Sch”).

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

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

The parameter #Con is the number of schools in the consideration set for each household. This captures the possibility that most parents only realistically consider 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.") specify the number of observed and unobserved student characteristics that affect outcomes. The sixth parameter is the dimension of the amenity vector over which households have preferences. In most of the specifications we assume that it less then or equal to the number of observed characteristics and that the rows of Θ^U form a linear subspace of the rows of $\tilde{\Theta}$, as required by Proposition 1.

The seventh parameter determines ρ, the correlation between any pair of random variables $(\Theta_{k\ell}, \Theta_{jm})$ from which each $(\Theta_{k\ell}, \Theta_{jm}$ 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).

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 $\Theta^U \neq R\tilde{\Theta}$ for any R).

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 $(X_s^U \beta^U)$ to each school's average outcome (which captures the potential bias from unobservable sorting) on the full vector of school-averages of observed characteristic, X_s. 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 $\overline{X}_s$ 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_i^U, κ_i, and $\{\varepsilon_{is}\}$ from the distributions described above to calculate the WTP of each household for each school.[41] 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

[41] To minimize the statistical "chatter" introduced by the particular $\tilde{\Theta}$ matrix that we happened to draw, we drew ten different $\tilde{\Theta}$ 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.

using the simplex method combined with sparse matrix techniques. [42]

A4.2 Simulation Results

The simulation results are presented in Appendix Table A2. 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 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. This is probably a conservative choice. Finally, the correlation ρ between the random variables of which the taste weight matrices Θ and Θ^U are multivariate draws is assumed to be .25.

The first takeaway from Row (1) is that the 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).[43]

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 increases with 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 reduces

[42] 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.

[43] In other simulations not shown, we directly examine the impact of increasing the variance of ε_{is}. We find that increasing $Var(\varepsilon_{is})$ 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 $Var(\varepsilon_{is}) \to \infty$, 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.

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.[44] 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.

Finally, Row (9) displays the results of a specification in which all of the $\Theta_{k\ell}$ and $\Theta k\ell^U$ 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 Appendix Table A2 indicate that the 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

[44] 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 Θ^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!

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 A7 using the North Carolina administrative data.

Note that all the specifications in Appendix Table A2 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.

Appendix Table A3 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 WTP 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 A2, but allowing the elements of X_i^U to only predict WTP for the last amenity, A_5, while the elements of 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, it comes as no surprise that 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 WTP 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 WTP for an amenity is unaffected by the entire vector 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 WTP for the last amenity (A_5) is only by one of the ten components of the unobserved vector X_i^U. In the third scenario, $X_{i,10}^U$ affects WTP for A_5 only. In the fourth scenario, $X_{i,10}^U$ 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 $X_{i,10}^U$ predicts WTP for A_5 only, and 97% when $X_{i,10}^U$ predicts WTP for all five amenities (compared to 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 WTP for just a few additional amenities for which X_i does not predict preferences.

A5 Estimation of Model Parameters

In this section we discuss estimation of the coefficients B, G_1, and G_2. The estimation strategy depends on the outcome, so we consider the outcomes in turn.

A5.1 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: $Z_s = [X_s; Z_{2s}]$. Z_{2s} consists of school and neighborhood characteristics for which direct measures are available, such as student/teacher ratio, city size, and school type. X_s 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 X_s differs across specifications that use different X vectors. G_1 and G_2 are the corresponding subsets of the coefficients in G.

We replace X_s with $\bar{X}_s$, where $\bar{X}_s$ is the average of X_i computed over all available students from the school.[45] We estimate G_1 and G_2 by applying least squares regression to

$$Y_{si} - X_{si}\hat{B} = \overline{X}_s G_1 + Z_{2s} G_2 + v_{s,i}$$

using the appropriate panel weights from the surveys.

A5.2 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.[46]

[45] A substantial number of students who appear in the base year of the surveys can be used to construct $\bar{X}_s$ but cannot be used to estimate (A5.1) 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 7, 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.

[46] 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

$\xi_{s,i,t}$ 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 16 with Y_{si} defined to be $W_{s,i}$. After substituting for $W_{s,i}$, the wage equation is

$$W_{s,i,t} = X_{s,i}B + X_sG_1 + Z_{2s}G_2 + v_{s,i} + e_{s,i,t} + \xi_{s,i,t}.$$

We estimate B by OLS with school fixed effects included.[47]

Let $\tilde{W}_{s,i,t} \equiv W_{s,i,t} - X_i\hat{B}$. We estimate G_1 and G_2 by applying OLS to

$$\tilde{W}_{s,i,t} = \bar{X}_sG_1 + Z_{2s}G_2 + v_{s,i} + e_{s,i,t} + \xi_{s,i,t} \quad (42)$$

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

A5.3 High School Graduation and College Enrollment

The methods outlined in Appendix Sections A5.1 and A5.2 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},

$$HSGRAD_{s,i} = 1(X_iB + X_sG_1 + Z_{2s}G_2 + +v_{s,i} > 0) \quad (43)$$

We replace X_s with $\bar{X}_s$ and estimate the equation

$$HSGRAD_{s,i} = 1(X_iB + \bar{X}_sG_1 + Z_2G_2 + (X_s - \bar{X}_s)G_1 + v_{s,i} > 0) \quad (44)$$

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

A6 Decomposing the Variance in Educational Attainment and Wages

In this section we discuss an analysis of variance based on equation that can be used to place a lower bound on the importance of factors that are common to students from the same school.[48] As

and Solon (2003), Haider (2001), and Meghir and Pistaferri (2004).

[47] 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 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}.

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

with parameter estimation, the details of our procedure depend upon the outcome. We begin with years of postsecondary education.

A6.1 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 (A6),

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

and

$$Y_s = X_s B + X_s G_1 + Z_{2s} G_2 + v_s.$$

Thus, one may express the outcome variance as[49]

$$Var(Y_i) = [Var((X_i - X_s)B) + Var(v_{s,i} - v_s)] + \quad (45)$$

$$[Var(X_s B) + 2Cov(X_s B, X_s G_1) + 2Cov(X_s B, Z_{2s} G_2) + Var(X_s G_1) + \quad (46)$$

$$2Cov(X_s G_1, Z_{2s} G_2) + Var(Z_{2s} G_2) + Var(v_s)] \quad (47)$$

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 $Var(X_s B)$ can be calculated as $\widehat{Var}(X_i B) - \widehat{Var}((X_i - X_s)B)$. One can also estimate the components $Var(X_s G_1)$,$Var(Z_{2s} G_2)$ of the school/community contribution and the common terms $2Cov(X_s B, X_s G_1)$, $2Cov(X_s B, Z_{2s} G_2)$ and $2Cov(X_s G_1, Z_{2s} G_2)$ using the estimates of B, G_1, G_2 and the data $\bar{X}_s$ and Z_{2s}. $Var(v_s)$ can be calculated as

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

[49]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 $Cov(X_s, v_s) = 0$ and $Cov(Z_{2s}, v_s) = 0$, which are implied by our definition of $[G_1, G_2]$ and v_s (see Section 5).

A6.2 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 A8, 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 $X_{s,i}B$, X_sG_1, $Z_{2s}G_2$, $v_{s,i}$, and v_s to $Var(W_{si})$ proceeds as in previous subsection.

A6.3 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

$$Var(Y_i) = [Var((X_i - X_s)B) + 1] + \tag{48}$$

$$[Var(X_sB) + 2Cov(X_sB, X_sG_1) + 2Cov(X_sB, Z_{2s}G_2) + Var(X_sG_1) + \tag{49}$$

$$2Cov(X_sG_1, Z_{2s}G_2) + Var(Z_{2s}G_2) + Var(v_s)] \tag{50}$$

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.

A6.4 Calculation of Standard Errors

We calculate bootstrap standard errors for each of our point estimates and bound estimates based on re-sampling schools with replacement, with 150 replications. To preserve the size distribution of the samples of students from particular schools, we divide the sample into 5 school sample size classes and resample schools within class.

A7 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 4 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 unobservable 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 A4 presents the results of this exercise. For comparison, the first column of Panel A presents the variance decomposition described in Section 6 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.[50] 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, $Var(Z_{2s}G_2)$, by 0.88%, and understates true variance fraction for the lower bound that may include common shocks, $Var(Z_{2s}G_2 + v_s)$, 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 distribution in our simulation test, we do not report any NELS88 results that group students by their grade 10 school.

[50] 10th grade schools in NELS88 are the schools in which the original 8th grade NELS sample are observed in the first follow-up survey.

A8 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.

A9 Other Applications: Estimating Teacher Value-Added

This section examines how our central insight that group-averages of observed individual characteristics can control for group-averages of unobserved individual characteristics can be extended to contexts in which group assignments are determined by a central administrator rather in a decentralized competitive equilibrium. The particular context we consider is one in which a school principal is assigning students to classrooms based on a combination of observed and unobserved (to the econometrician) student inputs, where the goal is to estimate each teacher's value-added to test score achievement.

A9.1 Sorting of Students Across Class Rooms

Assume for now 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 A_c. Consider first the case in which none of amenities reflect the demographic make up of the class and are endogenous to the principal's assignment decisions, so that the amenity vector A_c can be considered exogenous to the principal's student-to-classroom allocation problem. Instead, 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 classroom technology in the room, the time in the day that the class is held, or the difficulty level of the class. As noted in Section 9, exogeneity of the amenity vector may be a reasonable assumption in some high school and college contexts in which students submit course preferences and a schedule-making algorithm assigns students to classrooms.

We can then adapt the utility function featured in equation (2) to model the payoff that the principal obtains from assigning student i to class c (simply replace all s subscripts with c subscripts). As before, X_i is 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, 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. The Θ 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. Θ might also reflect a desire to placate parents or students, where students/parents with certain values of X_i or 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. Similarly, the idiosyncratic match value 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. Thus, we model parent and student preferences as affecting choice through their impact on principal preferences.[51]

Suppose, as in Appendix Section A3, that the space of classroom amenities is continuous, that $\varepsilon_{ic} = 0 \; \forall \; (i,c)$, and that the distributions of both student characteristics and student-weighted classroom amenities $A_{c(i)}$ are jointly normal. Recall that the efficiency implications of alternative allocations only depend on the summary taste vector $\lambda_i \equiv (X_i\Theta + X_i^U\Theta^U + \kappa)$. Let $f(\lambda_i)$ denote the PDF of λ_i across the student population, and let $U(\lambda, A) = \lambda A$ denote the rewritten utility function. Then

[51] Rothstein (2009) provide a useful classroom assignment model in which principals assign students to classrooms based on student characteristics that are observable to both the principal and the econometrician X_i and student characteristics that are only available to the principal (part of X_i^U). He discusses bias in VAM models that include X_i and possible some other controls. He does not study the potential for X_c to control for X_c^U.

we can write the principal's problem as:

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

where we have restricted our attention to linear allocations of the form $A_{c(i)} = \Psi\lambda_i$. Note that the constraints represent the same feasibility conditions as those from the school choice problem solved in Appendix Section A3, and pareto optimality remains a necessary condition for the objective function $\int_{\underline{\lambda}}^{\overline{\lambda}} U(\lambda, \Psi\gamma) f(\lambda) d\lambda$ to be maximized. Indeed, the equivalence of the two problems is essentially a manifestation of the first and second welfare theorems. Consequently, the solution $\Psi = \Sigma_{\lambda'}^{-1/2}(\Sigma_{\gamma'}^{1/2}\Sigma_A\Sigma_{\lambda'}^{1/2})\Sigma_{\lambda'}^{-1/2}$ will also be the same. Furthermore, if the spanning condition $\Theta^U = R\tilde{\Theta}$ is satisfied for some matrix R, X_c will be a linear function of X_c^U.

However, in the elementary and middle school contexts, it seems particularly likely that some elements of A_c could reflect the student makeup of the class. 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. The classroom sorting problem differs from the school/neighborhood sorting problem in that the principal would internalize the effect that allocating a student to c has on A_c, while parents would take A_s as given. We have not yet solved a classroom assignment problem with endogenous amenities.

A9.2 Implications for Estimation of Teacher Value Added

Suppose that the true classroom contribution to a given student i's test scores can be captured by $Z_c\Gamma + Z_{c(i),i}^U\Gamma^U$, mirroring equation (16). As before, partition the vector of observed classroom characteristics into two parts $Z_c = [X_c, Z_{2c}]$, where X_c captures classroom averages of observed student characteristics, and Z_{2c} represents other observed classroom characteristics. [52] Consider the classroom version of our estimating equation (29):

$$Y_i = X_i\beta + X_cG_1 + Z_{2c}G_2 + v_{ci}, \tag{51}$$

When past test scores are elements of X_i and a design matrix $D_{c(i)}$ indicating which classrooms were taught by which teachers is included in Z_{2c}, equation (51) represents a standard teacher value-added specification.[53]

Suppose that Proposition 1 can be extended to classroom choice setting (as proven in the exogenous amenities case) and that the corresponding spanning condition is satisfied, so that X_c and X_c^U are linearly dependent. Suppose in addition that the principal's perception of teacher quality is

[52] We assume here that teacher quality is not classroom-specific, as in most teacher value-added models.
[53] Z_{2c} might also include a set of indicators for the teacher's experience level.

noisy, so that D_c is not collinear with A_c (and therefore not collinear with X_c). Then our analysis in Section 5.3 suggests that $G_2 = \Gamma_2 + \Pi_{Z_c^U Z_{2c}} \Gamma^U$. Since Z_{2c} includes the teacher design matrix $D_{c(i)}$, we see that including classroom averages of student characteristics X_c in teacher value-added regressions will purge estimates of individual teachers' value-added from any bias from non-random student sorting on either observables or unobservables. Any remaining bias $\Pi_{Z_c^U Z_{2c}} \Gamma^U$ stems from the possible correlation between the assignment of the chosen teacher to the classroom and other aspects of the classroom environment.

However, suppose that all unobserved classroom factors that are inequitably distributed across teachers are either being used as a basis for student allocation to classrooms or are directly included as other controls in Z_c. Then the analysis in Section 5.3.1 reveals that including classroom averages of observed student characteristics will also purge teacher value-added estimates G_2 of any omitted variables bias driven by inequitable access to advantageous classroom environments (the subvector of $\Pi_{Z_c^U Z_{2c}}$ corresponding to the teacher design matrix D_c will equal 0).

Of course, our simulations suggest that the effectiveness of the control function approach depends on observing reasonably large samples of students with each teacher. And in practice there may be classroom factors ignored by students and principals that do not even out across teachers. While these caveats should be kept in mind, 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. (2014), Kane and Staiger (2008)).

Appendix Tables

Table A1: Principal Components Analysis of the Vector of School Average Observable Characteristics X_s

Panel A: Fraction of Total Variance in X_s Explained by Various Numbers of Principal Components

		NLS72		NELS88 gr8		ELS2002	
		Baseline	Full	Baseline	Full	Baseline	Full
		(1)	(2)	(3)	(4)	(5)	(6)
(1)	# of Variables in X_s	32	34	39	49	40	51
	# Factors Needed to Explain:						
(2)	75% of Total X_s Var.	7	7	7	9	6	8
(3)	90% of Total X_s Var.	12	12	13	16	11	14
(4)	95% of Total X_s Var.	15	15	17	20	14	19
(5)	99% of Total X_s Var.	20	21	22	26	20	25
(6)	100% of Total X_s Var.	23	24	27	32	26	33

Panel B: Fraction of Variance in the Regression Index $X_s\hat{G}_1$ Explained by Various Numbers of Principal Components

		NLS		NELS gr8		ELS	
		Baseline	Full	Baseline	Full	Baseline	Full
		(1)	(2)	(3)	(4)	(5)	(6)
(1)	# of Variables in X_s	32	34	39	49	40	51
	# Factors Needed to Explain:						
(2)	75% of $Var(X_sG_1)$	3	3	6	5	2	5
(3)	90% of $Var(X_sG_1)$	8	7	10	10	5	11
(4)	95% of $Var(X_sG_1)$	10	9	13	13	7	15
(5)	99% of $Var(X_sG_1)$	14	15	19	20	14	22
(6)	100% of $Var(X_sG_1)$	23	24	27	32	26	33

See Section A2 for details.

Table A2: Monte Carlo Simulation Results: Cases in which the Spanning Condition in Proposition 1 is Satisfied ($\Theta^U = R\Theta$ For Some R)

Row	# Stu.	# Sch.	# Con.	# Ob.	# Un.	# Am.	Θ 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

Θ Corr: Correlation in Θ_{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 $X_s^U \beta^U$ 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 $X_s^U \beta^U$ explained by the control function $\overline{X}_s$ (sample school averages $\overline{X}_s$ computed using 10/20/40 students)

Table A3: 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 X_i and X_i^U	Depends on all elements of X_i and X_i^U	0.972	0.368	0.501	0.640
(2)	Depends on each element of X_i, independent of all X_i^U	Depends on each element of X_i^U, independent of X_i	0.369	0.260	0.282	0.313
(3)	Depends on all elements of X_i and X_i^U	Depends on each element of X_i^U, independent of X_i	0.561	0.296	0.361	0.440
(4)	Depends on all elements of X_i and X_i^U	Depends on only on $X_{i,10}^U$	0.961	0.367	0.490	0.635
(5)	Depends on all elements of X_i and X_i^U except $X_{i,10}^U$	Depends only on $X_{i,10}^U$	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, Θ Corr = 0.25 (See Table A2 for definitions of parameters).

Table A4: 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) $Var(v_{si} - v_s)$	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 $2*Cov(X_sB, X_sG_1 + Z_{2s}G_2)$	0.0165	0.0175	0.0170	0.0187	0.175
School-Avg. Student-Level/ School Char. Covariance $2*Cov(X_sG_1, Z_{2s}G_2)$	-0.0166	-0.0047	0.0061	-0.0053	-0.0054
School-Avg. Student-Level $Var(X_sG_1)$	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 $Var(Z_{2s}G_2 + v_s)$	0.1742	0.1631	0.164	0.1959	0.1626

The column "Full NC Sample" reports variance decompositions based on the full North Carolina sample. They are the same as the estimates reported for NC sample in Appendix Table A8.

The other columns report estimates based on draws of samples of students from the North Carolina schools to match the distributions of sample sizes per school from the NLS72, NELS88 grade 8, NELS88 grade 10, or ELS2002 samples (respectively).

To remove the chatter produced by a single draw from these sampling schemes, we report averages of estimates for each of 100 samples drawn from each sampling scheme.

Table A5: 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 of X_s)*
School Pct. Minority
School Characteristics (Treated as elements of Z_{2s})
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 (Treated as elements of Z_{2s})
Urbanicity Indicators (Detailed), Indicators for U.S. Census Region

*School characteristics treated as elements of X_s 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.
School averages of all individual-level variables are also included in each specification.

Table A6: 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 of X_s)*
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 of Z_{2s})
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 (Treated as elements of Z_{2s})
Urbanicity Indicators (Urban/Suburban/Rural), Indicators for U.S. Census Region

*School characteristics treated as elements of X_s 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.
School averages of all individual-level variables are also included in each specification.

Table A7: 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 of X_s)*
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 of Z_{2s})
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 (Treated as elements of Z_{2s})
Urbanicity Indicators (Detailed), Indicators for U.S. Census Region Neighborhood Crime Level Category (Sch. Administrator Survey)

*School characteristics treated as elements of X_s 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.
School averages of all individual-level variables are also included in each specification.

Table A8: 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 (Treated as elements of Z_{2s})
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 (Treated as elements of Z_{2s})
Urbanicity Indicator Variables (12 Categories)

School averages of all individual-level variables are also included in each specification.

Table A9: Decomposition of Variance in Latent Index Determining High School Graduation from the NC, NELS88, and ELS2002 Datasets (Baseline and Full Specifications)

	NC		NELS88 gr8		ELS2002	
Fraction of Variance	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.915	0.919	0.830	0.836	0.874	0.881
$Var(Y_i - Y_s)$	(0.016)	(0.015)	(0.019)	(0.017)	(0.016)	(0.016)
Observable Student-Level (Within):	0.124	0.213	0.162	0.292	0.134	0.221
$Var((X_i - X_s)B)$	(0.005)	(0.006)	(0.010)	(0.015)	(0.035)	(0.037)
Unobservable Student-Level (Within)	0.791	0.706	0.668	0.543	0.740	0.660
$Var(v_{si} - v_s)$	(0.014)	(0.013)	(0.019)	(0.017)	(0.030)	(0.031)
Between School:						
Total	0.085	0.081	0.170	0.164	0.126	0.119
$Var(Y_s)$	(0.016)	(0.015)	(0.019)	(0.017)	(0.016)	(0.016)
Observable Student-Level:	0.018	0.033	0.073	0.109	0.037	0.060
$Var(X_s B)$	(0.002)	(0.003)	(0.009)	(0.012)	(0.005)	(0.008)
Student-Level/ School-Level Covariance	0.016	0.010	0.025	0.007	0.025	0.006
$2 * Cov(X_s B, X_s G_1 + Z_{2s} G_2)$	(0.003)	(0.005)	(0.019)	(0.019)	(0.010)	(0.011)
School-Avg. Student-Level/ School Char. Covariance	-0.017	-0.008	0.007	0.004	0.001	-0.002
$2 * Cov(X_s G_1, Z_{2s} G_2)$	(0.006)	(0.004)	(0.007)	(0.005)	(0.012)	(0.012)
School-Avg. Student-Level	0.018	0.009	0.037	0.029	0.028	0.029
$Var(X_s G_1)$	(0.005)	(0.004)	(0.013)	(0.007)	(0.014)	(0.012)
School Char.	0.018	0.012	0.011	0.006	0.025	0.024
$Var(Z_{2s} G_2)$	(0.008)	(0.004)	(0.019)	(0.004)	(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)	(0.007)	(0.003)	(0.002)	(0.001)

The table reports fractions of the total variance of the latent index that determines high school graduation.

The rows labels indicate the variance component.

Bootstrap standard errors based on resampling at the school level are in parentheses.

Appendix Sections 5 and 6 discuss estimation of model parameters and the variance decompositions.

The columns headed NC refers to a variance decomposition that uses the 9th grade school as the group variable for schools in North Carolina.

NELS88 gr8 is based on the the NELS88 sample and refers to a decomposition that uses the 8th grade school as the group variable.

ELS2002 is based on the ELS2002 sample and refers to a decomposition that uses the 10th grade school as the group variable.

For each data set the variables in the baseline model and the full model are specified in Web Appendix Tables A5 - A8

Table A10: Decomposition of Variance in Latent Index Determining Enrollment in a Four-Year College from the NLS72, NELS88, and ELS2002 Datasets (Baseline and Full Specifications)

	NLS72		NELS88 gr8		ELS2002	
Fraction of Variance	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.857	0.857	0.776	0.774	0.785	0.791
$Var(Y_{is} - Y_s)$	(0.012)	(0.012)	(0.016)	(0.017)	(0.015)	(0.014)
Observable Student-Level (Within):	0.176	0.354	0.192	0.316	0.184	0.330
$Var((X_{si} - X_s)B)$	(0.029)	(0.017)	(0.010)	(0.012)	(0.016)	(0.013)
Unobservable Student-Level (Within)	0.681	0.503	0.584	0.458	0.600	0.461
$Var(v_{si} - v_s)$	(0.027)	(0.015)	(0.015)	(0.013)	(0.016)	(0.012)
Between School:						
Total	0.143	0.143	0.224	0.226	0.215	0.209
$Var(Y_s)$	(0.012)	(0.012)	(0.016)	(0.017)	(0.015)	(0.014)
Observable Student-Level:	0.042	0.062	0.010	0.143	0.079	0.127
$Var(X_s B)$	(0.004)	(0.006)	(0.010)	(0.012)	(0.007)	(0.009)
Student-Level/ School-Level Covariance	0.037	0.032	0.057	0.027	0.071	0.039
$2 * Cov(X_s B, X_s G_1 + Z_{2s} G_2)$	(0.006)	(0.008)	(0.012)	(0.014)	(0.009)	(0.011)
School-Avg. Student-Level/ School Char. Covariance	0.000	-0.002	0.004	0.005	-.003	-0.002
$2 * Cov(X_s G_1, Z_{2s} G_2)$	(0.004)	(0.004)	(0.005)	(0.004)	(0.008)	(0.006)
School-Avg. Student-Level	0.026	0.020	0.023	0.021	0.022	0.015
$Var(X_s G_1)$	(0.006)	(0.005)	(0.005)	(0.004)	(0.006)	(0.004)
School Char.	0.026	0.019	0.018	0.015	0.024	0.018
$Var(Z_{2s} G_2)$	(0.005)	(0.004)	(0.006)	(0.005)	(0.007)	(0.006)
Unobservable School-Level	0.012	0.013	0.021	0.014	0.022	0.013
$Var(v_s)$	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.003)

The table reports fractions of the total variance of the latent index that determines enrollment in a 4-year college two years after high school graduation.

The rows labels indicate the variance component.

Bootstrap standard errors based on resampling at the school level are in parentheses.

NLS72 refers to a variance decomposition that employs NLS72 data and uses the 12th grade school as the group variable.

See the note to Table A9 for additional details.

Table A11: 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
$Var(Y_{is} - Y_s)$	(0.007)	(0.008)	(0.019)	(0.017)	(0.022)	(0.021)
Observable Student-Level (Within):	0.154	0.280	0.140	0.174	0.212	0.224
$Var((X_{si} - X_s)B)$	(0.007)	(0.007)	(0.010)	(0.009)	(0.011)	(0.013)
Unobservable Student-Level (Within)	0.749	0.624	0.697	0.660	0.617	0.605
$Var(v_{si} - v_s)$	(0.007)	(0.008)	(0.020)	(0.019)	(0.024)	(0.025)
Between School:						
Total	0.096	0.096	0.163	0.166	0.171	0.171
$Var(Y_s)$	(0.007)	(0.008)	(0.019)	(0.017)	(0.022)	(0.021)
Observable Student-Level:	0.041	0.058	0.045	0.055	0.061	0.065
$Var(X_sB)$	(0.003)	(0.004)	(0.007)	(0.008)	(0.005)	(0.005)
Student-Level/ School-Level Covariance	0.031	0.023	0.033	0.028	0.033	0.029
$2 * Cov(X_sB, X_sG_1 + Z_{2s}G_2)$	(0.003)	(0.006)	(0.009)	(0.011)	(0.008)	(0.009)
School-Avg. Student-Level/ School Char. Covariance	0.024	0.016	-0.002	0.001	-.003	0.000
$2 * Cov(X_sG_1, Z_{2s}G_2)$	(0.003)	(0.003)	(0.009)	(0.010)	(0.009)	(0.009)
School-Avg. Student-Level	0.012	0.008	0.033	0.029	0.029	0.028
$Var(X_sG_1)$	(0.003)	(0.002)	(0.013)	(0.010)	(0.012)	(0.011)
School Char.	0.017	0.010	0.039	0.041	0.039	0.040
$Var(Z_{2s}G_2)$	(0.002)	(0.002)	(0.010)	(0.010)	(0.011)	(0.011)
Unobservable School-Level	0.005	0.004	0.014	0.011	0.011	0.009
$Var(v_s)$	(0.003)	(0.002)	(0.012)	(0.009)	(0.018)	(0.017)

The table reports fractions of the total variance of years of postsecondary education, permanent wages controlling for year of post secondary education, and permanent wages not controlling for years of post secondary education.

Bootstrap standard errors based on re-sampling at the school level are in parentheses.

See the note to Table A9 for additional details.