Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects

Joseph G. Altonji
Yale University and NBER

Richard K. Mansfield*
Cornell University

May 1, 2016

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: Department of Economics and School of Industrial and Labor Relations, Cornell University, 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, and Shintaro Yamiguchi 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 also thank the Russell Sage Foundation and the Yale 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 studied whether group outcomes differ because the groups in fluence 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] Suppose that Y_{si} is determined according to[4]

$$Y_{si} = [X_i\beta + x_i^U] + [Z_s\Gamma + z_s^U]\ . \tag{1}$$

[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, which we discuss in brief below. Meghir et al. (2011) 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 to attend more distant schools. 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.

The vector X_i is a set of student and family characteristics observed by the econometrician (with corresponding productivities β), while $x_i^U \equiv X_i^U \beta^U$ is a scalar index that combines the outcome contributions of unobserved student and family characteristics X_i^U. Together, $[X_i, X_i^U]$ represent the *complete* set of student and family characteristics that have a causal impact on student i's educational attainment. Analogously, the row vector Z_s is a set of school and neighborhood characteristics observed by the econometrician (with corresponding productivities Γ), while $z_s^U \equiv Z_s^U \Gamma^U$ is a scalar index that combines the effects of unobserved school and neighborhood characteristics. Together, $[Z_s, Z_s^U]$ capture the complete set of school and neighborhood level influences common to students who live in s, so that the school/neighborhood treatment effect is given by $[Z_s\Gamma + z_s^U]$.

Sorting leads the school average of X_i^U, denoted X_s^U, to vary across s. This contaminates estimates of Γ and fixed effect estimates of the school treatment effect $Z_s\Gamma + z_s^U$. 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 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 neighborhod/school choice and providing conditions under which the mapping from X_s to X_s^U is exact. We provide further conditions (most notably an additively separable specification of utility) under which this mapping from X_s to X_s^U is *linear*. When these conditions are satisfied, including X_s in a linear regression of the outcome Y_{si} fully controls for sorting on X_s^U.

As we make precise in Proposition 1 below, X_i and X_i^U need not affect preferences for all of the amenities A_s. Partition X_i^U into a subset X_{1i}^U that is correlated X_i and a subset X_{2i}^U that is not correlated with X_s. Roughly speaking, the key requirement is that (1) X_i and/or X_{1i}^U affect preferences for all amenities that any elements of X_{2i}^U shifts preferences for, and (2) that X_i has enough elements to span this amenity space. The theoretical analysis assumes that group sizes are sufficiently large so that random variation in group choice does not affect the group averages. It also assumes that the number of groups is considerably larger than the number of amenity factors agents consider when

evaluating each group.

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 parent'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 leads 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.

While the control function approach based on the group averages X_s potentially solves the sorting-on-unobservables problem, X_s controls for too much. These observed group averages 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, without further assumptions, our estimator will only place 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 (the National Longitudinal Study of 1972 (NLS72), the National Educational Longitudinal Survey of 1988 (NELS88), and the Educational Longitudinal Survey of 2002 (ELS2002)), along with 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 starting at a school system and associated neighborhood at the 10th quantile in the distribution of school contributions instead of a 50th or 90th quantile system (a more intuitive scale).

[5] 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 W_i variables in the analysis below.

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/neighborhoods. However, the average impact of moving to a superior school on binary outcomes such as high school 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 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. It would increase the permanent component of adult wages by 19 percent (in NLS72). A one-standard deviation shift in school/neighborhood quality would raise wages by about 7 percent. Note that our estimates are derived from a static model of what is in fact a dynamic process. ??The most conservative interpretation is that our estimates represent lower bounds the cumulative effects of growing up in different school systems/neighborhoods.

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 (e.g., Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2006), among others.), labor supply functions (e.g., Metcalf (1974) and Altonji (1982)), distinguishing between uncertainty and heterogeneity in earnings (e.g., Cunha et al. (2005)), and even estimating neighborhood effects (Bayer and Ross (2009)).[6] Our application is unusual in that the control function involves group aggregates that reflect individual choices rather than relationships among different choices by the same agent.

Second, we draw on the rich theoretical and empirical literature on equilibrium sorting and matching models across several fields. Browning et al. (2014) and Chiappori and Salanie (forthcoming) provide recent surveys of the extensive literature on marriage and matching more generally. A central concern of this literature is who marries who–the sorting of marriage partners with heterogenous characteristics. A number of recent papers analyse labor market sorting based on firm and worker quality, including Lise et al. (2013) and Melo (2015). Lindenlaub (2013) presents a closed form solution to the sorting equilibrium of a labor market in which jobs differ on a continuum in

[6] Our econometric approach is only loosely related to the large literature on the use of control functions to estimate triangular systems with continuous or discrete treatment variables. In that literature, model assumptions relating to how the endogenous treatment variable and outcome of interest are determined imply that a function of the endogenous variable and an instrument or set of instruments can control for the source of endogeneity in the equation for Y. See Imbens (2007) for a survey in the context of nonadditive models, and Kasy (2011) for necessary and sufficient conditions for the existence of a control function. In our case, there is no instrument, but the sorting model implies a relationship between observable and unobservable group averages.

the skill vectors they require and workers differ on a continuum in the skill vectors they supply. The match between consumers and products (which could be locations with various characteristics) is a central concern of the hedonic demand literature, including the important contributions of Rosen (1974), Ekeland et al. (2004), and Heckman et al. (2010) among others.

Most directly relevant is the large literature on sorting across neighborhoods and schools that grew out of Tiebout (1956), particularly Epple and Platt (1998) and Epple and Sieg (1999). Epple and Platt's model features one dimension of neighborhood quality and two dimensions of heterogeneity across households–income and tastes for the public good. They show that in equilibrium the distributions of income and tastes both shift with the level of the public good in a location. This implies a mapping between income in a location and tastes in a location—the same type of mapping that we exploit. They also show that house prices are monotonic in location quality.[7] Bayer and Ross (2009) consider the implications of Epple and Platt's analysis for dealing with sorting on unobservables when estimating the effects of school and neighborhood characteristics on outcomes. They assume neighborhood quality depends on a vector of observed characteristics (Z_s in our notation) and a one dimensional unobservable. They use housing prices to construct a control function for the unobservable. They recognize that both the control function and Z_s are endogenous in the outcome equation because of sorting on X_i^U.[8] Unfortunately, the estimation scheme that they propose to address the issue is invalid in the presence of unobserved heterogeneity in location preferences and multiple unobserved location amenities.[9]

Third, the multinomial choice formulation that we use to characterize the school/location choice problem is standard in the consumer choice literature. It assumes that preferences for observed and unobserved location characteristics depend on both observed and unobserved student/parent attributes, as in McFadden et al. (1978), McFadden (1984) and Berry (1994) and many subsequent papers. Bayer et al. (2007) use a similar specification to estimate models of housing demand in which the estimation of preferences for observed and unobserved characteristics of schools and neighborhoods is a central objective. We do not estimate preferences. Our contribution is to show that the 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

[7] See Bayer and Timmins (2005) for an analysis of the equilibrium properties of a model similar to that of Epple and Platt (1998).

[8] The idea that the choice of a location, an occupation, a firm, or a school may reveal information about individuals provides motivation for the use of "fixed effects" estimation in a variety of contexts. For example, Fu and Ross (2013) use neighborhood fixed effects to control for worker heterogeneity when estimating the effect on wage rates of agglomeration at workplace locations.

[9] Our estimation strategy is closely related to the correlated random effects approach (Mundlak (1978), Chamberlain (1980), Chamberlain et al. (1984)). In that literature a function of the vector of observations on X_i from members of group s is used to control for correlation between X_i and the group error term. In many applications the mean $X_{s(i)}$ is used. However, in that literature, much of the focus is on estimating the effects of person specific variables, such as β in our application, while accounting for correlation with a common group error. In our application, the focus is on the group effect, a model of sorting provides the justification for the use of X_s as a control, and β is not identified. Despite the similarity in titles, our analysis is also 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.

averages of unobservables in the estimation of group treatment effects.

The empirical part of the paper adds to a vast literature on school and neighborhood effects that we cannot do justice to here.[10] Our analysis of sorting is directly relevant to the large number that use regression models of the form of (1). A few 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 student outcomes. Similarly, the Moving To Opportunity (MTO) 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 find much evidence that moving to a low-poverty neighborhood improves economic outcomes. However, Chetty et al. (2015) revisit the MTO experiment using Internal Revenue Service data on later outcomes, including earnings, college attendance, and single parenthood. Their treatment-on-the-treated estimates indicate that children who move to a lower poverty neighborhood when they are under age 13 experience large gains in annual income in their mid-twenties, while those who move after age 13 experience no gain or a loss. Their estimates of treatment effects on adult earnings also increase with the number of years of exposure to a lower poverty neighborhood.[11] Using a sibling differences approach that also exploits high quality data from tax records, Chetty and Hendren (In Progress) identify county level neighborhood effects on earnings that are larger than but qualitatively consistent with our results. Aaronson (1998) finds substantial effects of the census track level poverty rate and high school dropout rate on dropout rates and years of education using a sibling differences design and PSID data.[12]

Deming et al. (2014), in 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

[10] Jencks and Mayer (1990) provide a comprehensive review of earlier studies from in economics and sociology. They conclude that there is no strong evidence for neighborhood effects. However, some of the studies they summarize do find effects. More recent reviews include Sampson et al. (2002), Durlauf (2004), and Harding et al. (2011). Duncan and Murnane (2011) contains several recent papers on school and neighborhood effects, with references to the literature. Meghir et al. (2011) 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.The papers we discuss in the text provide references to many other recent contributions to the literature on school and neighborhood effects.

[11] Aliprantis (2011) stresses the limitations of the MTO study for uncovering the full distribution of school system and neighborhood effects on children. Aliprantis and Richter (2013) focus on the adults in MTO. They combine experimental variation in the costs of moving to neighborhoods of different poverty levels with an ordered discrete choice model of the decision to move. They find that neighborhood quality has a substantial effect on the outcomes of adults. They also conclude that the overall effect of the experiment on adult economic outcomes was small because the induced improvements in neighborhood quality were small for most.

[12] In contrast to Aaronson (1998), Plotnick and Hoffman (1995) do not find neighborhood effects on postsecondary education using a sibling difference design with PSID data on sister pairs. Aaronson provides evidence that Plotnick and Hoffman's choice of sample and neighborhood quality measures leads to weaker results. He also finds that neighborhood effects are smaller for postsecondary education than for high school graduation.

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 briefly discusses other applications of our methodology, including the assessment of 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.

Each location $s \in \{1,...,S\}$ can be characterized by a vector of K underlying latent amenities $A_s \equiv [A_{1s},\ldots,A_{Ks}]'$.[13]

We adopt a money-metric representation of the expected utility the parents of student i receive from choosing school/neighborhood s, so that the utility function $U_i(s)$ can be interpreted as the family's consumer surplus from their choice. We assume $U_i(s)$ takes the following linear form:

$$U_i(s) = \Upsilon_i A_s + \varepsilon_{si} - P_s. \tag{2}$$

In the above equation $\Upsilon_i \equiv [\Upsilon_{1i},\ldots,\Upsilon_{Ki}]$ 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 ε_{si} 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 elements, while X_i^U has L^U elements. Then

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

we obtain:

$$\Upsilon_i = X_i\Theta + X_i^U\Theta^U + W_i\,, \tag{3}$$

where Θ (Θ^U) is an $L \times K$ ($L^U \times K$) matrix whose ℓk-th entry captures the extent to which the willingness to pay for the k-th element of the amenity vector A_s varies with the ℓ-th element of X_i (X_i^U). We sometimes refer to the elements of Θ and Θ^U as WTP slopes or WTP coefficients. The $1 \times K$ vector W_i captures the components of i's taste for the K amenities in A_s that are uncorrelated with $[X_i, X_i^U]$. Since $[X_i, X_i^U]$ is the complete set of student attributes that determine Y_{si}, the elements of W_i influence school choice but have no direct effect on student outcomes.

Substituting equation (3) into equation (2), we obtain:

$$U_i(s) = (X_i\Theta + X_i^U\Theta^U + W_i)A_s + \varepsilon_{si} - P_s \tag{4}$$

In the absence of restrictions on the elements of Θ and Θ^U, this formulation of utility allows for a very general pattern of relationships between different student characteristics (observable or unobservable) and tastes for different school/neighborhood amenities, subject to the additive separability assumed in (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 $[X_i, X_i^U, W_i]$ and the values of ε_{si}, $s = 1, ..., S$. 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. The amenities may also include aspects of the demographic composition of the school/neighborhood. Some (such as spending per pupil) may be influenced by demographic composition. Thus, some of the amenities are outcomes of the sorting equilibrium.

The parents of student i choose the school s if net utility $U_i(s)$ is the highest among the S options. That is,

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

Parents behave competitively in the sense that prices and A_s are taken as given, and choice is unrestricted. In equilibrium the values of some elements of A_s may in fact depend on the averages of X_i and X_i^U for the parents who choose s, but parents ignore the externalities that they are imposing on others.

3 The Link Between Group Observables and Group Unobservables

In Section 3.1 we state and prove Proposition 1, which concerns the relationship between X_s^U and X_s implied by the above choice model. In Section 3.2 we discuss the proposition and the

assumptions that underlie it.

3.1 Proposition 1: X_s^U Is a Linear Function of X_s

Before stating Proposition 1, we need to define more notation. Decompose X_i^U into its projection on X_i and the orthogonal component $\tilde{X}_i^U$:[14]

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

Use (5) to rewrite (3) as $\Upsilon_i = X_i\tilde{\Theta} + \tilde{X}_i^U \Theta^U + W_i$, where $\tilde{\Theta} = [\Theta + \Pi_{X^U X}\Theta^U]$. In the rewritten form, all three components of Υ_i are mutually orthogonal. We are now prepared to present the main proposition of the paper.

Proposition 1: *Assume the following assumptions hold:*

A1: *Preferences are given by (4).*

A2: *Parents take $P(A_s)$ and A_s as given when choosing location, and face a common choice set.*

A3: *The idiosyncratic preference components ε_{si} have a mean of* 0 *and are independent of X_i, X_i^U, W_i* , and *A_s for all s.*

A4: *$E(X_i|\Upsilon_i)$ and $E(X_i^U|\Upsilon_i)$ are linear in Υ_i.*

A5: (Spanning Assumption) *The row space of the WTP coefficient matrix $\tilde{\Theta}$ spans the row space of the WTP coefficient matrix Θ^U relating tastes for A* to X_i^U . That is,

$$\Theta^U = R\tilde{\Theta} \tag{6}$$

for some $L^U \times L$ matrix R.

Then the expectation X_s^U is linearly dependent on the expectation X_s. Specifically,

$$X_s^U = X_s[\Pi_{X^U X} + Var(X_i)^{-1} R' Var(\tilde{X}_i^U)] \tag{7}$$

3.1.1 Proof of Proposition 1:

Equation (2) states that the utility of each location s depends on X_i, X_i^U, W_i only through Υ_i. This fact and independence of ε_{si} from X_i, X_i^U, and W_i, imply that

$$\Pr(s(i) = s|X_i, X_i^U, W_i, \Upsilon_i) = \Pr(s(i) = s|\Upsilon_i) \tag{8}$$

[14] We use the symbol Π_{DQ} to denote the vector or matrix 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 $Q = X_i$.

where $\Pr(.)$ is the probability function. The above fact and Bayes rule imply that[15]

$$f(X_i|\Upsilon_i, s(i)=s) = f(X_i|\Upsilon_i) \tag{9}$$
$$f(X_i^U|\Upsilon_i, s(i)=s) = f(X_i^U|\Upsilon_i)\ . \tag{10}$$

These equations then imply that $E[X_i|\Upsilon_i, s(i)=s] = E[X_i|\Upsilon_i]$ and $E[X_i^U|\Upsilon_i, s(i)=s] = E[X_i^U|\Upsilon_i]$. Consequently, using the Law of Iterated Expectations, we have:

$$X_s^U \equiv E[X_i^U|s(i)=s] = E[E(X_i^U|\Upsilon_i, s(i)=s)|s(i)=s] = E[E(X_i^U|\Upsilon_i)|s(i)=s] \tag{11}$$
$$X_s \equiv E[X_i|s(i)=s] = E[E(X_i|\Upsilon_i, s(i)=s)|s(i)=s] = E[E(X_i|\Upsilon_i)|s(i)=s]. \tag{12}$$

Next we find expressions for $E[X_i^U|\Upsilon_i]$ and $E[X_i|\Upsilon_i]$, which appear in the above equations. Since by construction $\tilde{X}_i^U$ is uncorrelated with X_i, and W_i is uncorrelated with both X_i and $\tilde{X}_i^U$,

$$Cov(\Upsilon_i', \tilde{X}_i^U) = Cov(\Theta^{U'}\tilde{X}_i^U, \tilde{X}_i^U) = \Theta^{U'}Var(\tilde{X}_i^U) \tag{13}$$
$$Cov(\Upsilon_i', X_i) = Cov(\tilde{\Theta}'X_i, X_i) = \tilde{\Theta}'Var(X_i) \tag{14}$$

Since from assumption A4 $E[X_i|\Upsilon_i]$ and $E[X_i^U|\Upsilon_i]$ are linear in Υ_i, $E[\tilde{X}_i^U|\Upsilon_i]$ is also linear in Υ_i. Consequently, assumption A4, equations (13)-(14), and basic regression theory imply that

$$E[\tilde{X}_i^U|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1}Cov(\Upsilon_i', \tilde{X}_i^U) = \Upsilon_i Var(\Upsilon_i)^{-1}\Theta^{U'}Var(\tilde{X}_i^U) \tag{15}$$
$$E[X_i|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1}Cov(\Upsilon_i', X_i) = \Upsilon_i Var(\Upsilon_i)^{-1}\tilde{\Theta}'Var(X_i). \tag{16}$$

Next, if we use the spanning assumption A5 to replace $\Theta^{U'}$ with $\tilde{\Theta}'R'$ in (15), and then use the expression for $E[X_i|\Upsilon_i]$ from 48, we obtain:

$$\begin{aligned} E[\tilde{X}_i^U|\Upsilon_i] &= \Upsilon_i Var(\Upsilon_i)^{-1}\tilde{\Theta}'R'Var(\tilde{X}_i^U)) \\ &= \Upsilon_i Var(\Upsilon_i)^{-1}\tilde{\Theta}'Var(X_i)Var(X_i)^{-1}R'Var(\tilde{X}_i^U)) \\ &= E[X_i|\Upsilon_i]Var(X_i)^{-1}R'Var(\tilde{X}_i^U). \end{aligned} \tag{17}$$

To find $E[X_i^U|\Upsilon_i]$ first take expectations of both sides of (5) conditional on Υ_i:

$$E[X_i^U|\Upsilon_i] = E[X_i|\Upsilon_i]\Pi_{X^UX} + E[\tilde{X}_i^U|\Upsilon_i]. \tag{18}$$

[15] One can write the conditional density $f(X_i|\Upsilon_i, s_i=s)$ as

$$\begin{aligned} f(X_i|\Upsilon_i, s_i=s) &= \frac{\Pr(s(i)=s|X_i,\Upsilon_i)f(\Upsilon_i|X_i)}{\Pr(s(i)=s|\Upsilon_i)f(\Upsilon_i)}f(X_i) \\ &= \frac{\Pr(s(i)=s|\Upsilon_i)f(\Upsilon_i|X_i)}{\Pr(s(i)=s|\Upsilon_i)f(\Upsilon_i)}f(X_i) \\ &= f(X_i|\Upsilon_i) \end{aligned}$$

where the first equality is Bayes rule, the second equality uses (8), and the third follows from cancellation of terms and Bayes rule. The same line of argument establishes that $f(X_i^U|\Upsilon_i, s(i)=s) = f(X_i^U|\Upsilon_i)$.

Substitution for $E[\tilde{X}_i^U|\Upsilon_i]$ using (17) leads to

$$E[X_i^U|\Upsilon_i] = E[X_i|\Upsilon_i](\Pi_{X^UX} + Var(X_i)^{-1}R'Var(\tilde{X}_i^U)). \tag{19}$$

The final step is to take expectations of both sides of the above equation conditional on $s(i) = s$ and employ equations (11) and (12). Doing so leads to

$$X_s^U = X_s[\Pi_{X^UX} + Var(X_i)^{-1}R'Var(\tilde{X}_i^U)].$$

This completes the proof.

3.2 Discussion of Proposition 1

Proposition 1 lays out the conditions under which X_s^U, the between group component of the vector of individual-level unobservables, will be an exact linear function of its observable counterpart X_s.[16] Remarkably, the dependence between the group averages X_s^U and X_s arises even when the vector X_i^U is uncorrelated with the vector X_i at the individual level. Note also that if unobservable characteristics do not affect amenity preferences (i.e. individuals do not sort based on unobservables), so that $\Theta^U = 0$, then $R = 0$. When $R = 0$, (7) states that $X_s^U = X_s\Pi_{X^UX}$ and $\tilde{X}_s^U = 0$. As we discuss in Section 8.5, this fact means that if sorting is driven by X_i but not X_i^U, one can estimate the variance in group treatment effects $Var(Z_s\Gamma + z_s^U)$.

Note that Proposition 1 is a statement about the expectations X_s and X_s^U. Thus, it concerns the averages of X_{si} and X_{si}^U when the number of individuals is large relative to the number of choices. With a finite number of individuals per group, random variation associated with W_i and ε_{si} will cause group averages at a point in time to deviate from their expectations. This could weaken the link between group averages of observable and unobservable characteristics. Monte Carlo simulations in Section 4 indicate that the procedure works fairly well with samples of 20-40 individuals per group. Note also that implementation of X_s as a control function for X_s^U requires that the number of groups in the sample must be larger than the number of elements in X_s (and implicitly the number of factors in A_s). Otherwise, one cannot estimate the coefficient vector on X_s.

The next two subsections discuss the assumptions underlying Proposition 1.

3.2.1 Discussion of Assumptions A1-A4

Assumption A1 about preferences is fairly general given that both X_i and X_i^U can include non-linear terms.

Assumption A2 simply says that households take characteristics of neighborhoods as given. As

[16] In Altonji and Mansfield (2014), we 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. Perhaps surprisingly, equation (7) in Proposition 1 holds for the continuous case.

we mentioned above, this is fully consistent with the possibility that A_s depends on who chooses s in equilibrium. If some of the neighborhood amenities are functions of resident characteristics, the distribution of amenities will be endogenous. There might be multiple equilibria. However, Proposition 1 follows entirely from utility maximization. The linear dependence between X_s and X_s^U will hold in any equilibrium of the model.

Assumption A2 also imposes that households face a common set of choices. In the next section we discuss monte carlo simulations that demonstrate that our control function also works well when different households face choice sets that are overlapping subsets of the full set of schools.

Furthermore, it is a statement about the expectations X_s and X_s^U. Thus, it concerns the averages of X_{si} and X_{si}^U when the number of households is large relative to the number of choices. With a finite number of students per school, random variation associated with W_i and ε_{si} will cause school averages at a point in time to deviate from their expectations. This could weaken the link between school averages of observable and unobservable characteristics.

The independence assumption A3 seems minor given that ε_{si} can be defined to be uncorrelated with X_i, X_i^U and W_i without loss of generality. A sufficient condition for the linearity in expectations assumption A4 to hold is that the joint distribution belongs to the continuous elliptical class. Examples include the multivariate normal, the multivariate t, the Laplace, and the multivariate exponential power family.[17] However, in our application X_i contains a number of discrete variables, so this sufficient condition will not be satisfied.

Proposition 2 in online Appendix A3 establishes that if A4 fails, then an approximation error term appears in equation (7) for X_s^U. The approximation error consists of the average for s of a linear function of the difference between $E(X_i|\Upsilon_i)$ and $E(X_i^U|\Upsilon_i)$ (respectively) and the best least square linear predictions of X_i and X_i^U given Υ_i. As we discuss in Section 6.1, this could lead to upward bias in the less conservative of our two estimators of the variance of school/neighborhood effects. Note, though, that because X_s^U appears in the outcome equation through the index $X_s^U\beta^U$, any upward bias depends on a weighted index of the approximation error terms for each element of X_s^U, with the elements of β^U as the weights. This may lead to some cancellation of the approximation errors.

We now turn to the spanning assumption A5.

3.2.2 When Will the Spanning Assumption A5 Hold?

The key restriction on preferences in Proposition 1 is the spanning assumption (A5). It requires the coefficient vectors Θ^U relating tastes for amenities to the elements of X_i^U to be linear combinations of the coefficient vectors $\tilde{\Theta}$ relating tastes for amenities to the observables X_i and/or elements of X_i^U that are correlated with X_i. Given the importance and subtlety of this spanning condition, we further develop the intuition underlying the condition and highlight cases in which it fails to hold.

Reconsider the more general function formulation used in the introduction. Let $A^X \subseteq A$ represent

[17] Elliptical continuous distributions have density functions that are constant over ellipsoids. Gómez et al. (2003) survey some of the properties of these distributions.

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_i^U correlated with X_i. Likewise, $A^{X^U} \subseteq A$ represents the subset of amenities that affect the distribution of unobservable school averages X_s^U. 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, $X_s^U = f^U(A^{X^U})$. We can write $X_s^U = g(X_s)$ if we can write $X_s^U = f^U(f^{-1}(X_s))$, where $g(X_s) = f^U(f^{-1}(X_s))$. Thus, jointly sufficient conditions are

Assumption A5.1: *f is invertible, so that we can write $A^X = f^{-1}(X_s)$*

Assumption A5.2: *$A^{X^U} \subseteq A^X$, so that the amenity space that X_s spans is the relevant amenity space that drives the variation in X_s^U (i.e. the range of f^{-1} must encompass the domain of f^U).*

While these conditions are not necessary, they suggest two fundamental ways that the spanning condition $\Theta^U = R\tilde{\Theta}$ can fail.[18] The first way, which leads A5.1 to fail, is that the vector X_i may affect tastes for more amenities than its own number of elements. That is $\dim(A^X) > L$ where $\dim(A^X)$ is the number of elements in A^X. In this case, the function $f(*)$ is not invertible.[19] In the case of the additively separable utility function from (4), $\dim(A^X)$ is equal to 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 income, and the amenity space consisted of two imperfectly correlated factors: schools' quality of teachers and quality of athletic facilities. 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.

The validity of A5.1 depends on the number and breadth of coverage of variables in X_i. It 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 $\dim(A^X) \leq L$. 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 $\dim(A^X) > L$. The evidence presented in Section 8.6 and online Appendix A2 is fully consistent with $\dim(A^X) < L$ in our application.

What about Assumption 5.2? Partition X_i^U into a subset X_{1i}^U that is correlated with X_i and a subset X_{2i}^U that is not correlated with X_i. Assumption 5.2 will fail if X_{2i}^U affects preferences for an

[18] Invertibility of $f(A^X)$ is not a necessary condition. It is possible that $X_s = f(A^X)$ is one-to-many, meaning that the same value of A^X leads to multiple values of X_s. In this case the key is that one can still write $A^X = h(X_s)$, where $h(.) = f^{-1}(.)$ in the one-to-one case. The mapping from A^{X^U} to X_s^U need not be one-to-one either. However, there must be a mapping $X_s^U = f(A^{X^U}, X_s) = f(h(X_s), X_s) = g(X_s)$ that is one-to-one or one-to-many.

[19] More specifically, what is relevant for invertibility is not the number of elements of X_i (denoted L) per se but the number of independent taste factors that these L 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.

amenity that neither X_i nor X_{1i}^U affect preferences for.[20]

Revisiting one of the examples from the introduction helps illustrate how the assumption can fail. In that example parental education is the only observable and student athleticism is the only unobservable. Parental education and student athleticism are assumed to be uncorrelated, so student athleticism is an X_{2i}^U variable rather than an X_{1i}^U. Furthermore, parental education does not affect WTP for athletic facilities in the neighborhood, while student athleticism does. Athletic facility quality is an element of A^{X^U} but not A^X, so that $A^{X^U} \not\subset A^X$. Assumption 5.2 would fail. Consequently, variation in athletic facility quality would drive between-neighborhood variation in average student athleticism that average parental education would not capture. Online Appendix A1 goes through further examples that illustrate when the spanning condition will and will not be satisfied.

Assumption A5.2 is a statement about unobservables and thus is not testable without more structure than we impose. But one can assess the assumption through the following thought process. First, draw on the literature to identify the factors, both observed and unobserved, that are most important for the outcome. Next consider each unobserved variable and ask whether it is likely to be uncorrelated with all of the observed variables. Also ask whether it is likely to be the only determinant of WTP for some amenity that influences location choice. If the answer to both questions is "no" for all of the elements of X_i^U, then Assumption A5.2 is plausible.

This line of reasoning leads us to believe that A5.2 is plausible in an application such as ours in which X_i contains a rich and diverse set of variables that are likely to matter for student outcomes. Consider, for example, the priority that a child's parents and broader family places on academic learning and educational attainment. One would expect this unobservable to boost willingess to pay for peer groups and community and school characteristics that foster achievement, such as enrichment programs. However, parents' education (observed in all 4 data sets), parents' desired years of education, parental school involvement (observed in ELS2002 and NELS:88), and grandparents' education (observed in ELS:2002) are likely to be correlated with the priority parents place on education. They are also likely to directly affect willingness to pay for a similar set of education-related school and neighborhood characteristics. To take another example, taste for/proficiency in music may affect academic performance and influence willingness to pay for schools and communities with good music programs and music venues. But parental education and parental income are likely to be correlated with a child's proficiency in music (through home investments). They also may influence WTP for opportunities in music. One can make similar arguments about other unobservables (e.g. wealth (unobserved) vs. income (observed).[21]

[20] More mathematically, the unobservable vector X_i^U affects WTP for certain amenities that no element in X_i predicts WTP for, so that $A^{X^U} \not\subset A^X$. In the case of the additively separable utility function from equation (4), $A^{X^U} \subseteq A^X$ 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 $\tilde{\Theta}_{lk}$.

[21] One could in principle observe school averages W_s of the individual variables W_i that influence location preferences but do not influence outcomes. The control function variables should include not only X_s but also W_s if X_s alone is inadequate for the spanning condition to hold. Proposition 1 can easily be modified to account for the presence of W_s. We

Online Appendix A4 derives an analytical formula for the component of X_s^U that cannot be predicted by X_s when the spanning assumption is violated (and thus may be a source of bias in our lower bound estimates of the variance in school/neighborhood treatment effects). The variance in this component depends on the following five factors: a) the joint distribution of amenities; b) the joint distribution of the WTP index Υ_i; c) the matrix Θ^U mapping unobserved individual characteristics into willingness to pay for particular amenities; d) the joint distribution of the residual component of unobserved outcome-relevant student characteristics $\tilde{X}_i^U$ and e) the joint distribution of the unobserved outcome-irrelevant (but school choice-relevant) student characteristics W_i.

Given the complicated manner in which each of these five factors enters the expression for the unexplained component of X_s^U, there does not appear to be any straightforward way to place a bound on the variance in this error component.

4 Monte Carlo Evidence on the Performance of X_s as a Control Function

In this section we present monte carlo simulation results that examine the properties of our control function approach across a number of key dimensions. We start by examining how well X_s controls for X_s^U with finite samples of students per school and when choice sets of parents differ. In the initial designs the spanning condition (A5) is satisfied. We then turn to simulations in which the spanning condition fails. A full description of our simulation methodology and results is contained in online Appendix A5 and online Appendix Tables A8 and A9. Here we provide a brief summary.

We do not attempt to fully characterize the performance of our estimator.[22] Instead, our simulations center around a stylized test case that is calibrated to represent a plausible description of the school/neighborhood choice context.

The first key result is that the control function can work extremely well even in settings where 1) there is only a moderate number of groups to join, and 2) only a subset of these are considered by any given individual. In all of our simulations in such settings at least 99.5% of the variance in group-average values of the unobservable index x_s^U is absorbed by controlling for X_s ($R^2 = .995$). In all cases the residual variance in x_s^U not accounted for by X_s is negligible – less than .08% of the individual-level outcome variance $Var(Y_i)$. This is true even though the designs we consider feature very strong sorting on unobservables: x_s^U accounts for between 10% and 14% of $Var(Y_i)$ in all but one case.

The second key result is that the control function also works well even when group-averages of

do not use W_s variables in our empirical analysis. Note also that if one observed an element of A^{X^U} that drives sorting on X_{2i}^U, one could also include this observed amenity as part of the control function.

[22] A full characterization is a daunting task given the large number of parameters that determine the full spatial equilibrium sorting of students to schools. The parameters include those characterizing the joint distribution of the individual characteristics affecting choice $[X_i, X_i^U, W_i]$, 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.

the observables X_s are constructed using small samples of group members rather than the full school population. The R^2 exceeds 0.93 in all but one case. And despite the large fraction of the variance in outcomes that is due to sorting in all the designs we consider, the unexplained sorting variance is between 0.4% and 0.8% of the outcome variance even when samples of 20 are used to construct X_s. Given our reliance on such small samples in the three panel survey datasets used in the empirical analysis below, we revisit the issue in Section 7 and online Appendix A8. There we use the North Carolina administrative data to directly assess the effect of using smaller samples of students to construct X_s for some of the outcomes and characteristics we actually consider Our main results are relatively insensitive to restricting school sample sizes to match the distribution of sample sizes observed in the NLS72, NELS88, and ELS2002 datasets.

The third result is that the control function approach is quite robust to 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 elements of X_i. This is arguably the most plausible case when rich data on students and parents are available.

5 The Econometric Model of Educational Attainment and Wage Rates

We start by elaborating on the underlying model of student outcomes presented in the introduction. In Section 5.2, we show how sorting and omitted school neighborhood characteristics affects estimates of neighborhood/school effects based on OLS estimation of that model. In Section 5.3, we show that the OLS estimates in combination with Proposition 1 are sufficient to place a lower bound on the variance of school and neighborhood effects given the production function (20) below.

5.1 The Model of Outcomes

In our application the outcomes are high school graduation, attendance at a four-year college, a measure of years of postsecondary education, and the permanent wage rate. The outcome Y_{si} of student i whose family has chosen the school and surrounding neighborhood s is determined according to

$$Y_{si} = X_i\beta + x_i^U + Z_s\Gamma + z_s^U + \eta_{si} + \xi_{si} \; . \tag{20}$$

For binary outcomes such as college attendance, Y_{si} is the latent variable that determines attendance. As discussed above, the student's outcome contribution can be summarized by the index $(X_i\beta + x_i^U)$, where $x_i^U \equiv X_i^U\beta^U$ is a scalar index summarizing the contributions of unobserved student characteristics X_i^U, and 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. Since X_i and X_i^U may include non-linear functions, the linear in parameters specification for Y_{is} is without much loss of generality.

Analogously, the average school/neighborhood outcome contribution is captured by the index $Z_s\Gamma + z_s^U$ where $z_s^U \equiv Z_s^U\Gamma^U$ is a scalar index summarizing the contributions of unobserved school

and neighborhood characteristics. The vector 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^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). Note that Z_s and Z_s^U may include averages of X_i and X_i^U, respectively, which capture peer effects.

The unobserved scalar index η_{si} captures variation in school/neighborhood contributions across students 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 of the factors that determine η_{si} may represent the within-school components of Z_s.

The component ξ_{si} 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, z_s^U and η_{si}. 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 ξ_{si} as $\xi_s + \xi_i$, where ξ_s is common to all students at school s and ξ_i is idiosyncratic. ξ_s is 0 for high school graduation. More generally, the productivity parameters β and Γ and the indices x_i^U, z_s^U, η_{si} and ξ_{si} depend implicitly upon the specific outcome under consideration as well as the time period in the case of wages.

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).[23]. 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^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, note that if a randomly selected student attended school s^1 rather than s^0, the expected difference in his/her outcome would be $(Z_{s^1}\Gamma + z_{s^1}^U) - (Z_{s^0}\Gamma + z_{s^0}^U)$. We wish to quantify differences across schools/neighborhoods in $Z_s\Gamma + z_s^U$. In the case of college attendance and permanent wage rates, the difference in expected outcomes will also reflect the difference between ξ_{s^1} and ξ_{s^0}, which are common to those who attend s^1 or s^0 but are determined after high school is completed.[24]

One could generalize the above model for Y_{si} to allow the effects of school characteristics to depend on individual attributes by adding interactions of Z_s and/or z_s^U with individual attributes

[23] See Todd and Wolpin (2003) and Cunha et al. (2006)

[24] The outcomes of a specific student i will also differ across schools/neighborhoods because the values of the idiosyncratic terms η_{si} will differ.

X_i and/or X_i^U. Indeed, the preference weights on amenities that represent school characteristics depend on X_i and X_i^U in the choice model, as would be the case if parents choose locations with the match to their child's needs in mind. Allowing for non-separability in outcome model does not break the linear relationship between X_s and X_s^U. However, it would imply that the distribution of school treatment effects varies with X_i and X_i^U. We discuss the issues involved in footnote 32 while describing our empirical methodology, but focus on the homogenous effects case in this paper.

5.2 The Bias in OLS Estimates of School Effects

In this section we discuss the slope parameters and error components that OLS recovers when outcomes are regressed on only the observed student-level and school-level variables, X_i and Z_s.

To facilitate the analysis, first partition Z_s into $[X_s, Z_{2s}]$, where X_s consists 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. Section 7.3 provides a discussion of which variables should be included in X_s and Z_{2s}, respectively.

Next, project the index of unobserved school inputs z_s^U onto X_s and Z_{2s}:

$$z_s^U = X_s \Pi_{z_s^U, X_s} + Z_{2s} \Pi_{z_s^U, Z_{2s}} + \widetilde{z_s^U}. \tag{21}$$

Similarly, project η_{si} on the student level variables:

$$\eta_{si} = X_i \Pi_{\eta_{si}, X_i} + \tilde{\eta}_{si}. \tag{22}$$

Next, in order to more clearly demonstrate the impact of student sorting as separate from simple omitted variables bias, we project $x_i^U \equiv X_i^U \beta^U$ onto the space of observable variables in two steps. First, we regress x_i^U on the student-level observable vector X_i only:

$$x_i^U = X_i \Pi_{x_i^U X_i} + \tilde{x}_i^U \ . \tag{23}$$

The coefficient matrix $\Pi_{x_i^U X_i}$ captures the relationship in the full population between the unobserved student-level contribution to Y_i and observed student-level characteristics. It contributes to standard omitted variables bias in estimation of the coefficient vector on X_i even in the absence of non-random student sorting to schools. In the second step, we project the residual from the first-step, $\tilde{x}_i^U$, onto both the student-level and school-level vectors of observables (X_i and Z_s):

$$\tilde{x}_i^U = X_i \Pi_{\tilde{x}_i^U X_i} + X_s \Pi_{\tilde{x}_i^U X_s} + Z_{2s} \Pi_{\tilde{x}_i^U Z_{2s}} + \varepsilon_{si}^{\tilde{x}^U} \ , \tag{24}$$

where $\varepsilon_{si}^{\tilde{x}^U}$ is an error component. 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 given the regression equation (23) 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 may be revealing high residual taste for education. This unobserved characteristic might also improve their kids' outcomes regardless of school, thus belonging in X_i^U and contributing to $\tilde{x}_i^U$.

Substituting the projections (21), (22), (23), and (24) for Z_{si}^U, x_i^U, and η_{si} into (20), we obtain:

$$Y_{si} = X_i B + X_s G_1 + Z_{2s} G_2 + v_s + (v_{si} - v_s), \quad \textit{where} \tag{25}$$

$$B \equiv [\beta + \Pi_{\eta_{si}^U X_i} + (\Pi_{x_i^U X_i} + \Pi_{\tilde{x}_i^U X_i})] \tag{26}$$

$$G_1 \equiv [\Gamma_1 + \Pi_{z_s^U X_s} + \Pi_{\tilde{x}_i^U X_s}] \tag{27}$$

$$G_2 \equiv [\Gamma_2 + \Pi_{z_s^U Z_{2s}} + \Pi_{\tilde{x}_i^U Z_{2s}}] \tag{28}$$

$$v_s \equiv \widetilde{z_s^U} + \varepsilon_s^{\tilde{x}^U} + \xi_s \tag{29}$$

$$v_{si} - v_s \equiv \xi_i + (\varepsilon_{si}^{\tilde{x}_i^U} - \varepsilon_s^{\tilde{x}_i^U}) + \tilde{\eta}_{si} \tag{30}$$

The expressions for G_1, G_2 and v_s in (27), (28) and (29) 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 $\varepsilon_s^{\tilde{x}_i^U}$, which captures the part of the average unobservable student contribution that is not related to observed school-level characteristics or average student-level characteristics. The terms $X_s \Pi_{\tilde{x}_i^U X_s}$, $Z_{2s} \Pi_{\tilde{x}_i^U Z_{2s}}$ and $\varepsilon_s^{\tilde{x}_i^U}$ capture sorting. They are not school/neighborhood effects, since a child who was reallocated to a school with a higher value of these components could not expect an increase in test scores[25]. Without further assumptions about how students sort into schools, regression and variance decomposition techniques cannot be used to identify or even bound the contribution of schools/neighborhoods to student outcomes. In the next section, we show that the assumptions laid out in Proposition 1 are sufficient to place a lower bound on the variance of school and neighborhood effects given the production function (20) above.

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

Section 3 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. We now show that the relationship between X_s and X_s^U implies restrictions on G_2 and v_s that allow the recovery of

[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.

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. Equations (5) and (7) from Proposition 1 and (24) together reveal that $\Pi_{\tilde{x_i^U} X_s} = Var(X_i)^{-1} R' Var(\tilde{X}_i^U)\beta^U$, $\Pi_{\tilde{x_i^U} Z_{2s}} = 0$, and $\varepsilon_s^{\tilde{X}} = 0$.[26]

Thus,

$$G_2 \equiv \Gamma_2 + \Pi_{z_s^U Z_{2s}} \tag{31}$$

$$v_s \equiv \widetilde{z_s^U} + \xi_s \, . \tag{32}$$

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 $Var(Z_{2s}G_2)$ and $Var(v_s)$ 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. Since this component will be absorbed by $X_s\hat{G}_1$, school/neighborhood peer effects associated with X_s^U also will be excluded from the estimate of school/neighborhood effects.

Third, (27) reveals that X_s will also absorb part of the unobserved school contribution z_s^U via $\Pi_{z_s^U X_s}$. To see why, note that X_s spans the space of X_s^U *because* the amenity vector, A_s, is the source of variation in both X_s and X_s^U. Given that parents are likely to value the contributions of schools to student outcomes, many of the characteristics contributing to z_s^U that affect school quality are likely to be reflected in A_s. Hence, while the inclusion of X_s in the estimated specification 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.[27]

On the other hand, components of $Z_{2s}\Gamma_2 + z_s^U$ that are either not valued or not fully known (or

[26] Post multiplying both sides of (5) by β^U and taking expectations conditional on $s(i) = s$ establishes that

$$x_s^U = X_s \Pi_{X^U X} \beta^U + \tilde{x}_s^U .$$

This fact and (7) from Proposition 1 (after multiplying both sides by β^U) implies that $\tilde{x}_s^U = X_s Var(X_i)^{-1} R' Var(\tilde{X}_i^U)\beta^U$. Comparing this result with the equation for $\tilde{X_s^U}$ implied by taking expectations of both sides of (24) conditional on $s(i) = s$ establishes the claims in the text.

[27] Nor can we estimate the effect of a school level variable in the unlikely event that it is perfectly determined by X_s through the political process.

knowable) by parents at the time the school/neighborhood is chosen will not be elements of A_s, although they may be correlated with A_s. 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 $\Pi_{z_s^U Z_{2s}}$ in the expression for G_2 in (31) 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 index z_s^U. Thus, estimating the vector of causal effects Γ_2 associated with the school characteristics Z_{2s} 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 every unobserved school characteristic that contributes to the index z_s^U and is correlated with Z_{2s} is either an amenity considered by individuals at the time of choice or is perfectly predicted by the vector of amenities. Furthermore, suppose the spanning assumption is satisfied so that A_s is a function of X_s. This implies that X_s also perfectly determines the part of z_s^U that is correlated with Z_{2s}. In this case, the residual variation in z_s^U will be orthogonal to Z_{2s}. As a result, $\Pi_{z_s^U Z_{2s}} = 0$, and $\hat{G}_2$ will be an unbiased estimator of Γ_2.

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 $\hat{G}_2$.[28] 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 Mechanics of Measuring School and Neighborhood Effects

6.1 Variance Decomposition

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

$$Y_i = X_i B + X_s G_1 + Z_{2s} G_2 + v_{si}, \tag{33}$$

[28] See Meghir et al. (2011) for a recent discussion of some of the issues in estimating the effects of particular school characteristics. They highlight the vector of omitted school characteristics that determines z_s^U as a key source of bias.

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). We can decompose the variance of Y_i into observable and unobservable components of both within- and between- school variation via

$$\hat{Var}(Y_i) = \hat{Var}(Y_i - Y_s) + \hat{Var}(Y_s) \tag{34}$$

$$= [\hat{Var}((X_i - X_s)B) + \hat{Var}(v_{si} - v_s)] +$$

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

$$2\hat{Cov}(X_sG_1, Z_{2s}G_2) + \hat{Var}(Z_{2s}G_2) + \hat{Var}(v_s)]. \tag{35}$$

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

The first lower bound estimator is $\hat{Var}(Z_{2s}G_2 + v_s)$. Due to the presence of X_s in (33) it will be purged of any effects of student sorting (observable or unobservable). Thus, it isolates only school/neighborhood factors. The component v_s includes z_s^U, the unpredicted component of the school/neighborhood contribution. However, for post secondary outcomes such as college enrollment and permanent wage rates v_s will also include ξ_s. Recall that ξ_s is an index of 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. The effect is likely to be second order for permanent wages. This is because local shocks that persist for less than 6 years will not bias the method of moments estimator that we use. Furthermore, we pointed out in Section 3.2 that v_s will also contain an approximation error if the linearity assumption A4 is violated. This could lead to upward bias in our estimates of variance of school/neighborhood effects.

Consequently, we also consider a second, more conservative lower bound estimator: $\hat{Var}(Z_{2s}G_2)$. This estimator only attributes to schools/neighborhooods the part of the residual between-school variation that could be predicted based on observable characteristics of the schools at the time students were attending. $\hat{Var}(Z_{2s}G_2)$ excludes true school quality variation that is orthogonal to observed characteristics, but also excludes any truly idiosyncratic local shocks that occur after graduation.[29]

Online Appendices A6 and A7 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.

[29]The approximation error might also bias G_2, but we think this is likely to be minor given that we are controlling for X_s.

6.2 Interpreting the 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 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.[30] 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.[31] 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-

[30] 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.

[31] This interpretation is consistent with the evidence on moving in our data. In the ELS2002 base year survey, parents report the number of years they have lived in the current neighborhood. 22.5% report 3 years or less, 29.91% report 4 to 7 years, 14.32% report 8 to 10 years, and 42.3% report more than 10 years. Parents also report the number of times the student changed schools, not counting natural transitions resulting from grade advancement (e.g., from the elementary school building to the middle school building). The values are 43.31% for no changes, 24% for 1 change, 12.5% for 2 changes, 9.94% for 3 changes, 5.28% for 4 changes, and 4.99% for 5 changes.

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:

$$\begin{aligned} E[\hat{Y}^{90} - \hat{Y}^{10}] = \quad & \frac{1}{I}\sum_i \Phi\left(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} + 1.28(\widehat{Var}(Z_{2s}G_2))^{.5}]}{(1+\widehat{Var}(v_s))^{.5}}\right) \\ & - \frac{1}{I}\sum_i \Phi\left(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} - 1.28(\widehat{Var}(Z_{2s}G_2))^{.5}]}{(1+\widehat{Var}(v_s))^{.5}}\right) \end{aligned} \tag{36}$$

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:

$$\begin{aligned} E[\hat{Y}^{90} - \hat{Y}^{10}] = \quad & \frac{1}{I}\sum_i \Phi\left(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} + 1.28(\widehat{Var}(Z_{2s}G_2 + v_s))^{.5}]}{(1)}\right) \\ & - \frac{1}{I}\sum_i \Phi\left(\frac{[X_i\hat{B} + X_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} - 1.28(\widehat{Var}(Z_{2s}G_2 + v_s))^{.5}]}{(1)}\right) \end{aligned} \tag{37}$$

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 $Z_{2s}G_2$, or $(Z_{2s}G_2 + v_s)$ will depend on the values of a student's observable characteristics, X_iB. Thus, we report average impacts for certain subpopulations of interest as well.[32]

[32]Recall that we have ruled out interactions between X_i or X_i^U and Z_s or z_s^U in the production of Y_i. To see how such nonseparabilities can be addressed, first 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. Next consider the interaction between an observed school characteristic X_{sli} and the unobserved index z_s^U. This will show up as variation across schools in $Cov(Y_{si}, Y_{si'}|X_{lsi}, X_{lsi'})$ as well as variation across schools in $Var(Y_{si'}|X_{sli})$. It might be possible to learn about the importance of $X_{sli}z_s^U$ from such moments. Similarly, interactions between x_{si}^U and elements of Z_s would influence $Var(Y_{si}|Z_s)$. With a nonseparable education production

7 Data and Variable Selection

7.1 Overview of Data Sources

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 ELS2002. Second, each source provides a representative sample of American high schools or 8th grades and samples of students are selected within each school. Both public and private schools are represented.[33] Enough students are sampled from each school to permit construction of estimates of the school means of a large array of student-specific variables and to provide sufficient within-school variation to support the variance decomposition described above. Third, each survey administered questionnaires to school administrators 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, years of college, and and wage rates.

While these common properties are very helpful, differences in the surveys complicates 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_i student-level measures that may not appear in the other datasets. Section 7.3 details the process by which we chose what to include in X_i, X_s, and Z_{2s}, and Table 1 provides a list.

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 indicate that samples of this size may reduce to some degree 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.

function 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 could measure the average performance of a school over the distribution of student 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 between the two schools, and then average over all students.

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

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. Online Appendix Table A10 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, high school graduation is the only outcome we observe. And the set of observable characteristics is not as diverse as in the panel surveys, though it is surprisingly rich for administrative data.

We restrict our samples 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.[34] We then impute values for the other explanatory variables to preserve the sample size, since no one other variable is critical to our analysis.[35] Finally, we use panel weights. The appropriate weights depend on the analysis. See online Appendix A9 for the details.

7.2 Outcome Measures

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.[36] 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

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

[35] This results in sample sizes for the four-year college enrollment analyses of: 12,257 from 903 schools for NLS72, 11,937 from 942 schools for NELS88, 12,168 from 686 schools for ELS2002. The sample sizes and number of schools for the high school graduation analyses are 12,307 and 943 for NELS88, 12,096 and 686 for ELS2002, and 283,157 and 338 for North Carolina respectively. The analysis of years of postsecondary education uses 12,229 observations from 902 schools from NLS72, and the wage analysis uses 4,932 individuals with 9,864 wage observations from 901 schools. We include mother's education combined with a missing indicator for mother's education when performing imputation, along with school averages of all the key characteristics above. Appendix Tables A11-A18 report percent imputed for each variable.

[36] 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.

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 dropout rate. 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.

7.3 Selection of X_i, X_s, and Z_{2s}

X_i should include variables that directly affect the outcome and/or are correlated with unobserved student level characteristics that affect the outcome. 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 the 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.

For purposes of the control function, X_s should contain aggregates of X_i. If one has school level averages of student level variables for which one does not have individual level data, then these aggregates should also be included (there are no such variables is the data sets we use).

What should be in Z_{2s}? Observed school and neighborhood characteristics that could plausibly influence the socioeconomic outcome of interest.

What should **not** be in Z_{2s}? Z_{2s} should exclude variables that are simple aggregates of parent/student traits that might also affect willingness to pay for neighborhood characteristics and thus lead to sorting. These are X_s variables regardless of whether the source is aggregates of the student micro data, Census data or administrative data from the schools.

School level variables that are determined both by school policy/efficacy and by the characteristics of the students fall in a grey area. In ELS2002, we include Frequency of Fights at the school in X_s in our full specification. This variable is determined by school and neighborhood quality and by the unobserved characteristics of students. To the extent school policy and the skill of teachers and the administration have a big effect on fighting, we are being conservative in our estimates of school effects.

We also have a separate measure of school security policies. This belongs in Z_{2s} even if the policies in part are a response to the characteristics of students.

The other example we wish to highlight is average daily attendance percentage. Daily attendance reflects both characteristics of the students and school quality. Suppose Proposition 1 fails, and X_s is not sufficient to control for unobserved student body characteristics that directly influence school attendance and education outcomes. Then including daily attendance in Z_{2s} rather than in X_s might bias our estimates school/neighborhood upward. On the other hand, including it in X_s would lead us to understate school/neighborhood effects if school policy/quality has a big effect on attendance. In the end we opted to exclude average percent daily attendance from the model.

The same issues apply to test scores measured during high school. Test scores are determined by school quality and by student characteristics. We include them in X_{2s} in the "full" specification. We never include them in Z_{2s}. This is conservative.

Table 1 lists the final choices of individual-level and school-level explanatory measures used in each dataset. Online Appendix Tables A11 - A18 provide the mean, standard deviation, and percent of observations imputed for each individual-level and school-level characteristic for each of our four datasets.

8 Results

We now turn to the results. Along with the point estimates, we report bootstrap standard error estimates based on re-sampling schools with replacement, with 500 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

The full variance decompositions described in Section 6 are provided for each of our outcomes in online Appendix Tables A19, A20, and A21. Panel A of Table 2 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 include $Var(v_s)$ (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 unambiguously attributable to schools/neighborhoods than either NC or ELS2002 ($\sim$ 1% relative to $\sim$ 2-3%). One possible explanation for this finding is that NELS88 school-level observables (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 $\hat{Var}(Z_2G_2)$ relative to $\hat{Var}(v_s)$, 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. Online Appendix Table 19 provides a full variance decomposition that shows the critical role that individual-specific factors play. 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 choosing a school at the 90th percentile of the distribution of school/neighborhood contributions instead of a school at the 10th 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 3 presents results for the decomposition of the latent index determining enrollment in a four-year college. Comparing the baseline specifications from NLS72, NELS88, and ELS2002 (Columns 1, 3, and 5), we observe a surprising consistency in both of the 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 between 1.5 and 1.9 percent, and the estimates that include the residual v_s dropping to between 2.9 and 3.2 percent.

Panel B of Table 3 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 four-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 3 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 2 and 3 are based on starting the full distribution of students at a 10th quantile school versus starting them at 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 2 and 3 conceal heterogeneity in the relative impact of alternative 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 four-year college, the sensitivity to school quality is higher for groups with values of $X_i\hat{B}$ that place them closer to a probability of 0.5. 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 4 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 4 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 5 reports a corresponding set of results for enrollment in a four-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 $Z_{2s}G_2 + v_s$ would increase the college 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. In 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 $Z_{2s}G_2 + v_s$ would increase enrollment rates at four-year colleges by 16.7 percentage points (from 72.8% to 89.6%). More generally, across datasets the lower bound estimates excluding and including v_s 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, many students 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 6 displays the lower bound estimates of the impact of 10th-to-90th and 10th-to-50th shifts in school/neighborhood 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 .31 years, while including standardized tests among the observable characteristics reduces this estimate to .20 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 .45 and .33 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 $(100e^{0.157} - 100)$. 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/neighborhood quality increases wages by about 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.

Chetty and Hendren (In Progress) find that 20 years in a 1 standard deviation better neighborhood raises the log of adult earnings by about 10%. When we include v_s we find that a one standard deviation shift in school/neighborhood raises permanent wage rates by 0.069 (0.069=0.177/(2*1.28). Several factors contribute to the difference between the studies.. First, families are mobile and we only condition on attendance in that same high school (or in 8th grade in the the case of NELS88).

Consequently, our estimate represents the effect of a substantially shorter period of exposure than 20 years. This fact alone could easily reconcile the studies. Second, if neighborhood/school quality raises employment and hours as well as wage rates, then the wage effect will be smaller than the effects on earnings. Third, our estimates are likely to be a lower bound. On the other hand, the fact that Chetty and Hendren work at the county level is likely to reduce their estimates relative to our school level estimates. This because the standard deviation of county level effects abstracts from within county heterogeneity in school/neighborhood quality.

8.5 Alternative Estimators

In this subsection we compare our lower bound estimates above with two alternative estimators of school and neighborhood effects more commonly observed in the literature.

First, in Online Appendix Tables A1 - A2 we report estimates of $Var(X_sG_1 + Z_{2s}G_2 + v_s)$, or equivalently $Var(Y_s - X_sB)$. By including X_sG_1, this estimate reintroduces peer effects that operate through school averages of observable or unobservable student characteristics as well as other unobserved school inputs that are predictable based on X_s given Z_{2s}. But X_sG_1 also includes the component $X_s\Pi_{\tilde{X_s^U}X_s}\beta^U$, which reflects student sorting on unobservable characteristics. If there is no sorting on X_i^U, then the sorting component $X_s\Pi_{\tilde{X_s^U}X_s}\beta^U = 0$, and $Var(X_sG_1 + Z_{2s}G_2 + v_s) = Var(Z_s\Gamma + z_s^U + \xi_s)$. This is the true variance in school/neighborhood treatment effects. When unobservables do contribute to sorting, then $Var(X_sG_1 + Z_{2s}G_2 + v_s)$ will generally overstate the variance in school/neighborhood treatment effects.[37]

Indeed, across all of the specifications and outcomes for the panel surveys these estimates are noticeably larger than our lower bound estimates. For example, for the full specification in NELS88, the sorting-on-observables estimator attributes 5.2% of the variance in the latent index that determines high school graduation to schools/neighborhoods, compared to 2.5% for the lower bound estimate of $Var(Z_{2s}G_2)$. The associated effect of a 10th-to-90th quantile shift in the school/neighborhood quality on graduation is .10 (relative to .07 for the lower bound estimate).[38] For enrollment in a four-year college, the corresponding school/neighborhood variance fractions for the ELS full specification is 4.3% (versus 3.1% for the lower bound estimate), which corresponds to a 10th-to-90th shift in the probability of enrollment of .205 (versus .170). The only case in which $\hat{Var}(X_sG_1 + Z_{2s}G_2 + v_s)$ is not substantially higher than $\hat{Var}(Z_{2s}G_2 + v_s)$ is for the high school grad-

[37] From (27), $Var(X_sG_1 + Z_{2s}G_2 + v_s) = Var(Z_s\Gamma + z_s^U + \xi_s + X_s\Pi_{\tilde{X_i^U}X_s}\beta^U)$. If the covariances between $X_s\Pi_{\tilde{X_i^U}X_s}$ and the components of the school treatment effect $Z_s\Gamma + z_s^U + \xi_s$ are sufficiently negative, then one can find $Var(X_sG_1 + Z_{2s}G_2 + v_s) < Var(Z_s\Gamma + z_s^U + \xi_s)$. In this case, which we consider unlikely, even $Var(X_sG_1 + Z_{2s}G_2 + v_s)$ would understate the true contribution of schools/neighborhoods to the variance in outcomes.

[38] The effects a 10-to-90th shift in $X_sG_1 + Z_{2s}G_2 + v_s$ are constructed as

$$E[\hat{Y}^{90} - \hat{Y}^{10}] = \frac{1}{I}\sum_i \Phi\left(\frac{[X_i\hat{B} + \overline{X}_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} + 1.28(\widehat{Var}(X_{1s}G_1 + Z_{2s}G_2 + v_s))^{.5}]}{(1)}\right)$$
$$- \frac{1}{I}\sum_i \Phi\left(\frac{[X_i\hat{B} + \overline{X}_s\hat{G}_1 + \overline{Z_{2s}\hat{G}_2} - 1.28(\widehat{Var}(X_{1s}G_1 + Z_{2s}G_2 + v_s))^{.5}]}{(1)}\right). \quad (38)$$

uation outcome in the North Carolina administrative data, where $\hat{Var}(X_sG_1)$ is nearly offset by a strong negative covariance between $X_s\hat{G}_1$ and $Z_{2s}\hat{G}_2$.

Second, Online Appendix Table A3 reports estimates of $Var(Z_{2s}G_2)$ and $Var(Z_{2s}G_2+v_s)$ from a four-year college enrollment specification in which the school-averages X_s are omitted. A small fraction of the variance previously absorbed by the control function is now captured by $Z_{2s}\hat{G}_2$, while the bulk of it now enters the between-school residual $\hat{v}_s$. Thus, $\hat{Var}(Z_{2s}G_2)$ increases slightly relative to our main college enrollment estimates in Table 3, while $\hat{Var}(Z_{2s}G_2+v_s)$ increases substantially, to the point where they nearly match the sorting-on-observables estimates $\hat{Var}(X_sG_1+Z_{2s}G_2+v_s)$ reported in the previous paragraph. Columns 3 and 4 of Online Appendix Table A4 report corresponding estimates for years of postsecondary education, while columns 1 and 2 report estimates from a specification in which B is estimated in a first stage in which school fixed effects are included, and then $Y_s - X_s\hat{B}$ is regressed on Z_{2s} to recover $\hat{G}_2$ and $\hat{v}_s$. Each of these specifications exhibits substantially higher estimates of $Var(Z_{2s}G_2+v_s)$.

Taken together, the results from these alternative estimators suggest that our lower bound estimates, while more conservative than other existing estimators, still seem to capture a substantial portion of the variation in the contributions of schools/neighborhoods.

8.6 Empirical Evidence on the Spanning Condition

In Online Appendix A2 we investigate the factor structure of X_s using two approaches. First, we use principal components analysis to compute the eigenvalues and eigenvectors of $\hat{Cov}(X_s)$, the estimated covariance matrix of X_s. While $Cov(X_s)$ must be positive semidefinite, $\hat{Cov}(X_s)$ need not be positive semidefinite given sampling error and the fact that our sample is unbalanced. In practice we obtain small negative values for some of the eigenvalues. We interpret these estimates as corresponding to eigenvalues that are in fact 0 or very close to 0. We find that for each of our three survey datasets the number of positive eigenvalues is less than L, indicating that $\hat{Cov}(X_s)$ 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 dimension of the amenity vector A^X, this supports our assumption that the dimension of $A^X \leq L$. Indeed, we further show that in each dataset an even smaller number of latent factors (generally around 10) can explain 90% of the sum of the variances of the elements of X_s, suggesting that the variation in student composition across schools is driven primarily by a small number of amenity factors. Bootstrap confidence interval estimates of the number needed to explain 90% of the variances are fairly tight. The number of latent factors required to explain a given percentage of the sum of the variances of the elements of X_s is larger in the full specification, which contains more variables. This would be expected in the presence of sampling error in $\widehat{Cov}(X_s)$. However, it might also indicate that there are in fact a few additional amenity factors that play a very small role in driving sorting (and thus have very small eigenvalues) and are picked up by the additional elements of X_s in the full specification.

Our second approach draws on the literature on testing for the number of factors or the matrix rank, including Lewbel (1991), Cragg and Donald (1997), Robin and Smith (2000), Bai and Ng (2002) and Kleibergen and Paap (2006). The test of the rank of a matrix proposed by Kleibergen and Paap (2006) fits our application well. The test involves a singular value decomposition of $\widehat{Cov}(X_s)$, and can accommodate arbitrary forms of heteroskedasticity and correlation at the school level. We perform tests of the null hypothesis of $\text{rank}(Cov(X_s)) = j$ against the alternative that $Cov(X_s) > j$. For all three data sets and specifications, we cannot reject the null hypothesis for values of j well below L. See the Online Appendix.

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.[39] 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, a higher 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 classroom allocation problem aligns with that of the decentralized choice problem considered in Section 2. The price vector $P(A_c)$ is the shadow price associated with the capacity constraint of c. However, in the elementary and middle school contexts, it seems likely that 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-

[39] See Altonji and Mansfield (2014), appendix A9 for additional discussion of the teacher value added case.

random student sorting on unobservables and observables. Furthermore, as we note in the Online Appendix, it may also reduce omitted variables bias from non-random assignment of teachers to other unobserved outcome-relevant classroom environmental factors such as course difficulty level (e.g. basic versus honors) 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.[40]

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

In this paper we provide conditions under which the tactic of controlling for group averages of observed individual-level characteristics can control perfectly for group averages of unobservables. This insight leads to a way to estimate a lower bound on the contribution of group effects to individual outcomes. We also examine the conditions under which causal effects of particular observed group characteristics can be estimated. Going forward, we view the central message of the paper

[40] In principle, one could adopt the model of group choice and the control function approach to the analysis of the effects of years of schooling, dosage levels, or other endogenous choice problems that have a natural ordering. Let s denote number of years of schooling. Each schooling level has an associated set of characteristics A_s governing the pecuniary and non-pecuniary return to choosing level s. A_s is weighted by X_i and X_i^U. This leads to a relationship between X_s and X_s^U that could serve as the basis for a control function for X_s^U. However, as pointed out at the beginning of Section 3.2, there must be at least as many levels of s as there are elements of X_s. Otherwise s will not vary conditional on X_s unless restrictions are available that reduce the dimension of the index of X_s required to control for X_s^U. Essentially, there are fewer degrees of freedom (the number of levels) than there are parameters in coefficient vector on X_s (G_1 in our outcome equation). We leave a full analysis of the possibilities to future research.

to be that the features of the distribution of observables in a group contains information about the distribution of unobservables in the group—not that the relationship between the observed and unobserved group averages is necessarily linear. We would like to know if a variant of Proposition 1 carries over to more general specifications of preferences than the class that work with. We would also like to know how the choice mechanism affects sorting on observables and unobservables. In particular, does a version of Proposition 1 carry over to two sided selection problems, such the sorting of students across universities or workers across firms?

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? 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 and 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 for wage rates are qualitatively consistent with those of Chetty et al. (2015) and Chetty and Hendren (In Progress) for earnings and with Aaronson's (1998) findings for high school dropouts, although the magnitudes are hard to compare for a number of reasons.

There is much to do on the empirical side. We briefly discussed the possibility of using an outcome model that allows for interactions between observed and unobserved student characteristics and observed and unobserved neighborhood characteristics. The details need to be worked out. The application of our approach to distinguishing true group effects from sorting in other applications, such as hospital quality, teacher productivity, and doctor quality should be explored.

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

Aliprantis, Dionissi, and Francisca G-C Richter (2013) 'Evidence of neighborhood effects from mto: Lates of neighborhood quality'

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

_ (2014) 'Group-average observables as controls for sorting on unobservables when estimating group treatment effects: the case of school and neighborhood effects.' Technical Report, National Bureau of Economic Research

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'

Bai, Jushan, and Serena Ng (2002) 'Determining the number of factors in approximate factor models.' *Econometrica* 70(1), 191–221

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 Ross (2009) 'Identifying individual and group effects in the presence of sorting: A neighborhood effects application.' Technical Report, University of Connecticut, Department of Economics

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 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)

Chamberlain, Gary (1980) 'Analysis of covariance with qualitative data.' *The Review of Economic Studies* 47(1), 225–238

Chamberlain, Gary et al. (1984) 'Panel data.' *Handbook of Econometrics* 2, 1247–1318

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

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz (2015) 'The effects of exposure to better

neighborhoods on children: New evidence from the moving to opportunity experiment.' Technical Report, National Bureau of Economic Research

Chiappori, Pierre-Andre, and Bernard Salanie (forthcoming) 'The econometrics of matching models.' *Journal of Economic Literature*

Cragg, John G, and Stephen G Donald (1997) 'Inferring the rank of a matrix.' *Journal of econometrics* 76(1), 223–250

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)

Durlauf, Steven N (2004) 'Neighborhood effects.' *Handbook of regional and urban economics* 4, 2173–2242

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

Ekeland, Ivar, James J. Heckman, and Lars Nesheim (2004) 'Identification and estimation of hedonic models.' *Journal of Political Economy* 112.S1, S60–S109

Epple, Dennis, and Glenn J Platt (1998) 'Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes.' *Journal of urban Economics* 43(1), 23–51

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

Fu, Shihe, and Stephen L Ross (2013) 'Wage premia in employment clusters: How important is worker heterogeneity?' *Journal of Labor Economics* 31(2), 271–304

Gómez, Eusebio, Miguel A Gómez-Villegas, and J Miguel Marín (2003) 'A survey on continuous elliptical vector distributions.' *Revista matemática complutense* 16(1), 345–361

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

Harding, David, Lisa Gennetian, Christopher Winship, Lisa Sanbonmatsu, and Jeffrey Kling (2011) 'Unpacking neighborhood influences on education outcomes: Setting the stage for future research.' *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances: Rising Inequality, Schools, and Children's Life Chances* p. 277

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

Imbens, Guido W (2007) 'Nonadditive models with endogenous regressors.' *Econometric Society Monographs* 43, 17

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, and Susan E Mayer (1990) 'The social consequences of growing up in a poor neighborhood.' *Inner-city poverty in the United States* 111, 186

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

Kasy, Maximilian (2011) 'Identification in triangular systems using control functions.' *Econometric Theory* 27(03), 663–671

Kleibergen, Frank, and Richard Paap (2006) 'Generalized reduced rank tests using the singular value decomposition.' *Journal of econometrics* 133(1), 97–126

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

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

Lewbel, Arthur (1991) 'The rank of demand systems: theory and nonparametric estimation.' *Econometrica* pp. 711–730

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, Steven Rivkin et al. (2011) 'Econometric methods for research in education.' *Handbook of the Economics of Education* 3, 1–87

Melo, Rafael Lopez de (2015) 'Firm wage differentials and labor market sorting: Reconciling theory and evidence.' *Working paper. University of Chicago*

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

Mundlak, Yair (1978) 'On the pooling of time series and cross section data.' *Econometrica: journal of the Econometric Society* pp. 69–85

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

Robin, Jean-Marc, and Richard J Smith (2000) 'Tests of rank.' *Econometric Theory* 16(02), 151–175

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*

Sampson, Robert J, Jeffrey D Morenoff, and Thomas Gannon-Rowley (2002) 'Assessing" neighborhood effects": Social processes and new directions in research.' *Annual review of sociology* pp. 443–478

Tiebout, Charles M (1956) 'A pure theory of local expenditures.' *The journal of political economy* pp. 416–424

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: Variables Used in Baseline and Full (in Italics) Specifications, by Dataset

Description of Variable(s)	NLS72	NELS88	ELS2002	NC
Student Characteristics				
Race Indicators, 1(Female), 1(Immigrant)	X	X	X	X
Student Ability				
Math Standardized Score, Reading Standardized Score	X	X	X	X
1(Gifted at Math), 1(Gifted at Reading)				X
Student Behavior				
Hrs./Wk. Spent on Homework		X	X	X
Hrs./Wk. Spent on Leisure Reading, Hrs./Wk. Spent Watching TV		X	X	X
Hrs./Wk. Spent on Computer			X	
1(Physical Fight This Year), Parents Often Check Homework		X	X	
Family Background				
Standardized SES, Number of Siblings	X	X	X	
Indicators for Presence of Biological Parents	X	X	X	
Father's Yrs. of Ed., Mother's Yrs. of Ed.	X	X	X	X
Moth. Yrs. Ed. Missing	X	X	X	X
Average of Grandparents' Education			X	
Log(Family Income), 1(English Spoken at Home)	X	X	X	
Indicators for Parental Religion	X	X	X	
1(Parents are Married)		X	X	
1(Immigrant Father), 1(Immigrant Mother)		X	X	
Indicators for Father's Occupation Group		X	X	
Indicators for Mother's Occupation Group		X	X	
Home Environ. Indicators (1st Prin. Comp.)	X	X	X	
Parental Sch. Involv. Indicators (1st Prin. Comp.)		X	X	
1(Eligible for Free/Reduced Price Lunch)				X
1(Currently Limited English Proficiency), 1(Ever LEP)				X
Parental Expectations				
Mother's Desired Yrs. Of Ed., Father's Desired Yrs. Of Ed.		X	X	
School Characteristics (Treated as elements of X_s)*				
School Pct. Minority	X	X	X	
School Pct. Free/Reduced Price Lunch		X	X	X
School Pct. LEP, *School Pct. Special Ed.*		X	X	
School Pct Remedial Reading, School Pct. Remedial Math		X	X	
Frequency of Fights (Administrator's Impression)			X	
School Characteristics (Treated as elements of Z_{2s})				
1 (Catholic School), 1 (Private Non-Catholic School)	X	X	X	
Total School Enrollment, Student-Teacher Ratio	X	X	X	X
Log(Min. Teacher Salary)		X	X	
% Tch. Turnover, % of Teachers w/ Master's Degrees or More	X	X	X	X
% of Teachers w/ Master's Degrees or More	X	X	X	X
% of Teachers w/ Certification			X	
School Teacher Pct. Minority	X	X	X	
1(Minimum Competency Test Exists)			X	
1(Gifted Program Exists), 1(Collectively Bargained Contract)		X		
1(Tracking System), Age of School Building	X			
Distance to 4-year College, Distance to Community College	X			
Teacher Evaluation Mechanism Indicators (1st Principal Component)			X	
Teacher Incentives Indicators (1st Principal Component)			X	
School Security Policy Indicators (1st, 2nd Principal Components)			X	
School Security Implementation Indicators (1st & 2nd Prin. Comps.)		X	X	
Sch. Environ. Indicators (1st and 2nd Prin. Comps.)			X	
Sch. Facilities Indicators (Admin. Survey, 1st & 2nd Prin. Comps.)			X	
Teacher Access to Tech. Indicators (Admin. Survey, 1st Prin. Comp.)			X	
Magnet School, Charter School, Sch. Tch. % Highly Qualified				X
Neighborhood Characteristics (Treated as elements of Z_{2s})				
Urbanicity Indicators	X	X	X	X
Indicators for U.S. Census Region	X	X	X	
Neighborhood Crime Level Category (Sch. Admin. Survey)			X	

*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 contributions of schools/neighborhoods.
School averages of all student-level variables are also included in each specification. The school population average is used where available (see the "School Characteristics (Treated as elements of X_s)" category in this table); otherwise the average among sampled students is used in its place.

Table 2: 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
$Var(Z_{2s}G_2)$	(0.007)	(0.005)	(0.006)	(0.005)	(0.010)	(0.010)
LB w/ unobs	0.049	0.036	0.028	0.016	0.036	0.025
$Var(Z_{2s}G_2+v_s)$	(0.013)	(0.008)	(0.009)	(0.006)	(0.010)	(0.010)

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 $Var(Z_{2s}G_2)$	(0.021)	(0.015)	(0.013)	(0.012)	(0.011)	(0.011)
LB w/ unobs: 10th-90th	0.174	0.152	0.098	0.075	0.083	0.070
Based on $Var(Z_{2s}G_2+v_s)$	(0.024)	(0.017)	(0.017)	(0.014)	(0.011)	(0.011)
LB no unobs: 10th-50th	0.056	0.044	0.033	0.025	0.040	0.038
Based on $Var(Z_{2s}G_2)$	(0.012)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
LB w/ unobs: 10th-50th	0.096	0.083	0.055	0.041	0.048	0.039
Based on $Var(Z_{2s}G_2+v_s)$	(0.014)	(0.010)	(0.010)	(0.008)	(0.007)	(0.007)
Sample Mean	0.760	0.760	0.838	0.838	0.917	0.917

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 used in the baseline and full models are specified in 1.

The full variance decompositions underlying these estimates are presented in Web Appendix Table A19.

Appendix Sections A6 and A7 discuss estimation of model parameters and the variance decompositions. Section 6.3 discusses estimation of the 10-50 and 10-90 differentials.

Table 3: 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	0.024	0.018
$Var(Z_{2s}G_2)$	(0.006)	(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
$Var(Z_{2s}G_2+v_s)$	(0.008)	(0.006)	(0.008)	(0.007)	(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	0.138	0.118	0.127	0.112	0.155	0.132
Based on $Var(Z_{2s}G_2)$	(0.013)	(0.012)	(0.017)	(0.016)	(0.018)	(0.017)
LB w/ unobs: 10th-90th	0.168	0.153	0.188	0.155	0.216	0.172
Based on $Var(Z_{2s}G_2+v_s)$	(0.017)	(0.016)	(0.021)	(0.019)	(0.022)	(0.019)
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)	(0.007)	(0.008)	(0.008)
LB w/ unobs: 10th-50th	0.077	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)	(0.010)	(0.009)
Sample Mean	.265	.265	.310	.310	.418	.418

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

The notes to Table 2 apply, except that Table 3 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 4: 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 $Var(Z_{2s}G_2)$	(0.028)	(0.022)	(0.024)	(0.026)	(0.019)	(0.021)
LB w/ unobs	0.242	0.229	0.176	0.159	0.146	0.144
Based on $Var(Z_{2s}G_2+v_s)$	(0.031)	(0.024)	(0.030)	(0.030)	(0.019)	(0.021)
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)	(0.007)	(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 $Var(Z_{2s}G_2+v_s)$	(0.015)	(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 $Var(Z_{2s}G_2)$	(0.021)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)
LB w/ unobs	0.176	0.152	0.098	0.084	0.094	0.084
Based on $Var(Z_{2s}G_2+v_s)$	(0.024)	(0.017)	(0.018)	(0.017)	(0.014)	(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 $Var(Z_{2s}G_2)$	(0.027)	(0.020)	(0.022)	(0.020)	(0.017)	(0.017)
LB w/ unobs	0.235	0.206	0.159	0.125	0.120	0.099
Based on $Var(Z_{2s}G_2+v_s)$	(0.031)	(0.022)	(0.028)	(0.024)	(0.017)	(0.017)
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 $Var(Z_{2s}G_2)$	(0.013)	(0.009)	(0.006)	(0.005)	(0.006)	(0.005)
LB w/ unobs	0.102	0.084	0.040	0.025	0.037	0.016
Based on $Var(Z_{2s}G_2+v_s)$	(0.015)	(0.010)	(0.008)	(0.006)	(0.006)	(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 2 for row and column definitions

Table 5: 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)	(0.008)
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)	(0.016)	(0.008)
XB: 90th Quantile						
LB no unobs	0.191	0.182	0.160	0.128	0.166	0.128
Based on $Var(Z_{2s}G_2)$	(0.018)	(0.017)	(0.022)	(0.019)	(0.020)	(0.019)
LB w/ unobs	0.234	0.234	0.236	0.187	0.231	0.167
Based on $Var(Z_{2s}G_2+v_s)$	(0.024)	(0.024)	(0.026)	(0.024)	(0.024)	(0.020)
Black						
LB no unobs	0.132	0.109	0.125	0.111	0.145	0.121
Based on $Var(Z_{2s}G_2)$	(0.014)	(0.012)	(0.017)	(0.015)	(0.017)	(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)	(0.021)	(0.019)	(0.021)	(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)	(0.017)
LB w/ unobs	0.134	0.127	0.132	0.102	0.195	0.162
Based on $Var(Z_{2s}G_2+v_s)$	(0.015)	(0.013)	(0.014)	(0.013)	(0.021)	(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 $Var(Z_{2s}G_2)$	(0.017)	(0.015)	(0.021)	(0.019)	(0.020)	(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)	(0.025)	(0.023)	(0.025)	(0.022)

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

The notes to Table 4 apply, except that Table 5 reports results for enrollment in a four-year college two years after graduation, and the NLS72 is one of the data sets.

Table 6: 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	0.005	0.002	0.039	0.041	0.039	0.040
$Var(Z_{2s}G_2)$	(0.002)	(0.002)	(0.010)	(0.011)	(0.012)	(0.012)
LB w/ unobs	0.010	0.006	0.052	0.052	0.050	0.049
$Var(Z_{2s}G_2+v_s)$	(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	0.308	0.197	0.152	0.157	0.155	0.157
Based on $Var(Z_{2s}G_2)$	(0.057)	(0.045)	(0.021)	(0.020)	(0.021)	(0.021)
LB w/unobs: 10th-90th	0.445	0.334	0.177	0.177	0.175	0.173
Based on $Var(Z_{2s}G_2+v_s)$	(0.065)	(0.049)	(0.028)	(0.026)	(0.028)	(0.027)
LB no unobs: 10th-50th	0.154	0.098	0.076	0.079	0.077	0.078
Based on $Var(Z_{2s}G_2)$	(0.028)	(0.022)	(0.011)	(0.010)	(0.011)	(0.010)
LB w/unobs: 10th-50th	0.222	0.167	0.088	0.088	0.087	0.087
Based on $Var(Z_{2s}G_2+v_s)$	(0.032)	(0.025)	(0.014)	(0.013)	(0.014)	(0.013)
Sample Mean	.27	.27	.31	.31	.37	.37

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 Table 1 for the variables in the baseline model and the full model. The full variance decompositions are in Appendix Table A21. Web Appendix Sections A6 and A7 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: $rank(\Theta^U) \leq rank(\tilde{\Theta}) = dim(A)$

Suppose that the matrices $\tilde{\Theta} = \Theta + \Pi_{X^U X}\Theta^U$ and Θ^U, are each full rank. For example:

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

Then we can write $\Theta^U = R\tilde{\Theta}$, 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 $\tilde{\Theta}$ are unchanged from Case 1:

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

Then we can write $\Theta^U = R\tilde{\Theta}$, 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 $\tilde{\Theta}$ is at least as large as the row rank of Θ^U. Any scenario satisfying $rank(\Theta^U) \leq rank(\tilde{\Theta}) = dim(A)$ will satisfy the spanning condition in Proposition 1.

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

Suppose instead that neither X_1 nor X_2 predicts willingness to pay for A_2. Further, suppose that neither X_1 nor X_2 is correlated with any elements of X^U that predict willingness to pay for A_2. This implies that the second column of $\tilde{\Theta}$ is a zero vector:

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

Since $\tilde{\Theta}$ is now rank-deficient, there is no matrix R such that $R\tilde{\Theta} = \Theta^U$. 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 (and are each uncorrelated with X^U, so that $\Pi_{X^U X} = 0$, a rank-deficiency will also occur:

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

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 $R\tilde{\Theta} = \Theta^U$. For any choice of R, in each row of $R\tilde{\Theta}$ 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 $\tilde{\Theta}$ 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 $\tilde{\Theta}$.

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

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, X_1^U only predicts WTP for A_2, and X_1 and X_1^U are uncorrelated:

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

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

Indeed, suppose that we alter $\tilde{\Theta}$ and Θ^U so that both X_1 and X_1^U affect WTP for both amenities (but in different proportions):

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

There is no scalar R such that $R\tilde{\Theta} = \Theta^U$, 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 $\tilde{\Theta}$. 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 (W may affect taste for A_2, so that A_2 is still relevant for school choice):

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

Then for $R = 2$, $R\tilde{\Theta} = \Theta^U$, and the spanning condition is satisfied. Note that the row space of $\tilde{\Theta}$ is a subspace of the row space of Θ^U, despite the fact that both $\tilde{\Theta}$ 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 $\tilde{\Theta}$ span the same (or a superspace) of the amenity subspace spanned by the rows of Θ^U.

A2 Testing Whether X_s Spans the Amenity Space A^X

As discussed in Section 3.2.2, Assumption 5.1 is one of the two key sufficient conditions for the spanning assumption, Assumption 5, to hold. Assumption 5.1 requires that the vector of observables X_i captures enough independent factors determining families' preferences over group amenities that X_s can span the space of amenities (denoted A^X) for which X_i affects tastes, either through direct effects on willingness to pay or indirectly through correlation between X_i and elements of X_i^U. For the particular linear specification of utility featured in (2), this condition is tantamount to requiring that $rank(\tilde{\Theta}) \geq \dim(A_s^X)$.

The restriction $rank(\tilde{\Theta}) \geq \dim(A_s^X)$ restricts $rank(Cov(X_s))$, which forms the basis for our test. To see this, note that taking expectations of both sides of (48) conditional on s implies that

$$X_s = \Upsilon_s Var(\gamma_i)^{-1} \tilde{\Theta}' Var(X_i),$$

where $\Upsilon_s \equiv E(\Upsilon_i | s_i = s)$ is the average of the willingness to pay vector for those who choose s. Thus X_s is a linear combination of Υ_s. Recall that the length of Υ_s is K, the number of valued amenities. Consequently, if $L > K$, then the L elements of X_s are all linear combinations of the smaller number of components of the average willingness to pay vector Υ_s. But this implies that $Cov(X_s)$ will be rank deficient, with $rank(Cov(X_s)) = K$. In fact, if WTP for some of the K amenities is not influenced by X_i, then some of the columns of $\tilde{\Theta}$ will be 0. In this case, $rank(Cov(X_s)) = \dim(A^X) < K$ further reducing the rank of $Cov(X_s)$. This is a testable condition.

More generally, suppose Assumption 5.1 is nearly satisfied, so that a small number of amenity factors drive the vast majority of the variation in X_s, but elements of X_i slightly influence tastes for several other amenities. Our simulations in section 4 suggest that such minor departures from the Assumptions 5.1 and 5.2 have little impact on the ability of X_s to effectively control for the

unobservable between-school variation X_s^U. But 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 $Cov(X_s)$[41], and then directly perform the principal components analysis on the estimated covariance matrix.[42]

The results are in online Appendix Table A5. Panel A reports, for each dataset we use, the number of principal components necessary to explain 75%, 90%, 95%, 99%, and 100% of the sum $\sum_{\ell=1}^{L} Var(X_{s\ell})$ of the variances of the standardized values of the L characteristics in X_s, 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 $rank(\tilde{\Theta}) \geq \dim(A_s^X)$, implying that Assumption 5.1 for the spanning condition $\Theta^U = R\tilde{\Theta}$ 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 $L = 51$ 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 $\sum_{\ell=1}^{L} Var(X_{s\ell})$ or $Var(X_s\hat{G}_1)$. In the ELS2002 full specification, only 19 and 15 factors are needed to explain 95% of the variation in $\sum_{\ell=1}^{L} Var(X_{s\ell})$ 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). The number of latent factors required to explain a given percentage of the sum the variances of the elements of X_s is larger in the full specification, which contains more variables. This would be expected in the presence of sampling error in $\widehat{Cov}(X_s)$. However, it might also indicate that there are in fact additional amenity factors that play a very small role in driving sorting (and thus have very small eigenvalues) that can be picked up by the additional elements of X_s in the full specification.

[41] 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 $\hat{Cov}(X_s) = \hat{Cov}(X_i) - \hat{Cov}(X_i - X_s)$.

[42] When constructing our control function in our main estimating equations we augment the vector $\overline{X}_s$ that comes from directly aggregating student level variables X_i with school-level aggregates directly reported by the school administrators (e.g. percent minority), since these are likely to measure the true school population average X_s with minimal error. However, when performing the principal components analysis of X_s, we do not include these additional measurements that come directly from schools, since they are likely to be nearly collinear with $\overline{X}_s$, and could cause us to find spurious evidence of rank deficiency in $Cov(X_s)$.

Note, though, that because we only observe small samples of students in each school in our panel surveys and only have a sample of schools, the covariance matrix $\hat{Cov}(X_s)$ that is decomposed by PCA is merely a consistent estimate of the population covariance matrix $Cov(X_s)$, and thus contains sampling error. The assumption underlying the spanning condition pertains to the rank of the population matrix $Cov(X_s)$. We address this issue in two ways. First, Panel A and B of online Appendix Table A5 report bootstrap confidence interval estimates of the number needed to explain the specified percentages of $\sum_l^L Var(X_{sl})$ and $Var(X_s\hat{G}_1)$. They are fairly tight.

Second, we also implement the formal test of rank proposed by Kleibergen and Paap (2006). Building on Cragg and Donald (1997) and Robin and Smith (2000), this test exploits the fact that a rank deficient matrix will have a subset of its singular values equal to 0, and tests whether the smallest singular values are farther from zero than one would expect based on sampling error.[43] The test compares the null hypothesis that $rank(Cov(X_s)) = q$, for some $q < L$, against the alternative that $rank(Cov(X_s)) > q$. Thus, Table A6 report the p-value from this test for each possible rank $1,\ldots,L-1$ for each of our panel survey datasets for our baseline specification. Table A7 displays the corresponding p-values across datasets for our full specification.

One advantage of this test is that it can accommodate both heteroskedasticity and autocorrelation among the error components. However, while the tests that cluster at the school-level allow for the most general correlation structure, they sometimes fail to converge in our samples (indicated by "NaN" in Tables A6 and A7). Consequently, for each dataset we display p-values both from tests that are robust to heteroskedasticity but assume zero autocorrelation as well as those that cluster at the school-level and are robust to both heteroskedasticity and autocorrelation.

Across tests and datasets, the results are broadly quite consistent with the PCA results reported above. In particular, not only do the tests consistently fail to reject rank values well below the number of observables, but in fact the p-values generally converge to values indistinguishable from 1 as the numbers of factors being tested nears the number of principal components identified in Table A5. In sum, the Kleibergen/Paap tests reveal a complete absence of evidence with which to reject the null hypothesis that a much smaller number of factors are driving sorting on the vector of observable characteristics X_i.

[43] Specifically, Kleibergen and Paap (2006) show that if the vectorized form of the covariance matrix estimator has a normal limiting distribution, then the limiting distribution of an orthogonal transformation of the smallest singular values of this matrix is also normal. Their rank statistic thus consists of a quadratic form of this orthogonal transformation with respect to the inverse of its covariance matrix, and hence follows a χ^2 limiting distribution. Bai and Ng (2002) provide an alternative approach.

A3 The Relationship between X_s^U and X_s when $E(X_i|\Upsilon_i)$ and $E(X_i^U|\Upsilon_i)$ are Nonlinear

Decompose $E[\tilde{X}_i^U|\Upsilon_i]$ and $E[X_i|\Upsilon_i]$ as

$$E[\tilde{X}_i^U|\Upsilon_i] = E^*[\tilde{X}_i^U|\Upsilon_i] + e_i^{\tilde{X}^U}(\Upsilon_i) \tag{39}$$

$$E[X_i|\Upsilon_i] = E^*(X_i|\Upsilon_i) + e_i^X(\Upsilon_i) \tag{40}$$

where the vectors $E^*[\tilde{X}_i^U|\Upsilon_i]$ and $E^*[X_i|\Upsilon_i]$ are the linear least squares projections of $\tilde{X}_i^U$ and X_i on Υ_i and the error terms $e_i^{\tilde{X}^U}$ and $e_i^{X^U}$ are uncorrelated with Υ_i.

Proposition 2: *Assume that Assumptions A1, A2, A3, and A4 hold.*

Then the expectation X_s^U *is*

$$\begin{aligned} X_s^U = {} & X_s[\Pi_{X^U X} + Var(X_i)^{-1}R' Var(\tilde{X}_i^U)] \\ & -E[e_i^X(\Upsilon_i)|s(i)=s][Var(X_i)^{-1}R'Var(\tilde{X}_i^U)] + E[e_i^{\tilde{X}^U}(\Upsilon_i)|s_i=s] \end{aligned} \tag{41}$$

A3.1 Proof of Proposition 2:

The key steps of the proof are identical to first steps of the proof of Proposition 1 that lead to (11) and (12). These say that

$$\begin{aligned} X_s^U &\equiv E[X_i^U|s(i)=s] = E[[E(X_i^U|\Upsilon_i)]|s(i)=s] \\ X_s &\equiv E[X_i|s(i)=s] = E[[E(X_i|\Upsilon_i)]|s(i)=s]. \end{aligned}$$

Next we find expressions for $E[X_i^U|\Upsilon_i]$ and $E[X_i|\Upsilon_i]$ involving $E^*[\tilde{X}_i^U|\Upsilon_i]$ and $E^*[X_i|\Upsilon_i]$ and $e_i^{\tilde{X}^U}$ and e_i^X By definition of a linear projection,

$$E^*[\tilde{X}_i^U|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1}\Theta^{U'}Var(\tilde{X}_i^U) \tag{42}$$

$$E^*[X_i|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1}\tilde{\Theta}' Var(X_i). \tag{43}$$

Assumption A4 says that $\Theta^U = R\tilde{\Theta}$. Substituting for $\Theta^{U'}$in (42) and using (43) for $E^*[X_i|\Upsilon_i]$ leads to

$$\begin{aligned} & E^*[\tilde{X}_i^U|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1}\tilde{\Theta}'R'Var(\tilde{X}_i^U)) \\ & = \Upsilon_i Var(\Upsilon_i)^{-1}\tilde{\Theta}'Var(X_i)Var(X_i)^{-1}R'Var(\tilde{X}_i^U)) \\ & = E^*[X_i|\Upsilon_i]Var(X_i)^{-1}R'Var(\tilde{X}_i^U). \end{aligned} \tag{44}$$

Using

$$E[X_i^U|\Upsilon_i] = E[X_i|\Upsilon_i]\Pi_{X^U X} + E[\tilde{X}_i^U|\Upsilon_i]. \tag{45}$$

and (39), (40) and (44), we obtain:

$$E[X_i^U|\Upsilon_i] = [E^*[X_i|\Upsilon_i] + e_i^X]\Pi_{X^U X} + E^*[X_i|\Upsilon_i]Var(X_i)^{-1}R'Var(\tilde{X}_i^U) + e_i^{\tilde{X}^U}. \tag{46}$$

The final step is to take expectations of both sides of the above equation conditional on $s(i) = s$ and use (11) and (12). Doing so leads to

$$\begin{aligned} X_s^U &= E[E^*[X_i|\Upsilon_i] + e^{X_i}|s_i = s][\Pi_{X^U X} + Var(X_i)^{-1}R'Var(\tilde{X}_i^U)] \\ &\quad -E(e_i^X|s_i = s)[Var(X_i)^{-1}R'Var(\tilde{X}_i^U)] + E[e_i^{\tilde{X}^U}|s_i = s]. \\ &= X_s[\Pi_{X^U X} + Var(X_i)^{-1}R'Var(\tilde{X}_i^U)] \\ &\quad -E(e_i^X|s_i = s)[Var(X_i)^{-1}R'Var(\tilde{X}_i^U)] + E[e_i^{\tilde{X}^U}|s_i = s] \end{aligned}$$

where the second and third terms combine to form an approximation error. This completes the proof.

A4 Deriving an Analytical Formula for X_s^U when the Spanning Assumption (A5) Is Not Satisfied

We begin by introducing new notation that will be necessary to generalize Proposition 1 in the case where Assumption (A5) is not satisfied.

Partition X_i^U into a subset X_{1i}^U that is correlated with X_i and a subset X_{2i}^U that is not correlated with X_i. Let L denote the number of elements of X_i, L^{1U} denote the number of elements of X_{1i}^U, and let L^{2U} denote the number of elements of X_{2i}^U. Recall that Assumption 5.2 will fail if X_{2i}^U affects preferences for an amenity that neither X_i nor X_{1i}^U affect preferences for.

Denote by A^{U2} the subvector of A that is not contained in A^X. Similarly, let K_1 be the number of amenities in A^X and let K_2 capture the number of amenities in A^{U2}. Then consider writing the taste matrix Θ^U as:

$$\Theta^U = \left\{ \begin{array}{cc} \Theta_{11}^U & \Theta_{12}^U \\ \Theta_{21}^U & \Theta_{22}^U \end{array} \right\} = \left\{ \begin{array}{cc} \Theta_{11}^U & 0 \\ \Theta_{21}^U & \Theta_{22}^U \end{array} \right\}$$

Where Θ_{11}^U is $L^{U1} \times K_1$, Θ_{21}^U is $L^{U2} \times K_1$, Θ_{12}^U is $L^{U1} \times K_2$, and Θ_{22}^U is $L^{U2} \times K_2$. Note that since X_{1i}^U does not affect WTP for any amenities in A^{U2}, $\Theta_{12}^U = 0$. Similarly, consider writing the taste matrix Θ as:

$$\Theta = \left\{ \begin{array}{cc} \Theta_1 & \Theta_2 \end{array} \right\} = \left\{ \begin{array}{cc} \Theta_1 & 0 \end{array} \right\}$$

Where Θ_1 is $L \times K_1$ and $\Theta_2 = 0$ is $L \times K_2$.

We can then write $\tilde{\Theta}$ as:

$$\tilde{\Theta} = \left\{ \begin{array}{cc} \tilde{\Theta}_1 & \tilde{\Theta}_2 \end{array} \right\} = \left\{ \begin{array}{cc} \Theta_1 + \Pi^1_{X^U X} \Theta^U_{11} & 0 \end{array} \right\}$$

Consider replacing assumption (A5) with the following assumptions, (A6) and (A7):

- (A6): There exists an $L^{U1} \times L$ matrix R_1 such that $\Theta^U_{11} = R_1 \tilde{\Theta}_1$.
- (A7): There exists an $L^{U2} \times L$ matrix R_2 such that $\Theta^U_{21} = R_2 \tilde{\Theta}_1$

We can also define the $L^U \times L$ matrix R as:

$$R = \left\{ \begin{array}{c} R_1 \\ R_2 \end{array} \right\}$$

Given these definitions and additional assumptions, we are now ready to develop a more general expression for $E[\tilde{X}^U_i | s(i) = s]$. We begin by generalizing the expression for $E[\tilde{X}^U_i | \Upsilon_i]$. Note first that since $E[X_i|\Upsilon_i]$ and $E[X^U_i|\Upsilon_i]$ are linear in Υ_i (from Assumption (A4), $E[\tilde{X}^U_i|\Upsilon_i]$ is also linear in Υ_i. Basic regression theory then implies that

$$E[\tilde{X}^U_i|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1} Cov(\Upsilon'_i, \tilde{X}^U_i) \tag{47}$$

$$E[X_i|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1} Cov(\Upsilon'_i, X_i). \tag{48}$$

Next, recall that we can write Υ_i as:

$$\Upsilon_i = X_i \tilde{\Theta} + \tilde{X}^U_i \Theta^U + W_i$$

where X_i, $\tilde{X}^U_i$, and W_i are mutually uncorrelated by construction. This leads to the following expression for $Cov(\Upsilon_i, \tilde{X}^U_i)$:

$$\begin{aligned}
Cov(\Upsilon'_i, \tilde{X}^U_i) &= Cov(\Theta^{U'} \tilde{X}^{U\prime}_i, \tilde{X}^U_i) = Cov(\left\{ \begin{array}{cc} \Theta^{U\prime}_{11} & \Theta^{U\prime}_{21} \\ \Theta^{U\prime}_{12} & \Theta^{U\prime}_{22} \end{array} \right\} \left\{ \begin{array}{c} \tilde{X}^{U\prime}_{i1} \\ \tilde{X}^{U\prime}_{i2} \end{array} \right\}, \left\{ \begin{array}{cc} \tilde{X}^U_{i1} & \tilde{X}^U_{i2} \end{array} \right\}) \\
&= \left\{ \begin{array}{cc} Cov(\Theta^{U\prime}_{11}\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{1i}) + Cov(\Theta^{U\prime}_{21}\tilde{X}^{U\prime}_{2i}, \tilde{X}^U_{1i}) & Cov(\Theta^{U\prime}_{11}\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{2i}) + Cov(\Theta^{U\prime}_{21}\tilde{X}^{U\prime}_{2i}, \tilde{X}^U_{2i}) \\ Cov(\Theta^{U\prime}_{12}\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{1i}) + Cov(\Theta^{U\prime}_{22}\tilde{X}^{U\prime}_{2i}, \tilde{X}^U_{1i}) & Cov(\Theta^{U\prime}_{12}\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{2i}) + Cov(\Theta^{U\prime}_{22}\tilde{X}^{U\prime}_{2i}, \tilde{X}^U_{2i}) \end{array} \right\} \\
&= \left\{ \begin{array}{cc} \Theta^{U\prime}_{11} Var(\tilde{X}^U_{1i}) + \Theta^{U\prime}_{21} Cov(\tilde{X}^{U\prime}_{2i}, \tilde{X}^U_{1i}) & \Theta^{U\prime}_{11} Cov(\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{2i}) + \Theta^{U\prime}_{21} Var(\tilde{X}^U_{2i}) \\ \Theta^{U\prime}_{22} Cov(\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{2i}) & \Theta^{U\prime}_{22} Var(\tilde{X}^U_{2i}) \end{array} \right\} \\
&= \left\{ \begin{array}{cc} \tilde{\Theta}'_1 R'_1 Var(\tilde{X}^U_{1i}) + \tilde{\Theta}'_1 R'_2 Cov(\tilde{X}^{U\prime}_{2i}, \tilde{X}^U_{1i}) & \tilde{\Theta}'_1 R'_1 Cov(\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{2i}) + \tilde{\Theta}'_1 R'_2 Var(\tilde{X}^U_{2i}) \\ \Theta^{U\prime}_{22} Cov(\tilde{X}^{U\prime}_{1i}, \tilde{X}^U_{2i}) & \Theta^{U\prime}_{22} Var(\tilde{X}^U_{2i}) \end{array} \right\}
\end{aligned}$$

Where the last line imposes (A6), (A7) and $\Theta^U_{12} = 0$.

Similarly, we have:

$$Cov(\Upsilon_i, X_i) = Cov(\tilde{\Theta}' X_i, X_i) = \tilde{\Theta}' Var(X_i) = \tag{49}$$

$$\left\{ \begin{array}{c} \tilde{\Theta}_1' \\ \tilde{\Theta}_2' \end{array} \right\} Var(X_i) = \left\{ \begin{array}{c} \Theta_1' + \Theta_{11}^{U'} \Pi_{X^U X}^{1\prime} \\ 0 \end{array} \right\} Var(X_i) \tag{50}$$

Plugging in the formulas for $Cov(\Upsilon_i', \tilde{X}_i^U)$ and $Cov(\Upsilon_i', \tilde{X}_i)$ into 47 and 48 , we obtain:

$$E[\tilde{X}_i^U | \Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1} \left\{ \begin{array}{ll} \tilde{\Theta}_1' R_1' Var(\tilde{X}_{1i}^U) + \tilde{\Theta}_1' R_2' Cov(\tilde{X}_{1i}^U, \tilde{X}_{2i}^U) & \tilde{\Theta}_1' R_1' Cov(\tilde{X}_{1i}^U, \tilde{X}_{2i}^U) + \tilde{\Theta}_1' R_2' Var(\tilde{X}_{2i}^U) \\ \Theta_{22}^{U\prime} Cov(\tilde{X}_{1i}^U, \tilde{X}_{2i}^U) & \Theta_{22}^{U\prime} Var(\tilde{X}_{2i}^U) \end{array} \right\} \tag{51}$$

$$E[X_i | \Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1} \left\{ \begin{array}{c} \tilde{\Theta}_1' \\ 0 \end{array} \right\} Var(X_i). \tag{52}$$

Using (52), we can rewrite (51) as:

$$E[\tilde{X}_i^U | \Upsilon_i] = E[X_i | \Upsilon_i] Var(X_i)^{-1} \left\{ \begin{array}{cc} R_1' & R_2' \end{array} \right\} \left\{ \begin{array}{ll} Var(\tilde{X}_{1i}^U) & Cov(\tilde{X}_{1i}^U, \tilde{X}_{2i}^U) \\ Cov(\tilde{X}_{2i}^U, \tilde{X}_{1i}^U) & Var(\tilde{X}_{1i}^U) \end{array} \right\} \tag{53}$$

$$+ \Upsilon_i Var(\Upsilon_i)^{-1} \left\{ \begin{array}{ll} 0 & 0 \\ \Theta_{22}^{U\prime} Cov(\tilde{X}_{1i}^U, \tilde{X}_{2i}^U) & \Theta_{22}^{U\prime} Var(\tilde{X}_{2i}^U) \end{array} \right\}$$

$$= E[X_i | \Upsilon_i] Var(X_i)^{-1} R' Var(\tilde{X}_i^U) + \Upsilon_i Var(\Upsilon_i)^{-1} \left\{ \begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array} \right\} Var(\tilde{X}_i^U) \tag{54}$$

Plugging back into the original iterated expectations formula and taking expectations at the school level, we recover:

$$\tilde{X}_s^U = X_s Var(X_i)^{-1} R' Var(\tilde{X}_i^U) + \Upsilon_s Var(\Upsilon_i)^{-1} \left\{ \begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^U \end{array} \right\} Var(\tilde{X}_i^U) \tag{55}$$

Note that in equilibrium $E[\Upsilon_i | s(i) = s]$ will depend on the full joint distribution of amenities and the joint distribution of Υ_i. With a finite number of students and schools and with idiosyncratic student-school match components in preferences (ε_{is}), there exists no closed-form solution for the equilibrium mapping between the amenity vector A_s and school averages of the WTP for amenities Υ_s.

However, we can gain additional insight by re-considering the continuous version of the model analyzed in Altonji-Mansfield (2014). In that context we assumed a continuum of schools and therefore a continuous joint distribution of amenity vectors. In Appendix A3 of Altonji-Mansfied (2014), we solve for an explicit unique equilibrium mapping between A_s and Υ_s under the assumptions that a) $[X_i, X_i^U, W_i]$ and $A_{s(i)}$ are each jointly normally distributed (with variance matrices Σ_Υ and Σ_A respectively), b) and the equilibrium allocation takes a linear form: $A_{s(i)} = \Psi \Upsilon_i'$. The unique

equilibrium mapping takes the form:

$$\Psi = \Sigma_{\Upsilon'}^{-1/2}(\Sigma_{\Upsilon'}^{1/2}\Sigma_A\Sigma_{\Upsilon'}^{1/2})\Sigma_{\Upsilon'}^{-1/2} \tag{56}$$

Note that the spanning condition (A5) is not necessary to derive the equilibrium relationship in equation 56.

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

$$\Upsilon_i = (\Psi^{-1}A_{s(i)})'. \tag{57}$$

In the continuous version of the model, every individual at the same school has the same value of Υ_i. Thus, we also obtain:

$$E(\Upsilon_i|s = s(i)) \equiv \Upsilon_s = (\Psi^{-1}A_{s(i)})'. \tag{58}$$

Substituting equation (58) into the formula in the previous section, we obtain:

$$\tilde{X}_s^U = X_s Var(X_i)^{-1}R'Var(\tilde{X}_i^U) + A'_{s(i)}\Psi^{-1}Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U) \tag{59}$$

This shows more clearly that the variances and covariances involving X_i, $\tilde{X}_i^U$ and W_i play a role, and that $Var(A_s)$ plays a role.

However, note that the variance of $A'_{s(i)}\Psi^{-1}Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U)$ is not the variance of the component of X_s^U that is not controlled for by X_s. This is because the two terms in equation (59) for X_s^U co-vary.

Next, recall the composition of Υ_i:

$$\Upsilon_i = X_i\tilde{\Theta} + \tilde{X}_i^U\Theta^U + W_i \tag{60}$$

Substituting equation (60) into equation (55), we obtain:

$$\tilde{X}_s^U = X_s\{Var(X_i)^{-1}R'Var(\tilde{X}_i^U) + [Xs\Theta + \tilde{X}_{1s}^U\Theta_1^U + \tilde{X}_2^U\Theta_2^U + W_s]Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U) \tag{61}$$

Now suppose that in addition to Assumption (A4), we assume that $E[W_i|\Upsilon_i]$ is also linear in Υ_i, so that:

$$E[W_i|\Upsilon_i] = \Upsilon_i Var(\Upsilon_i)^{-1}Cov(\Upsilon_i', W_i). \tag{62}$$

If we take iterated expectations of equations (48), (47), and (62) conditional on school $s(i)$ and replace Υ_s with $(\Psi^{-1}A_{s(i)})'$, we obtain:

$$\tilde{X}_s^U = A_{s(i)}{}'\Psi^{-1}Var(\Upsilon_i)^{-1}\Theta^{U'}Var(\tilde{X}_i^U) \tag{63}$$

$$X_s = A_{s(i)}{}'\Psi^{-1}Var(\Upsilon_i)^{-1}\tilde{\Theta}'Var(X_i) \tag{64}$$

$$W_s = A_{s(i)}{}'\Psi^{-1}Var(\Upsilon_i)^{-1}Var(W_i) \tag{65}$$

Collecting terms involving X_s and substituting equations (63) and (65) into (61) yields:

$$\tilde{X}_s^U = X_s\{Var(X_i)^{-1}R'Var(\tilde{X}_i^U) + \tilde{\Theta}'Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U)\} \tag{66}$$

$$+A_{s(i)}{}'\Psi^{-1}Var(\Upsilon_i)^{-1}\Theta^{U'}Var(\tilde{X}_i^U)Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U) \tag{67}$$

$$+A_{s(i)}{}'\Psi^{-1}Var(\Upsilon_i)^{-1}Var(W_i)Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U) \tag{68}$$

Even this doesn't let us decompose $Var(\tilde{X}_s^U)$ into a term involving X_s and an uncorrelated residual piece, because $A_{s(i)}$ will be correlated with X_s.

But consider projecting the amenity subvectors $A_s^{X\prime}$ and $A_s^{U2\prime}$ onto X_s:

$$A_s^{X\prime} = X_s\Pi_{A^X X_s} + \tilde{A}_s^{X\prime} = X_s\Pi_{A^X X_s} \tag{69}$$

$$A_s^{U2\prime} = X_s\Pi_{A^{U2}X_s} + \tilde{A}_s^{U2\prime} \tag{70}$$

where $\Pi_{A^X X_s}$ is an $L \times K_1$ projection matrix, $\Pi_{A^X X_s}$ is an $L \times K_2$ projection matrix, and $\tilde{A}'_{s(i)}$ $\tilde{A}_s^{U2\prime}$ are the residuals from these projections. Note that $\tilde{A}_s^{X\prime} = 0$ as long as $\tilde{\Theta}_1$ is full rank (essentially Assumption A5.1 adapted to the linear utility case).

This implies:

$$\tilde{X}_s^U = X_s[Var(X_i)^{-1}R'Var(\tilde{X}_i^U) + \tilde{\Theta}'Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U) + M]$$

$$+\{0, \tilde{A}_s^{U2\prime}\}\Psi^{-1}Var(\Upsilon_i)^{-1}[\Theta^{U'}Var(\tilde{X}_i^U) + Var(W_i)]\Theta^U Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U) \tag{71}$$

where the matrix M is

$$M = \{\Pi_{A^X X_s}, \Pi_{A^{U2}X_s}\}\Psi^{-1}Var(\Upsilon_i)^{-1}[\Theta^{U'}Var(\tilde{X}_i^U) + Var(W_i)]\Theta^U Var(\Upsilon_i)^{-1}\left\{\begin{array}{cc} 0 & 0 \\ 0 & \Theta_{22}^{U\prime} \end{array}\right\}Var(\tilde{X}_i^U) \tag{72}$$

While cumbersome, the second term in equation (71 provides an expression for the component of

$\tilde{X}_s^U$ that cannot be predicted by X_s (and thus may be a source of bias in our lower bound estimates of the variance in school/neighborhood treatment effects). The variance in this component depends on the following five factors: a) the full joint distribution of amenities (through Ψ); b) the joint distribution of the WTP index Υ_i (entering via the covariance matrix $Var(\Upsilon_i)$); c) the matrix Θ^U mapping unobserved individual characteristics into willingness to pay for particular amenities; d) the joint distribution of the residual component of unobserved outcome-relevant student characteristics (entering via the covariance matrix $Var(\tilde{X}_i^U)$; and e) the joint distribution of the unobserved outcome-irrelevant (but school choice-relevant) student characteristics (entering via the covariance matrix $Var(W_i)$.

Given the complicated manner in which each of these five factors enters the second term in equation (71), there does not appear to be any straightforward way to place an bound on the variance in this error component.

A5 Monte Carlo Evidence on the Properties of the Control Function Estimator

This section describes a set of monte carlo simulations designed to explore the performance of our control function estimator across a number of key dimensions. As we noted in section 4, a full characterization of these finite-sample properties is not feasible. Instead, we focus on sensitivity to deviations in a set of key parameters from a stylized test case that is rich enough to reveal the strengths and weaknesses of our approach. In the first set of simulations, we restrict attention to cases in which the conditions of Proposition 1 are satisfied in an infinite population, and consider the sensitivity of the performance of the control function approach in removing bias from sorting on unobservables to various parameters capturing the structure of tastes, amenities, school sizes, and survey sampling design. Then, in a second set of simulations, we fix the parameters considered in the first set of simulations at a set of baseline values, and examine the sensitivity of our approach to violations of the key spanning condition in Proposition 1 that vary in nature and degree. Section A5.1 lays out the simulation methodology, while section A5.2 presents and interprets the results.

A5.1 Methodology

The stylized test case we consider is one in which:

1. The elements of $[X_i, X_i^U, W_i]$ are jointly normally distributed; the elements of W_i are independent of each other and $[X_i, X_i^U]$, and each pair of characteristics in $[X_i, X_i^U]$ features a .25 correlation.[44]
2. The latent amenity vectors A_s are normally distributed with a .25 correlation between any pair of amenities across schools.

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

3. The matrices of taste parameters Θ and Θ^U represent draws from a multivariate normal distribution in which (a) $corr(\Theta_{k\ell}, \Theta_{jm}) \equiv \rho$ if $j = k$ or $\ell = m$, and 0 otherwise, (b) $corr(\Theta^U_{k\ell}, \Theta^U_{jm}) = \rho$ if $j = k$ or $\ell = m$, and 0 otherwise, and (c) $corr(\Theta_{k\ell}, \Theta^U_{jm}) = \rho$ if $\ell = m$, and 0 otherwise.
4. The variances of the elements of A_s, $[X_i, X_i^U, W_i]$, and $\varepsilon_{i,s}$ (i.i.d. draws from a normal distribution) 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.
5. There are no school/neighborhood effects, so that $Y = X_i\beta + x_i^U$, where $x_i^U \equiv X_i^U\beta^U$. Consequently, our estimating equation also omits the school level controls Z_{2s} that are not averages of student characteristics. These simplifications allow us to focus attention exclusively on the extent to which a vector of group averages of observable individual characteristics can absorb between-school variation in the outcome contributions of unobservable individual characteristics.
6. All the observable and unobservable characteristics in X_i and X_i^U are equally important in determining the outcome, so that each characteristic features the same (unit) variance, $\beta_\ell = 1 \; \forall \; \ell$, and $\beta_\ell^U = 1 \; \forall \; \ell$.

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 (e.g. ethnic composition, crime, principal quality) drive most of the systematic sorting of students to schools. Given restrictions 1-6, we complete the model by choosing particular sets of seven remaining parameters. The first parameter is students per school. For simplicity, we impose that each school has capacity equal to a common student/school ratio.[45] The student/school ratio is denoted "# Stu" in online Appendix Table A8. The second parameter is the total number of school/neighborhood combinations available (denoted "# Sch").

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

[45] 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.

form a linear subspace of the rows of $\tilde{\Theta}$, as required by Proposition 1.

The seventh parameter determines ρ, introduced in the definition of our stylized test case, which governs the correlation between pairs of random variables from which each $(\Theta_{k\ell}, \Theta_{jm})$ or $\Theta^U_{k\ell}, \Theta^U_{jm})$ is a draw. If ρ is high, then student characteristics that have a strong positive effect on willingness to pay for one amenity factor will also tend to have a relatively strong positive effect on WTP for other amenities. And if ρ is high, then amenities that are strongly weighted by one characteristic are likely to be strongly weighted by other characteristics. That is, WTP for some amenity factors may generally be particularly sensitive to student characteristics.

In addition, in a second set of simulations we hold fixed these seven parameters at their baseline values, and consider 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). These simulations consider robustness of the control function approach to changes in the structure of the three matrices that determine whether a one-to-one mapping from a vector of group-average unobservables to a vector of group-average observables exists at the population level: (1) the projection matrix $\Pi_{X^U X}$, which captures the degree to which individual-level unobservables project onto the space of individual-level observables, (2) the taste matrix Θ, which captures the degree to which each of the student-level observables affects tastes for each of the school/neighborhood amenities, and (3) the corresponding taste matrix for unobservable student characteristics, Θ^U.

We have two related metrics for evaluating the effectiveness of our control function approach. The first is the fraction of the between-group variance in the outcome contribution of unobservable individual-level characteristics $(Var(x^U_s) \equiv Var(X^U_s \beta^U)$ 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^U_s, on the vector X_s. In cases where the conditions of Proposition 1 are satisfied, the R^2 should converge to 1 as the number of students per school gets large. However, the rate at which it does so is important for the efficacy of the control function approach.

The second metric is $[(1-R^2)Var(x^U_s)]/Var(Y_i)$, which is the fraction of the total variance in the outcome Y_i accounted for by residual variance of x^U_s not accounted for by X_s. In the results tables presented in the next subsection, we denote our measures "R-sq" and "Resid" (short for "residual sorting variance fraction").

We present values of R-squared and the residual sorting variance fraction from specifications where the full population of students is used to calculate the school averages of observables $\overline{X}_s$ that compose the control function (denoted "R-sq (All)" and "Resid (All)", respectively), as well as values from specifications in which random samples of 10, 20, or 40 students from each school are used to calculate $\overline{X}_s$ (these values are denoted "R-sq (10/20/40)" and "Resid (10/20/40)", respectively, in our tables).

We draw X_i, X_i^U, W_i, and $\{\varepsilon_{is}\}$ from the distributions described above to calculate the WTP of each household for each school.[46] Since our method does not require observation of the equilibrium price function $P(A)$, rather than iterating on an excess demand function to find the equilibrium matching, we instead exploit the fact that a perfectly competitive market will always lead to a pareto efficient allocation. The problem of allocating students to schools to maximize total consumer surplus can be written as a linear programming problem, and solved quickly at relatively large scale using the simplex method combined with sparse matrix techniques.[47]

A5.2 Simulation Results

The simulation results are presented in online Appendix Table A8. 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 within-row and within-column correlation ρ among the elements of the random matrices from which the taste weight matrices Θ and Θ^U are drawn is assumed to be .25.

The first takeaway from Row (1) is that the control function approach is extremely 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: 99.8 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 9). Furthermore, the control function only leaves two hundredths of a percent of the variance in the outcome Y_i that can be attributed to residual between-school sorting (Column 10).

The second insight from Row (1) is that the performance of the control function may suffer somewhat when estimation is based on small subsamples of students at each school. We see that the R-squared falls from .998 to .896 when school averages are merely approximated based on samples of 10 students (top entry in column (11)). Increasing the sample size to 20 students per school (middle entry in column (11)) raises the R-squared to .941, while increasing it further to 40 students per school (bottom entry in column (11)) raises the R-squared to .967. Column (12) shows that the fraction of the outcome variance consisting of residual between-school sorting unabsorbed by the control function is .013/.007/.004 when 10/20/40 student samples, respectively, are used to construct the vector of school averages, X_s.

Rows (2) and (3) illustrate the impact of adapting the specification in Row (1) by decreasing or

[46] To minimize the statistical "chatter" introduced by the particular Θ and Θ^U matrices that we happened to draw, we drew ten different sets of Θ and Θ^U matrices from the prescribed distributions, ran the simulations for each parameter set for each of these sets of matrices, and then averaged the results across the ten iterations within each parameter set.

[47] 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.

increasing the number of individuals per group. Decreasing school sizes from 1000 to 500 decreases the R-squared from .998 to .997, while increasing from 1000 to 2000 increases the R-squared to .999 (column (9)). Perhaps not surprisingly, more individuals per school has almost no impact on the effectiveness of the control function if the larger number of individuals are not used to construct the group averages of individual characteristics, $\overline{X}_s$. In columns (11) and (12), the R-squared values and residual sorting variance fraction when samples of 10, 20 and 40 students are used to construct X_s (R-sq(10/20/40)) 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 minimal 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 the problems posed by using small samples of students from each school to construct $\overline{X}_s$ (column (11)). Similarly, 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 negligibly. The R-squared is effectively unchanged when the full population of students is used to construct X_s, but drops modestly from Row (1) to Row (5) when samples of 10, 20, or 40 students are used instead. Nonetheless, the high R-squared and low variance of the residual sorting component in Row (5) reveals that our approach works well even if households only consider a relatively small number 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.[48] Doubling the number of elements of X_i and X_i^U increases the R-squared from .9983 in Row (1) to .9996, and decreases the fraction of outcome variance attributable to the residual sorting component to one-hundredth of a percentage point. 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. Column 11 shows that when only 10 students are used to construct sample school averages, doubling the control set from 10 to 20 characteristics increases the R-squared from .896 to .939. 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 very slightly reduces the effectiveness of the control function, dropping the R-squared from .9983 in Row (1) to .9947. 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. However, Row (8), when compared to

[48] 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 W_i to compensate. While the control function absorbs a slightly smaller *fraction* of the between-school variance of the regression index of 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 outcome-relevant unobservables is small. Thus, there is very little potential bias to be absorbed.

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 changes as 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 $\Theta^U_{k\ell}$ elements are drawn independently ($\rho = 0$). Compared to Row (1), the R-squared for the full population falls slightly (.9983 to .9953), and the R-squared when samples of (10/20/40) are used to construct X_s falls more substantially, from (.90/.94/.97) to (.72/.82/.89). However, removing correlation among the elements of Θ also reduces the amount of sorting on unobservables to be explained, since the school averages of the various unobservables become more weakly correlated with one another, so that their contributions to student outcomes tend to cancel each other out. Consequently, the fraction of between school outcome variation that can be attributed to residual school-level differences in unobservable student characteristics that is unpredictable based on the vector of school-average observables X_s remains quite small.

Overall, the results in online Appendix Table A8 indicate that the control function approach could potentially work extremely 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.[49] The simulations suggest that the control function works well even when only a small sample of individuals is observed in each group. In online Appendix A8, we use the North Carolina administrative data to directly assess the effect of using smaller samples of students to construct X_s for some of the outcomes and characteristics we actually consider. We find that our main results are relatively insensitive to restricting school sample sizes to match the distribution of sample sizes observed in the NLS72, NELS88, and ELS2002 datasets.

A5.2.1 Performance of the Control Function When the Spanning Condition Fails

Note that all the specifications in online Appendix Table A8 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 to which the addition of group-averages of individual characteristics can substantially reduce bias from sorting on unobservables, even if it cannot completely eliminate the bias. Online Appendix Table A9 considers a number of such scenarios.

Recall from the discussion in Section 3.1 that $\tilde{\Theta}$ can be represented as the sum $\tilde{\Theta} = \Theta + \Pi_{X^U X}\Theta^U$. Thus, in general, the mapping from X^U_s to X_s is generated partly because observed characteristics

[49] In other simulations available upon request, we have also examined the impact of altering the variance of ε_{is}. We find that increasing $Var(\varepsilon_{is})$ reduces the between school variance in both X_i and X^U_i symmetrically, but does not erode the effectiveness of X_s as a control for X^U_s. Intuitively, as $Var(\varepsilon_{is}) \to \infty$, idiosyncratic tastes fully drive choice, and the between school variation in X_i and X^U_i disappears, so that there is no more sorting problem to address.

X_i and unobserved characteristics X_i^U directly affect WTP for overlapping sets of amenities (which creates a degree of overlap in the row spaces of Θ and Θ^U), and partly because X_i indirectly predicts WTP for the amenities for which X_i^U predicts WTP through the correlation between X_i and X_i^U (thereby creating further overlap in the row spaces of Θ and Θ^U). The spanning condition ($\Theta^U = R\tilde{\Theta}$ for some matrix $L^U \times L$ matrix R) is satisfied whenever these two pathways, working in combination, produce a preference matrix $\tilde{\Theta}$ whose row space is a linear superspace of the row space of Θ^U.

Thus, before investigating the impact of violations of the spanning condition, we illustrate the importance of both pathways by considering specifications in which one or the other pathway is shut down. Row (1) is identical to Row (1) of online Appendix Table A8, and represents the baseline case against which the other specifications are compared. Row (2) considers the case in which the entire vector of unobservable characteristics X_i^U is independent of the vector of obervables X_i, so that $\Pi_{X^U X}$ converges to the zero matrix as school sizes become large. However, X_i and X_i^U predict tastes for a common set of amenities ($A_1 - A_5$), so that Θ has (full) rank K and the row space of Θ^U is a linear subspace of the row space of Θ. The results in Row (2) suggest that the control function approach still works quite well when large populations of students at each school are available (R-squared of .972), but suffers somewhat when school averages are constructed using subsamples of 10, 20 or 40 students: R-squared values of .60/.69/.78 (column 10), with substantial residual bias from sorting on unobservables left uncaptured by the control function $\overline{X}_s$ (column 11).

Row (3) considers the opposite case in which the spanning condition is satisfied only through the indirect pathway that operates via the correlation between X_i and X_i^U. Specifically, the observables and unobservables affect tastes for disjoint sets of amenities ($\{A_1, \ldots, A_4\}$ and $\{A_5\}$ respectively), so that the row space of Θ^U is orthogonal to the row space of Θ, but each element of X_i is correlated .25 with each element of X_i^U, so that $\Pi_{X^U X}$ is full rank and the row space of Θ^U is a linear subspace of the row space of $\Pi_{X^U X}\Theta^U$. The results in Row (3) are quite similar to those in Row (2): strong when large samples are used to construct school averages, weaker otherwise. Rows (2) and (3) combined illustrate that the two pathways by which a mapping between X_s and X_s^U may be generated are each sufficient in isolation to produce desirable finite sample properties with large samples of students per school, but it is the blend of both pathways to spanning that produced the surprisingly strong finite sample results in online Appendix Table A8.

The remaining rows of online Appendix Table A9 consider cases in which the spanning condition fails (the row space of Θ^U is not a linear subspace of the row space of $\tilde{\Theta} = \Theta + \Pi_{X^U X}\Theta^U$). Row (4) presents results from the the worst-case scenario: the entire vector of unobservable characteristics is independent of the entire vector of observable characteristics ($\Pi_{X^U X} = 0$), and the unobservable characteristics only predict WTP for an amenity (A_5) that the observable characteristics do not affect taste for (they exclusively weight $A_1 - A_4$). Thus, Θ and Θ^U have orthogonal row spaces as well. Since the group averages of the observables and unobservables are functions of disjoint sets of amenities, it comes as no surprise that only 32% of the variance in X_s^U is predictable given X_s, even when the universe of students at each school is observed (column 8).[50]

[50]The limited explanatory power we do obtain derives from correlation between A_5 and $A_1 - A_4$.

Row (5) alters the scenario from Row (4) 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: 62% 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 affected by one of the ten components of the unobserved vector X_i^U. In Row (6), $X_{i,10}^U$ affects WTP for A_5 only. In Row (7), $X_{i,10}^U$ predicts willingness to pay for all amenities A_1 to A_5. Rows (6) and (7) reveal that our control function performs quite well in these scenarios: it absorbs around 96% of the variation in X_s^U in each case.

Finally, Rows (8) and (9) replicate the scenarios in Rows (6) and (7) but allow each of the unobservable characteristics except the one affecting taste for A_5 ($X_{i,10}^U$) to exhibit a .25 correlation with each of the observed characteristics, so that both $\Pi_{X^U X}\Theta^U$ and Θ would be linear superspaces of Θ^U in the absence of the last unobservable, $X_{i,10}^U$. The performance of the control function for these specifications is every bit as strong as in the baseline specification in Row (1). This suggests that a violation of the spanning condition in Proposition 1 need not produce appreciable bias if it is driven by only a small number of characteristics that weakly affect school/neighborhood choices.

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: namely, cases in which just a few components of the subvector of X_i^U that is orthogonal to X_i affect WTP for just a few additional amenities for which X_i does not affect WTP.

A6 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.

A6.1 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_i, 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.[51] We estimate G_1 and G_2 by applying least squares regression to

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

using the appropriate panel weights from the surveys.

A6.2 Permanent Wage Rates

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

$$Y_{sit} = Y_{si} + e_{sit} + \varsigma_{sit}.$$

In the above equation Y_{si} is i's "permanent" log wage (given that he/she attended high school s) as of the time by which most students have completed education and spent at least a couple of years in the labor market, which we take to be 1979 in the case of NLS72. e_{sit} is a random walk component that evolves as a result of luck in the job search process or within a company, or perhaps changes in motivation or productivity due to health and other factors. We normalize e_{sit} to be 0 in 1979.[52] ς_{sit} includes measurement error and relatively short term factors that have little influence on the lifetime earnings of an individual. The determination of the permanent wage is given by (20). After substituting for Y_{si}, the wage equation is

$$Y_{sit} = X_iB + X_sG_1 + Z_{2s}G_2 + v_{si} + e_{sit} + \varsigma_{sit}.$$

We estimate B by OLS with school fixed effects included.[53]

Let $\tilde{Y}_{sit} \equiv Y_{sit} - X_i\hat{B}$. We estimate G_1 and G_2 by applying OLS to

$$\tilde{Y}_{sit} = \bar{X}_sG_1 + Z_{2s}G_2 + v_{si} + e_{sit} + \varsigma_{sit} \tag{73}$$

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

[51] 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 (A6.1) because some variables, such as test scores, are missing, or because the students are not included in the follow-up surveys that provide the measure of Y_{si}. As we discuss in 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.

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

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

A6.3 High School Graduation and College Enrollment

The methods outlined in online Appendix A6.1 and A6.2 need to be adapted for binary measures such as high school graduation and college attendance. Consequently, for high school graduation we reinterpret Y_{si} to be the latent variable that determines the indicator for whether a student graduates, $HSGRAD_{si}$. That is,

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

Or, after substituting for Y_{si},

$$HSGRAD_{si} = 1(X_i B + X_s G_1 + Z_{2s} G_2 + v_{si} > 0) \tag{74}$$

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

$$HSGRAD_{si} = 1(X_i B + \bar{X}_s G_1 + Z_2 G_2 + (X_s - \bar{X}_s) G_1 + v_{si} > 0) \tag{75}$$

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

A7 Decomposing the Variance in Educational Attainment and Wages

In this section we discuss an analysis of variance based on equation (35) that can be used to place a lower bound on the importance of factors that are common to students from the same school.[54] As with parameter estimation, the details of our procedure depend upon the outcome. We begin with years of postsecondary education.

A7.1 Years of Postsecondary Education

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

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

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

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

Then, from (33) we obtain

$$(Y_{si} - Y_s) = (X_i - X_s)B + (v_{si} - v_s)$$

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

and

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

Thus, one may express the outcome variance as[55]

$$Var(Y_i) = [Var((X_i - X_s)B) + Var(v_{si} - v_s)] + \tag{76}$$

$$[Var(X_s B) + 2Cov(X_s B, X_s G_1) + 2Cov(X_s B, Z_{2s} G_2) + Var(X_s G_1) + \tag{77}$$

$$2Cov(X_s G_1, Z_{2s} G_2) + Var(Z_{2s} G_2) + Var(v_s)] \tag{78}$$

Given an estimate of B, $Var((X_i - X_s)B)$ can be estimated using its corresponding sample variance, $Var((X_i - \overline{X}_s)B)$. $Var(v_{si} - v_s)$ can then be estimated as $\widehat{Var}(Y_{si} - Y_s) - \widehat{Var}((X_i - X_s)B)$, 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 $Var(X_s G_1)$ and $Var(Z_{2s} G_2)$ 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

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

A7.2 Permanent Wage Rates

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

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

where $Cov(\varsigma_{sit}, \varsigma_{sit'})$ is assumed to be 0 given that the observations are seven years apart and $Cov(e_{sit}, e_{sit'}) = 0$ from normalizing e_{sit} to be 0 in 1979. We use the sample estimate of $Cov(Y_{sit}, Y_{sit'})$ as our estimate of $Var(Y_{si})$. We estimate this covariance by subtracting out the global mean for Y_{sit}, calculating the wage product $(Y_{sit})(Y_{sit'})$ for each individual, and taking a weighted average across all the individuals in the sample using the weights discussed in online Appendix A9, adjusting for

[55] The equation below imposes $Cov(X_{si} B, v_{si} - v_s) = 0$, which is implied by our definition of B and $v_{si} - v_s$. 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).

degrees of freedom. Similarly, we estimate the between-school component of the permanent wage, $Var(Y_s)$, 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

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

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

We estimate the corresponding within school component using

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

Given $\widehat{Var}(Y_{si})$, $\widehat{Var}(Y_{si} - Y_s)$, $\widehat{Var}(Y_s)$, $\hat{G}_1$, $\hat{G}_2$, and $\hat{B}$, estimation of the contributions of X_iB, X_sG_1, $Z_{2s}G_2$, v_{si}, and v_s to $Var(Y_{si})$ proceeds as in previous subsection.

A7.3 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 $Var(v_{si} - v_s)$ to one, and define the total variance of the latent variable to be

$$Var(Y_i) = [Var((X_i - X_s)B) + 1] + \tag{79}$$

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

$$2Cov(X_sG_1, Z_{2s}G_2) + Var(Z_{2s}G_2) + Var(v_s)] \tag{80}$$

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.

A7.4 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 500 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.

A8 Using the North Carolina Data to Assess 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 diminish 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-estimating 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 A10 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 the 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 recomputing these estimates for subsamples of the North Carolina population featuring the same distributions of school-specific sample sizes as in NLS72, ELS2002, grade 8 schools in NELS88 and grade 10 schools in NELS88.[56] Focusing first on Column 2, we see that the use of small student samples at each school may actually produce a relatively small amount of bias in our NLS72 results. 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 0.48%. These translate to over/under estimates of the impact of a 10th-90th quantile shift in school quality on the probability of graduation of .0198 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

[56] 10th grade schools in NELS88 are the schools in which the original 8th grade NELS sample are observed in the first follow-up survey.

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 see that the NELS88 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 3.8 percentage points and 2.2 percentage points, respectively. Due the poor performance of the NELS88 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.

A9 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.

A10 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.

A10.1 Sorting of Students Across Class Rooms

Assume for now that the administrator has already determined which teachers to allocate to which courses for which periods of the day, so that a classroom c can be effectively captured by a vector of amenity values A_c. Consider first the case in which none of amenities reflect the demographic makeup of the class, 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 capture 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 W_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 and Θ^U parameter matrices might capture a principal's belief about which types of students are most likely to benefit from a better teacher or difficulty level. Θ and Θ^U might also reflect the 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_{ic} might capture, for example, the desire to fulfill a particular family'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.[57]

Let $\mathscr{I}$ represent the set of students to be allocated, and let $\mathscr{C}$ represent the set of available classrooms (each of which has an associated teacher). The principal's problem is to choose the mapping $c : \mathscr{I} \rightarrow \mathscr{C}$ from students to classrooms that maximizes the sum of student utilities, subject to the constraints that each classroom cannot exceed its capacity and every student (or perhaps student-subject combination at the high school level) can only be assigned to one classroom:

$$\begin{aligned} &\max_{c:\mathscr{I}\rightarrow\mathscr{C}} \sum_{i\in\mathscr{I}} U_{ic(i)} \\ &s.t. \sum_{c'} \mathbb{1}(c(i) = c') = 1 \; \forall \; i \\ &s.t. \sum_{i'} \mathbb{1}(c(i') = c) = \overline{C}_c \; \forall \; c \in \mathscr{C} \end{aligned} \tag{81}$$

where $\mathbb{1}(c(i) = c')$ indicates that student i is assigned to classroom c', and $\overline{C}_{c'}$ is the capacity of classroom c'.

This optimization problem can be recast as a binary integer programming problem:

$$\begin{aligned} &\max_{\mathbf{d}} \; \mathbf{a} * \mathbf{d} \\ &s.t. \; \mathbf{M}_i * \mathbf{d} = 1 \; \forall \; i \in \mathscr{I} \\ &s.t. \; \mathbf{N}_c * \mathbf{d} = \overline{C}_c \; \forall \; c \in \mathscr{C} \\ &s.t. \; \mathbf{d} \in \{0, 1\} \end{aligned} \tag{82}$$

Here $\mathbf{a}$ consists of a $1 \times (I * C)$ row vector of the student utility values associated with each potential student-classroom combination:

$$\mathbf{a} = \begin{pmatrix} U_{11} & \dots & U_{I1} & U_{12} & \dots & U_{I2} & \dots & U_{1C} & \dots & U_{IC} \end{pmatrix}$$

[57] Rothstein (2009) provides 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 possibly other controls. He does not consider the potential for X_c to control for X_c^U.

d consists of a $(I*C)\times 1$ vector of potential student-classroom assignments:

$$\mathbf{d} = \begin{pmatrix} d_{11} \\ \vdots \\ d_{I1} \\ d_{12} \\ \vdots \\ d_{I2} \\ \vdots \\ d_{1C} \\ \vdots \\ d_{IC} \end{pmatrix}$$

where $d_{ic'} = \mathbb{1}(c(i) = c')$ is an indicator for whether student i is assigned to classroom c'.

$\mathbf{M}_i$ consists of a $1 \times I*C$ row vector capturing the number of classrooms to which each student (or student-subject combination) is assigned (restricted here to be $1\ \forall\ i$):

$$\mathbf{M}_i = \left(\underbrace{\overbrace{0\ldots0}^{i-1}\ 1\ \overbrace{0\ldots0}^{I-i}\ldots\overbrace{0\ldots0}^{i-1}\ 1\ \overbrace{0\ldots0}^{I-i}}_{\text{repeated C times}} \right)$$

$\mathbf{N}_c$ consists of a $1 \times I*C$ row vector capturing the number of students assigned to classroom c (restricted to be less than or equal to the classroom capacity $\overline{C}_c$):

$$\mathbf{N}_c = \left(\overbrace{0\ldots0}^{(c-1)*I}\ \underbrace{1\ldots1}_{I}\ \overbrace{0\ldots0}^{(C-c)*I} \right).$$

Koopmans and Beckmann (1957) show that the solution to this binary integer program problem can be sustained by a one-sided set of prices for classrooms $\{P_c\}$. This means that the optimal assignment for each individual is also the solution to his/her utility maximization problem:

$$c(i) = \arg\max_c \tilde{U}_{ic} - P_c \equiv U_{ic} \tag{83}$$

Notice that the structure of this utility maximization problem is isomorphic to that of the decentralized school choice problem from Section 2. Consequently, 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.

A10.2 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 $Z_c\Gamma + z_c^U + \eta_{ci}$, mirroring equation (20). 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.[58] Consider the classroom version of our estimating equation (33):

$$Y_i = X_i\beta + X_cG_1 + Z_{2c}G_2 + v_{ci}, \tag{84}$$

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 (84) represents a standard teacher value-added specification.[59]

Suppose that Proposition 1 can be extended to the 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 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 $G_2 = \Gamma_2 + \Pi_{Z_c^U Z_{2c}}$. 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 $\Pi_{Z_c^U Z_{2c}}$ 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 moderately 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

[58] We assume here that teacher quality is not classroom-specific, as in most teacher value-added models.

[59] Z_{2c} might also include a set of indicators for the teacher's experience level.

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: Estimates of the Contribution of School Systems and Neighborhoods to High School Graduation Decisions Under the Assumption that Only Observables X_i Drive Sorting

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)
No Unobs. Sort.	0.050	0.038	0.072	0.048	0.064	0.052
$Var(X_sG_1 + Z_{2s}G_2 + v_s)$	(0.017)	(0.011)	(0.009)	(0.009)	(0.011)	(0.010)

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)
No Unobs. Sort.: 10th-90th	0.177	0.155	0.156	0.132	0.111	0.100
Based on $Var(X_sG_1 + Z_{2s}G_2 + v_s)$	(0.026)	(0.017)	(0.015)	(0.012)	(0.020)	(0.008)
No Unobs. Sort.: 10th-50th	0.098	0.085	0.093	0.076	0.068	0.060
Based on $Var(X_sG_1 + Z_{2s}G_2 + v_s)$	(0.016)	(0.010)	(0.024)	(0.008)	(0.012)	(0.006)

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 label "No Unobs. Sort. " reports $Var(X_sG_1 + Z_{2s}G_2 + v_s)$, which captures the variance in true school/neighborhood contributions under the assumption that sorting is driven only by X_i.

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 and full models are specified in Table 1.

The full variance decompositions underlying these estimates are presented in Web Appendix Table A19.

Appendix Sections A6 and A7 discuss estimation of model parameters and the variance decompositions. Section 6.3 discusses estimation of the 10-50 and 10-90 differentials.

Table A2: Estimates of the Contribution of School Systems and Neighborhoods to Four Year College Enrollment Decisions Under the Assumption that Only Observables X_i Drive Sorting

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)
No Unobs. Sort.	0.064	0.049	0.067	0.055	0.065	0.043
$Var(X_sG_1 + Z_{2s}G_2 + v_s)$	(0.012)	(0.006)	(0.009)	(0.007)	(0.007)	(0.005)

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)
No Unobs. Sort.: 10th-90th	0.220	0.190	0.245	0.213	0.258	0.205
Based on $Var(X_sG_1 + Z_{2s}G_2 + v_s)$	(0.015)	(0.016)	(0.020)	(0.014)	(0.018)	(0.013)
No Unobs. Sort.: 10th-50th	0.098	0.087	0.112	0.099	0.121	0.098
Based on $Var(X_sG_1 + Z_{2s}G_2 + v_s)$	(0.006)	(0.007)	(0.007)	(0.006)	(0.008)	(0.006)

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

The notes to Table A1 apply, except that Table A2 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 A3: Lower Bound Estimates of the Contribution of School Systems and Neighborhoods to Four Year College Enrollment Decisions (Naive OLS Specification: School-Averages X_s omitted from estimating equation)

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	0.030	0.020	0.023	0.021	0.028	0.020
$Var(Z_{2s}G_2)$	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
LB w/ unobs	0.060	0.046	0.061	0.050	0.061	0.042
$Var(Z_{2s}G_2+v_s)$	(0.009)	(0.007)	(0.007)	(0.006)	(0.008)	(0.006)

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	0.147	0.119	0.140	0.130	0.164	0.136
Based on $Var(Z_{2s}G_2)$	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)
LB w/ unobs: 10th-90th	0.211	0.182	0.229	0.201	0.245	0.199
Based on $Var(Z_{2s}G_2+v_s)$	(0.016)	(0.014)	(0.014)	(0.013)	(0.017)	(0.015)
LB no unobs: 10th-50th	0.068	0.056	0.067	0.062	0.079	0.066
Based on $Var(Z_{2s}G_2)$	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
LB w/ unobs: 10th-50th	0.095	0.083	0.105	0.093	0.116	0.095
Based on $Var(Z_{2s}G_2+v_s)$	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)

"Naive OLS Specification" refers to a specification in which school-averages of individual characteristics $\overline{X}_s$ are omitted from the estimating equation (or equivalently, the coefficient vector G_1 is constrained to be equal to 0).

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

The notes to Table 2 apply, except that Table A3 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 A4: Lower Bound Estimates of the Contribution of School Systems and Neighborhoods to Completed Years of Postsecondary Education in NLS72 data (Naive OLS Specification: School-Averages X_s omitted from estimating equation)

Panel A: Fraction of Variance Attributable to School/Neighborhood Quality				
Lower Bound	Yrs. Postsec. Ed. Fixed Effects		Yrs. Postsec. Ed. No Fixed Effects	
	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)
LB no unobs	0.010	0.006	0.008	0.004
$Var(Z_{2s}G_2)$	(0.002)	(0.002)	(0.001)	(0.001)
LB w/ unobs	0.045	0.029	0.040	0.026
$Var(Z_{2s}G_2+v_s)$	(0.007)	(0.006)	(0.006)	(0.005)

Panel B: Effects on Years of Postsecondary Education of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile				
Lower Bound	Yrs. Postsec. Ed. Fixed Effects		Yrs. Postsec. Ed. No Fixed Effexts	
	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)
LB no unobs: 10th-90th	0.431	0.339	0.396	0.299
Based on $Var(Z_{2s}G_2)$	(0.048)	(0.036)	(0.046)	(0.034)
LB w/unobs: 10th-90th	0.923	0.747	0.875	0.703
Based on $Var(Z_{2s}G_2+v_s)$	(0.074)	(0.079)	(0.069)	(0.074)
LB no unobs: 10th-50th	0.216	0.169	0.198	0.149
Based on $Var(Z_{2s}G_2)$	(0.024)	(0.018)	(0.023)	(0.017)
LB w/unobs: 10th-50th	0.461	0.373	0.437	0.351
Based on $Var(Z_{2s}G_2+v_s)$	(0.037)	(0.040)	(0.035)	(0.037)

"Naive OLS Specification" refers to a specification in which school-averages of individual characteristics $\overline{X}_s$ are omitted from the estimating equation (or equivalently, the coefficient vector G_1 is constrained to be equal to 0).

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

Panel A 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 Table Table 1 for the variables in the baseline model and the full model. The full variance decompositions are in Appendix Table A21. Web Appendix Sections A6 and A7 discuss estimation of model parameters and the variance decompositions.

Table A5: 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 [7,8]	7 [8,8]	7 [7,8]	9 [8,9]	6 [6,7]	8 [7,8]
(3)	90% of Total X_s Var.	12 [11,12]	12 [12,13]	13 [11,13]	16 [14,15]	11 [11,12]	14 [14,15]
(4)	95% of Total X_s Var.	15 [14,15]	15 [14,15]	17 [14,16]	20 [18,19]	14 [14,15]	19 [17,19]
(5)	99% of Total X_s Var.	20 [18,19]	21 [17,18]	22 [19,21]	26 [23,25]	20 [18,20]	25 [23,25]
(6)	100% of Total X_s Var.	24 [21,23]	25 [18,19]	27 [23,26]	32 [29,31]	26 [23,25]	33 [29,31]

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 [3,5]	3 [3,6]	6 [3,7]	5 [5,8]	2 [2,3]	5 [4,7]
(3)	90% of $Var(X_sG_1)$	8 [5,9]	7 [5,10]	10 [6,11]	10 [9,14]	5 [3,7]	11 [8,14]
(4)	95% of $Var(X_sG_1)$	10 [8,13]	9 [7,11]	13 [9,14]	13 [12,17]	7 [5,11]	15 [11,17]
(5)	99% of $Var(X_sG_1)$	14 [13,17]	15 [10,15]	19 [13,19]	20 [19,24]	14 [11,16]	22 [17,23]
(6)	100% of $Var(X_sG_1)$	24 [21,23]	25 [18,19]	27 [23,26]	32 [29,31]	26 [23,25]	33 [29,31]

See Section A2 for details.

Table A6: Estimating the Number of Latent Amenities ($dim(A_s)$): Kleibergen and Paap (2006) Heteroskedasticity-Robust and Cluster Robust Tests of the Rank of the X_s Covariance Matrix (Baseline Specification Results)

		Dataset (Number of Variables in X_s)					
		NLS72 (32)		NELS88 gr8 (39)		ELS2002 (40)	
		Het. Only	Cluster	Het. Only	Cluster	Het. Only	Cluster
# Fact.		(1)	(2)	(3)	(4)	(5)	(6)
H_0	H_A	P-val	P-val	P-val	P-val	P-val	P-val
0	1+	0	NaN	0	NaN	0	NaN
1	2+	0	NaN	0	NaN	0	NaN
2	3+	0	.483	0	NaN	0	NaN
3	4+	0	.332	0	NaN	0	NaN
4	5+	0	.137	0	NaN	0	NaN
5	6+	0	.096	0	NaN	0	NaN
6	7+	0	.049	0	NaN	0	NaN
7	8+	0	.066	0	NaN	0	NaN
8	9+	0	.230	0	NaN	0	NaN
9	10+	0	.270	0	.485	0	NaN
10	11+	0	.210	0	.401	0	NaN
11	12+	0	.199	0	.370	0	NaN
12	13+	0	.211	.001	.389	0	NaN
13	14+	.016	.354	.001	.368	.047	NaN
14	15+	.278	.485	.009	.309	.532	NaN
15	16+	.834	.641	.139	.253	.942	NaN
16	17+	.995	.944	.557	.349	.993	NaN
17	18+	.999	.950	.718	.349	.999	NaN
18	19+	1	.991	.879	.576	1	NaN
19	20+	1	.996	.984	.705	1	NaN
20	21+	1	.990	.998	.747	1	NaN
21	22+	1	.994	.999	.865	1	NaN
22	23+	1	.999	1	.867	1	NaN
23	24+	1	.999	1	.902	1	NaN
24	25+	1	1	1	.918	1	NaN
25	26+	1	1	1	.990	1	.499
26	27+	1	1	1	.986	1	.580
27	28+	1	1	1	.991	1	.690
28	29+	1	1	1	.997	1	.701
29	30+	.998	.999	1	.999	1	.888
30	31+	.982	.978	1	.999	1	.973
31	32+	.921	.940	1	1	1	.991
32	33+	–	–	1	1	1	.997
33	34+	–	–	1	1	1	.999
34	35+	–	–	1	1	1	1
35	36+	–	–	1	1	1	1
36	37+	–	–	.999	.999	1	1
37	38+	–	–	.998	.998	1	1
38	39+	–	–	.985	.985	.998	1
39	40+	–	–	–	–	.886	1

Under the conditions laid out in Proposition 1 of the paper, the rank of the covariance of X_s reveals the number of amenity factors driving sorting. See Section A2 for details. Each element in the table reports a p-value from a test based on Kleibergen and Paap (2006) of the null that the rank of the covariance matrix of school-averages of observable student characteristics X_s is equal to value associated with the row label, against the alternative hypothesis that the rank exceeds this value. "Het. Only" refers to the heteroskedasticity-robust (but unclustered) version of the test. "Cluster" refers to the more general test that is robust to arbitrary correlation in sampling error within clusters. We cluster at the school level. Each test is implemented via the STATA ranktest.ado code provided by Kleibergen and Paap (2006).

'–' indicates that the entry corresponds to a case in which the hypothesized rank associated with the row is as large as or larger than the size of the covariance matrix whose rank is being tested (which corresponds to the number of variables in X_s for the dataset associated with the chosen column), thus obviating the need for a rank test.

'NaN' indicates that the entry corresponds to a case in which the Kleibergen-Paap rank test returned an error due to a non-positive definite covariance matrix.

Table A7: Estimating the Number of Latent Amenities ($dim(A_s)$): Kleibergen and Paap (2006) Heteroskedasticity-Robust and Cluster Robust Tests of the Rank of the X_s Covariance Matrix (Full Specification Results)

		Dataset (Number of Variables in X_s)					
		NLS72 (34)		NELS88 gr8 (49)		ELS2002 (51)	
		Het. Only	Cluster	Het. Only	Cluster	Het. Only	Cluster
# Fact.		(1)	(2)	(3)	(4)	(5)	(6)
H_0	H_A	P-val	P-val	P-val	P-val	P-val	P-val
0	1+	0	NaN	0	NaN	0	NaN
1	2+	0	NaN	0	NaN	0	NaN
2	3+	0	NaN	0	NaN	0	NaN
3	4+	0	NaN	0	NaN	0	NaN
4	5+	0	.471	0	NaN	0	NaN
5	6+	0	.341	0	NaN	0	NaN
6	7+	0	.199	0	NaN	0	NaN
7	8+	0	.185	0	NaN	0	NaN
8	9+	0	.336	0	NaN	0	NaN
9	10+	0	.347	0	NaN	0	NaN
10	11+	0	.351	0	NaN	0	NaN
11	12+	0	.275	0	NaN	0	NaN
12	13+	0	.187	0	NaN	0	NaN
13	14+	.001	.399	0	NaN	0	NaN
14	15+	.074	.693	0	NaN	0	NaN
15	16+	.451	.596	0	NaN	.001	NaN
16	17+	.918	.745	.002	NaN	.136	NaN
17	18+	.998	.925	.021	NaN	.632	NaN
18	19+	.999	.920	.139	NaN	.970	NaN
19	20+	1	.972	.445	.430	.996	NaN
20	21+	1	.998	.762	.377	.999	NaN
21	22+	1	.998	.967	.497	1	NaN
22	23+	1	.999	.998	.576	1	NaN
23	24+	1	1	.999	.590	1	NaN
24	25+	1	1	1	.725	1	NaN
25	26+	1	1	1	.697	1	.499
26	27+	1	1	1	.701	1	.580
27	28+	1	1	1	.636	1	.690
28	29+	1	1	1	.858	1	.701
29	30+	1	1	1	.944	1	.888
30	31+	1	1	1	.952	1	.973
31	32+	1	1	1	.996	1	.991
32	33+	.991	.996	1	.994	1	.997
33	34+	.996	.997	1	1	1	.999
34	35+	–	–	1	1	1	1
35	36+	–	–	1	1	1	1
36	37+	–	–	1	1	1	1
37	38+	–	–	1	1	1	1
38	39+	–	–	1	1	1	1
39	40+	–	–	1	1	1	1
40	41+	–	–	1	1	1	1
41	42+	–	–	1	1	1	1
42	43+	–	–	1	1	1	1
43	44+	–	–	1	1	1	1
44	45+	–	–	1	1	1	1
45	46+	–	–	1	1	1	1
46	47+	–	–	1	1	1	1
47	48+	–	–	.999	.998	1	1
48	49+	–	–	.993	.992	1	1
49	50+	–	–	–	–	.998	.998
50	51+	–	–	–	–	.919	.911

Under the conditions laid out in Proposition 1 of the paper, the rank of the covariance of X_s reveals the number of amenity factors driving sorting. See Section A2 for details. Each element in the table reports a p-value from a test based on Kleibergen and Paap (2006) of the null that the rank of the covariance matrix of school-averages of observable student characteristics X_s is equal to value associated with the row label, against the alternative hypothesis that the rank exceeds this value. "Het. Only" refers to the heteroskedasticity-robust (but unclustered) version of the test. "Cluster" refers to the more general test that is robust to arbitrary correlation in sampling error within clusters. We cluster at the school level. Each test is implemented via the STATA ranktest.ado code provided by Kleibergen and Paap (2006).

'–' indicates that the entry corresponds to a case in which the hypothesized rank associated with the row is as large as or larger than the size of the covariance matrix whose rank is being tested (which corresponds to the number of variables in X_s for the dataset associated with the chosen column), thus obviating the need for a rank test.

'NaN' indicates that the entry corresponds to a case in which the Kleibergen-Paap rank test returned an error due to a non-positive definite covariance matrix.

Table A8: 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	$\frac{Var(X_s^U B^U)}{Var(Y)}$	R-Sq (All)	Resid (All)	R-Sq (10/20/40)	Resid (10/20/40)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	1000	50	50	10	10	5	0.25	.122	0.9983	.0002	.896 .941 .967	.013 .007 .004
(2)	500	50	50	10	10	5	0.25	.123	0.9969	.0004	.896 .941 .968	.013 .007 .004
(3)	2000	50	50	10	10	5	0.25	.122	0.9991	.0001	.896 .940 .967	.013 .007 .004
(4)	1000	100	50	10	10	5	0.25	.122	0.9981	.0002	.881 .932 .963	.015 .008 .005
(5)	1000	50	10	10	10	5	0.25	.100	0.9981	.0001	.869 .927 .960	.013 .007 .004
(6)	1000	50	50	20	20	5	0.25	.122	0.9996	.0001	.939 .967 .983	.008 .004 .002
(7)	1000	50	50	10	10	10	0.25	.136	0.9947	.0007	.898 .939 .962	.014 .008 .005
(8)	1000	50	50	20	20	10	0.25	.135	0.9993	.0001	.946 .971 .984	.008 .004 .002
(9)	1000	50	50	10	10	5	0	.048	0.9953	.0002	.721 .824 .894	.013 .008 .005

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

$\frac{Var(X_s^U \beta^U)}{Var(Y_i)}$: Fraction of variance in the student-level outcome accounted for by between-school variation in the regression index of unobserved student characteristics

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 of both $\overline{X}_s$ and X_s^U are computed using all students)

Resid (All): Fraction of outcome variance accounted for by the residual component of the between-school variation in the regression index of unobserved student characteristics that cannot be predicted based on the vector of observed school-averages $\overline{X}_s$, $[(1-R^2)Var(X_s^U \beta^U)]/Var(Y_i)$ (sample averages of both $\overline{X}_s$ and X_s^U are 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 $\overline{X}_s$ are computed using 10/20/40 students, while school averages X_s^U are computed using all students.)

Resid (10/20/40): Fraction of outcome variance accounted for by the part of the between-school variation in the regression index of unobserved student characteristics that cannot be predicted based on the vector of observed school-averages $\overline{X}_s$ (sample averages $\overline{X}_s$ are computed using 10/20/40 students, while school averages X_s^U are computed using all students.)

Table A9: Monte Carlo Simulation Results: Sensitivity of Control Function Performance to the Spanning Condition in Proposition 1

Row	X/X^U Corr. Structure	WTP for A1-A4 Depends On	WTP for A5 Depends On	Assu. (A5) Satisfied	Assu. (A5.1) Satisfied	Assu. (A5.2) Satisfied	$\frac{Var(X_i^U B^U)}{Var(Y)}$	R-Sq (All)	Resid (All)	R-Sq (10/20/40)	Resid (10/20/40)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Corr = .25 for each pair of (obs. or unobs.) char.	All elements of X_i and X_i^U	All elements of X_i and X_i^U	Yes	Yes	Yes	.122	0.998	.0002	.896 .941 .967	.013 .007 .004
(2)	Elements of X^U independent of elements of X	All elements of X_i and X_i^U	All elements of X_i and X_i^U	Yes	No	Yes	.101	0.972	.0028	.596 .687 .775	.041 .032 .023
(3)	Corr = .25 for each pair of (obs. or unobs.) char.	All elements of X_i	All elements of X_i^U	Yes	Yes	No	.049	0.974	.0012	.699 .777 .837	.015 .011 .008
(4)	Elements of X^U independent of elements of X	All elements of X_i	All elements of X_i^U	No	No	No	.069	0.322	.047	.295 .306 .315	.049 .048 .047
(5)	Elements of X^U independent of elements of X	All elements of X_i and X_i^U	All elements of X_i^U	No	No	No	.098	0.621	.037	.463 .502 .539	.052 .049 .045
(6)	Elements of X^U independent of elements of X	All elements of X_i and X_i^U	$X_{i,10}^U$ only	No	No	No	.109	0.969	.003	.673 .747 .809	.036 .028 .021
(7)	Elements of X^U independent of elements of X	All obs. and unobs. char. except $X_{i,10}^U$	$X_{i,10}^U$ only	No	No	No	.095	0.962	.004	.666 .743 .806	.032 .024 .018
(8)	Corr = .25 for each pair of obs. or unobs. char. except $X_{i,10}^U$ (independent)	All elements of X_i and X_i^U	$X_{i,10}^U$ only	No	No	No	.117	0.998	.0002	.925 .958 .977	.0010 .005 .003
(9)	Corr = .25 for each pair of obs. or unobs. char. except $X_{i,10}^U$ (independent)	All obs. and unobs. char. except $X_{i,10}^U$	$X_{i,10}^U$ only	No	No	No	.131	0.998	.0002	.915 .954 .975	.0010 .005 .003

All specifications share the following parameter values: # Stu. = 1000, # Sch. = 50, # Con. = 50, # Ob = 10, # Un = 10, # Am. = 5, Θ Corr = 0.25 (See Table A8 for definitions of parameters).
The column labeled "X/X^U Corr. Structure" describes the correlation structure among and between the elements of the vectors of observed and unobserved individual characteristics X_i and X_i^U.
The columns labeled "WTP for A1-A4 Depends On" and "WTP for A5 Depends On" specifies which elements of the observable (X_i) and unobservable (X_i^U) characteristics predict willingness-to-pay for amenity factors 1-4 and amenity factor 5, respectively.
The columns labeled "Assu. A5/A5.1/A5.2 Satisfied" specify whether Assumptions A5, A5.1 and A5.2 are satisfied, respectively. In the context of a linear utility function, these assumptions are tantamount to assuming that the taste matrix Θ^U can be written as $\Theta^U = R\tilde{\Theta}$, $\Theta^U = R^A\Theta$, and $\Theta^U = R^B\Pi^{X^UX}\Theta^U$, for some matrix (matrices) R, R^A, and R^B, respectively. Assumption A5 is a necessary condition for Proposition 1 to hold, while Assumptions A5.1 and A5.2 are each sufficient conditions for Condition 2 to hold. See Section 3.2.2 for further discussion of these conditions.

Table A10: 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 A11: Summary Statistics for Student Characteristics in NLS72

Variable	% Imputed	Mean	Std. Dev.
Student Demographics			
1(Female)	0.00	.510	.500
1(Black)	0.00	.123	.328
1(Hispanic)	0.00	.043	.203
1(Asian)	0.00	.012	.108
Student Ability			
Std. Math Score	0.00	-.062	1.01
Std. Reading Score	0.00	-.059	1.01
Student Behavior			
[None]			
Family Background Characteristics			
SES Index	0.00	-.136	1.08
Number of Siblings	2.90	2.70	2.06
1(Both Parents Present)	43.17	.752	.432
1(Mother, Male Guardian)	43.17	.019	.136
1(Mother Only Present)	43.17	.126	.332
1(Father Only Present)	43.17	.039	.194
Father's Years of Educ.	0.74	12.46	2.47
Mother's Years of Educ.	0.00	12.22	2.07
1(Mother's Ed. Missing)	0.00	.005	.071
Log(Family Income)	19.98	10.82	.745
1(Eng. Spoken at Home)	0.46	.911	.285
1(Home Environ. Index)	3.33	.037	1.34
1(No Religion)	0.00	.053	.224
1(Eastern Religion)	0.00	.046	.209
1(Jewish)	0.00	.023	.149
1(Catholic)	0.00	.300	.458
1(Oth. Christian Relig.)	0.00	.196	.397
1(Fath. Occ.: Service)	22.21	.108	.310
1(Fath. Occ.: Security/Military)	22.21	.054	.225
1(Fath. Occ.: Farmer/Laborer)	22.21	.294	.456
1(Fath. Occ.: Craftsman/Technician)	22.21	.208	.406
1(Fath. Occ.: Manager)	22.21	.128	.334
1(Fath. Occ.: Owner)	22.21	.069	.254
1(Fath. Occ.: Professional)	22.21	.137	.344
1(Moth. Occ.: Sales)	18.42	.033	.180
1(Moth. Occ.: Service)	18.42	.060	.238
1(Moth. Occ.: Clerical)	18.42	.148	.355
1(Moth. Occ.: Professional)	18.42	.092	.289
1(Moth. Occ.: Other)	18.42	.095	.293
Parental Beliefs/Desires			
[None]			
Outcomes			
1(Enrolled at a 4-Yr. Coll.)	0.00	.263	.440

Table A12: Summary Statistics for School Characteristics in NLS72

Variable	% Imputed	Mean	Std. Dev.
School Characteristics			
% Minority Students	1.87	.200	.265
1(Catholic School)	3.52	.065	.246
1(Private School)	3.52	.003	.058
% of Teachers with Masters' Deg.	1.03	.405	.210
Teacher Turnover Rate	0.27	.086	.091
Total School Enrollment	0.86	1364	892
Student-to-Teacher Ratio	1.51	20.30	4.36
% of Minority Teachers	2.61	.096	.162
1(Tracking System Exists)	17.80	.752	.432
Age of School Building	1.32	21.85	17.48
Neighborhood Characteristics			
Distance to 4-Year College	4.61	19.97	26.57
Distance to Community College	4.64	18.82	26.4
1(South Region)	0.00	.345	.475
1(Midwest Region)	0.00	.260	.438
1(West Region)	0.00	.173	.378
1(Small Town)	0.00	.290	.454
1(Medium-Sized City)	0.00	.084	.277
1(Suburb of Medium-Sized City)	0.00	.045	.208
1(Large City)	0.00	.109	.311
1(Suburb of Large City)	0.00	.096	.295
1(Huge City)	0.00	.097	.296
1(Suburb of Huge City)	0.00	.082	.285

Table A13: Summary Statistics for Student Characteristics in NELS88

Variable	% Imputed	Mean	Std. Dev.
Student Demographics			
1(Female)	0.00	.515	.500
1(Black)	0.00	.106	.307
1(Hispanic)	0.00	.131	.338
1(Asian)	0.00	.069	.253
1(Immigrant)	6.80	.070	.255
Student Ability			
Std. Math Score (8th grd.)	0.00	.091	1.02
Std. Reading Score (8th grd.)	0.00	.074	1.01
Student Behavior			
Parent checks HW	0.36	.439	.496
# Weekly HW Hours	5.71	6.03	5.24
# Weekly Reading Hours	4.28	2.24	2.68
# Weekly TV Hours	14.15	21.73	10.81
1(Often Missing Pencil)	4.20	.219	.413
1(Fought at School)	1.45	.202	.401
Family Background Characteristics			
SES Index	0.00	.003	1.03
Number of Siblings	0.46	2.30	1.59
1(Both Parents Present)	0.84	.681	.466
1(Mother, Male Guardian)	0.00	.098	.297
1(Mother Only Present)	0.00	.148	.355
1(Father Only Present)	0.00	.044	.205
Father's Years of Educ.	6.38	13.30	2.98
Mother's Years of Educ.	0.00	12.85	2.41
1(Mother's Ed. Missing)	0.00	.023	.148
Log(Family Income)	9.67	10.87	.945
1(Eng. Spoken at Home)	0.87	.871	.335
1(Moth. Is Immigrant)	7.66	.145	.352
1(Fath. Is Immigrant)	8.62	.138	.345
1(Parents Married)	7.70	.793	.405
1(No Religion)	0.00	.024	.154
1(Eastern Religion)	0.00	.048	.214
1(Jewish)	0.00	.014	.119
1(Catholic)	0.00	.308	.462
1(Oth. Christian Relig.)	0.00	.308	.462
1(Home Environ. Index)	6.49	-.013	1.45
1(Fath. Occ.: Service)	24.39	.103	.304
1(Fath. Occ.: Security/Military)	24.39	.044	.206
1(Fath. Occ.: Farmer/Laborer)	24.39	.256	.437
1(Fath. Occ.: Craftsman/Technician)	24.39	.194	.396
1(Fath. Occ.: Dentist/Lawyer/Etc.)	24.39	.062	.241
1(Fath. Occ.: Accountant/Nurse/Etc.)	24.39	.102	.302
1(Fath. Occ.: Manager)	24.39	.130	.337
1(Fath. Occ.: Owner)	24.39	.083	.276
1(Moth. Occ.: Sales)	11.23	.056	.231
1(Moth. Occ.: Service)	11.23	.137	.344
1(Moth. Occ.: Clerical)	11.23	.219	.414
1(Moth. Occ.: Teacher)	11.23	.073	.261
1(Moth. Occ.: Accountant/Nurse/Etc.)	11.23	.092	.289
1(Moth. Occ.: Other)	11.23	.259	.438
Parental Sch. Engage. Index	10.79	-.045	1.54
Parental Beliefs/Desires			
Moth. Desired Educ. for Child	12.63	16.3	2.07
Fath. Desired Educ. for Child	16.09	16.23	2.11
Outcomes			
1(Enrolled at a 4-Yr. Coll.)	0.00	.329	.470

Table A14: Summary Statistics for School Characteristics in NELS88

Variable	% Imputed	Mean	Std. Dev.
School Characteristics			
% Minority Students	1.51	.223	.288
% Limited English Proficient	1.31	.070	.084
% Receiving Free/Reduced Price Lunch	1.49	.233	.231
% in Special Ed.	1.31	.063	.052
% in Remedial Reading	1.19	.099	.127
% in Remedial Math	1.19	.072	.101
Admin's Perceived Sch. Problems Index	1.16	3.10	.681
1(Catholic School)	0.00	.093	.290
1(Private School)	0.00	.072	.258
% of Teachers with Masters' Deg.	3.75	.478	.248
Total School Enrollment	1.05	665.9	373.1
Student-to-Teacher Ratio	1.05	17.74	5.06
% of Minority Teachers	2.92	.108	.183
Log(Minimum Teacher Salary)	2.51	9.76	.180
1(Collectively Bargained Contracts)	1.49	.561	.496
1(Gifted Program Exists)	1.05	.658	.474
Admin.'s Reported Security. Policies Index (1)	1.36	.098	1.14
Admin.'s Reported Security. Policies Index (2)	1.36	-.035	1.07
Neighborhood Characteristics			
1(Urban Neighborhood)	0.00	.253	.434
1(Suburban Neighborhood)	0.00	.429	.495
1(South Region)	0.00	.353	.478
1(Midwest Region)	0.00	.266	.442
1(West Region)	0.00	.196	.397

Table A15: Summary Statistics for Student Characteristics in ELS2002

Variable	% Imputed	Mean	Std. Dev.
Student Demographics			
1(Female)	0.00	.520	.500
1(Black)	0.00	.129	.335
1(Hispanic)	0.00	.141	.348
1(Asian)	0.00	.091	.287
1(Immigrant)	10.78	.100	.300
Student Ability			
Std. Math Score	0.00	.125	1.01
Std. Reading Score	0.00	.108	1.01
Student Behavior			
Parent checks HW	14.43	.343	.475
# Weekly HW Hours	3.72	10.98	9.07
# Weekly Reading Hours	4.06	2.80	4.10
# Weekly Computer Hours	3.92	2.20	1.71
# Weekly TV Hours	4.01	22.73	12.21
1(Often Missing Pencil)	1.71	.166	.372
1(Fought at School)	0.85	.124	.329
Family Background Characteristics			
SES Index	0.00	.099	1.04
Number of Siblings	17.22	2.26	1.51
1(Both Parents Present)	10.44	.615	.487
1(Mother, Male Guardian)	10.44	.118	.322
1(Mother Only Present)	10.44	.174	.379
1(Father Only Present)	10.44	.057	.231
Father's Years of Educ.	9.24	13.95	2.72
Mother's Years of Educ.	0.00	13.65	2.33
1(Mother's Ed. Missing)	0.00	.033	.179
Avg. Grandparents' Educ.	23.77	12.32	1.82
Log(Family Income)	21.01	10.97	.957
1(Eng. Spoken at Home)	13.32	.889	.317
1(Moth. Is Immigrant)	11.38	.215	.411
1(Fath. Is Immigrant)	12.33	.214	.410
1(Parents Married)	10.85	.755	.430
1(No Religion)	18.55	.031	.172
1(Eastern Religion)	18.55	.072	.259
1(Jewish)	18.55	.012	.108
1(Catholic)	18.55	.364	.481
1(Oth. Christian Relig.)	18.55	.184	.387
1(Home Environ. Index)	13.35	.042	1.39
1(Fath. Occ.: Service)	30.74	.108	.310
1(Fath. Occ.: Security/Military)	30.74	.047	.211
1(Fath. Occ.: Farmer/Laborer)	30.74	.228	.419
1(Fath. Occ.: Craftsman/Technician)	30.74	.179	.383
1(Fath. Occ.: Dentist/Lawyer/Etc.)	30.74	.069	.254
1(Fath. Occ.: Accountant/Nurse/Etc.)	30.74	.141	.348
1(Fath. Occ.: Manager)	30.74	.163	.370
1(Fath. Occ.: Owner)	30.74	.060	.238
1(Fath. Occ.: Other)	30.74	.004	.064
1(Moth. Occ.: Sales)	21.10	.047	.212
1(Moth. Occ.: Service)	21.10	.142	.349
1(Moth. Occ.: Clerical)	21.10	.176	.381
1(Moth. Occ.: Teacher)	21.10	.081	.273
1(Moth. Occ.: Accountant etc.)	21.10	.171	.376
1(Moth. Occ.: Other)	21.10	.230	.421
Parental Sch. Engage. Index	20.71	.012	1.55
Parental Beliefs/Desires			
Moth. Desired Educ. for Child	15.90	16.76	2.38
Fath. Desired Educ. for Child	23.04	16.74	2.45
Outcomes			
1(Enrolled at a 4-Yr. Coll.)	0.00	.422	.493

Table A16: Summary Statistics for School Characteristics in ELS2002

Variable	% Imputed	Mean	Std. Dev.
School Characteristics			
% Minority Students	1.53	.331	.306
% Limited English Proficient	4.71	.040	.083
% Receiving Free/Reduced Price Lunch	7.68	.231	.246
% in Special Ed.	5.98	.092	.088
% in Remedial Reading	17.81	.041	.072
% in Remedial Math	19.24	.057	.094
Admin's Perceived Sch. Problems Index	15.74	3.60	.817
1(Catholic School)	1.23	.133	.339
1(Private School)	1.84	.084	.278
% of Teachers with Masters' Deg.	33.72	.459	.216
Teacher Turnover Rate	28.01	.058	.061
Total School Enrollment	0.34	1254	807
Student-to-Teacher Ratio	2.67	16.6	4.09
% of Minority Teachers	37.99	.115	.185
Log(Minimum Teacher Salary)	20.01	10.2	.176
% of Teachers with Certification	3.35	92.4	17.48
Teacher Evaluation Policy Index	14.42	-.002	1.17
Teacher Incentive Pay Index (1)	13.25	-.007	1.46
Teacher Incentive Pay Index (2)	13.25	-.005	1.18
Teaching Technology Index	16.29	.029	1.71
1(High Stakes Competency Exam)	0.00	.994	.075
Observed Sch. Cleanliness/Disorder Index (1)	29.85	-.062	2.02
Observed Sch. Cleanliness/Disorder Index (2)	29.85	.006	1.40
Security Policy Implementation Index (1)	8.56	-.010	1.39
Security Policy Implementation Index (2)	8.56	-.008	1.08
Admin.'s Reported Security. Policies Index (1)	15.78	.019	1.55
Admin.'s Reported Security. Policies Index (2)	15.78	-.007	1.30
Admin.'s Impression of Fac. Quality Index (1)	19.31	-.004	2.36
Admin.'s Impression of Fac. Quality Index (2)	19.31	.004	1.08
Neighborhood Characteristics			
1(Rural within MSA)	0.24	.104	.305
1(Small Town)	0.24	.095	.294
1(Large Town)	0.24	.014	.117
1(Suburb of Medium City)	0.24	.085	.279
1(Suburb of Large City)	0.24	.277	.447
1(Medium City)	0.24	.163	.370
1(Large City)	0.24	.160	.367
1(South Region)	0.00	.367	.482
1(Midwest Region)	0.00	.256	.437
1(West Region)	0.00	.192	.394
Admin. Perception of N-Hood Crime	12.24	.422	.494

Table A17: Summary Statistics for Student Characteristics in North Carolina Administrative Data

Variable	% Imputed	Mean	Std. Dev.
Student Demographics			
1(Female)	0.00	.505	.500
1(Black)	0.00	.276	.447
1(Hispanic)	0.00	.059	.236
1(Asian)	0.00	.023	.149
Student Ability			
Std. Math Score (Grade 8)	13.0	.059	.990
Std. Reading Score (Grade 8)	13.0	.054	.979
Std. Math Score (Grade 7)	15.9	.061	.985
Std. Reading Score (Grade 7)	16.0	.057	.971
1(Gifted in Math)	15.8	.136	.343
1(Gifted in Reading)	15.8	.133	.339
Student Behavior			
1(Daily HW Hours $<$ 1)	17.3	.267	.442
1(Daily HW Hours $>=$ 1 and $<$ 3)	17.2	.463	.499
1(Daily HW Hours $>=$ 3)	17.3	.239	.426
1(Ignore Homework)	17.3	.013	.114
1(Daily TV Hours $<$ 1)	17.3	.226	.418
1(Daily TV Hours $\approx$ 2)	17.3	.270	.444
1(Daily TV Hours $\approx$ 3)	17.3	.222	.416
1(Daily TV Hours $>=$ 4 and $<=$ 5)	17.3	.160	.367
1(Daily TV Hours $>=$ 6)	17.3	.091	.287
1(Daily Free Reading Hours $<=$ 1/2)	17.2	.489	.500
1(Daily Free Reading Hours $\approx$ 1)	17.2	.215	.411
1(Daily Free Reading Hours $>$ 1 and $<=$ 2)	17.2	.110	.313
1(Daily Free Reading Hours $>=$ 2)	17.2	.055	.227
Family Background Characteristics			
1(Highest Parent Education = HS Graduate)	0.00	.221	.415
1(Highest Parent Education = Some College)	0.00	.131	.337
1(Highest Parent Education = Community College)	0.00	.163	.370
1(Highest Parent Education = 4-Yr College Graduate)	0.00	.223	.417
1(Highest Parent Education = Graduate School)	0.00	.104	.306
1(Free/Reduced Price Lunch Eligible)	0.00	.596	.491
1(Limited English Proficiency)	0.54	.027	.161
1(Ever Limited English Proficient)	0.00	.062	.242
Parental Beliefs/Desires			
[None]			
Outcomes			
1(High School Graduate)	0.00	.760	.427

Table A18: Summary Statistics for School Characteristics in North Carolina Administrative Data

Variable	% Imputed	Mean	Std. Dev.
School Characteristics			
# of Books Per Student	0.41	10.85	6.74
1(Magnet School)	0.00	.064	.244
1(Charter School)	0.00	.007	.083
% of Teachers with Advanced Degrees	0.79	.249	.079
% of Classrooms Taught by "High Quality" Teachers	0.03	.956	.060
Teacher Turnover Rate	0.87	.214	.081
Total School Enrollment	0.03	1323	581
Student-to-Teacher Ratio	0.03	15.5	2.02
Neighborhood Characteristics			
1(Remote Rural)	0.00	.028	.166
1(Distant Rural)	0.00	.160	.366
1(Fringe Rural)	0.00	.284	.451
1(Remote Town)	0.00	.006	.078
1(Distant Town)	0.00	.075	.263
1(Fringe Town)	0.00	.050	.218
1(Small Suburb)	0.00	.006	.076
1(Mid-Sized Suburb)	0.00	.049	.216
1(Large Suburb)	0.00	.096	.295
1(Small City)	0.00	.072	.259
1(Midsize City)	0.00	.086	.281

Table A19: 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
$Var(Y_i - Y_s)$	(0.014)	(0.013)	(0.018)	(0.016)	(0.016)	(0.016)
Observable Student-Level (Within):	0.124	0.213	0.162	0.292	0.134	0.221
$Var((X_i - X_s)B)$	(0.004)	(0.006)	(0.013)	(0.015)	(0.040)	(0.042)
Unobservable Student-Level (Within)	0.791	0.706	0.668	0.543	0.740	0.660
$Var(v_{si} - v_s)$	(0.013)	(0.011)	(0.019)	(0.017)	(0.037)	(0.037)
Between School:						
Total	0.085	0.081	0.170	0.164	0.126	0.119
$Var(Y_s)$	(0.014)	(0.013)	(0.018)	(0.016)	(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)	(0.002)	(0.010)	(0.011)	(0.006)	(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)	(0.018)	(0.019)	(0.009)	(0.012)
School-Avg. Student-Level/						
School Char. Covariance	-0.017	-0.008	0.007	0.004	0.001	-0.002
$2 * Cov(X_s G_1, Z_{2s} G_2)$	(0.007)	(0.005)	(0.008)	(0.006)	(0.013)	(0.013)
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)	(0.010)	(0.009)	(0.013)	(0.012)
School Char.	0.018	0.012	0.011	0.006	0.025	0.024
$Var(Z_{2s} G_2)$	(0.008)	(0.005)	(0.006)	(0.005)	(0.010)	(0.010)
Unobservable School-Level	0.031	0.026	0.017	0.010	0.010	0.001
$Var(v_s)$	(0.006)	(0.005)	(0.007)	(0.004)	(0.002)	(0.000)

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 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 **??** - **??**

Table A20: 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
$Var(Y_{is} - Y_s)$	(0.018)	(0.012)	(0.016)	(0.016)	(0.016)	(0.015)
Observable Student-Level (Within):	0.176	0.354	0.192	0.316	0.184	0.330
$Var((X_{si} - X_s)B)$	(0.082)	(0.017)	(0.010)	(0.013)	(0.031)	(0.024)
Unobservable Student-Level (Within)	0.681	0.503	0.584	0.458	0.600	0.461
$Var(v_{si} - v_s)$	(0.070)	(0.015)	(0.016)	(0.014)	(0.026)	(0.019)
Between School:						
Total	0.143	0.143	0.224	0.226	0.215	0.209
$Var(Y_s)$	(0.018)	(0.012)	(0.016)	(0.016)	(0.016)	(0.015)
Observable Student-Level:	0.042	0.062	0.010	0.143	0.079	0.127
$Var(X_sB)$	(0.006)	(0.006)	(0.010)	(0.012)	(0.007)	(0.010)
Student-Level/						
School-Level Covariance	0.037	0.032	0.057	0.027	0.071	0.039
$2 * Cov(X_sB, X_sG_1 + Z_{2s}G_2)$	(0.008)	(0.008)	(0.011)	(0.014)	(0.009)	(0.012)
School-Avg. Student-Level/						
School Char. Covariance	0.000	-0.002	0.004	0.005	-.003	-0.002
$2 * Cov(X_sG_1, Z_{2s}G_2)$	(0.005)	(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
$Var(X_sG_1)$	(0.006)	(0.005)	(0.005)	(0.005)	(0.007)	(0.005)
School Char.	0.026	0.019	0.018	0.015	0.024	0.018
$Var(Z_{2s}G_2)$	(0.006)	(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)	(0.006)	(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 A19 for additional details.

Table A21: Decomposition of Variance in Years of Post-Secondary Education and Adult Log Wages using NLS72 (Baseline and Full Specifications)

	Yrs. Postsec. Ed.		Perm. Wages No Post-sec Ed.		Perm. Wages w/ Post-sec Ed.	
Fraction of Variance	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.904	0.904	0.837	0.834	0.829	0.829
$Var(Y_{is} - Y_s)$	(0.008)	(0.008)	(0.020)	(0.019)	(0.020)	(0.019)
Observable Student-Level (Within):	0.154	0.280	0.140	0.174	0.212	0.224
$Var((X_{si} - X_s)B)$	(0.007)	(0.008)	(0.014)	(0.016)	(0.016)	(0.016)
Unobservable Student-Level (Within)	0.749	0.624	0.697	0.660	0.617	0.605
$Var(v_{si} - v_s)$	(0.009)	(0.008)	(0.022)	(0.022)	(0.022)	(0.022)
Between School:						
Total	0.096	0.096	0.163	0.166	0.171	0.171
$Var(Y_s)$	(0.008)	(0.008)	(0.020)	(0.019)	(0.020)	(0.019)
Observable Student-Level:	0.041	0.058	0.045	0.055	0.061	0.065
$Var(X_s B)$	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.008)
Student-Level/ School-Level Covariance	0.031	0.023	0.033	0.028	0.033	0.029
$2 * Cov(X_s B, X_s G_1 + Z_{2s} G_2)$	(0.006)	(0.006)	(0.020)	(0.010)	(0.011)	(0.009)
School-Avg. Student-Level/ School Char. Covariance	0.001	0.002	-0.002	0.001	-0.003	0.000
$2 * Cov(X_s G_1, Z_{2s} G_2)$	(0.002)	(0.004)	(0.012)	(0.011)	(0.012)	(0.011)
School-Avg. Student-Level	0.012	0.008	0.033	0.029	0.029	0.028
$Var(X_s G_1)$	(0.003)	(0.002)	(0.010)	(0.009)	(0.010)	(0.009)
School Char.	0.005	0.002	0.039	0.041	0.039	0.040
$Var(Z_{2s} G_2)$	(0.002)	(0.002)	(0.012)	(0.011)	(0.012)	(0.012)
Unobservable School-Level	0.005	0.004	0.014	0.011	0.011	0.009
$Var(v_s)$	(0.002)	(0.002)	(0.012)	(0.011)	(0.011)	(0.011)

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 A19 for additional details.