Forthcoming, American Economic Review

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
University of Colorado-Boulder and NBER *

January 10, 2018

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 multiple datasets, we find that a 90th versus 10th percentile school/neighborhood increases the high school graduation probability and college enrollment probability by at least 0.04 and 0.11 and permanent wages by 13.7%.

*Altonji: Department of Economics, Yale University, PO Box 208264, New Haven CT 06520-8264. joseph.altonji@yale.edu. Mansfield: Department of Economics, University of Colorado-Boulder, 04B Economics Building, Boulder, CO 80309. richard.mansfield@colorado.edu. We thank Steven Berry, Gary Chamberlain, Greg Duncan, Phil Haile, Hidehiko Ichimura, Amanda Kowalski, Costas Meghir, Richard Murnane, Jonah Rockoff, Douglas Staiger, Jonathan Skinner, Shintaro Yamiguchi and 4 anonymous referees as well as seminar participants at CEMFI, Cornell, Dartmouth, Duke, the Federal Reserve Banks of Chicago and Cleveland, McMaster, the NBER Economics of Education Conference, Melbourne, Monash, Northwestern, the Paris School of Economics, the Stockholm School of Economics, the London School of Economics, Stanford, U. of Colorado-Boulder, U. of Colorado-Denver, U. Illinois Urbana-Champaign, U. Illinois-Chicago, U. of New South Wales, U. of Sydney, UTS, U. of Western Ontario, and Yale for helpful comments and discussions. This research uses data from the National Center for Education Statistics as well as from the 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 influence individual outcomes or because the groups have attracted 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 for the variance in the contributions of group-level treatments to individual outcomes. We also provide conditions under which causal effects of particular observed group characteristics can be estimated. A key message of the paper is that in some cases one can address the effects of sorting on unobservables using multivariate regression, without having to fully specify and estimate the model of group choice in addition to the model of outcomes.

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?

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.[2] Suppose that Y_{si} is determined by

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

The vector $\mathbf{X_i}$ is a set of student and family characteristics observed by the econometrician (with corresponding productivities $\boldsymbol{\beta}$), while $x_i^U \equiv \mathbf{X_i^U}\boldsymbol{\beta^U}$ is a scalar index that combines the contributions

[1] Ash et al. (2012) provide an overview of the issues involved in assessing hospitals. Doyle Jr et al. (2015) also discuss the issues and provide a short literature survey. See Chetty et al. (2014) and Rothstein (2014) for discussions and references related to the estimation of teacher value-added.

[2] 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.

of unobserved student and family characteristics $\mathbf{X_i^U}$ to the outcome. Together, $[\mathbf{X_i}, \mathbf{X_i^U}]$ represent the *complete* set of student and family characteristics that have a causal impact on student i's educational attainment. Analogously, the row vector $\mathbf{Z_s}$ is a set of school and neighborhood characteristics observed by the econometrician (with corresponding productivities $\mathbf{\Gamma}$), while $z_s^U \equiv \mathbf{Z_s^U}\mathbf{\Gamma}^U$ is a scalar index that combines the effects of unobserved school and neighborhood characteristics. Together, $[\mathbf{Z_s}, \mathbf{Z_s^U}]$ 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 $[\mathbf{Z_s}\mathbf{\Gamma} + z_s^U]$.

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

Our key insight follows directly from the parent's school/neighborhood choice decision—as schools get large, 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 $\mathbf{A_s}$) that parents consider when making their school choices. Thus, the school averages $\mathbf{X_s}$ and $\mathbf{X_s^U}$ will be different vector-valued functions of the same common set of amenities: $\mathbf{X_s} = \mathbf{f}(\mathbf{A_s})$ and $\mathbf{X_s^U} = \mathbf{f^U}(\mathbf{A_s})$. The functions $\mathbf{f}$ and $\mathbf{f^U}$ 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: $\mathbf{A_s} = \mathbf{f^{-1}}(\mathbf{X_s})$. But this implies that the vector of school averages of unobserved characteristics can also be written as a function of observed characteristics: $\mathbf{X_s^U} = \mathbf{f^U}(\mathbf{f^{-1}}(\mathbf{X_s}))$. This function of $\mathbf{X_s}$ can serve as a control function for $\mathbf{X_s^U}$ when estimating group effects.

We formalize this intuition by introducing a multidimensional spatial equilibrium model of neighborhood/school choice and providing conditions under which the mapping from $\mathbf{X_s}$ to $\mathbf{X_s^U}$ is exact. Under our full set of assumptions (most notably an additively separable specification of utility) the mapping from $\mathbf{X_s}$ to $\mathbf{X_s^U}$ is *linear*. When these conditions are satisfied, including $\mathbf{X_s}$ in a linear regression of the outcome Y_{si} fully controls for $\mathbf{X_s^U}$.

While this control function approach potentially solves the sorting-on-unobservables problem, the observed group averages $\mathbf{X_s}$ control for too much. They will absorb peer effects that depend on $\mathbf{X_s}$ and/or $\mathbf{X_s^U}$. They will also absorb a part of the unobserved school/neighborhood quality component that is both orthogonal to the observed school characteristics and 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. However, the fact that controlling for the group averages eliminates bias from sorting implies that the causal effects ($\mathbf{\Gamma}$) of particular school inputs or policies (in $\mathbf{Z}_s$) can be point identified in situations where bias from omitted neighborhood/school characteristics in z_s^U is not a problem or can be addressed through a complementary instrumental variables scheme.

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). To make our estimates easier to interpret, we also convert each variance estimate into a lower bound estimate of the expected 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.

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 7.9 percentage points. In the NELS88 and ELS2002 the corresponding estimates are 4.8 and 4.1 percentage points, respectively, although these may be less reliable due to sampling error in school average characteristics. The North Carolina, NELS88 and ELS 2002 estimates are 15.3, 12.8, and 6.2 percentage points, respectively, when we also consider unobserved differences in quality. We estimate large average impacts despite the fact that our most conservative estimates only attribute 1-2% 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 11-14 percentage points across all three survey datasets. It would increase the permanent component of adult wages by at least 13.7 percent (in NLS72). A one-standard deviation shift in school/neighborhood quality would raise wages by at least 5.1 percent, or .16 standard deviations. 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 on 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. (2015), among others), labor supply functions (e.g., Altonji (1982)), distinguishing between uncertainty and heterogeneity in earnings (e.g., Cunha et al. (2005)), and even estimating neighborhood effects (Bayer and Ross (2009)).[3] 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.[4]

Second, we draw on the rich theoretical and empirical literature on equilibrium sorting and matching models across several fields, including marriage markets (Browning et al. (2014) and Chiappori and Salani (2016)), labor markets (Lise et al. (2013), Melo (2015), and Lindenlaub (2017)), and product markets (Rosen (1974), Ekeland et al. (2004), and Heckman et al. (2010)).

Most directly relevant is the large literature on sorting across neighborhoods and schools that grew out of Tiebout (1956), particularly Epple and Platt (1998), Epple and Sieg (1999) and Bayer and Timmins (2005). 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. 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 ($\mathbf{Z_s}$ 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 $\mathbf{Z_s}$ are endogenous in the outcome equation because of sorting on $\mathbf{X_i^U}$.[5] However, 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.

Third, our formulation of the school/location choice problem is standard in the consumer choice

[3] 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 our case, there is no instrument, but the sorting model implies a relationship between observable and unobservable group averages.

[4] Our estimation strategy is also 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 $\mathbf{X_i}$ from members of group s is used to control for correlation between $\mathbf{X_i}$ and the group error term. In many applications the mean $\mathbf{X_s}$ is used. However, in that literature, much of the focus is on estimating the effects of person specific variables, such as $\boldsymbol{\beta}$ 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 $\mathbf{X_s}$ as a control, and $\boldsymbol{\beta}$ is not identified. Our analysis is also completely distinct from that of Altonji et al. (2005). They 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.

[5] 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, including Fu and Ross (2013) in the neighborhood context.

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. 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 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.[6] Our analysis of sorting is directly relevant to the large number of papers that study group effects using regression models with both individual and group-level characteristics. A few recent papers have employed experimental or quasi-experimental strategies to isolate the contribution of either schools or neighborhoods to longer run student outcomes. Oreopoulos (2003) and Jacob (2004) use quasi-random assignment of neighborhood in the wake of housing project closings to estimate the magnitude of neighborhood effects on student outcomes. Similarly, the Moving To Opportunity (MTO) experiment, evaluated in Kling et al. (2007), randomly assigned housing vouchers that required movement to a lower poverty 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. (2016) 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. Using a sibling differences approach that also exploits high quality data from tax records, Chetty and Hendren (2015) 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 tract-level poverty rate and high school dropout rate on dropout rates and years of education using a sibling differences design and PSID data.

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 for students from low quality urban schools. Angrist et al. (2016) also use admissions lotteries and find positive effects of attending a Boston charter high school on test scores and attendance at four-year colleges relative to two-year colleges. On the other hand, Cullen et al. (2006) use a similar identification strategy with lotteries in Chicago Public Schools and find little

[6]Jencks and Mayer (1990) provide a comprehensive review of earlier studies from 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), Harding et al. (2011) and Graham (2016). Duncan and Murnane (2011) contains several recent papers on school and neighborhood effects. Meghir et al. (2011) discuss alternative approaches to estimating school treatment effects and the effects of particular school inputs.

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. In Section 4 we elaborate on the model for long run outcomes presented above, and show that OLS estimates combined with restrictions implied by our model of sorting are sufficient to place a lower bound on the variance of school and neighborhood effects. Section 5 describes how we use this result to estimate lower bounds on school and neighborhood effects. Section 6 describes the four datasets we use to estimate the model of outcomes. Section 7 presents our results. Section 8 closes the paper with a brief summary and discussion of other applications of our methodology, including the assessment of teacher value added.

2 A Multinomial Model of School Choice and Sorting

In this section we present a model of how families choose school systems and associated neighborhoods. Throughout the paper, matrices, vectors, and matrix- or vector-valued functions are in bold. The "prime" symbol denotes matrix or vector transposes.

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 $V_i(s)$ can be interpreted as the family's consumer surplus from their choice. We assume $V_i(s)$ takes the following linear form:

$$V_i(s) = \mathbf{W_i A_s} + \varepsilon_{si} - P_s. \tag{2}$$

In the above equation $\mathbf{A_s} \equiv [A_{1s}, \ldots, A_{Ks}]'$ represents a vector of K underlying latent amenities that characterize each location $s \in \{1, ..., S\}$. $\mathbf{W_i} \equiv [W_{1i}, \ldots, W_{Ki}]$ 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 willingness to pay (hereafter denoted WTP) for particular amenities across families onto the families' observable ($\mathbf{X_i}$) and unobservable ($\mathbf{X_i^U}$) characteristics. In particular, suppose that $\mathbf{X_i}$ has L elements, while $\mathbf{X_i^U}$ has L^U elements. Then we obtain:

$$\mathbf{W_i} = \mathbf{X_i \Theta} + \mathbf{X_i^U \Theta^U} + \mathbf{Q_i \Theta^Q}. \tag{3}$$

where $\mathbf{\Theta}$ ($\mathbf{\Theta^U}$) 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 $\mathbf{A_s}$ varies with the ℓ-th element of $\mathbf{X_i}$

($\mathbf{X_i^U}$). We sometimes refer to the elements of $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ as WTP coefficients. The $1 \times L^Q$ vector $\mathbf{Q_i}$ consists of additional individual characteristics (observed and unobserved) that affect WTP in accordance with the WTP coefficient matrix $\mathbf{\Theta^Q}$. $\mathbf{Q_i}$ is defined to be uncorrelated with $[\mathbf{X_i}, \mathbf{X_i^U}]$. Since $[\mathbf{X_i}, \mathbf{X_i^U}]$ is the complete set of student attributes that determine Y_{si}, the characteristics that contribute to $\mathbf{Q_i}$ influence school choice but have no direct effect on student outcomes.

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

$$V_i(s) = (\mathbf{X_i}\mathbf{\Theta} + \mathbf{X_i^U}\mathbf{\Theta^U} + \mathbf{Q_i}\mathbf{\Theta^Q})\mathbf{A_s} + \varepsilon_{si} - P_s \tag{4}$$

In the absence of restrictions on the elements of $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$, 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 $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$ 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. Some of the elements of $\mathbf{A_s}$ may depend on school/neighborhood characteristics $\mathbf{Z_s}$ that influence educational attainment and labor market outcomes. The amenities may also include or depend on aspects of the demographic composition of the school/neighborhood. Some determinants of amenities (such as spending per pupil) may be influenced by demographic composition. Thus, some of the amenities are influenced by the sorting equilibrium.

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

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

Parents behave competitively in the sense that prices and $\mathbf{A_s}$ are taken as given, and choice is unrestricted. In equilibrium the values of some elements of $\mathbf{A_s}$ may in fact depend on the averages of $\mathbf{X_i}$ and $\mathbf{X_i^U}$ 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 Proposition 1, which concerns the relationship between $\mathbf{X_s^U}$ and $\mathbf{X_s}$ implied by the above choice model. The proof is supplied in Appendix A1. In Section 3.2 we discuss the proposition and the assumptions that underlie it.

3.1 Proposition 1: $\mathbf{X_s^U}$ is a linear function of $\mathbf{X_s}$

Before stating Proposition 1, we need to define more notation. Decompose $\mathbf{X_i^U}$ into its projection on $\mathbf{X_i}$ and the orthogonal component $\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}$:[7]

$$\mathbf{X_i^U} = \mathbf{X_i \Pi_{X^U X}} + \tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U} \tag{5}$$

Use (5) to rewrite (3) as $\mathbf{W_i} = \mathbf{X_i}\tilde{\mathbf{\Theta}} + \tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}\mathbf{\Theta^U} + \mathbf{Q_i}\mathbf{\Theta^Q}$, where $\tilde{\mathbf{\Theta}} = [\mathbf{\Theta} + \mathbf{\Pi_{X^U X}}\mathbf{\Theta^U}]$. In the rewritten form, all three components of $\mathbf{W_i}$ 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_s and $\mathbf{A_s}$ 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* $\mathbf{X_i}$, $\mathbf{X_i^U}$, $\mathbf{Q_i}$, and $\mathbf{A_s}$ *for all s.*

A4: $\mathbf{E}(\mathbf{X_i}|\mathbf{W_i})$ *and* $\mathbf{E}(\mathbf{X_i^U}|\mathbf{W_i})$ *are linear in* $\mathbf{W_i}$.

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

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

for some $L^U \times L$ matrix $\mathbf{R}$.

Then the expectation $\mathbf{X_s^U}$ is linearly dependent on the expectation $\mathbf{X_s}$. Specifically,

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

3.2 Discussion of Proposition 1

Proposition 1 lays out the conditions under which $\mathbf{X_s^U}$, the between group component of the vector of individual-level unobservables, will be an exact linear function of its observable counterpart $\mathbf{X_s}$.[8] Remarkably, the dependence between the group averages $\mathbf{X_s^U}$ and $\mathbf{X_s}$ arises even when the vector $\mathbf{X_i^U}$ is uncorrelated with the vector $\mathbf{X_i}$ at the individual level. Note also that if $\mathbf{\Theta^U} = \mathbf{0}$ so that unobservable characteristics do not affect amenity preferences (i.e. individuals do not sort based on unobservables), then $\mathbf{R} = \mathbf{0}$. When $\mathbf{R} = \mathbf{0}$, (7) states that $\mathbf{X_s^U} = \mathbf{X_s \Pi_{X^U X}}$, which implies that $\tilde{\mathbf{X}}_\mathbf{s}^\mathbf{U} =$

[7]We use the symbol $\mathbf{\Pi_{DV}}$ to denote the vector or matrix of the partial regression coefficients relating a dependent variable or vector of dependent variables $\mathbf{D}$ to a vector of explanatory variables $\mathbf{V}$, holding the other variables that appear in the regression constant. In the case of $\mathbf{\Pi_{X^U X}}$, $\mathbf{D} = \mathbf{X_i^U}$ and $\mathbf{V} = \mathbf{X_i}$.

[8]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 large enough to be well approximated by a continuum of neighborhoods that create a continuous joint distribution of amenities $\mathbf{A}$. Perhaps surprisingly, equation (7) in Proposition 1 holds for the continuous case.

$\mathbf{0}$, where $\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{s}}$ is the expectation of $\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}$ for group s. As we discuss in Section 7.5, this fact means that if sorting is driven by $\mathbf{X_i}$ but not $\mathbf{X^U_i}$, one can estimate the variance in group treatment effects $Var(\mathbf{Z_s}\mathbf{\Gamma} + z^U_s)$.

Note that the effectiveness of $\mathbf{X}_s$ as a control function for $\mathbf{X}^U_s$ stems from the fact that sorting creates a mapping from $\mathbf{X}_s$ to $\mathbf{A}_s$ and from $\mathbf{A}_s$ to $\mathbf{X}^U_s$, leading to (7). The key is that both $\mathbf{X}_i$ and $\mathbf{X}^U_i$ shift preferences for $\mathbf{A}_s$, not that $\mathbf{X}_i$ affects the outcome Y_i. Because $\mathbf{Q}_i$ also shifts preferences for $\mathbf{A}_s$, one may easily extend Proposition 1 so that the control vector, say $\mathbf{C}^*_s$, includes both $\mathbf{X}_s$ and school averages of observed elements of $\mathbf{Q}_i$. In addition, because $\mathbf{X}^U_s$ is a function of $\mathbf{A}_s$ in equilibrium, $\mathbf{C}^*_s$ might directly include flexible functions of some of the amenities $\mathbf{A}_s$ themselves for which measures are available. The price P_s depends in part on $\mathbf{A}_s$ and could be included, as Bayer and Ross (2009) suggest. In principle $\mathbf{C}^*_s$ could even include functions of observed school policies or productive inputs (elements of $\mathbf{Z}_s$ below) as long as they were fully determined by $\mathbf{A}_s$ (possibly through school composition $\mathbf{X}_s$). If observed school policies or inputs contain a pure school component, then including them in the control function will lead to even more conservative lower bounds on the variance in treatment effects across schools.[9] In our empirical work we only include in our control function observed student variables that plausibly affect student outcomes. Thus, to conserve notation we continue to refer to the control function as $\mathbf{X}_s$ rather than $\mathbf{C}^*_s$, but this is without loss of generality.

Note also that Proposition 1 is a statement about the expectations $\mathbf{X_s}$ and $\mathbf{X^U_s}$. Thus, it concerns the averages of $\mathbf{X_i}$ and $\mathbf{X^U_i}$ when the number of individuals is large relative to the number of choices. With a finite number of individuals per group, random variation 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 Online Appendix A6 indicate that the control function also works well even when the expectation $\mathbf{X_s}$ is estimated using sample averages (which we denote $\hat{\mathbf{X}}_{\mathbf{s}}$) based on small samples of group members rather than the full school population. In Section 6 and Online Appendix A11, we use the North Carolina administrative data to directly assess the effect of using smaller samples of students to estimate $\mathbf{X_s}$ for some of the outcomes and characteristics we actually consider. Our main results are relatively insensitive to working with samples from North Carolina schools that match the distribution of sample sizes in the NLS72, NELS88, and ELS2002 datasets.

The use of $\mathbf{X_s}$ as a control function for $\mathbf{X^U_s}$ requires the number of groups in the sample to be larger than the number of elements in $\mathbf{X_s}$ (and implicitly the number of factors in $\mathbf{A_s}$). Otherwise, one cannot estimate the coefficient vector on $\mathbf{X_s}$.

The next two subsections discuss the assumptions underlying Proposition 1.

[9] However, we strongly suspect that school inputs and policies are not fully determined by student composition. And we argue below that the spanning assumption A5 is likely to be a good approximation in our application, so that $\mathbf{X}_s$ fully controls for $\mathbf{X}^U_s$ by itself. Consequently, in our empirical work the control function does not contain any observed school policies and resources.

3.2.1 Discussion of Assumptions A1-A4

Assumption A1, which relates to the specification of preferences, is fairly general given that both $\mathbf{X_i}$ and $\mathbf{X_i^U}$ can include nonlinear terms.

Assumption A2 says that households take characteristics of neighborhoods as given. As we mentioned above, this is fully consistent with the possibility that $\mathbf{A_s}$ 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 $\mathbf{X_s}$ and $\mathbf{X_s^U}$ will hold in any equilibrium of the model.

Assumption A2 also imposes that households face a common set of choices. In Online Appendix A6 we present monte carlo simulation results that examine the properties of our control function approach across a number of key dimensions. The simulations indicate that the control function works well even when different households face choice sets that are overlapping subsets of the full set of schools.

The common choice set assumption implicitly assumes a static environment. If different families were making school/neighborhood choices at different points in time and face substantial moving costs, then school/neighborhood populations would consist of mixtures of families who made school/neighborhood choices in different periods.

Extending the analysis to a fully dynamic choice model is beyond the scope of the paper. However, consider a simple overlapping generations-style model where the joint distribution of amenity factors evolves across periods and school/neighborhood populations combine members of different decision periods ("generations"). Next, note that the mapping from $\mathbf{X}_s^U$ to $\mathbf{X}_s$ in (7) does not depend on the distribution of amenities $\mathbf{A_s}$. Consequently, it can easily be shown that an analogue of Proposition 1 still holds as long as the taste parameter matrices $\mathbf{\Theta}$ and $\mathbf{\Theta}^U$ and the joint distribution of individual characteristics $[\mathbf{X}_i, \mathbf{X}_i^U, \mathbf{Q}_i]$ are fixed across these "generations".

The independence assumption A3 for the idiosyncratic preference components ε_{si} is strong, but the ε_{si} can be defined to be uncorrelated with $\mathbf{X_i}, \mathbf{X_i^U}$ and $\mathbf{Q_i}$ without loss of generality. Furthermore, the presence of $\mathbf{Q}_i$ and $\mathbf{X}_i^U$ means that we are allowing for unobserved random variation in preferences for location characteristics $\boldsymbol{A}_s$ separately from ε_{si}.

Given the linear relationship between $\mathbf{W_i}$ and $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$ in A1, a sufficient condition for the linearity in expectations assumption A4 to hold is that the joint distribution of $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$ belongs to the continuous elliptical class. Examples include the multivariate normal, the multivariate t, the Laplace, and the multivariate exponential power family (Gómez et al. (2003)). However, in our application $\mathbf{X_i}$ contains a number of discrete variables, so this sufficient condition will not be satisfied.

Proposition 1A in online Appendix A4 establishes that if A4 fails, then an approximation error term appears in equation (7) for $\mathbf{X_s^U}$. The approximation error consists of the average for s of a linear function of the differences between $\mathbf{E}(\mathbf{X_i}|\mathbf{W_i})$ and $\mathbf{E}(\mathbf{X_i^U}|\mathbf{W_i})$ (respectively) and the best least square linear predictions of $\mathbf{X_i}$ and $\mathbf{X_i^U}$ given $\mathbf{W_i}$. As we discuss in Section 5.2, the approximation error would primarily contribute to the school/neighborhood error component v_s of the outcome model (12) introduced in the next section. This would lead to upward bias in the less conservative of our two estimators of the variance of school/neighborhood effects. Note, though, that variation in the differences between $\mathbf{E}(\mathbf{X_i}|\mathbf{W_i})$ and $\mathbf{E}(\mathbf{X_i^U}|\mathbf{W_i})$ (respectively) and the best linear predictors $\mathbf{E}^*(\mathbf{X_i}|\mathbf{W_i})$ and $\mathbf{E}^*(\mathbf{X_i^U}|\mathbf{W_i})$ will be small for values of $\mathbf{W_i}$ that are near the population mean $\mathbf{W_i}$. In the data, the variation across s in $\mathbf{X_i}$ is much smaller than the population variance. Presumably the same is true of $\mathbf{X_i^U}$ and thus $\mathbf{W_i}$. Consequently, variation in the averages $\mathbf{E}[\mathbf{E}(\mathbf{X_i}|\mathbf{W_i}) - \mathbf{E}^*(\mathbf{X_i}|\mathbf{W_i})|s(i) = s]$ and $\mathbf{E}[\mathbf{E}(\mathbf{X_i^U}|\mathbf{W_i}) - \mathbf{E}^*(\mathbf{X_i^U}|\mathbf{W_i})|s(i) = s]$ (respectively) are likely to be small. Furthermore, note that because $\mathbf{X_s^U}$ appears in the outcome equation through the index $\mathbf{X_s^U}\boldsymbol{\beta}^\mathbf{U}$, any upward bias depends on a weighted index of the approximation error terms for each element of $\mathbf{X_s^U}$, with the elements of $\boldsymbol{\beta}^\mathbf{U}$ 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 $\boldsymbol{\Theta}^\mathbf{U}$ relating tastes for amenities to the elements of $\mathbf{X_i^U}$ to be linear combinations of the coefficient vectors $\tilde{\boldsymbol{\Theta}}$ relating tastes for amenities to the observables $\mathbf{X_i}$ and/or elements of $\mathbf{X_i^U}$ that are correlated with $\mathbf{X_i}$. 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 $\mathbf{A^X} \subseteq \mathbf{A}$ represent the subset of amenities that affect the distribution of observable school averages $\mathbf{X_s}$. An amenity will be included in $\mathbf{A^X}$ if WTP for the amenity is affected by $\mathbf{X_i}$ and/or by elements of $\mathbf{X_i^U}$ that are correlated with $\mathbf{X_i}$. Likewise, $\mathbf{A^{X^U}} \subseteq \mathbf{A}$ represents the subset of amenities that affect the distribution of unobservable school averages $\mathbf{X_s^U}$. The between-school variation in $\mathbf{X_i}$ will only be driven by $\mathbf{A^X}$, so that $\mathbf{X_s} = \mathbf{f}(\mathbf{A^X})$ for some vector-valued function $\mathbf{f}$. Similarly, $\mathbf{X_s^U} = \mathbf{f^U}(\mathbf{A^{X^U}})$. We can write $\mathbf{X_s^U} = \mathbf{g}(\mathbf{X_s})$ if we can write $\mathbf{X_s^U} = \mathbf{f^U}(\mathbf{f}^{-1}(\mathbf{X_s}))$, where $\mathbf{g}(\mathbf{X_s}) = \mathbf{f^U}(\mathbf{f}^{-1}(\mathbf{X_s}))$. Thus, jointly sufficient conditions are

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

Assumption A5.2: $\mathbf{A^{X^U}} \subseteq \mathbf{A^X}$*, so that the set of amenities that drive variation in* $\mathbf{X_s}$ *contains the set of amenities that drive variation in* $\mathbf{X_s^U}$ *(i.e. the range of* $\mathbf{f^{-1}}$ *must encompass the domain of* $\mathbf{f^U}$*).*

While these conditions are not necessary, they suggest two fundamental ways that the spanning condition $\boldsymbol{\Theta}^\mathbf{U} = \mathbf{R}\tilde{\boldsymbol{\Theta}}$ can fail.[10] The first way, which leads A5.1 to fail, is that the vector $\mathbf{X_i}$ may

[10] Invertibility of $\mathbf{f}(\mathbf{A^X})$ is not a necessary condition. It is possible that the mapping from $\mathbf{A^X}$ to $\mathbf{X_s}$ is one-to-many, meaning that the same value of $\mathbf{A^X}$ leads to multiple values of $\mathbf{X_s}$. In this case the key is that one can still write $\mathbf{A^X}$

affect tastes for more amenities than the number of elements in $\mathbf{X_i}$. That is $\dim(\mathbf{A^X}) > L$ where $\dim(\mathbf{A^X})$ is the number of elements in $\mathbf{A^X}$. In this case, the function $\mathbf{f}(*)$ is not invertible.[11] In the case of the additively separable utility function from (4), $\dim(\mathbf{A^X})$ is equal to the row rank of $\tilde{\mathbf{\Theta}}$. To see how A5.1 might fail, suppose that the only observable characteristic were parental education and that the amenity space consisted of two imperfectly correlated factors: schools' quality of teachers and quality of athletic facilities. Even if parental education 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 education. We would need to observe a second individual characteristic, such as parental income, in order to satisfy the spanning condition.

The validity of A5.1 depends on the number and breadth of coverage of variables in $\mathbf{X_i}$. It is testable. The model implies a factor structure for the vector $\mathbf{X_s}$, where the number of factors is determined by the row rank of $\tilde{\mathbf{\Theta}}$ (See online Appendix A3). A finding that the number of factors that determine $\mathbf{X_s}$ is smaller than the dimension of $\mathbf{X_i}$ is consistent with the assumption that $\dim(\mathbf{A^X}) \leq L$. A finding that the number of factors is at least as large as the dimension of $\mathbf{X_i}$ is also technically consistent with the assumption, but would strongly suggest that $\dim(\mathbf{A^X}) > L$. The evidence presented in Section 7.6 and online Appendix A3 is fully consistent with $\dim(\mathbf{A^X}) < L$ in our application.

What about Assumption 5.2? Partition $\mathbf{X_i^U}$ into a subvector $\mathbf{X_{1i}^U}$ that is correlated with $\mathbf{X_i}$ and a subvector $\mathbf{X_{2i}^U}$ that is not correlated with $\mathbf{X_i}$. Assumption 5.2 will fail if $\mathbf{X_{2i}^U}$ affects preferences for an amenity for which neither $\mathbf{X_i}$ nor $\mathbf{X_{1i}^U}$ affect preferences.

To illustrate how Assumption 5.2 can fail, we modify the example above so that parental education and immigrant status are the only $\mathbf{X_i}$ variables. Suppose that student athleticism is the only unobservable and that it is uncorrelated with both parental education and immigrant status. And suppose that neither parental education nor immigrant status affect WTP for athletic facilities in the neighborhood, while student athleticism does. Then student athleticism is an $\mathbf{X_{2i}^U}$ variable rather than an $\mathbf{X_{1i}^U}$. Athletic facility quality would be an element of $\mathbf{A^{X^U}}$ but not $\mathbf{A^X}$, so that $\mathbf{A^{X^U}} \not\subset \mathbf{A^X}$. Assumption 5.2 would fail. Consequently, variation in athletic facility quality would drive between-neighborhood variation in average student athleticism that average parental education and immigrant status would not capture. Online Appendix A2 provides additional examples of 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

$= \mathbf{h}(\mathbf{X_s})$, where $\mathbf{h}(.) = \mathbf{f}^{-1}(.)$ in the one-to-one case. The mapping from $\mathbf{A^{X^U}}$ to $\mathbf{X_s^U}$ need not be one-to-one either. However, there must be a mapping $\mathbf{X_s^U} = \mathbf{f}(\mathbf{A^{X^U}}, \mathbf{X_s}) = \mathbf{f}(\mathbf{h}(\mathbf{X_s}), \mathbf{X_s}) = \mathbf{g}(\mathbf{X_s})$ that is one-to-one or many-to-one.

[11]More specifically, what is relevant for invertibility is not the number of elements of $\mathbf{X_i}$ (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 they affected willingness to pay for each amenity in the same relative proportions. Then adding father's education to $\mathbf{X_i}$ would not make $\mathbf{f}(*)$ invertible if it were not already when only mother's education was included in $\mathbf{X_i}$.

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 $\mathbf{X_i^U}$, 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 $\mathbf{X_i}$ 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 place on academic learning and educational attainment. One would expect this unobservable to boost willingness 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 NELS88), and grandparents' education (observed in ELS2002) 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)).

We already mentioned that group averages of elements of $\mathbf{Q}_i$ should be included in the control function if they are observed and $\mathbf{X_s}$ alone is inadequate for the spanning condition to hold. One could also include an observed element of $\mathbf{A^{X^U}}$ that drives sorting on $\mathbf{X_{2i}^U}$ in the control function. In applications in which $\mathbf{X_i}$ is limited, these additions may be needed for the control function to be adequate.

Online Appendix A5 derives an analytical formula for the component of $\mathbf{X_s^U}$ that cannot be predicted by $\mathbf{X_s}$ 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 $\mathbf{W_i}$; c) the matrix $\mathbf{\Theta^U}$ 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 $\mathbf{\tilde{X}_i^U}$ and e) the joint distribution of the unobserved outcome-irrelevant (but school choice-relevant) student characteristics $\mathbf{Q_i}$.

Given the complicated manner in which each of these five factors enters the expression for the unexplained component of $\mathbf{X_s^U}$, there does not appear to be any straightforward way to place a bound on the variance in this error component. Online Appendix A6 presents monte carlo simulations for cases in which the spanning condition (A5) fails. The control function approach is quite robust to violations of the spanning condition in which just a few outcome-relevant unobservables in $\mathbf{X_i^U}$ affect WTP for just a few additional amenities that are not weighted by any elements of $\mathbf{X_i}$. This is arguably the most plausible case when rich data on students and parents are available.

4 The Econometric Model of Educational Attainment and Wage Rates

We begin this section by elaborating on the underlying model of student outcomes presented in the introduction. Next, in Section 4.2 we show how sorting and omitted school and neighborhood characteristics affect estimates of neighborhood/school effects based on OLS estimation of that model. Then in Section 4.3 we show that the OLS estimates of the student outcomes model in combination with Proposition 1 and Assumption 6 below are sufficient to place a lower bound on the variance of school and neighborhood effects.

4.1 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

$$Y_{si} = \mathbf{X_i}\boldsymbol{\beta} + x_i^U + \mathbf{Z_s}\boldsymbol{\Gamma} + z_s^U + \eta_{si} + \xi_{si} \ . \tag{8}$$

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 $\mathbf{X_i}\boldsymbol{\beta} + x_i^U$, where $x_i^U \equiv \mathbf{X_i^U}\boldsymbol{\beta}^\mathbf{U}$ is a scalar index summarizing the contributions of unobserved student characteristics $\mathbf{X_i^U}$, and the row vector $[\mathbf{X_i}, \mathbf{X_i^U}]$ is an exhaustive set of child and family characteristics that have a causal impact on student i's outcome. Since $\mathbf{X_i}$ and $\mathbf{X_i^U}$ may include non-linear functions, the linear-in-parameters specification for Y_{si} is without much loss of generality.

Analogously, the average school/neighborhood outcome contribution in school/neighborhood s is captured by the index $\mathbf{Z_s}\boldsymbol{\Gamma} + z_s^U$ where $z_s^U \equiv \mathbf{Z_s^U}\boldsymbol{\Gamma}^\mathbf{U}$ is a scalar index summarizing the contributions of unobserved school and neighborhood characteristics. The vector $\mathbf{Z_s}$ captures the influence of observed school/neighborhood-level characteristics (which in our empirical work do not vary among students within a school). The vector $\mathbf{Z_s^U}$ represents the remaining unobserved school/neighborhood influences that vary between school attendance areas (e.g. quality of the school principal or the local crime rate). Note that $\mathbf{Z_s}$ and $\mathbf{Z_s^U}$ may include averages of $\mathbf{X_i}$ and $\mathbf{X_i^U}$, respectively, that capture peer effects.

The unobserved scalar index η_{si} captures variation in school/neighborhood contributions among students within a school attendance area and within a school itself (e.g. trustworthiness of immediate neighbors or distinct course tracks at the school). Some of the factors that determine η_{si} may represent the within-school components of $\mathbf{Z_s}$.

The component ξ_{si} captures other influences on student i's outcome that are determined after secondary school but are not predictable given $\mathbf{X_i}$, x_i^U, $\mathbf{Z_s}$, 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

$\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ 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 $\mathbf{X_i}$ and $\mathbf{Z_s}$ at a single point in time. Thus, some components of $\mathbf{X_i}$ associated with student inputs such as student aptitude will have been determined in part by parental inputs from earlier periods such as parent income (Todd and Wolpin (2003) and Cunha et al. (2006)). Such links make it difficult to interpret the coefficient associated with a given component of $\mathbf{X_i}$, because once we have conditioned on the other components, we have removed many of the avenues through which the component determines Y_i. Consequently, we do not attempt to estimate the productivity parameters $\boldsymbol{\beta}$ or $\boldsymbol{\beta}^{\mathbf{U}}$, 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 $\boldsymbol{\Gamma}$ stemming from correlations between $\mathbf{Z_s}$ 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 treatment 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 $(\mathbf{Z_{s^1}}\boldsymbol{\Gamma}+z_{s^1}^U)-(\mathbf{Z_{s^0}}\boldsymbol{\Gamma}+z_{s^0}^U)$. We wish to quantify differences across schools/neighborhoods in $\mathbf{Z_s}\boldsymbol{\Gamma}+z_s^U$. 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.[12]

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 $\mathbf{Z_s}$ and/or z_s^U with individual attributes $\mathbf{X_i}$ and/or $\mathbf{X_i^U}$. Indeed, the preference weights on amenities that represent school characteristics depend on $\mathbf{X_i}$ and $\mathbf{X_i^U}$ 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 the outcome model does not break the linear relationship between $\mathbf{X_s}$ and $\mathbf{X_s^U}$. However, it would imply that the distribution of school treatment effects varies with $\mathbf{X_i}$ and $\mathbf{X_i^U}$. We focus on the homogenous effects case in this paper. Agrawal et al. (in progress) are extending the analysis to a model of Y_{si} with interactions.

4.2 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, $\mathbf{X_i}$ and $\mathbf{Z_s}$.

To facilitate the analysis, first partition $\mathbf{Z_s}$ into $[\mathbf{X_s},\mathbf{Z_{2s}}]$, where $\mathbf{X_s}$ consists of school-averages of observable student characteristics, and $\mathbf{Z_{2s}}$ 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 $\boldsymbol{\Gamma}\equiv[\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2]$ analogously. Section 6.3 provides a discussion of which

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

variables should be included in $\mathbf{X_s}$ and $\mathbf{Z_{2s}}$, respectively.

Recall that ξ_{si} is defined to be unrelated to $[\mathbf{X_i}, \mathbf{X_s}, \mathbf{Z_{2s}}]$. Decompose the other unobserved components in the production function (8), x_i^U, z_s^U and η_{si}, into their projection onto $[\mathbf{X}_i, \mathbf{X}_s, \mathbf{Z}_{2s}]$ and the orthogonal components $\tilde{x}_i^U$, $\tilde{z}$, and $\tilde{\eta}_{si}$:

$$x_i^U = \mathbf{X_i}\mathbf{\Pi}_{x_i^U\mathbf{X_i}} + \mathbf{X_s}\mathbf{\Pi}_{x_i^U\mathbf{X_s}} + \mathbf{Z_{2s}}\mathbf{\Pi}_{x_i^U\mathbf{Z_{2s}}} + \tilde{x}_i^U \tag{9}$$

$$z_s^U = \mathbf{X_s}\mathbf{\Pi}_{z_s^U\mathbf{X_s}} + \mathbf{Z_{2s}}\mathbf{\Pi}_{z_s^U\mathbf{Z_{2s}}} + \tilde{z}_s^U \tag{10}$$

$$\eta_{si} = \mathbf{X_i}\mathbf{\Pi}_{\eta_{si}\mathbf{X_i}} + \tilde{\eta}_{si}. \tag{11}$$

Substituting the projections (9), (10), and (11) for x_i^U, z_s^U, and η_{si} into (8), we obtain:

$$Y_{si} = \mathbf{X_i}\mathbf{B} + \mathbf{X_s}\mathbf{G_1} + \mathbf{Z_{2s}}\mathbf{G_2} + v_s + (v_{si} - v_s), \quad \textit{where} \tag{12}$$

$$\mathbf{B} \equiv [\boldsymbol{\beta} + \mathbf{\Pi}_{x_i^U\mathbf{X_i}} + \mathbf{\Pi}_{\eta_{si}\mathbf{X_i}}] \tag{13}$$

$$\mathbf{G_1} \equiv [\mathbf{\Pi}_{x_i^U\mathbf{X_s}} + \mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U\mathbf{X_s}}] \tag{14}$$

$$\mathbf{G_2} \equiv [\mathbf{\Pi}_{x_i^U\mathbf{Z_{2s}}} + \mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U\mathbf{Z_{2s}}}] \tag{15}$$

$$v_s \equiv \tilde{x}_s^U + \tilde{z}_s^U + \xi_s \tag{16}$$

$$v_{si} - v_s \equiv (\tilde{x}_i^U - \tilde{x}_s^U) + \tilde{\eta}_{si} + \xi_i \tag{17}$$

The expressions for $\mathbf{G_1}, \mathbf{G_2}$ and v_s in (14), (15) and (16) reveal that the observable school components $\mathbf{X_s}\mathbf{G_1}$ and $\mathbf{Z_{2s}}\mathbf{G_2}$ and the unobservable residual component v_s all reflect a mixture of school effects and student composition biases. Specifically, $\mathbf{X_s}\mathbf{G_1}$ and $\mathbf{Z_{2s}}\mathbf{G_2}$ will reflect $\mathbf{X_s}\mathbf{\Pi}_{x_i^U\mathbf{X_s}}$ and $\mathbf{Z_{2s}}\mathbf{\Pi}_{x_i^U\mathbf{Z_{2s}}}$, respectively, which capture differences across schools in x_i^U that are predictable by $\mathbf{X_s}$ and $\mathbf{Z_{2s}}$ conditional on $\mathbf{X_i}$. The unpredicted between-school component v_s will reflect $\tilde{x}_s^U$, 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 $\mathbf{X_s}\mathbf{\Pi}_{x_i^U\mathbf{X_s}}$, $\mathbf{Z_{2s}}\mathbf{\Pi}_{x_i^U\mathbf{Z_{2s}}}$ and $\tilde{x}_s^U$ 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.[13] 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. However, in the next section we show that the assumptions laid out in Proposition 1, plus an additional assumption, are sufficient to place a lower bound on the variance of school and neighborhood effects given the production function (8) above.

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

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

Section 3 provides conditions under which the school-average values of student observables $\mathbf{X_s}$ and unobservables $\mathbf{X_s^U}$ are linearly dependent, as summarized in Proposition 1. In this subsection we present two additional Propositions that illustrate the value of Proposition 1 for characterizing the distribution of school/neighborhood treatment effects. In particular, Proposition 2 shows that the relationship between $\mathbf{X_s}$ and $\mathbf{X_s^U}$ implies restrictions on $\mathbf{G_2}$ and v_s. Proposition 3 shows that Propositions 1 and 2, when combined with an additional plausible assumption, allow the recovery of a lower bound estimate of the contribution of schools (and groups more generally) to individual outcomes. We also present the more demanding conditions under which unbiased estimates of the causal effects of particular group-level characteristics can be recovered.

Proposition 2: *Assume that assumptions A1-A5 from Proposition 1 hold.*
Then equations (15)-(16) simplify to:

$$\mathbf{G_2} = \boldsymbol{\Gamma}_2 + \boldsymbol{\Pi}_{z_s^U \mathbf{Z_{2s}}} \tag{18}$$
$$v_s = \tilde{z}_s^U + \xi_s \tag{19}$$

An expanded version of Proposition 2 that includes formulae for $\mathbf{B}$ and $\mathbf{G_1}$ is stated and proved in Online Appendix A7.

We see that when the conditions of Proposition 1 are satisfied, the inclusion of $\mathbf{X_s}$ in $\mathbf{Z_s}$ purges both $\mathbf{G_2}$ and v_s of biases from student sorting, so that $Var(\mathbf{Z_{2s}G_2})$ and $Var(v_s)$ only reflect true school/neighborhood contributions and, in the case of v_s, later common shocks. However, the sum $Var(\mathbf{Z_{2s}G_2}) + Var(v_s)$ is likely to understate the full variance in school contributions $Var(\mathbf{Z_s}\boldsymbol{\Gamma} + z_s^U)$ for three reasons. The first and obvious one is that the causal effect of $\mathbf{X_s}$ on outcomes, $\mathbf{X_s}\boldsymbol{\Gamma}_1$, will contribute to $\mathbf{G}_1$ and therefore 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 $\mathbf{X_s^U}$ 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 also be absorbed by $\mathbf{X_sG_1}$, school/neighborhood peer effects associated with $\mathbf{X_s^U}$ will be excluded from the estimate of school/neighborhood effects.

Third, (14) reveals that $\mathbf{X_s}$ will also absorb part of the unobserved school contribution z_s^U via $\boldsymbol{\Pi}_{z_s^U \mathbf{X_s}}$. To see why, note that $\mathbf{X_s}$ spans the space of $\mathbf{X_s^U}$ *because* the amenity vector, $\mathbf{A_s}$, is the source of variation in both $\mathbf{X_s}$ and $\mathbf{X_s^U}$. 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 $\mathbf{A_s}$. Hence, while the inclusion of $\mathbf{X_s}$ in the estimated specification removes sorting bias, it also absorbs some of the variation in z_s^U associated with underlying amenity factors for which $\mathbf{X_i}$ affects taste. Furthermore, if some elements of the school-level observables $\mathbf{Z_{2s}}$ also serve directly as amenities in $\mathbf{A_s}$ or perfectly determine them, then these elements determine $\mathbf{X_s}$. These relationships need not be linear, but if they are, then we cannot identify the corresponding elements of the vector $\mathbf{G_2}$.[14]

On the other hand, components of $\mathbf{Z_{2s}\Gamma_2} + z_s^U$ that are either not directly valued or only partially known by parents at the time the school/neighborhood is chosen will not be elements of $\mathbf{A_s}$, although they may be correlated with $\mathbf{A_s}$. Parents probably are not perfectly informed about specific school quality determinants such as student/teacher ratio and in any event care about broader qualities of schools rather than specific inputs. The broader qualities are the amenities in the model. Parents would know some variables, such as whether a school is Catholic, but in many cases would have chosen locations prior to high school based on education options more generally. Such specific components of $\mathbf{Z_{2s}}$ will not be collinear with $\mathbf{X_s}$.

The fact that $\mathbf{Z_{2s}G_2}$ and v_s exclude three components of $\mathbf{Z_{2s}\Gamma_2} + z_s^U$ is the reason to expect that $Var(\mathbf{Z_{2s}G_2})$ and $Var(\mathbf{Z_{2s}G_2} + v_s)$ will both understate $Var(\mathbf{Z_{2s}\Gamma_2} + z_s^U)$. However, it is theoretically possible that the covariance of the excluded components with the rest of $\mathbf{Z_{2s}\Gamma_2} + z_s^U$ is sufficiently negative that $Var(\mathbf{Z_{2s}G_2}) + Var(v_s)$ and even $Var(\mathbf{Z_{2s}G_2})$ exceeds $Var(\mathbf{Z_{2s}\Gamma_2} + z_s^U)$. To rule this out, we assume

$$\text{A6.1:}\quad Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}]) + 2Cov(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}]) + Var(\tilde{z}_s^U) \geq 0$$

or

$$\text{A6.2:}\ Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}]) + 2Cov(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}]) - Var(\xi_s) \geq 0. \tag{20}$$

The following proposition provides formal justification for using $Var(\mathbf{Z_{2s}G_2})$ and even $Var(\mathbf{Z_{2s}G_2} + v_s)$ as lower bound estimates of the variance in school/neighborhood treatment effects.

Proposition 3: *If assumptions A1-A5 from Proposition 1 and A6.1 hold, then* $Var(\mathbf{Z_{2s}G_2}) \leq Var(\mathbf{Z_s\Gamma} + z_s^U)$. *If assumptions A1-A5 and A6.2 hold, then* $Var(\mathbf{Z_{2s}G_2} + v_s) \leq Var(\mathbf{Z_s\Gamma} + z_s^U)$.

The proof is in Online Appendix A8. Note that since $Var(\tilde{z}_s^U)$ and $Var(\xi_s)$ are non-negative, Assumption 6.2 is strictly stronger than Assumption 6.1. By the same token, since v_s is uncorrelated with $\mathbf{Z_{2s}}$ by construction, $Var(\mathbf{Z_{2s}G_2})$ will always produce a more conservative lower bound than $Var(\mathbf{Z_{2s}G_2} + v_s)$.

While A6.1 and A6.2 are technically necessary to interpret $Var(\mathbf{Z_{2s}G_2})$ and $Var(\mathbf{Z_sG_2} + v_s)$ as lower bound estimators of the school/neighborhood effect variance, in practice we believe that even the stronger assumption A6.2 is very likely to hold if the common shocks component $Var(\xi_s)$ is 0

[14] Nor can we estimate the effect of a school level variable in the unlikely event that the political process leads it to be an exact linear function of $\mathbf{X_s}$.

(as in high school graduation) or is not too large.[15] In particular, when $Var(\xi_s)$ is 0 the stronger assumption A6.2 is violated only if the covariance term is sufficiently negative to outweigh the variance term. That can only happen when the peer and unobserved school inputs that project onto $\mathbf{X_s}$ are both strongly negatively correlated with those that project onto $\mathbf{Z_{2s}}$ *and* account for a much smaller share of the total school effect variance. Recall that a large share of the variation in $\mathbf{X_s}(\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X}_s})$ is likely to represent observed and unobserved peer inputs.A strong negative correlation between peer inputs and productive school resources or policies will generally require that the student and parent attributes that contribute to a strong peer environment also predict *lower* valuation of superior school inputs. This is at odds with both empirical evidence and casual observation. Online Appendix A8 fleshes out this argument and provides additional theoretical, statistical, and empirical justifications for A6.1 and A6.2.

4.3.1 Identification of $\mathbf{\Gamma}_2$

The existence of $\mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}$ in the expression for $\mathbf{G_2}$ in (18) reveals that even when the conditions of Proposition 1 are satisfied, $\mathbf{G_2}$ still reflects omitted variables bias driven by correlations between $\mathbf{Z_{2s}}$ and the unobserved school characteristics index z_s^U. Thus, estimating the vector of causal effects $\mathbf{\Gamma_2}$ associated with the school characteristics $\mathbf{Z_{2s}}$ will in general still require a vector of instruments, an extension that we discuss in the conclusion.

However, the sorting model in Section 2 also sheds light on the circumstances in which $\mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}} = \mathbf{0}$, so that $\hat{\mathbf{G}}_\mathbf{2}$ is an unbiased estimator of $\mathbf{\Gamma}_2$. In particular, suppose that every unobserved school characteristic that contributes to the index z_s^U and is correlated with $\mathbf{Z_{2s}}$ 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 $\mathbf{A_s}$ is a function of $\mathbf{X_s}$. This implies that $\mathbf{X_s}$ also perfectly determines the part of z_s^U that is correlated with $\mathbf{Z_{2s}}$. In this case, the residual variation in z_s^U will be orthogonal to $\mathbf{Z_{2s}}$. As a result, $\mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}} = \mathbf{0}$, and $\hat{\mathbf{G}}_\mathbf{2}$ will be an unbiased estimator of $\mathbf{\Gamma_2}$.

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 $\mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}} = \mathbf{0}$ in our empirical work. Thus, we do not attempt to interpret the individual coefficients estimated by $\hat{\mathbf{G}}_\mathbf{2}$.[16] 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.[17]

[15] If one regards the common shocks as part of the school/neighborhood treatment, then $Var(\xi_s)$ becomes part of $Var(z_s^U)$ in A6.1 and disappears from A6.2, strengthening both inequalities.

[16] 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.

[17] Of course, as with any control variable, controlling for $\mathbf{X_s}$ could make the bias in $\hat{\mathbf{G}}_\mathbf{2}$ as an estimator of $\mathbf{\Gamma_2}$ worse. This could happen if the regression relationship between the components of $\mathbf{Z_{2s}}$ and z_s^U controlling for sorting is stronger than and opposite in sign to the relationship between $\mathbf{Z_{2s}}$ and z_s^U that is unrelated to sorting.

5 Mechanics of Measuring School and Neighborhood Effects

5.1 Estimating the Model Parameters

Online Appendix A9 describes the process by which the coefficients $\mathbf{B}$, $\mathbf{G_1}$, and $\mathbf{G_2}$ and the error component variances $Var(v_s)$ and $Var(v_{si} - v_s)$ are estimated. In all cases we use sample school averages of observable characteristics $\hat{\mathbf{X}}_\mathbf{s}$ in place of the population averages $\mathbf{X_s}$, which are not observed. We have modified our estimation procedures from previous drafts of this paper to correct for small sample bias in estimates of $Var(v_s)$ associated with the use of degrees of freedom to estimate $\mathbf{B}$, $\mathbf{G_1}$, and $\mathbf{G_2}$. For the continuous outcome years of postsecondary education we treat v_s as a random effect and estimate the model by restricted maximum likelihood (REML). REML accounts for degrees of freedom in estimating $Var(v_s)$ and $Var(v_{si} - v_s)$, while maximum likelihood does not. For the log wage outcome we have panel data on each sample member. We treat both v_s and $(v_{si} - v_s)$ as random effects and allow for an additional person-specific transitory component of wages that may be serially correlated for up to 7 years. We estimate by REML. Due to computational difficulties, we do not incorporate weights into the REML estimation procedure for either continuous outcome.[18]

For binary outcomes such as high school graduation, we reinterpret Y_{si} to be the latent variable that determines the indicator for whether a student graduates, $HSGRAD_{si} = 1(Y_{si} > 0)$. We assume errors are normally distributed and work with a random effects probit model. The scale of Y_{si} is chosen by setting $Var(v_{si} - v_s)$ to 1. The REML estimator does not exist for binary outcomes, so we use a two-step procedure. In the first step, we use maximum likelihood to estimate the parameters $\mathbf{B}$, $\mathbf{G_1}$, $\mathbf{G_2}$ and $Var(v_s)$ of the random effects probit model. In the second step, we correct the maximum likelihood estimates of $Var(v_{si} - v_s)$ and $Var(v_s)$ for downward bias due to lost degrees of freedom in estimating the slope parameters. The correction is based upon formulas derived for continuous outcomes for relationships between the unbiased REML estimators and the corresponding ML estimators of $Var(v_s)$ and $Var(v_{si} - v_s)$. We assume that these formulas also approximately hold in the probit case once scale is accounted for. We then use them to obtain simple formulas to bias adjust the ML estimators. See Supplemental Appendix A9.3.1 for the derivation and formulas.

5.2 Variance Decomposition

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

$$Y_{si} = \mathbf{X_i B} + \mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_{si}, \tag{21}$$

[18]However, to minimize the impact of omitting weights from model estimation, sampling weights are reincorporated when computing each of the observable components of our variance decomposition and also when taking averages over the population of the impact of 10th-to-90th quantile shifts in school quality.

where $\mathbf{X_s}$ is a vector of school-averages of student characteristics, and $\mathbf{Z_{2s}}$ is a vector of observed school characteristics (such as school size or student-teacher ratio). We can decompose $Var(Y_{si})$ into observable and unobservable components of both within- and between- school variation via

$$\widehat{Var}(Y_{si}) = \widehat{Var}(Y_{si} - Y_s) + \widehat{Var}(Y_s) \tag{22}$$

$$= [\widehat{Var}((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + \widehat{Var}(v_{si} - v_s)] + \\ [\widehat{Var}(\mathbf{X_s B}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{X_s G_1}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{X_s G_1}) + \\ 2\widehat{Cov}(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{Z_{2s} G_2}) + \widehat{Var}(v_s)]. \tag{23}$$

Drawing on the analysis above, particularly Proposition 3, we introduce two alternative lower bound estimators of the contribution of school/neighborhood choice to student outcomes.

The first lower bound estimator is $\widehat{Var}(\mathbf{Z_{2s} G_2} + v_s)$. Due to the presence of $\mathbf{X_s}$ in (21) it will be purged of any effects of student sorting (observable or unobservable). Thus, it only captures 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. This bias is likely to be second order for permanent wages because the effect on our random effects estimator of v_s of local shocks that persist for less than the 7 years between sample members' wage observations will be muted. But 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: $\widehat{Var}(\mathbf{Z_{2s} G_2})$. 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. $\widehat{Var}(\mathbf{Z_{2s} G_2})$ excludes true school quality variation that is orthogonal to observed characteristics, but also excludes any truly idiosyncratic local shocks that occur after graduation.[19]

The estimators of the variance and covariance terms in (23) account for sampling error in the regression coefficient estimators. For example, our estimator of $Var(\mathbf{Z_{2s} G_2})$ is

$$\widehat{\mathbf{Var}}(\mathbf{Z_{2s} G_2}) = [\frac{1}{N}\sum_i (\mathbf{Z_{s(i)}} \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_2 \mathbf{Z}'_{\mathbf{s(i)}}) - \frac{1}{N}\sum_i \mathbf{Z_{s(i)}} \widehat{\mathbf{Var}}(\hat{\mathbf{G}}_2) \mathbf{Z}'_{\mathbf{s(i)}}].$$

In practice, we incorporate student weights into the sums in the above equation. See Online Appendix A9.4 for more detail. Online Appendix A10 discusses how the empirical variance decom-

[19]The linear approximation error component might also bias $\mathbf{G_2}$, but we think this is likely to be minor given that we are controlling for $\mathbf{X_s}$.

position is implemented.

5.3 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, although Section 3.2.1 briefly discusses an extension of Proposition 1 to a dynamic model. 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 $\mathbf{X_i}$ and $\mathbf{X_i^U}$. Thus, if prior schooling inputs affect WTP for school/neighborhood amenities, our control function argument suggests that 10th grade school averages of $\mathbf{X_i}$ will absorb all between-school variation in prior school contributions to $\mathbf{X_i^U}$. In this case, the residual variance contributions $Var(\mathbf{Z_{2s}G_2})$ or $Var(\mathbf{Z_{2s}G_2}+v_s)$ that we identify will represent a lower bound on the contributions of only the high schools and their surrounding neighborhoods to our outcomes.

Now consider the opposite extreme: moving costs are prohibitive, and each family makes a one time choice about where to settle down when they begin to have children. Suppose that the observed characteristics $\mathbf{X_i}$ are unaffected by early schooling, as is the case in our baseline specification discussed in Section 6.3. In this scenario, the residual variance contributions $Var(\mathbf{Z_{2s}G_2})$ or $Var(\mathbf{Z_{2s}G_2}+v_s)$ that we identify will represent a lower bound on the variation in contributions to our later outcomes of entire sequences of schools (elementary, middle, and high) and entire childhoods of neighborhood exposure. In reality, of course, moving costs are substantial but not prohibitive, so that our estimates probably reflect a mix of elementary school and high school contributions, with a stronger weight on high school contributions.[20] 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 a 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.

[20]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, 29.9% report 4 to 7 years, 14.3% 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 43.3% for no changes, 24.0% for 1 change, 12.5% for 2 changes, 9.9% for 3 changes, 5.3% for 4 changes, and 5.0% for 5 changes.

5.4 Measuring the Effects of Shifts in School/Community Quality

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

The more comprehensive estimates use $\widehat{Var}(\mathbf{Z_{2s}G_2}+v_s)$ to calculate the 10th-90th shifts, while the more conservative estimates use $\widehat{Var}(\mathbf{Z_{2s}G_2})$. For binary outcomes, we estimate the effect of the shift in $\mathbf{Z_{2s}G_2}$ via:

$$\begin{aligned} E[\hat{Y}^{90}-\hat{Y}^{10}] = \quad & \frac{1}{I}\sum_i w_i \Phi\left(\frac{[\mathbf{X_i\hat{B}}+\overline{\mathbf{\hat{X}_s\hat{G}_1}}+\overline{\mathbf{Z_{2s}\hat{G}_2}}+1.28(\widehat{Var}(\mathbf{Z_{2s}G_2}))^{.5}]}{(1+\widehat{Var}(v_s))^{.5}}\right) \\ - \; & \frac{1}{I}\sum_i w_i \Phi\left(\frac{[\mathbf{X_i\hat{B}}+\overline{\mathbf{\hat{X}_s\hat{G}_1}}+\overline{\mathbf{Z_{2s}\hat{G}_2}}-1.28(\widehat{Var}(\mathbf{Z_{2s}G_2}))^{.5}]}{(1+\widehat{Var}(v_s))^{.5}}\right), \end{aligned} \tag{24}$$

where I is the sample size, Φ is the CDF of a standard normal, and $\overline{\mathbf{\hat{X}_s\hat{G}_1}}$ and $\overline{\mathbf{Z_{2s}\hat{G}_2}}$ are the means across sampled schools of $\mathbf{\hat{X}_s\hat{G}_1}$ and $\mathbf{Z_{2s}\hat{G}_2}$, and w_i are individual-level weights. This average effectively integrates over the distribution of $\mathbf{X_iB}+v_{si}$, but uses the empirical distribution of $\mathbf{X_iB}$ since it is observed instead of imposing normality. Note that the scale of the latent index Y_{si} is unobserved, so we have normalized $Var(v_{si}-v_s)$ to 1.

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

$$\begin{aligned} E[\hat{Y}^{90}-\hat{Y}^{10}] = \quad & \frac{1}{I}\sum_i w_i \Phi\left(\frac{[\mathbf{X_i\hat{B}}+\overline{\mathbf{\hat{X}_s\hat{G}_1}}+\overline{\mathbf{Z_{2s}\hat{G}_2}}+1.28(\widehat{Var}(\mathbf{Z_{2s}G_2}+v_s))^{.5}]}{(1)}\right) \\ - \; & \frac{1}{I}\sum_i w_i \Phi\left(\frac{[\mathbf{X_i\hat{B}}+\overline{\mathbf{\hat{X}_s\hat{G}_1}}+\overline{\mathbf{Z_{2s}\hat{G}_2}}-1.28(\widehat{Var}(\mathbf{Z_{2s}G_2}+v_s))^{.5}]}{(1)}\right) \end{aligned} \tag{25}$$

We also report lower bound estimates of the impact of a 10th-to-50th percentile shift in school/neighborhood quality (a 1.28 standard deviation shift). Researchers often use a 1 standard deviation shift in treatment variable when assessing treatment effect sizes. Another natural benchmark would be based on the distribution of shifts in school/neighborhood quality resulting from moves that parents make, but we have not found a way to estimate this with our data.

For the binary outcomes, the impact of a shift in $\mathbf{Z_{2s}G_2}$ or $(\mathbf{Z_{2s}G_2} + v_s)$ will depend on the values of a student's observable characteristics, $\mathbf{X_iB}$. Thus, we report average impacts for certain subpopulations of interest as well.[21]

6 Data and Variable Selection

6.1 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.[22] 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. 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, number of completed years of college, and wage rates.

While these common properties are very helpful, differences in the surveys complicates efforts to compare results across time. In Altonji and Mansfield (2011) we only used variables that are available and consistently measured across all three data sets. 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 $\mathbf{X_i}$ student-level measures that may not appear in the other datasets. Section 6.3 details the process by which we chose what to include in $\mathbf{X_i}$, $\mathbf{X_s}$, and $\mathbf{Z_{2s}}$, and Table 1 provides a list.

The one major drawback associated with the three panel surveys is that the number of students sampled per school is only about 18 in NLS72, 24 in NELS88 and 20 in ELS2002. The simulation results presented in online Appendix A6 indicate that samples of this size may reduce to some

[21] With a nonseparable education production function the ordering of school/neighborhoods and the size of the effects depend on $\mathbf{X_i}$. See Agrawal et al. (in progress).

[22] 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.

degree the ability of sample school averages of observable characteristics $\hat{\mathbf{X}}_{\mathbf{s}}$ to serve as an effective control function for variation in average unobservable student contributions across schools.

Consequently, we also exploit administrative data from North Carolina on the universe of public schools and public school students (including charter schools) in the state. Since the North Carolina data contains information on every student at each school, it does not suffer from the same small subsample problem as the panel surveys. Furthermore, we can use the North Carolina data to assess the potential for bias in our survey-based estimates more directly. Specifically, we draw samples of students from North Carolina schools using either the NLS72, NELS88, or ELS2002 sampling schemes and re-estimate the model for high school graduation using these samples. By comparing the results derived from such samples to the true results based on the universe of students in North Carolina, we can determine which if any of the survey datasets is likely to produce reliable results. Online Appendix Tables A10 and A11 report the results of this exercise. They show that using school sample sizes whose distributions match the NLS72, NELS88, or ELS2002 distributions produces only relatively minor biases, generally making $\widehat{Var}(\mathbf{Z_{2s}G_2})$ and $\widehat{Var}(\mathbf{Z_{2s}G_2} + v_s)$ *more* conservative 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.[23] We then impute values for the other explanatory variables to preserve the sample size, since no other single variable is critical to our analysis.[24] Finally, we use panel weights. The appropriate weights depend on the analysis. See Online Appendix A12 for the details.

6.2 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.[25] It is not available in the North Carolina data. For NELS88

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

[24] This results in sample sizes (rounded to the nearest 10) for the four-year college enrollment analyses of: 12,260 from 900 schools for NLS72, 12,390 from 960 schools for NELS88, and 12,170 from 690 schools for ELS2002. The sample sizes and number of schools for the high school graduation analyses are 12,310 and 940 for NELS88, 12,100 and 690 for ELS2002, and 284,090 and 340 for North Carolina respectively. The analysis of years of postsecondary education uses 12,230 observations from 900 schools from NLS72, and the wage analysis uses 4,930 individuals with 9,860 wage observations from 900 schools. All individuals present in the base year are used to compute $\hat{\mathbf{X}}_{\mathbf{s}}$. 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 A12-A19 report percent imputed for each variable.

[25] 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.

and ELS2002 the measure of high school graduation is an indicator for whether a student has a high school diploma (not including a GED) as of two years after the high school graduation year of his/her cohort. For the North Carolina data, the measure is an indicator for whether the student is classified as graduated for the official state reporting requirement. Notice, though, that since ELS2002 first surveys students in 10th grade, it misses a substantial fraction of the early dropouts. Indeed, in NELS88, about one third of the 16 percent who eventually drop out do so before the first follow up survey in the middle of 10th grade. The North Carolina data considers students as eligible for official dropout statistics if they are enrolled in a North Carolina school at the beginning of 9th grade, so there is little scope for underestimating the 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 permanent wage analysis, our estimation procedure requires that we include only respondents who report wages in both 1979 and 1986 .

6.3 Selection of Control Function Variables, $\mathbf{Z_{2s}}$ and $\mathbf{X}_i$

First, we discuss $\mathbf{X_s}$. As we pointed out in Section 2, the control function variables (referred to as $\mathbf{C_s^*}$ in that section) should contain means of individual variables that influence school/neighborhood choice. We do not want to rule out the possibility that variables that affect Y_i also influence choice. Consequently, $\mathbf{C}_s^*$ includes $\mathbf{X}_s$, the group means of all variables in $\mathbf{X}_i$. In our application, we do not have $\mathbf{Q}_i$ variables that shift location preferences but not outcomes. Consequently, $\mathbf{C}_{s^*}$ is limited to $\mathbf{X}_s$ and does not contain any $\mathbf{Q}_s$ variables. Note that school level averages of student level variables that affect choice should be included in the control function even if one does not have individual level data. There are no such variables in the data sets we use.

What should be in $\mathbf{Z_{2s}}$? Observed school and neighborhood characteristics that could plausibly influence the socioeconomic outcome of interest. School policies, such as school security policies, belong in $\mathbf{Z_{2s}}$ even if the policies are in part a response to the characteristics of students. Under Proposition 1, the coefficients on the school policies are not affected by sorting bias. Any individual student who switches schools will be subject to the full difference across schools in the policy.

School level variables that are determined both by school policy/efficacy and the behavior of the students fall in a grey area. In ELS2002, we include *Frequency of Fights* at the school in control function variables in our full specification (described below) rather than in $\mathbf{Z_{2s}}$. This variable is determined by both school and neighborhood quality and by observed and unobserved characteristics of students. Consider the case in which *Frequency of Fights* is the sum of an unobserved school policy variable, say $Z^U_{fight,s}$, and some elements of $\mathbf{X_s}$ and $\mathbf{X_s^U}$. Provided the spanning condition holds, the $\mathbf{G_2}$ coefficient corresponding to *Frequency of Fights* will not be contaminated by sorting bias. It will be identified by variation in $Z^U_{fight,s}$ that is orthogonal to $\mathbf{X_s}$ and the included $\mathbf{Z_{2s}}$ variables. However, because part of the variation across schools in *Frequency of Fights* is driven by the *direct effect* of $\mathbf{X^U_s}$ on fighting, one would overstate the importance of differences in school policy toward fighting if one were to treat *Frequency of Fights* as a $\mathbf{Z_{2s}}$ variable when measuring the variance of school treatment effects.[26] The fact that the policy variable $Z^U_{fight,s}$ is in part a function of $\mathbf{X_s}$ and/or $\mathbf{X^U_s}$ is not a problem. 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. The same issues apply to test scores measured during high school. Test scores are determined by school quality and by student characteristics. We never include them in $\mathbf{Z_{2s}}$. This is conservative.

What should **not** be in $\mathbf{Z_{2s}}$? $\mathbf{Z_{2s}}$ 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 $\mathbf{X_s}$ variables regardless of whether the measures are aggregates of the student micro data, Census data or administrative data from the schools. Even if sampling error in $\hat{\mathbf{X}}_\mathbf{s}$ breaks multicollinearity among the components of $\mathbf{X_s}$ that exists when the spanning condition holds, elements of $\hat{\mathbf{X}}_s$ should not be included in $\mathbf{Z_{2s}}$ because they pick up effects of x^U_s. In contrast, it may be prudent in some applications to treat group characteristics that are strongly related to the amenities that drive sorting as part of the control function rather than as part of $\mathbf{Z_{2s}}$ if $\mathbf{X_s}$ is too limited to plausibly control for sorting. The cost is likely to be a more conservative lower bound.

$\mathbf{X_i}$ 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 $\mathbf{X_i}$ measures of student behavior (e.g., fighting, hours/week spent on homework), parental expectations, and student academic ability (standardized test scores) and includes the corresponding school averages in $\mathbf{X_s}$. Such measures may be influenced directly by school inputs, so including them in $\mathbf{X_i}$ 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 from $\mathbf{X_i}$ and $\mathbf{X_s}$ 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 used as the control function.

Table 1 lists the final choices of individual-level and school-level explanatory measures used in each dataset. Online Appendix Tables A12 - A19 provide the mean, standard deviation, and percent of observations imputed for each individual-level and school-level characteristic for each of our four datasets.

[26]Let $Frequency\_of\_Fights_s = Z^U_{fight,s}a_1 + \mathbf{X_s a_2} + \mathbf{X^U_s a_3}$ where a_1, $\mathbf{a_2}$, and $\mathbf{a_3}$ are coefficients or coefficient vectors. Let G_{2FF} be the coefficient on $Frequency\_of\_Fights_s$. The variance $(G_{2FF})^2 var(Z^U_{fight,s}a_1 + \mathbf{X}_s\mathbf{a}_2 + \mathbf{X}^U_s\mathbf{a}_3)$ is likely to overstate $(G_{2FF})^2 var(Z^U_{fight,s}a_1)$, which is the contribution of the policy variation in school policy toward fighting.

7 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. We bootstrap the entire estimation procedure, including imputation of missing data, estimation of model parameters, variance decompositions, and treatment effects. 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.

7.1 High School Graduation

The full variance decompositions described in Section 5 are provided for each of our outcomes in Online Appendix Tables A20, A21, and A22. 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 rationale 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 $\mathbf{Z_{2s}}$ (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 5.1 percent of the total student-level variance in the latent index can be attributed exclusively to school system and neighborhood contributions. Since the set of observed individual-level measures contained in $\mathbf{X_i}$ is limited in the North Carolina data, it is possible that the WTP coefficient matrix $\tilde{\boldsymbol{\Theta}}$ associated with our control function of school-averages $\mathbf{X_s}$ does not span $\boldsymbol{\Theta}^U$ and thus 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 $\mathbf{X_i}$ 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 5.1 percent to 3.8 percent of the latent index variance.

NELS88 features a slightly smaller fraction of the variance attributable to schools/neighborhoods than NC in the baseline specification (5.0% versus 5.1%) and a slightly larger fraction in the full specification (4.4% versus 3.8%). The values for ELS2002 are smaller in both baseline and full cases (3.2% and 2.1%). Aside from sampling error, the fact that the three data sets start at different points (8th grade, 9th grade, and 10th grade) and that NELS88 and ELS2002 have richer controls and school level variables should lead to some differences across the data sets. The relatively sparse set of student-level variables for North Carolina (even in the full specification) may be a factor. It might lead v_s to contain some between-school variation in student unobservables x_s^U that is unabsorbed by the control function (the spanning condition in Proposition 1 fails). By contrast, ELS2002 has the richest set of both student-level and school-level observables, so that there is very little residual school-level variation that cannot be captured by either the control function $\mathbf{X_s}$ or the school-level observables $\mathbf{Z_{2s}}$.

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 A20 provides a full variance decomposition that shows the critical role that individual-specific factors play. However, to get a better sense of the difference that an effective school system and neighborhood can make, we use these two alternative lower bound variance estimates with (24) and (25) 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. The estimates are in Panel B of Table 2. They correspond to 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 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 17.8 percentage points relative to a school at the 10th quantile (from 67.4% to 85.2%) The corresponding estimates are 13.5 percentage points for NELS88 (78.1% to 91.6%) and 7.6 percentage points for ELS2002 (86.6% 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 15.3, 12.8 and 6.2 percentage points in NC, NELS88, and ELS2002, respectively. Notice further that the latter estimates are quite large despite the fact that the fractions of variance upon which they are based is small: 3.8, 4.4, and 2.1 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 a 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.

Note that differences in the treatment effect estimates across data sets reflect both differences in the variances of the school/neighborhood component and differences in the sample average graduation rates across the datasets. The graduation rate is 77 percent in the North Carolina data, 83 percent in NELS88, and 90 percent in ELS2002. As a result, a shift of the same magnitude will induce a greater increase in graduation rate in the North Carolina and NELS88 data than in ELS2002 because there are fewer students near the decision margin in ELS2002. Intuitively, as the sample average converges to 100 percent graduation almost the entire population is well above the decision threshold. Consequently, school-related shifts in the latent index that determines graduation become less relevant.

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 from high school.

7.2 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 consistency in both of the lower bound estimates of the school/neighborhood contribution across datasets and generations. In the baseline case, estimates that exclude the between-school residual v_s attribute at least 1.7 to 2.7 percent of the outcome variance to schools/neighborhoods, while estimates that include v_s attribute 4.8 to 5.2 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 1.4 and 1.8 percent, and the estimates that include the residual v_s dropping to between 3.7 and 3.8 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. Note 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 far from 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 11.1 and 11.6 percentage points (Row 1, Columns 2, 4, and 6 of Panel B). Specifically, a 10th to 90th quantile shift in the school/neighborhood component $\mathbf{Z_{2s}G_2}$ increases enrollment rates from 21.2% to 32.7% in NLS72, from 25.7% to 36.8% in NELS88, and from 31.0% to 42.6% in ELS2002. Including the residual between-school component boosts the range of estimates to 16.6 to 18.8 percentage points. For this specification a 10th-to-50th quantile shift has an average estimated impact ranging from 7.7 and 9.0 percentage points. Estimates are larger for the baseline set of controls.

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.

7.3 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/neighborhood versus starting them at a 90th quantile school/neighborhood. 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_{si} 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 $\mathbf{X_i\hat{B}}$ that place them closer to an outcome 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 $\mathbf{X_i\hat{B}}$ places them at the 10th quantile of the $\mathbf{X_i\hat{B}}$ distribution (Rows 1 and 2), and students at the 90th quantile of the $\mathbf{X_i\hat{B}}$ distribution (Rows 3 and 4). For the North Carolina sample and the full specification (Column 2), the lower bound estimate that includes the between-school residual component v_s suggests a 23.3 percentage point increase for students at the 10th quantile (42.8% to 66.1%). The increase is only 5.8% for students at the 90th quantile (91.9% to 97.7%). For NELS88 grade 8 (Column 4), the numbers are very similar at the 10th quantile but smaller at the 90th quantile. The lower bound estimates that include v_s are 24.1 percentage points (54.2 to 78.3) at the 10th quantile and 2.6 percentage points (96.7% to 99.4%) at the 90th quantile. The ELS2002 estimate is 12.8% at the 10th percentile of $\mathbf{X_i\hat{B}}$ and only 0.9% at the 90th, reflecting the lower dropout rate in ELS2002. The results for all three data sets 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 by 15.3 percentage points from 68.3% to 83.6%. The corresponding increase for white students with single mothers who did not attend college is 21.0 percentage points (49.5% to 70.5%), while the increase for white students with both parents present, at least one of whom completed college, is 7.9 percentage points (87.8% to 95.7%). Across all specifications, the estimated increases in graduation rates in the NELS88 and ELS samples are between 5.0 and 14.2 percentage points for black students and 5.7 and 20.8 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 $\mathbf{X_i\hat{B}}$ distribution are substantially less sensitive to school quality. This reflects 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 $\mathbf{Z_{2s}G_2} + v_s$ would increase the college enrollment rates of students at the 10th percentile of $\mathbf{X_i\hat{B}}$ from 2.0% to 9.0%. More generally, the lower bound estimates that exclude and include v_s are between 2.6 and 9.1 percentage points and between 3.6 and 15.0 percentage points, respectively, depending on the dataset and specification. In contrast, for students at the 90th percentile of $\mathbf{X_i\hat{B}}$ the ELS2002 estimate from the full specification suggests that a 10th-90th shift in $\mathbf{Z_{2s}G_2} + v_s$ would increase enrollment rates at four-year colleges by 18.1 percentage points (from 72.3% to 90.4%). More generally, across datasets the lower bound estimates excluding and including v_s for students at the 90th percentile of the $\mathbf{X_i\hat{B}}$ distribution are between 11.2 and 19.3 percentage points and 18.1 and 26.5 percentage points, respectively. The values for blacks and for whites with non-college-educated single mothers are about 1 percent and about 3 percent (respectively) below the results for the full sample. The values for whites with college educated parents are close to those for the 90th percentile of the $\mathbf{X_i\hat{B}}$ 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.

7.4 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 .22 years, which is about .12 standard deviations. Including standardized tests among the observable characteristics reduces this estimate to 0.02 years, which is not statistically significant. The lower bound is not informative in this case. Note, though, that since the NLS72 data are collected in 12th grade, the standardized test scores are particularly likely to reflect high school quality, making the full specification particularly conservative. Furthermore, adding in the contribution of v_s raises these estimates to .50 and .41 years respectively. 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 13.7 percent $(100e^{0.128} - 100)$. The 10-50 shifts are half as large at around 6.6 percent. Estimates that include v_s imply that a 10-90 shift in school/neighborhood quality increases wages by 0.20 log points or about 22 percent. For comparison, standard deviation of the log permanent wage component is 0.305. Thus, at least for the 1972 cohort, shifts in school/neighborhood quality seem to have important impacts on longer run outcomes of prime importance for worker welfare.

Chetty and Hendren (2015) find that 20 years in a one standard deviation better neighborhood raises the log of adult earnings by about 0.10. When we include v_s we find that a one standard deviation shift in school/neighborhood raises log permanent wage rates by 0.078. Several factors contribute to the modest difference in the estimates from the studies.[27]

7.5 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.

In Online Appendix Tables A1 - A2 we report estimates of $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$, or equivalently $Var(Y_s - \mathbf{X_s B})$. By including $\mathbf{X_s G_1}$, these estimates reintroduce 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 $\mathbf{X_s}$ given $\mathbf{Z_{2s}}$. But $\mathbf{X_s G_1}$ also includes the component $\mathbf{X_s \Pi}_{x_i^U \mathbf{X_s}}$, which reflects student sorting on unobservable characteristics. In Online Appendix A7, we show that $\mathbf{X_s \Pi}_{x_i^U \mathbf{X_s}} = 0$ when $\mathbf{X_i^U}$ does not affect location preferences and the assumptions of Proposition 1 hold, in which case $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s) = Var(\mathbf{Z_s \Gamma} + z_s^U + \xi_s)$. This is the true variance in school/neighborhood treatment effects (including common shocks). When unobservables do contribute to sorting, then $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$ will generally overstate the variance in school/neighborhood treatment effects.[28]

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 ELS2002, the sorting-on-observables estimator attributes 3.2% of the variance in the latent index that determines high school graduation to schools/neighborhoods, compared to 2.1% for the lower bound estimate of $Var(\mathbf{Z_{2s} G_2} + v_s)$. The associated effect of a 10th-to-90th quantile shift

[27] Sources of the difference in estimates include differences in the outcome measure (wage rate versus earnings), school/neighborhood geography (school versus county), birth cohort, period of exposure to a school/neighborhhood and the fact that our estimates are lower bounds.

[28] From (14), (18), and (19), $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s) = Var(\mathbf{Z_s \Gamma} + z_s^U + \xi_s + \mathbf{X_s \Pi}_{x_i^U \mathbf{X_s}} \boldsymbol{\beta}^{\mathbf{U}})$. If the covariances between $\mathbf{X_s \Pi}_{x_i^U \mathbf{X_s}}$ and the components of the school treatment effect $\mathbf{Z_s \Gamma} + z_s^U + \xi_s$ are sufficiently negative, then one can find $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s) < Var(\mathbf{Z_s \Gamma} + z_s^U + \xi_s)$. In this case, which we consider unlikely, even $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$ would understate the true contribution of schools/neighborhoods to the variance in outcomes.

in school/neighborhood quality on graduation is 0.077 (relative to 0.062 for the lower bound estimate).[29] For enrollment in a four-year college, the corresponding school/neighborhood variance fractions for the ELS full specification is 4.3% (versus 3.7% for the lower bound estimate). These variances correspond to 10th-to-90th shifts in the probability of enrollment of 0.204 (versus 0.188). $\widehat{Var}(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$ is higher than $\widehat{Var}(\mathbf{Z_{2s} G_2} + v_s)$ in every dataset and specification we consider.

In Online Appendix Table A3 we report estimates of $Var(\mathbf{Z_{2s} G_2})$ and $Var(\mathbf{Z_{2s} G_2} + v_s)$ from a four-year college enrollment specification in which the school-averages $\mathbf{X_s}$ are omitted. A small fraction of the variance previously absorbed by the control function is now captured by $\mathbf{Z_{2s}\hat{G}_2}$, while the bulk of it now enters the between-school residual $\hat{v}_s$. Thus, $\widehat{Var}(\mathbf{Z_{2s} G_2})$ increases slightly relative to our main college enrollment estimates in Table 3, while $\widehat{Var}(\mathbf{Z_{2s} G_2} + v_s)$ increases substantially, to the point that they typically exceed the sorting-on-observables estimates $\widehat{Var}(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$ reported in the previous paragraph. Online Appendix Table A4 report corresponding estimates for years of postsecondary education, and again exhibit substantially higher estimates of $Var(\mathbf{Z_{2s} G_2} + v_s)$.

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.

7.6 Empirical Evidence on the Spanning Condition

In Online Appendix A3 we explore the factor structure of $\mathbf{X_s}$ to test Assumption A5.1, which is a necessary condition for Assumption A5 and therefore Proposition 1. Recall that Assumption A5.1 is violated if the number of amenity factors driving sorting on $\mathbf{X_i}$ ($\mathbf{A^X}$) exceeds the number of observed characteristics that compose $\mathbf{X_i}$ ($Dim(\mathbf{X_i})$). We adopt two separate approaches. First, we use principal components analysis to compute the eigenvalues and eigenvectors of $\widehat{\mathbf{Var}}(\mathbf{X_s})$, the estimated covariance matrix of $\mathbf{X_s}$. While $\mathbf{Var}(\mathbf{X_s})$ must be positive semidefinite, $\widehat{\mathbf{Var}}(\mathbf{X_s})$ 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 $\widehat{\mathbf{Var}}(\mathbf{X_s})$ is rank deficient. This means that each element of $\mathbf{X_s}$ 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

[29] The effects of a 10-to-90th shift in $\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s$ are constructed as

$$E[\hat{Y}^{90} - \hat{Y}^{10}] = \frac{1}{I}\sum_i w_i \Phi\left(\frac{[\mathbf{X_i}\hat{\mathbf{B}} + \overline{\hat{\mathbf{X}}_\mathbf{s}\hat{\mathbf{G}}_\mathbf{1}} + \overline{\mathbf{Z_{2s}}}\hat{\mathbf{G}}_\mathbf{2} + 1.28(\widehat{Var}(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s))^{.5}]}{(1)}\right)$$
$$- \frac{1}{I}\sum_i w_i \Phi\left(\frac{[\mathbf{X_i}\hat{\mathbf{B}} + \overline{\hat{\mathbf{X}}_\mathbf{s}\hat{\mathbf{G}}_\mathbf{1}} + \overline{\mathbf{Z_{2s}}}\hat{\mathbf{G}}_\mathbf{2} - 1.28(\widehat{Var}(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s))^{.5}]}{(1)}\right). \quad (26)$$

dataset). Since the rank of $\mathbf{Var}(\mathbf{X_s})$ should reflect the dimension of the amenity vector $\mathbf{A^X}$, this supports our assumption that the dimension of $\mathbf{A^X} \leq L$. 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 $\mathbf{X_s}$, suggesting that the variation in student composition across schools is driven primarily by a small number of amenity factors. Bootstrap 90% 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 $\mathbf{X_s}$ is larger in the full specification, which contains more variables. This would be expected in the presence of sampling error in $\widehat{\mathbf{Var}}(\mathbf{X_s})$. 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 $\mathbf{X_s}$ 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 $\widehat{\mathbf{Var}}(\mathbf{X_s})$, and can accommodate arbitrary forms of heteroskedasticity and correlation at the school level. We perform tests of the null hypothesis of $\text{rank}(\mathbf{Var}(\mathbf{X_s})) = j)$ against the alternative that $\text{rank}(\mathbf{Var}(\mathbf{X_s})) > j$. For all three data sets and specifications, we cannot reject the null hypothesis for values of j well below L. See Online Appendix Tables A6 and A7.

8 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. 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 substantial 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.

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. In Online Appendix

A13, we consider identification of teacher value added. Teacher value added is one 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. We show that most of the analysis in Section 2 can be adapted to the teacher value added application, although we have not yet extended Proposition 1 to the case in which the administrator internalizes the effect that allocating a particular student to a classroom has on the other students. Nevertheless, our analysis suggests that the common practice of including classroom averages of student characteristics (such as in Chetty et al. (2014)) may play a potentially powerful role in purging value-added estimates of biases stemming from non-random student sorting on unobservables and observables.

Importantly, unlike the school application, where we only bound the variance of school effects, one can point identify teacher value added because the treatment of interest (the teacher) is observed in multiple groups or units (classrooms). The classrooms feature varying levels of other valued amenities (e.g. time of day) that cause within-teacher variation in group composition. Recent work by Fletcher et al. (2014) uses patient data matched to physicians to estimate the effects of physicians on health outcomes. They control for very detailed patient characteristics but not for the physician-specific averages of patient characteristics. Our analysis suggests that adding physician-clinic averages of patient characteristics for doctors who work in more than one setting would allay concerns about sorting on patient unobservables.

Consider also cases in which a particular group input, say spending per pupil, varies experimentally or quasi-experimentally across schools. This solves the problem that spending per pupil is correlated with other unobserved education inputs, such as peer quality. But as Caetano (2012) points out, individuals may re-sort in response to the change in spending, leading to bias in a difference-in-difference design. Our analysis suggests that one might be able to identify the causal effect of spending by combining an IV strategy with the use of school averages of individual characteristics to address the sorting problem.[30]

In principle, one could adapt 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. We leave an analysis of this possibility to future research.[31]

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. We are currently pursuing this in Agrawal et al. (in progress). Future research

[30]We have in mind an IV regression of Y_{it} on $\mathbf{X_{it}}$, $\mathbf{X_{st}}$ and $\mathbf{Z_{2st}}$ in the post policy intervention period using the pre-period value $\mathbf{X_{st-1}}$ and the exogenous policy change as the excluded instrumental variables. Similarly, when re-sorting is not considered likely but the policy intervention instrument depends in part on $\mathbf{X^U_{st-1}}$, using $\mathbf{X_{st}}$ to control for $\mathbf{X^U_{st}}$ might also lead to a valid IV strategy.

[31]Let s denote number of years of schooling. Each schooling level has an associated set of characteristics $\mathbf{A_s}$ governing the pecuniary and non-pecuniary return to choosing level s. $\mathbf{A_s}$ is weighted by $\mathbf{X_i}$ and $\mathbf{X^U_i}$. This leads to a relationship between $\mathbf{X_s}$ and $\mathbf{X^U_s}$ that could serve as the basis for a control function for $\mathbf{X^U_s}$. 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 $\mathbf{X_s}$. Otherwise s will not vary conditional on $\mathbf{X_s}$ unless restrictions are available that reduce the dimension of the index of $\mathbf{X_s}$ required to control for $\mathbf{X^U_s}$. Essentially, there are fewer degrees of freedom (the number of levels) than there are parameters in the coefficient vector on $\mathbf{X_s}$ ($\mathbf{G_1}$ in our outcome equation).

should also examine whether a variant of Proposition 1 carries over to more general specifications of preferences than the class that we work with, and to two-sided selection problems, such as the sorting of students across universities or workers across firms.

A1 Proof of Proposition 1

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

$$\Pr(s(i)=s|\mathbf{X_i},\mathbf{X_i^U},\mathbf{Q_i},\mathbf{W_i})=\Pr(s(i)=s|\mathbf{W_i}) \tag{27}$$

where Pr(.) is the probability function. The above fact and Bayes rule imply that[32]

$$f(\mathbf{X_i}|\mathbf{W_i},s(i)=s)=f(\mathbf{X_i}|\mathbf{W_i}) \tag{28}$$

$$f(\mathbf{X_i^U}|\mathbf{W_i},s(i)=s)=f(\mathbf{X_i^U}|\mathbf{W_i})\ . \tag{29}$$

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

$$\mathbf{X_s^U} \equiv \mathbf{E}[\mathbf{X_i^U}|s(i)=s]=\mathbf{E}[\mathbf{E}(\mathbf{X_i^U}|\mathbf{W_i},s(i)=s)|s(i)=s]=\mathbf{E}[\mathbf{E}(\mathbf{X_i^U}|\mathbf{W_i})|s(i)=s] \tag{30}$$

$$\mathbf{X_s} \equiv \mathbf{E}[\mathbf{X_i}|s(i)=s]=\mathbf{E}[\mathbf{E}(\mathbf{X_i}|\mathbf{W_i},s(i)=s)|s(i)=s]=\mathbf{E}[\mathbf{E}(\mathbf{X_i}|\mathbf{W_i})|s(i)=s]. \tag{31}$$

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

$$\mathbf{Cov}(\mathbf{W_i'},\mathbf{\tilde{X}_i^U})=\mathbf{Cov}(\mathbf{\Theta^{U'}}\mathbf{\tilde{X}_i^{U'}},\mathbf{\tilde{X}_i^U})=\mathbf{\Theta^{U'}}\mathbf{Var}(\mathbf{\tilde{X}_i^U}) \tag{32}$$

$$\mathbf{Cov}(\mathbf{W_i'},\mathbf{X_i})=\mathbf{Cov}(\mathbf{\tilde{\Theta}'}\mathbf{X_i'},\mathbf{X_i})=\mathbf{\tilde{\Theta}'}\mathbf{Var}(\mathbf{X_i}). \tag{33}$$

Since from assumption A4 $\mathbf{E}[\mathbf{X_i}|\mathbf{W_i}]$ and $\mathbf{E}[\mathbf{X_i^U}|\mathbf{W_i}]$ are linear in $\mathbf{W_i}$, $\mathbf{E}[\mathbf{\tilde{X}_i^U}|\mathbf{W_i}]$ is also linear in $\mathbf{W_i}$.

[32] One can write the conditional density $f(\mathbf{X_i}|\mathbf{W_i},s_i=s)$ as

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

where the first equality is Bayes rule, the second equality uses (27), and the third follows from cancellation of terms and Bayes rule. The same line of argument establishes that $f(\mathbf{X_i^U}|\mathbf{W_i},s(i)=s)=f(\mathbf{X_i^U}|\mathbf{W_i})$.

Consequently, assumption A4, equations (32)-(33), and basic regression theory imply that

$$\mathbf{E}[\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}|\mathbf{W_i}] = \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1}\mathbf{Cov}(\mathbf{W}_{\mathbf{i}}', \tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) = \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1}\mathbf{\Theta}^{\mathbf{U}'}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) \quad (34)$$

$$\mathbf{E}[\mathbf{X_i}|\mathbf{W_i}] = \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1}\mathbf{Cov}(\mathbf{W}_{\mathbf{i}}', \mathbf{X_i}) = \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1}\tilde{\mathbf{\Theta}}'\mathbf{Var}(\mathbf{X_i}). \quad (35)$$

Next, if we use the spanning assumption A5 to replace $\mathbf{\Theta}^{\mathbf{U}'}$ with $\tilde{\mathbf{\Theta}}'\mathbf{R}'$ in (34), and then use the expression for $\mathbf{E}[\mathbf{X_i}|\mathbf{W_i}]$ from (35), we obtain:

$$\begin{aligned}\mathbf{E}[\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}|\mathbf{W_i}] &= \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1}\tilde{\mathbf{\Theta}}'\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})\\ &= \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1}\tilde{\mathbf{\Theta}}'\mathbf{Var}(\mathbf{X_i})\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})\\ &= \mathbf{E}[\mathbf{X_i}|\mathbf{W_i}]\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}). \quad (36)\end{aligned}$$

To find $\mathbf{E}[\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}|\mathbf{W_i}]$ first take expectations of both sides of (5) conditional on $\mathbf{W_i}$:

$$\mathbf{E}[\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}|\mathbf{W_i}] = \mathbf{E}[\mathbf{X_i}|\mathbf{W_i}]\mathbf{\Pi}_{\mathbf{X^U X}} + \mathbf{E}[\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}|\mathbf{W_i}]. \quad (37)$$

Substitution for $\mathbf{E}[\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}|\mathbf{W_i}]$ using (36) leads to

$$\mathbf{E}[\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}|\mathbf{W_i}] = \mathbf{E}[\mathbf{X_i}|\mathbf{W_i}](\mathbf{\Pi}_{\mathbf{X^U X}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})). \quad (38)$$

The final step is to take expectations of both sides of the above equation conditional on $s(i) = s$ and employ equations (30) and (31). Doing so leads to

$$\mathbf{X}_{\mathbf{s}}^{\mathbf{U}} = \mathbf{X_s}[\mathbf{\Pi}_{\mathbf{X^U X}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})].$$

This completes the proof.

References

Aaronson, Daniel (1998) 'Using sibling data to estimate the impact of neighborhoods on children's educational outcomes.' *Journal of Human Resources* pp. 915–946

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer (2015) 'Identification properties of recent production function estimators.' *Econometrica* 83(6), 2411–2451

Agrawal, Mohit, Joseph Altonji, and Richard Mansfield (in progress) 'Quantifying family, school, and location effects in the presence of complementarities and sorting'

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, 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

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters (2016) 'Stand and deliver: Effects of bostons charter high schools on college preparation, entry, and choice.' *Journal of Labor Economics* 34(2), 275–318

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)

Caetano, Gregorio (2012) 'Neighborhood sorting and the valuation of public school quality.' *Unpublished paper, University of Rochester*

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 (2015) 'The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates.' Working Paper

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 (2016) 'The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment.' *The American Economic Review* 106(4), 855–902

Chiappori, Pierre-Andr, and Bernard Salani (2016) 'The econometrics of matching models.' *Journal of Economic Literature* 54(3), 832–61

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 A Kleiner (2015) 'Measuring returns to hospital care: Evidence from ambulance referral patterns.' *Journal of Political Economy* 123(1), 170–214

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

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

Graham, Bryan S (2016) 'Identifying and estimating neighborhood effects.' Technical Report, National Bureau of Economic Research

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

Harvey, Walter Robert (1977) *User's guide for LSML76: Mixed model least-squares and maximum likelihood computer program* (Ohio State University)

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

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 (2017) 'Sorting multidimensional types: Theory and application.' *The Review of Economic Studies* 84(2), 718–789

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*

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)	X	X	X	X
1(Immigrant)	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 $\mathbf{X_s}$)*				
School Pct. Minority	X	X	X	
School Pct. Free/Reduced Price Lunch		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	X	
School Characteristics (Treated as elements of $\mathbf{Z_{2s}}$)				
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/ 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
# of Library Books per Student				X
Neighborhood Characteristics (Treated as elements of $\mathbf{Z_{2s}}$)				
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 $\mathbf{X_s}$ 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 $\mathbf{X_s}$)" 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	0.010	0.011	0.006	0.012	0.009
$Var(\mathbf{Z_{2s}G_2})$	(0.008)	(0.004)	(0.006)	(0.007)	(0.010)	(0.009)
LB w/ unobs	0.051	0.038	0.050	0.044	0.032	0.021
$Var(\mathbf{Z_{2s}G_2} + v_s)$	(0.017)	(0.010)	(0.009)	(0.010)	(0.010)	(0.009)

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	0.104	0.079	0.064	0.048	0.047	0.041
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.019)	(0.010)	(0.013)	(0.022)	(0.013)	(0.013)
LB w/ unobs: 10th-90th	0.178	0.153	0.135	0.128	0.076	0.062
Based on $Var(\mathbf{Z_{2s}G_2} + v_s)$	(0.023)	(0.016)	(0.017)	(0.019)	(0.012)	(0.012)
LB no unobs: 10th-50th	0.055	0.042	0.035	0.026	0.026	0.022
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.011)	(0.006)	(0.008)	(0.012)	(0.008)	(0.008)
LB w/ unobs: 10th-50th	0.098	0.084	0.079	0.073	0.044	0.034
Based on $Var(\mathbf{Z_{2s}G_2} + v_s)$	(0.015)	(0.010)	(0.010)	(0.012)	(0.008)	(0.007)
Sample Mean	0.769	0.769	0.827	0.827	0.897	0.897

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 $Var(\mathbf{Z_{2s}G_2})$ and excludes the unobservable v_s while the row labeled "LB w/ unobs" reports $Var(\mathbf{Z_{2s}G_2} + v_s)$.

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 Online Appendix Table A20.

Online Appendices A9 and A10 discuss estimation of model parameters and the variance decompositions. Section 5.4 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	0.027	0.018	0.017	0.015	0.019	0.014
$Var(\mathbf{Z_{2s}G_2})$	(0.006)	(0.005)	(0.006)	(0.005)	(0.010)	(0.008)
LB w/ unobs	0.048	0.038	0.049	0.038	0.052	0.037
$Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.009)	(0.007)	(0.009)	(0.008)	(0.017)	(0.011)

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	0.140	0.115	0.123	0.111	0.138	0.116
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.016)	(0.013)	(0.018)	(0.017)	(0.020)	(0.020)
LB w/ unobs: 10th-90th	0.189	0.166	0.208	0.180	0.230	0.188
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.019)	(0.017)	(0.020)	(0.020)	(0.021)	(0.020)
LB no unobs: 10th-50th	0.065	0.054	0.059	0.053	0.067	0.057
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.007)	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)
LB w/ unobs: 10th-50th	0.086	0.077	0.096	0.084	0.108	0.090
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.008)	(0.007)	(0.009)	(0.009)	(0.010)	(0.009)
Sample Mean	.267	.267	.310	.310	.365	.365

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	0.144	0.120	0.109	0.090	0.082	0.086
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.023)	(0.015)	(0.024)	(0.041)	(0.023)	(0.028)
LB w/ unobs	0.247	0.233	0.233	0.241	0.134	0.128
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.027)	(0.022)	(0.030)	(0.037)	(0.021)	(0.024)
XB: 90th Quantile						
LB no unobs	0.060	0.031	0.019	0.011	0.013	0.006
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.013)	(0.005)	(0.004)	(0.005)	(0.004)	(0.002)
LB w/ unobs	0.100	0.058	0.039	0.027	0.021	0.009
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.017)	(0.008)	(0.005)	(0.005)	(0.004)	(0.002)
Black						
LB no unobs	0.105	0.079	0.067	0.052	0.053	0.050
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.019)	(0.010)	(0.015)	(0.024)	(0.016)	(0.017)
LB w/ unobs	0.179	0.153	0.142	0.138	0.086	0.075
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.023)	(0.016)	(0.018)	(0.021)	(0.014)	(0.015)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.141	0.109	0.098	0.076	0.066	0.057
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.018)	(0.010)	(0.022)	(0.021)	(0.012)	(0.011)
LB w/ unobs	0.242	0.210	0.208	0.202	0.108	0.085
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.023)	(0.016)	(0.028)	(0.019)	(0.011)	(0.010)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.058	0.041	0.029	0.020	0.021	0.016
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.013)	(0.006)	(0.006)	(0.009)	(0.007)	(0.006)
LB w/ unobs	0.098	0.079	0.060	0.052	0.034	0.024
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.017)	(0.009)	(0.008)	(0.009)	(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 $\mathbf{X_iB}$ is equal the estimated 10th (90th) quantile value of the $\mathbf{X_iB}$ distribution. See Section 7.3.

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	0.079	0.026	0.070	0.044	0.091	0.044
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.010)	(0.004)	(0.011)	(0.007)	(0.014)	(0.009)
LB w/ unobs	0.105	0.036	0.118	0.070	0.150	0.070
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.012)	(0.005)	(0.013)	(0.009)	(0.016)	(0.009)
XB: 90th Quantile						
LB no unobs	0.193	0.177	0.155	0.142	0.150	0.112
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.022)	(0.020)	(0.022)	(0.021)	(0.021)	(0.020)
LB w/ unobs	0.260	0.258	0.265	0.231	0.246	0.181
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.026)	(0.027)	(0.025)	(0.026)	(0.023)	(0.021)
Black						
LB no unobs	0.132	0.105	0.122	0.111	0.131	0.107
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.016)	(0.013)	(0.018)	(0.017)	(0.019)	(0.018)
LB w/ unobs	0.178	0.153	0.208	0.180	0.217	0.174
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.020)	(0.017)	(0.020)	(0.020)	(0.021)	(0.019)
White w/ Single Mother Who Did Not Attend College						
LB no unobs	0.110	0.094	0.098	0.081	0.126	0.110
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.016)	(0.014)	(0.018)	(0.017)	(0.021)	(0.021)
LB w/ unobs	0.148	0.137	0.166	0.131	0.210	0.178
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.020)	(0.018)	(0.021)	(0.021)	(0.023)	(0.022)
White w/ Both Parents, At Least One Completed College						
LB no unobs	0.182	0.154	0.149	0.138	0.154	0.129
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.020)	(0.018)	(0.022)	(0.021)	(0.022)	(0.022)
LB w/ unobs	0.246	0.224	0.254	0.225	0.256	0.209
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.025)	(0.023)	(0.025)	(0.025)	(0.024)	(0.023)

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	0.002	0.000	0.025	0.028	0.032	0.033
$Var(\mathbf{Z_{2s}G_2})$	(0.002)	(0.001)	(0.010)	(0.011)	(0.012)	(0.012)
LB w/ unobs	0.013	0.009	0.070	0.070	0.085	0.082
$Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.004)	(0.003)	(0.019)	(0.019)	(0.023)	(0.022)

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	0.216	0.018	0.121	0.128	0.125	0.128
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.064)	(0.064)	(0.021)	(0.022)	(0.021)	(0.022)
LB w/unobs: 10th-90th	0.502	0.407	0.203	0.203	0.203	0.200
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.077)	(0.076)	(0.031)	(0.030)	(0.031)	(0.031)
LB no unobs: 10th-50th	0.108	0.009	0.061	0.064	0.063	0.064
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.032)	(0.032)	(0.011)	(0.011)	(0.011)	(0.011)
LB w/unobs: 10th-50th	0.251	0.203	0.102	0.102	0.102	0.100
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.039)	(0.038)	(0.016)	(0.015)	(0.016)	(0.015)
Sample Mean	1.62	1.62	2.88	2.88	2.88	2.88

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 $Var(\mathbf{Z_{2s}G_2})$ and excludes the unobservable v_s while the row labeled "LB w/ unobs" reports $Var(\mathbf{Z_{2s}G_2}+v_s)$.

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 $[\widehat{Var}(\mathbf{Z_{2s}G_2}+v_s)]^{0.5}$ or $[\widehat{Var}(\mathbf{Z_{2s}G_2})]^{0.5}$ 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 Online Appendix Table A22. Online Appendices A9 and A10 discuss estimation of model parameters and the variance decompositions.

Supplementary Material: For Online Publication Only

Table of Contents:

Appendices:

Tables:

A2 Spanning Condition Examples

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

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

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

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

Then we can write $\mathbf{\Theta^U} = \mathbf{R}\tilde{\mathbf{\Theta}}$, where

$$\mathbf{R} = \left\{ \begin{matrix} 1 & 1 \\ 2 & -1 \end{matrix} \right\}$$

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

Now suppose that there are instead three outcome-relevant unobserved characteristics: $\mathbf{X^U} = [X_1^U, X_2^U, X_3^U]$, each of which affects WTP for the two amenities differentially. Suppose that $\mathbf{X}$ and $\tilde{\mathbf{\Theta}}$ are unchanged from Case 1:

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

Then we can write $\mathbf{\Theta^U} = \mathbf{R}\tilde{\mathbf{\Theta}}$, where

$$\mathbf{R} = \left\{ \begin{matrix} 1 & 1 \\ 2 & -1 \\ 1 & 0 \end{matrix} \right\}$$

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

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

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 $\mathbf{X^U}$ that predict willingness to pay for A_2. This implies that the second column of $\tilde{\mathbf{\Theta}}$ is a zero vector:

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

Since $\tilde{\mathbf{\Theta}}$ is now rank-deficient, there is no matrix $\mathbf{R}$ such that $\mathbf{R}\tilde{\mathbf{\Theta}} = \mathbf{\Theta^U}$. In particular, for any matrix $\mathbf{R}$, each entry in column 2 be zero, but the second column of $\mathbf{\Theta^U}$ 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 $\mathbf{X^U}$, so that $\mathbf{\Pi_{X^UX}} = 0$, a rank-deficiency will also occur:

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

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 $\mathbf{R}$ such that $\mathbf{R}\tilde{\mathbf{\Theta}} = \mathbf{\Theta^U}$. For any choice of $\mathbf{R}$, in each row of $\mathbf{R}\tilde{\mathbf{\Theta}}$ the second column will always be twice as large as the first column, but the second row of $\mathbf{\Theta^U}$ 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 $\tilde{\mathbf{\Theta}}$ is less than the row rank of $\mathbf{\Theta^U}$, then the row space of $\mathbf{\Theta^U}$ cannot possibly be a subspace of the row space of $\tilde{\mathbf{\Theta}}$.

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

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:

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

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

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

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

There is no scalar R such that $R\tilde{\mathbf{\Theta}} = \mathbf{\Theta^U}$, since any value of R will preserve the one-to-one ratio between the first and second entries in $\mathbf{\Theta}$, while $\mathbf{\Theta^U}$ 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 $\mathbf{\Theta^U}$ is not a subspace of the row space of $\tilde{\mathbf{\Theta}}$. 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 (Q may affect taste for A_2, so that A_2 is still relevant for school choice):

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

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

A3 Testing Whether dim($\mathbf{A^X}$) Is Less Than the Number of Elements of $\mathbf{X_s}$

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 $\mathbf{X_i}$ captures enough independent factors determining families' preferences over group amenities so that knowledge of $\mathbf{X_s}$ is sufficient to determine the value of the amenities (denoted $\mathbf{A_s^X}$) for which $\mathbf{X_i}$ affects tastes, either through direct effects on willingness to pay or indirectly through correlation between $\mathbf{X_i}$ and elements of $\mathbf{X_i^U}$. For the particular linear specification of utility featured in (2), this condition is tantamount to requiring that $rank(\tilde{\mathbf{\Theta}}) \geq \dim(\mathbf{A_s^X})$.

The restriction $rank(\tilde{\mathbf{\Theta}}) \geq \dim(\mathbf{A_s}^X)$ restricts $rank(\mathbf{Var}(\mathbf{X_s}))$, which forms the basis for our test. To see this, note that taking expectations of both sides of (35) conditional on s implies that

$$\mathbf{X_s} = \mathbf{W_s}\mathbf{Var}(\mathbf{W_i})^{-1}\tilde{\mathbf{\Theta}}'\mathbf{Var}(\mathbf{X_i}),$$

where $\mathbf{W_s} \equiv \mathbf{E}(\mathbf{W_i}|s_i = s)$ is the average of the willingness to pay vector for those who choose s. Thus $\mathbf{X_s}$ is a linear combination of $\mathbf{W_s}$. Recall that the length of $\mathbf{W_s}$ is K, the number of valued amenities. Consequently, if $L > K$, then the L elements of $\mathbf{X_s}$ are all linear combinations of the smaller number of components of the average willingness to pay vector $\mathbf{W_s}$. But this implies that $\mathbf{Var}(\mathbf{X_s})$ will be rank deficient, with $rank(\mathbf{Var}(\mathbf{X_s})) = K$. In fact, if WTP for some of the K amenities is not influenced by $\mathbf{X_i}$, then some of the columns of $\tilde{\mathbf{\Theta}}$ will be 0. In this case, $rank(\mathbf{Var}(\mathbf{X_s})) = \dim(\mathbf{A^X}) < K$ further reducing the rank of $\mathbf{Var}(\mathbf{X_s})$. 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 $\mathbf{X_s}$, but elements of $\mathbf{X_i}$ slightly influence tastes for several other amenities. Our simulations in section A6 suggest that such minor departures from the Assumptions 5.1 and 5.2 have little impact on the ability of $\mathbf{X_s}$ to effectively control for the unobservable between-school variation $\mathbf{X_s^U}$. But in such contexts, a small number of amenity factors should account for a very large fraction of the variation in $\mathbf{X_s}$, with only a very small amount of unexplained residual variation.

We test these predictions by performing principal components analysis (PCA) on $\mathbf{X_s}$. Because the sample school averages of observable characteristics $\hat{\mathbf{X}}_\mathbf{s}$ are noisy measures of the expected values $\mathbf{X_s} \equiv \mathbf{E}[\mathbf{X_i}|s(i) = s]$, we do not fit the PCA model to $\hat{\mathbf{X}}_\mathbf{s}$ directly. Instead, we estimate the underlying true covariance matrix $\mathbf{Var}(\mathbf{X_s})$,[33] and then directly perform the principal components analysis on the estimated covariance matrix.[34]

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 $\mathbf{X_s}$, 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 $\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1}$, 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 $rank(\tilde{\mathbf{\Theta}}) \geq \dim(\mathbf{A_s^X})$, implying that Assumption 5.1 for the spanning condition $\mathbf{\Theta^U} = \mathbf{R}\tilde{\mathbf{\Theta}}$ is satisfied in the datasets we use. Specifically, in each dataset, $\mathbf{Var}(\mathbf{X_s})$ 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 $\mathbf{X_s}$ (Panel A, Row 6, Column 6), compared to $L = 51$ elements of $\mathbf{X_s}$. Similarly, in the NELS88 full specification, only 32 factors fully explain the variance in the 49 factors of $\mathbf{X_s}$.

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 $Var(\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1})$. 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 $Var(\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1})$, 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 of the variances of the elements of $\mathbf{X_s}$ is larger in the full specification, which contains more variables. This would be expected in the presence of sampling error in $\widehat{\mathbf{Var}}(\mathbf{X_s})$. However, it might also indicate that there are in fact additional amenity factors that play a small role in driving sorting (and thus have small eigenvalues) that can be picked up by the additional elements of $\mathbf{X_s}$ in the full

[33]Specifically, we estimate $\widehat{\mathbf{Var}}(\mathbf{X_i})$ and $\widehat{\mathbf{Var}}(\mathbf{X_i} - \mathbf{X_s})$ by taking the sample (weighted) covariances of $\mathbf{X_i}$ and $\mathbf{X_i} - \hat{\mathbf{X}}_\mathbf{s}$, performing the requisite degrees-of-freedom adjustment, and then obtaining $\widehat{\mathbf{Var}}(\mathbf{X_s})$ via $\widehat{\mathbf{Var}}(\mathbf{X_s}) = \widehat{\mathbf{Var}}(\mathbf{X_i}) - \widehat{\mathbf{Var}}(\mathbf{X_i} - \mathbf{X_s})$.

[34]When constructing our control function in our main estimating equations we augment the vector $\hat{\mathbf{X}}_\mathbf{s}$ that comes from directly aggregating student level variables $\mathbf{X_i}$ 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 $\mathbf{X_s}$ with minimal error. However, when performing the principal components analysis of $\mathbf{X_s}$, we do not include these additional measurements that come directly from schools, since they are likely to be nearly collinear with $\hat{\mathbf{X}}_\mathbf{s}$, and could cause us to find spurious evidence of rank deficiency in $\mathbf{Var}(\mathbf{X_s})$.

specification.

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 $\widehat{\mathbf{Var}}(\mathbf{X_s})$ that is decomposed by PCA is merely a consistent estimate of the population covariance matrix $\mathbf{Var}(\mathbf{X_s})$, and thus contains sampling error. The assumption underlying the spanning condition pertains to the rank of the population matrix $\mathbf{Var}(\mathbf{X_s})$. We address this issue in two ways. First, Panel A and B of Online Appendix Table A5 report 90% bootstrap confidence interval estimates of the number needed to explain the specified percentages of $\sum_l^L Var(X_{sl})$ and $Var(\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1})$. 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.[35] The test compares the null hypothesis that $rank(\mathbf{Var}(\mathbf{X_s})) = q$, for some $q < L$, against the alternative that $rank(\mathbf{Var}(\mathbf{X_s})) > q$. 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 provide no evidence against the null hypothesis that the number of factors that drive sorting on the observables $\mathbf{X_i}$ is substantially small than dimension of $\mathbf{X_i}$.

[35]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.

A4 The Relationship between $\mathbf{X_s^U}$ and $\mathbf{X_s}$ when $\mathbf{E(X_i|W_i)}$ and $\mathbf{E(X_i^U|W_i)}$ are Nonlinear

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

$$\mathbf{E}[\mathbf{\tilde{X}_i^U}|\mathbf{W_i}] = \mathbf{E}^*[\mathbf{\tilde{X}_i^U}|\mathbf{W_i}] + e_i^{\tilde{X}^U} \tag{39}$$

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

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

Proposition 1A: *Assume that Assumptions A1, A2, A3, and A5 hold.*
Then the expectation $\mathbf{X_s^U}$ *is*

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

A4.1 Proof of Proposition 1A:

The key steps of the proof are identical to first steps of the proof of Proposition 1 that lead to (30) and (31). These say that

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

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

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

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

Assumption A5 says that $\mathbf{\Theta^U} = \mathbf{R}\mathbf{\tilde{\Theta}}$. Substituting for $\mathbf{\Theta^{U'}}$ in (42) and using (43) leads to

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

Using

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

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

$$\mathbf{E}[\mathbf{X_i^U}|\mathbf{W_i}] = [\mathbf{E}^*[\mathbf{X_i}|\mathbf{W_i}] + e_i^X]\mathbf{\Pi}_{X^U X} + \mathbf{E}^*[\mathbf{X_i}|\mathbf{W_i}]\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}) + e_i^{\tilde{X}^U}. \quad (46)$$

The final step is to take expectations of both sides of the above equation conditional on $s(i) = s$ and use (30) and (31). Doing so leads to

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

where the second and third terms combine to form an approximation error. This completes the proof. As we discuss in Section 5.2, the approximation error would contribute to the school/neighborhood error component v_s of the outcome model (12). This would lead to upward bias in the less conservative of our two estimators of the variance of school/neighborhood effects.

A5 Deriving an Analytical Formula for $\mathbf{X_s^U}$ when the Spanning Assumption (A5) Is Not Satisfied

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

Partition $\mathbf{X_i^U}$ into a subset $\mathbf{X_{1i}^U}$ that is correlated with $\mathbf{X_i}$ and a subset $\mathbf{X_{2i}^U}$ that is not correlated with $\mathbf{X_i}$. Let L denote the number of elements of $\mathbf{X_i}$, L^{U1} denote the number of elements of $\mathbf{X_{1i}^U}$, and let L^{U2} denote the number of elements of $\mathbf{X_{2i}^U}$. Recall that Assumption 5.2 will fail if $\mathbf{X_{2i}^U}$ affects preferences for an amenity that neither $\mathbf{X_i}$ nor $\mathbf{X_{1i}^U}$ affect preferences for.

Denote by $\mathbf{A^{U2}}$ the subvector of $\mathbf{A}$ that is not contained in $\mathbf{A^X}$. Similarly, let K_1 be the number of amenities in $\mathbf{A^X}$ and let K_2 capture the number of amenities in $\mathbf{A^{U2}}$. Write the taste matrix $\mathbf{\Theta^U}$ as:

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

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

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

where $\mathbf{\Theta_1}$ is $L \times K_1$ and $\mathbf{\Theta_2} = \mathbf{0}$ is $L \times K_2$.

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

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

where $\mathbf{\Pi}^1_{X^U X}$ represents the first L^{U1} columns of $\mathbf{\Pi}_{X^U X}$.

Consider replacing assumption (A5) with the following assumptions, (A5') and (A5"):

- (A5'): There exists an $L^{U1} \times L$ matrix $\mathbf{R_1}$ such that $\mathbf{\Theta^U_{11}} = \mathbf{R_1}\tilde{\mathbf{\Theta}}_1$.
- (A5"): There exists an $L^{U2} \times L$ matrix $\mathbf{R_2}$ such that $\mathbf{\Theta^U_{21}} = \mathbf{R_2}\tilde{\mathbf{\Theta}}_1$.

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

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

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

$$\mathbf{E}[\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}|\mathbf{W_i}] = \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1}\mathbf{Cov}(\mathbf{W_i}', \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}) \tag{47}$$

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

Next, recall that we can write $\mathbf{W_i}$ as:

$$\mathbf{W_i} = \mathbf{X_i}\tilde{\mathbf{\Theta}} + \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}\mathbf{\Theta^U} + \mathbf{Q_i}\mathbf{\Theta}^{Q}$$

where $\mathbf{X_i}$, $\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}$, and $\mathbf{Q_i}$ are mutually uncorrelated by construction. This leads to the following expression for $\mathbf{Cov}(\mathbf{W_i}', \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}})$:

$$\mathbf{Cov}(\mathbf{W_i}', \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}) = \mathbf{Cov}(\mathbf{\Theta^{U'}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}) = \mathbf{Cov}(\left\{ \begin{array}{cc} \mathbf{\Theta^{U\prime}_{11}} & \mathbf{\Theta^{U\prime}_{21}} \\ \mathbf{\Theta^{U\prime}_{12}} & \mathbf{\Theta^{U\prime}_{22}} \end{array} \right\} \left\{ \begin{array}{c} \tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{1i}} \\ \tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}} \end{array} \right\}, \left\{ \begin{array}{cc} \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}} & \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}} \end{array} \right\})$$

$$= \left\{ \begin{array}{cc} \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{11}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{1i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) + \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{21}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) & \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{11}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{1i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) + \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{21}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) \\ \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{12}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{1i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) + \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{22}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) & \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{12}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{1i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) + \mathbf{Cov}(\mathbf{\Theta^{U\prime}_{22}}\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) \end{array} \right\}$$

$$= \left\{ \begin{array}{ll} \mathbf{\Theta^{U\prime}_{11}}\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) + \mathbf{\Theta^{U\prime}_{21}}\mathbf{Cov}(\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) & \mathbf{\Theta^{U\prime}_{11}}\mathbf{Cov}(\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{1i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) + \mathbf{\Theta^{U\prime}_{21}}\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) \\ \mathbf{\Theta^{U\prime}_{22}}\mathbf{Cov}(\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) & \mathbf{\Theta^{U\prime}_{22}}\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) \end{array} \right\}$$

$$= \left\{ \begin{array}{ll} \tilde{\mathbf{\Theta}}'_{\mathbf{1}}\mathbf{R_1}'\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) + \tilde{\mathbf{\Theta}}'_{\mathbf{1}}\mathbf{R_2}'\mathbf{Cov}(\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) & \tilde{\mathbf{\Theta}}'_{\mathbf{1}}\mathbf{R_1}'\mathbf{Cov}(\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{1i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) + \tilde{\mathbf{\Theta}}'_{\mathbf{1}}\mathbf{R_2}'\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) \\ \mathbf{\Theta^{U\prime}_{22}}\mathbf{Cov}(\tilde{\mathbf{X}}^{\mathbf{U}\prime}_{\mathbf{2i}}, \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{1i}}) & \mathbf{\Theta^{U\prime}_{22}}\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{2i}}) \end{array} \right\}$$

Where the last two lines impose (A5'), (A5") and $\mathbf{\Theta^U_{12}} = \mathbf{0}$.

Similarly, we have:

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

$$\left\{ \begin{array}{l} \tilde{\mathbf{\Theta}}_1' \\ \tilde{\mathbf{\Theta}}_2' \end{array} \right\} \mathbf{Var}(\mathbf{X_i}) = \left\{ \begin{array}{l} \mathbf{\Theta}_1' + \mathbf{\Theta^{U'}_{11}} \mathbf{\Pi^{1\prime}_{X^U X}} \\ \mathbf{0} \end{array} \right\} \mathbf{Var}(\mathbf{X_i}) \tag{50}$$

Plugging in the formulas for $\mathbf{Cov}(\mathbf{W_i}', \tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}})$ and $\mathbf{Cov}(\mathbf{W_i}', \mathbf{X_i})$ into 47 and 48 , we obtain:

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

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

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

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

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

$$= \mathbf{E}[\mathbf{X_i}|\mathbf{W_i}]\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}) + \mathbf{W_i}\mathbf{Var}(\mathbf{W_i})^{-1} \left\{ \begin{array}{ll} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta^{U\prime}_{22}} \end{array} \right\} \mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}) \tag{54}$$

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

$$\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{s}} = \mathbf{X_s}\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}) + \mathbf{W_s}\mathbf{Var}(\mathbf{W_i})^{-1} \left\{ \begin{array}{ll} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta^U_{22}} \end{array} \right\} \mathbf{Var}(\tilde{\mathbf{X}}^{\mathbf{U}}_{\mathbf{i}}) \tag{55}$$

Note that in equilibrium $\mathbf{E}[\mathbf{W_i}|s(i) = s]$ will depend on the full joint distribution of amenities and the joint distribution of $\mathbf{W_i}$. 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 $\mathbf{A_s}$ and school averages of the WTP for amenities $\mathbf{W_s}$.

However, we can gain additional insight by re-considering the continuous version of the model analyzed in Altonji and 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 and Mansfield (2014), we solve for an explicit unique equilibrium mapping between $\mathbf{A_s}$ and $\mathbf{W_s}$ under the assumptions that a) $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$ and $\mathbf{A_{s(i)}}$ are each jointly normally distributed (with variance matrices for W_i and $A_{s(i)}$ of $\mathbf{\Sigma_W}$ and $\mathbf{\Sigma_A}$ respectively), b) the ε_{is} are 0, and c) the equilibrium allocation takes a linear form: $\mathbf{A_{s(i)}} = \mathbf{\Psi W_i}'$. The unique linear equilibrium mapping is

$$\mathbf{\Psi} = \mathbf{\Sigma_{W'}^{-1/2}}(\mathbf{\Sigma_{W'}^{1/2}}\mathbf{\Sigma_A}\mathbf{\Sigma_{W'}^{1/2}})\mathbf{\Sigma_{W'}^{-1/2}} \,. \tag{56}$$

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

Since every positive definite matrix is invertible, we can also express the vector $\mathbf{W_i}$ for any individual as a linear function of the amenity vector of their chosen school:

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

In the continuous version of the model with $\varepsilon_{is} = 0$, every individual at the same school has the same value of $\mathbf{W_i}$. Thus, we also obtain:

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

Substituting (58) into (61) leads to

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

This shows more clearly that the variances and covariances involving $\mathbf{X_i}$, $\mathbf{\tilde{X}_i^U}$ and $\mathbf{Q_i}$ play a role, and that $\mathbf{Var}(\mathbf{A_s})$ plays a role in determining the variation in $\mathbf{\tilde{X}_s^U}$ not accounted for by $\mathbf{X_s Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\mathbf{\tilde{X}_i^U})$.

However, note that the variance of $\mathbf{A_{s(i)}}'\mathbf{\Psi}^{-1}\mathbf{Var}(\mathbf{W_i})^{-1}\left\{\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta_{22}^{U\prime}} \end{array}\right\}\mathbf{Var}(\mathbf{\tilde{X}_i^U})$ is *not* the residual variance of $\mathbf{X_s^U}$ conditional on $\mathbf{X_s}$. This is because $\mathbf{X_s}$ and $\mathbf{A_s}$ co-vary, which leads the two terms in (59) for $\mathbf{\tilde{X}_s^U}$ to co-vary.

Next, recall the composition of $\mathbf{W_i}$:

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

Taking expectations of both sides of the above equation conditional on $s = s(i)$ one may substitute for $\mathbf{W_s}$ in (55). This leads to.

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

Now suppose that in addition to Assumption (A4), we assume that $\mathbf{E}[\mathbf{Q_i}|\mathbf{W_i}]$ is also linear in

$\mathbf{W_i}$, so that:

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

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

$$\tilde{\mathbf{X}}_{\mathbf{s}}^{\mathbf{U}} = \mathbf{A_{s(i)}}'\mathbf{\Psi}^{-1}\mathbf{Var}(\mathbf{W_i})^{-1}\mathbf{\Theta}^{\mathbf{U}'}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) \tag{63}$$

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

$$\mathbf{Q_s} = \mathbf{A_{s(i)}}'\mathbf{\Psi}^{-1}\mathbf{Var}(\mathbf{W_i})^{-1}\mathbf{\Theta}^{\mathbf{Q}'}\mathbf{Var}(\mathbf{Q_i}) \tag{65}$$

Collecting terms involving $\mathbf{X_s}$ and substituting equations (63) and (65) into (61) yields:

$$\tilde{\mathbf{X}}_{\mathbf{s}}^{\mathbf{U}} = \mathbf{X_s}\{\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) + \tilde{\mathbf{\Theta}}'\mathbf{Var}(\mathbf{W_i})^{-1}\left\{\begin{array}{cc}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Theta}_{22}^{\mathbf{U}'}\end{array}\right\}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})\} \tag{66}$$

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

$$+\mathbf{A_{s(i)}}'\mathbf{\Psi}^{-1}\mathbf{Var}(\mathbf{W_i})^{-1}\mathbf{\Theta}^{\mathbf{Q}'}\mathbf{Var}(\mathbf{Q_i})\mathbf{\Theta}^{\mathbf{Q}}\mathbf{Var}(\mathbf{W_i})^{-1}\left\{\begin{array}{cc}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Theta}_{22}^{\mathbf{U}'}\end{array}\right\}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) \tag{68}$$

Even this does not let us decompose $\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{s}}^{\mathbf{U}})$ into a term involving $\mathbf{X_s}$ and an uncorrelated residual piece. the reason is that $\mathbf{A_{s(i)}}$ will be correlated with $\mathbf{X_s}$.

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

$$\mathbf{A_s}^{\mathbf{X}\prime} = \mathbf{X_s}\mathbf{\Pi}_{\mathbf{A^X X_s}} + \tilde{\mathbf{A}}_{\mathbf{s}}^{\mathbf{X}\prime} = \mathbf{X_s}\mathbf{\Pi}_{\mathbf{A^X X_s}} \tag{69}$$

$$\mathbf{A_s}^{\mathbf{U2}\prime} = \mathbf{X_s}\mathbf{\Pi}_{\mathbf{A^{U2} X_s}} + \tilde{\mathbf{A}}_{\mathbf{s}}^{\mathbf{U2}\prime} \tag{70}$$

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

This implies:

$$\tilde{\mathbf{X}}_{\mathbf{s}}^{\mathbf{U}} = \mathbf{X_s}[\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) + \tilde{\mathbf{\Theta}}'\mathbf{Var}(\mathbf{W_i})^{-1}\left\{\begin{array}{cc}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Theta}_{22}^{\mathbf{U}'}\end{array}\right\}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) + \mathbf{M}]$$
$$+\{\mathbf{0}, \tilde{\mathbf{A}}_{\mathbf{s}}^{\mathbf{U2}\prime}\}\mathbf{\Psi}^{-1}\mathbf{Var}(\mathbf{W_i})^{-1}[\mathbf{\Theta}^{\mathbf{U}'}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})\mathbf{\Theta}^{\mathbf{U}} + \mathbf{\Theta}^{\mathbf{Q}\prime}\mathbf{Var}(\mathbf{Q_i})\mathbf{\Theta}^{\mathbf{Q}}]\mathbf{Var}(\mathbf{W_i})^{-1}\left\{\begin{array}{cc}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Theta}_{22}^{\mathbf{U}'}\end{array}\right\}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) \tag{71}$$

where the matrix $\mathbf{M}$ is

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

While cumbersome, the second term in (71) provides an expression for the component of $\tilde{\mathbf{X}}_\mathbf{s}^\mathbf{U}$ that cannot be predicted by $\mathbf{X_s}$ (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 $\mathbf{\Psi}$); b) the joint distribution of the WTP vector $\mathbf{W_i}$ (entering via the covariance matrix $\mathbf{Var}(\mathbf{W_i})$); c) the WTP coefficient matrix $\mathbf{\Theta^U}$ capturing the effect of $\mathbf{X}^U$ on 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 $\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})$; and e) the joint distribution of the unobserved outcome-irrelevant (but school choice-relevant) student characteristics weighted by effects on WTP (entering via $\mathbf{\Theta^{Q\prime}}\mathbf{Var}(\mathbf{Q_i})\mathbf{\Theta^Q}$).

Given the complicated manner in which each of these five factors enters the second term in (71), we do not see a straightforward way bound the variance of this error component.

A6 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. We do not attempt to fully characterize the performance of our estimator.[36] Instead, our simulations center around a stylized test case that is calibrated to represent a plausible description of the school/neighborhood choice context. We focus on sensitivity to deviations among a set of key parameters designed 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 A6.1 lays out the simulation methodology, while section A6.2 presents and interprets the results.

[36]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 $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$, the joint distribution of the amenities $\mathbf{A_s}$, and the distribution of the idiosyncratic tastes ε_{si}. The parameters also include the $\mathbf{\Theta}$, $\mathbf{\Theta^U}$, and $\mathbf{\Theta^Q}$ matrices that capture how observed and unobserved characteristics affect WTP.

A6.1 Methodology

The stylized test case we consider is one in which:

1. The elements of $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$ are jointly normally distributed; the elements of $\mathbf{Q_i}$ are independent of each other and $[\mathbf{X_i}, \mathbf{X_i^U}]$, and each pair of characteristics in $[\mathbf{X_i}, \mathbf{X_i^U}]$ features a .25 correlation.[37]
2. The latent amenity vectors $\mathbf{A_s}$ are normally distributed with a .25 correlation between any pair of amenities across schools.
3. The matrices of taste parameters $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ represent draws from a multivariate normal distribution in which (a) $corr(\Theta_{k\ell}, \Theta_{jm}) \equiv \rho_\Theta$ if $j = k$ or $\ell = m$, and 0 otherwise, (b) $corr(\Theta^U_{k\ell}, \Theta^U_{jm}) = \rho_\Theta$ if $j = k$ or $\ell = m$, and 0 otherwise, and (c) $corr(\Theta_{k\ell}, \Theta^U_{jm}) = \rho_\Theta$ if $\ell = m$, and 0 otherwise.
4. The number of elements of $\mathbf{Q}_i$ is equal to the number of elements of $\mathbf{A}_s$. $\mathbf{\Theta}^Q$ is the identity matrix.
5. The variances of the elements of $\mathbf{A_s}$, $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$, and $\varepsilon_{i,s}$ (i.i.d. draws from a normal distribution) are chosen to create interclass correlations for $\mathbf{X_i}$ and $\mathbf{X_i^U}$ of between .1 and .25 across specifications. These values are in line with the range observed across the datasets used in the empirical analysis.
6. There are no school/neighborhood effects, so that $Y = \mathbf{X_i}\boldsymbol{\beta} + x_i^U$, where $x_i^U \equiv \mathbf{X_i^U}\boldsymbol{\beta}^U$. Consequently, our estimating equation also omits the school level controls $\mathbf{Z_{2s}}$ 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.
7. All the observable and unobservable characteristics in $\mathbf{X_i}$ and $\mathbf{X_i^U}$ are equally important in determining the outcome, so that each characteristic features the same (unit) variance, $\beta_\ell = 1\ \forall\ \ell$, and $\beta^U_\ell = 1\ \forall\ \ell$.

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 1-7, we complete the model by choosing particular sets of seven remaining parameters. The first parameter is the number of students per school. For simplicity, we impose that each school has capacity equal to a common student/school ratio.[38] The student/school ratio is denoted "# Stu" in Online Appendix Table A8. The second

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

[38] 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.

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 (#Am). In most of the specifications we assume that it is less then or equal to the number of observed characteristics and that the rows of $\mathbf{\Theta^U}$ form a linear subspace of the rows of $\tilde{\mathbf{\Theta}}$, as required by Proposition 1.

The seventh parameter is ρ_Θ, 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_Θ 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. In addition, when ρ_Θ is high the 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 $\mathbf{\Theta^U} \neq \mathbf{R}\tilde{\mathbf{\Theta}}$ for any $\mathbf{R}$). 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 $\mathbf{\Pi_{X^UX}}$, which captures the degree to which individual-level unobservables project onto the space of individual-level observables, (2) the taste matrix $\mathbf{\Theta}$, 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, $\mathbf{\Theta^U}$.

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 ($Var(x^U_s) \equiv Var(\mathbf{X^U_s}\boldsymbol{\beta}^U)$) that can be predicted using group-averages of observable characteristics (after adjusting for the degrees of freedom absorbed by the vector of observables). This is the adjusted R^2 from a regression of the potential bias from unobservable sorting, x^U_s, on the vector $\mathbf{X_s}$. In cases where the conditions of Proposition 1 are satisfied, R^2_{adj} 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 $[(1-R^2_{adj})Var(x^U_s)]/Var(Y_{si})$, which is the fraction of the total variance in the outcome Y_{si} attributable to the variance of the residual component of x^U_s not accounted for by $\mathbf{X_s}$. In Appendix tables A8 and A9, we denote our measures "Adj-R-sq" and "Resid" (short for "residual sorting variance fraction").

We present values of adjusted 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 $\hat{\mathbf{X}}_\mathbf{s}$ that compose the control function (denoted "Adj-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 $\hat{\mathbf{X}}_\mathbf{s}$ (these values are denoted "Adj-R-sq (10/20/40)" and "Resid (10/20/40)", respectively, in our tables).

We draw $\mathbf{X_i}$, $\mathbf{X_i^U}$, $\mathbf{Q_i}$, and $\{\varepsilon_{is}\}$ from the distributions described above to calculate the WTP of each household for each school.[39] Since our method does not require observation of the equilibrium price function $P(\mathbf{A})$, 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.[40]

A6.2 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 5 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_Θ among the elements of the random matrices from which the taste weight matrices $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ 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 three hundredths of a percent of the variance in the outcome Y_{si} that can be attributed to residual between-school sorting (Column 10).

[39]To minimize the statistical "chatter" introduced by the particular $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ matrices that we happened to draw, we drew ten different sets of $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ 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.

[40]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.

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 adjusted R-squared falls from .998 to .869 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 adjusted R-squared to .926, while increasing it further to 40 students per school (bottom entry in Column (11)) raises the adjusted R-squared to .959. Column (12) shows that the fraction of the outcome variance consisting of residual between-school sorting unabsorbed by the control function is .016/.009/.005 when 10/20/40 student samples, respectively, are used to construct the vector of school averages, $\mathbf{X_s}$.

Rows (2) and (3) illustrate the impact of adapting the specification in Row (1) by decreasing or increasing the number of individuals per group. Decreasing school sizes from 1000 to 500 decreases the R-squared from .998 to .996, 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, $\mathbf{X_s}$. In Columns (11) and (12), the adjusted R-squared values and residual sorting variance fraction when samples of 10, 20 and 40 students are used to construct $\mathbf{X_s}$ 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 almost no 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 does not exacerbate problems posed by using small samples of students from each school to construct $\mathbf{X_s}$ (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 adjusted R-squared is effectively unchanged when the full population of students is used to construct $\mathbf{X_s}$, but drops slightly from Row (1) to Row (5) when samples of 10, 20, or 40 students are used instead. Nonetheless, the high adjusted 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.[41] Doubling the number of elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$ increases the adjusted R-squared from .9979 in Row (1) to .9993, and decreases the fraction of outcome variance attributable to the residual sorting component to two hundredths of

[41] 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 $\boldsymbol{\Theta}^{\mathbf{U}}$ smaller in magnitude and increasing the variance of $\mathbf{Q_i}$ 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.

a percentage point. This very 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 adjusted R-squared from .869 to .897. 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 adjusted R-squared from .9979 in Row (1) to .9933. The impact of doubling the number of amenities is also small when small samples of students are used to construct school averages. Comparing Row (8) to Row (6) reveals that the performance of the control function really depends on the dimension of the amenity space *relative* to the dimension of $\mathbf{X_s}$, rather than the absolute number of amenities: when $\mathbf{X_s}$ 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 ($\rho_\Theta = 0$). Compared to Row (1), the adjusted R-squared for the full population falls slightly (.9979 to .9941), but the adjusted R-squared when samples of (10/20/40) are used to construct $\mathbf{X_s}$ falls more substantially, from (.87/.93/.97) to (.65/.78/.87). However, removing correlation among the elements of $\boldsymbol{\Theta}$ also reduces the amount of sorting on unobservables to be explained. When the school averages of the various unobservables become more weakly correlated with one another, their contributions to student outcomes are more likely 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 $\mathbf{X_s}$ remains quite small (Row 9, Col. 11).

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 a moderate number of total groups to join, and 3) only a subset of these are considered by any given individual.[42] 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 A11, we use the North Carolina administrative data to directly assess the effect of using smaller samples of students to construct $\mathbf{X_s}$ 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.

[42] 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 $\mathbf{X_i}$ and $\mathbf{X_i^U}$ symmetrically, but does not erode the effectiveness of $\mathbf{X_s}$ as a control for $\mathbf{X_s^U}$. Intuitively, as $Var(\varepsilon_{is}) \to \infty$, idiosyncratic tastes fully drive choice, and the between school variation in $\mathbf{X_i}$ and $\mathbf{X_i^U}$ disappears, so that there is no more sorting problem to address.

A6.2.1 Performance of the Control Function When the Spanning Condition Fails

All the specifications in Online Appendix Table A8 consider cases in which the assumptions of Proposition 1 are satisfied, so we should expect the control function to perfectly absorb sorting on observables as the number of students per school gets sufficiently large. However, there also may be many contexts in which the set of observables is not sufficiently rich to make the 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 $\tilde{\mathbf{\Theta}}$ can be represented as the sum of $\mathbf{\Theta}$ and $\mathbf{\Pi}_{X^U X}\mathbf{\Theta}^{\mathbf{U}}$. The dependence on $\mathbf{\Theta}$ indicates that the mapping from $\mathbf{X}_{\mathbf{s}}^{\mathbf{U}}$ to $\mathbf{X}_{\mathbf{s}}$ is generated partly because observed characteristics $\mathbf{X}_{\mathbf{i}}$ and unobserved characteristics $\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}$ directly affect WTP for overlapping sets of amenities (which creates a degree of overlap in the row spaces of $\mathbf{\Theta}$ and $\mathbf{\Theta}^{\mathbf{U}}$). The term $\mathbf{\Pi}_{\mathbf{X}^{\mathbf{U}}\mathbf{X}}\mathbf{\Theta}^{\mathbf{U}}$ captures the part of the mapping that arises because $\mathbf{X}_{\mathbf{i}}$ indirectly predicts WTP for the amenities for which $\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}$ predicts WTP through the correlation between $\mathbf{X}_{\mathbf{i}}$ and $\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}$ (thereby creating further overlap in the row spaces of $\mathbf{\Theta}$ and $\mathbf{\Theta}^{\mathbf{U}}$). The spanning condition ($\mathbf{\Theta}^{\mathbf{U}} = \mathbf{R}\tilde{\mathbf{\Theta}}$ for some $L^U \times L$ matrix $\mathbf{R}$) is satisfied whenever these two pathways, working in combination, produce a preference matrix $\tilde{\mathbf{\Theta}}$ whose row space is a linear superspace of the row space of $\mathbf{\Theta}^{\mathbf{U}}$.

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 $\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}$ is independent of the vector of observables $\mathbf{X}_{\mathbf{i}}$, so that $\mathbf{\Pi}_{\mathbf{X}^{\mathbf{U}}\mathbf{X}}$ converges to the zero matrix as school sizes become large. However, $\mathbf{X}_{\mathbf{i}}$ and $\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}$ predict tastes for a common set of amenities ($A_1 - A_5$), so that $\mathbf{\Theta}$ has (full) rank K and the row space of $\mathbf{\Theta}^{\mathbf{U}}$ is a linear subspace of the row space of $\mathbf{\Theta}$. 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 (adjusted R-squared of .965), but suffers somewhat when school averages are constructed using subsamples of 10, 20 or 40 students: adjusted R-squared values fall to .49/.61/.72 (Column 10), with substantial residual bias from sorting on unobservables left uncaptured by the control function $\hat{\mathbf{X}}_{\mathbf{s}}$ (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 $\mathbf{X}_{\mathbf{i}}$ and $\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}$. Specifically, in row (3) the observables and unobservables affect tastes for disjoint sets of amenities ($\{A_1, \ldots, A_4\}$ and $\{A_5\}$ respectively). This means that the row space of $\mathbf{\Theta}^{\mathbf{U}}$ is orthogonal to the row space of $\mathbf{\Theta}$. However, each element of $\mathbf{X}_{\mathbf{i}}$ is correlated .25 with each element of $\mathbf{X}_{\mathbf{i}}^{\mathbf{U}}$, so that $\mathbf{\Pi}_{\mathbf{X}^{\mathbf{U}}\mathbf{X}}$ is full rank and the row space of $\mathbf{\Theta}^{\mathbf{U}}$ is a linear subspace of the row space of $\mathbf{\Pi}_{\mathbf{X}^{\mathbf{U}}\mathbf{X}}\mathbf{\Theta}^{\mathbf{U}}$. 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 $\mathbf{X_s}$ and $\mathbf{X_s^U}$ may be generated are each sufficient in isolation to produce desirable finite sample properties with large samples of students per school. But they also show that 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 $\mathbf{\Theta^U}$ is not a linear subspace of the row space of $\tilde{\mathbf{\Theta}} = \mathbf{\Theta} + \mathbf{\Pi_{X^UX}}\mathbf{\Theta^U}$). Row (4) presents results from the worst-case scenario: (a) the entire vector of unobservable characteristics is independent of the entire vector of observable characteristics ($\mathbf{\Pi}_{X^UX}$ converges to $\mathbf{0}$ as school sizes become large), and (b) 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, $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ 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 15% of the variance in $\mathbf{X_s^U}$ is predictable given $\mathbf{X_s}$, even when the universe of students at each school is observed (Column 8).[43]

Row (5) alters the scenario from Row (4) by allowing the unobservable characteristics $\mathbf{X_i^U}$ to predict WTP for amenities A_1 to A_4 in addition to A_5. The control function performs somewhat better: 52% of the variance in $\mathbf{X_s^U}$ is absorbed by the coefficients on $\mathbf{X_s}$.

These two scenarios are quite pessimistic, however. If WTP for an amenity is unaffected by the entire vector $\mathbf{X_i}$, then it seems likely that a subset of the unobservables may not predict WTP for this amenity either. Thus, we consider two additional scenarios in which WTP for the last amenity (A_5) is only affected by one of the ten components of the unobserved vector $\mathbf{X_i^U}$. 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 95% of the variation in $\mathbf{X_s^U}$ 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. In this case both $\mathbf{\Pi_{X^UX}}\mathbf{\Theta^U}$ and $\mathbf{\Theta}$ would be linear superspaces of $\mathbf{\Theta^U}$ 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 may 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 $\mathbf{X_i^U}$ that is orthogonal to $\mathbf{X_i}$ affect WTP for just a few additional amenities for which $\mathbf{X_i}$ does not affect WTP.

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

A7 Proof of Proposition 2

Let Δ denote the operator that takes deviations from school/neighborhood means, so that, for example, $\Delta \mathbf{X_i^U} \equiv (\mathbf{X_i^U} - \mathbf{X_s^U})$. Define $\mathbf{\Pi}_{\Delta \mathbf{X^U} \Delta \mathbf{X}}$ as the coefficient matrix from the following within-school regression:

$$\Delta \mathbf{X_i^U} = \Delta \mathbf{X_i} \mathbf{\Pi}_{\Delta \mathbf{X^U} \Delta \mathbf{X}} + \widetilde{\Delta \mathbf{X_i^U}} \tag{73}$$

Recall the projection equation (5): $\mathbf{X_i^U} = \mathbf{X_i} \mathbf{\Pi}_{\mathbf{X^U X}} + \tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}$. We now state and prove an expanded version of Proposition 2 that includes an expression for $\boldsymbol{B}$ and $\boldsymbol{G}_1$.

Proposition 2: *Assume that assumptions A1-A5 from Proposition 1 hold.*

Then equations (13)-(17) simplify to:

$$\begin{aligned}
\mathbf{B} &= \boldsymbol{\beta} + \mathbf{\Pi}_{\Delta \mathbf{X^U} \Delta \mathbf{X}} \boldsymbol{\beta}^U + \mathbf{\Pi}_{\eta_{si}^U \mathbf{X_i}} & (74)\\
\mathbf{G_1} &= [-\mathbf{\Pi}_{\Delta \mathbf{X^U} \Delta \mathbf{X}} + \mathbf{\Pi}_{\mathbf{X^U X}} + \mathbf{Var}(\mathbf{X_i})^{-1} \mathbf{R}' \mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})] \boldsymbol{\beta}^U + \mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}} & (75)\\
\mathbf{G_2} &= \mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}} & (76)\\
v_s &= \tilde{z}_s^U + \xi_s & (77)\\
v_{si} - v_s &= \Delta \tilde{x}_{si}^U + \tilde{\eta}_{si} + \xi_i & (78)
\end{aligned}$$

Proof: Recall the projection equation (9):

$$x_i^U = \mathbf{X_i} \mathbf{\Pi}_{x_i^U \mathbf{X_i}} + \mathbf{X_s} \mathbf{\Pi}_{x_i^U \mathbf{X_s}} + \mathbf{Z_{2s}} \mathbf{\Pi}_{x_i^U \mathbf{Z_{2s}}} + \tilde{x}_i^U$$

Note that this projection is the sum of a within-school and between-school projection:

$$\Delta x_{si}^U = \mathbf{X_i} \mathbf{\Pi}_1 + \mathbf{X_s} \mathbf{\Pi}_2 + \mathbf{Z_{2s}} \mathbf{\Pi}_3 + \Delta \tilde{x}_{si}^U = \Delta \mathbf{X_i} \mathbf{\Pi}_1 + \mathbf{X_s}[\mathbf{\Pi}_2 + \mathbf{\Pi}_1] + \mathbf{Z_{2s}} \mathbf{\Pi}_3 + \Delta \tilde{x}_{si}^U \tag{79}$$

$$x_s^U = \Delta \mathbf{X_i} \mathbf{\Pi}_4 + \mathbf{X_s} \mathbf{\Pi}_5 + \mathbf{Z_{2s}} \mathbf{\Pi}_6 + \tilde{x}_s^U = \Delta \mathbf{X_i} \mathbf{\Pi}_4 + \mathbf{X_s}[\mathbf{\Pi}_5 + \mathbf{\Pi}_4] + \mathbf{Z_{2s}} \mathbf{\Pi}_6 + \tilde{x}_s^U \tag{80}$$

Consider (79) first. Note that the deviation from group mean $\Delta \mathbf{X_i}$ and $\Delta \tilde{x}_{si}^U$ are orthogonal to both $\mathbf{X_s}$ and $\mathbf{Z_{2s}}$, so the projection of Δx_{si}^U on $\Delta \mathbf{X_i}$, $\mathbf{X_s}$, and $\mathbf{Z_{2s}}$ boils down to the projection of Δx_{si}^U on $\Delta \mathbf{X_i}$. Consequently,

$$\begin{aligned}
\mathbf{\Pi}_1 &= \mathbf{\Pi}_{\Delta \mathbf{X^U} \Delta \mathbf{X}} \boldsymbol{\beta}^U\\
\mathbf{\Pi}_2 + \mathbf{\Pi}_1 &= \mathbf{0} \text{ or } \mathbf{\Pi}_2 = -\mathbf{\Pi}_1\\
\mathbf{\Pi}_3 &= \mathbf{0}
\end{aligned}$$

where in the first equality $\mathbf{\Pi}_{\Delta \mathbf{X^U} \Delta \mathbf{X}}$ is the coefficient matrix from (73) and we have used the definition $x_i^U \equiv \mathbf{X}_i^U \boldsymbol{\beta}^U$.

Now, consider the between-school regression (80). By Proposition 1, $\mathbf{X_s^U} = \mathbf{X_s}[\mathbf{\Pi}_{\mathbf{X^U X}} + \mathbf{Var}(\mathbf{X_i})^{-1} \mathbf{R}' \mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})]$.

Post-multiplying both sides by $\boldsymbol{\beta}^{\mathbf{U}}$, we obtain:

$$\mathbf{X_s^U}\boldsymbol{\beta}^{\mathbf{U}} \equiv x_s^U = \mathbf{X_s}[\boldsymbol{\Pi}_{\mathbf{X^UX}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})]\boldsymbol{\beta}^{\mathbf{U}} \tag{81}$$

But since x_s^U can be perfectly predicted by $\mathbf{X_s}$, we have:

$$\boldsymbol{\Pi}_4 = \mathbf{0} \tag{82}$$
$$\boldsymbol{\Pi}_5 = [\boldsymbol{\Pi}_{\mathbf{X^UX}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})]\boldsymbol{\beta}^{\mathbf{U}} \tag{83}$$
$$\boldsymbol{\Pi}_6 = \mathbf{0} \tag{84}$$
$$\tilde{x}_s^U = 0 \tag{85}$$

Adding together $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_4$, $\boldsymbol{\Pi}_2$ and $\boldsymbol{\Pi}_5$, and $\boldsymbol{\Pi}_3$ and $\boldsymbol{\Pi}_6$ yields:

$$\boldsymbol{\Pi}_{x_i^U\mathbf{X_i}} = \boldsymbol{\Pi}_{\Delta\mathbf{X^U}\Delta\mathbf{X}}\boldsymbol{\beta}^U \tag{86}$$
$$\boldsymbol{\Pi}_{x_i^U\mathbf{X_s}} = [-\boldsymbol{\Pi}_{\Delta\mathbf{X^U}\Delta\mathbf{X}} + \boldsymbol{\Pi}_{\mathbf{X^UX}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})]\boldsymbol{\beta}^U \tag{87}$$
$$\boldsymbol{\Pi}_{x_i^U\mathbf{Z_{2s}}} = \mathbf{0} \tag{88}$$

Plugging equations (85)-(88) back into equations (13)- (17) yields:

$$\mathbf{B} = \boldsymbol{\beta} + \boldsymbol{\Pi}_{\Delta\mathbf{X^U}\Delta\mathbf{X}}\boldsymbol{\beta}^U + \boldsymbol{\Pi}_{\eta_{si}^U\mathbf{X_i}} \tag{89}$$
$$\mathbf{G_1} = [-\boldsymbol{\Pi}_{\Delta\mathbf{X^U}\Delta\mathbf{X}} + \boldsymbol{\Pi}_{\mathbf{X^UX}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})]\boldsymbol{\beta}^U + \boldsymbol{\Gamma}_1 + \boldsymbol{\Pi}_{z_s^U\mathbf{X_s}} \tag{90}$$
$$\mathbf{G_2} = \boldsymbol{\Gamma}_2 + \boldsymbol{\Pi}_{z_s^U\mathbf{Z_{2s}}} \tag{91}$$
$$v_s = \tilde{z}_s^U + \xi_s \tag{92}$$
$$v_{si} - v_s = \tilde{x}_{si}^U + \tilde{\eta}_{si} + \xi_i \tag{93}$$

This concludes the proof.

It is interesting to briefly discuss what happens if the linear conditional expectations assumption A4 fails. Then Proposition 1A in Online Appendix A4 establishes that

$$\mathbf{X_s^U}\boldsymbol{\beta}^{\mathbf{U}} \equiv x_s^U = \mathbf{X_s}[\boldsymbol{\Pi}_{\mathbf{X^UX}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})]\boldsymbol{\beta}^{\mathbf{U}} + \tilde{\boldsymbol{v}}_s^*\boldsymbol{\beta}^{\mathbf{U}}$$

where $\tilde{\boldsymbol{v}}_s^* = -\mathbf{E}[e_i^X|s_i = s][\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})] + \mathbf{E}[e_i^{\tilde{X}^U}|s_i = s]$. It is straightforward to show that a correlation between $\mathbf{Z_{2s}}$ and $\tilde{\boldsymbol{v}}_s^*\boldsymbol{\beta}^{\mathbf{U}}$ would alter the coefficient vector $\mathbf{G_2}$. Part of $\tilde{\boldsymbol{v}}_s^*\boldsymbol{\beta}^{\mathbf{U}}$ would also appear in v_s.

A7.1 Proof that under Assumptions A1-A5, $\mathbf{G_1} = \mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}$ in the Absence of Sorting on $\mathbf{X_i^U}$

Under A1-A5, Proposition 1 holds. The case in which $\mathbf{X_i^U}$ does not influence the choice of s corresponds to the case in which $\mathbf{\Theta}^U = \mathbf{0}$. A5 says that $\mathbf{\Theta^U} = \mathbf{R}\tilde{\mathbf{\Theta}}$. As we noted in Section 3.2, $\boldsymbol{R} = \mathbf{0}$ when $\mathbf{\Theta}^U = \mathbf{0}$. Thus, equation (7) from Proposition 1 implies immediately that $\mathbf{X_s^U} = \mathbf{X_s}\mathbf{\Pi_{X^U X}}$ when $\mathbf{\Theta}^U = \mathbf{0}$, where $\mathbf{\Pi_{X^U X}}$ is the coefficient matrix of the projection of $\mathbf{X_i^U}$ on $\mathbf{X_i}$ introduced in (5). This result, the fact that $\mathbf{Proj}(\mathbf{X_i^U}|\mathbf{X_i},\mathbf{X_s}) = \mathbf{Proj}(\mathbf{X_i^U}|\mathbf{X_i}-\mathbf{X_s},\mathbf{X_s})$ and the fact that $\mathbf{X_s}$ is orthogonal to $[\mathbf{X_i}-\mathbf{X}_s]$ together imply that $\mathbf{Proj}(\mathbf{X_i^U}|\mathbf{X_i},\mathbf{X_s})$ can be written as

$$\mathbf{Proj}(\mathbf{X_i^U}|\mathbf{X_i},\mathbf{X_s}) = [\mathbf{X_i}-\mathbf{X}_s]\mathbf{\Pi_{\Delta X^U \Delta X}} + \mathbf{X_s}\mathbf{\Pi_{X^U X}},$$

where $\mathbf{\Pi_{\Delta X^U \Delta X}}$ is the coefficient matrix of the regression of $\mathbf{X_i^U} - \mathbf{X_s^U}$ on $\mathbf{X_i} - \mathbf{X_s}$. By the law of iterated projections ,

$$\mathbf{X_i}\mathbf{\Pi_{X^U X}} \equiv \mathbf{Proj}(\mathbf{X_i^U}|\mathbf{X_i}) = \mathbf{Proj}(\mathbf{Proj}(\mathbf{X_i^U}|\mathbf{X_i},\mathbf{X_s})|\mathbf{X}_i) = \mathbf{X_i}\mathbf{\Pi_{\Delta X^U \Delta X}} + \mathbf{X}_i\mathbf{\Pi_{X_s X_i}}[\mathbf{\Pi_{X^U X}} - \mathbf{\Pi_{\Delta X^U \Delta X}}].$$

After rearranging terms, the above equation implies that for all $\boldsymbol{X}_i$,

$$\mathbf{X_i}\mathbf{\Pi_{X_s X_i}}[\mathbf{\Pi_{X^U X}} - \mathbf{\Pi_{\Delta X^U \Delta X}}] = \mathbf{X_i}[\mathbf{\Pi_{X^U X}} - \mathbf{\Pi_{\Delta X^U \Delta X}}]$$

Provided that $\mathbf{X_i}$ varies within groups, $\mathbf{X_i}\mathbf{\Pi_{X_s X_i}}$ is not equal to $\mathbf{X_i}$, in which case $\mathbf{\Pi_{X^U X}}$ must equal $\mathbf{\Pi_{\Delta X^U \Delta X}}$ for the equation to hold. Using $\mathbf{\Pi_{X^U X}} = \mathbf{\Pi_{\Delta X^U \Delta X}}$ and the fact that $\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})$ $= \mathbf{0}$ when $\mathbf{R} = \mathbf{0}$ to evaluate (75) establishes that $\mathbf{G_1} = \mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}$, as claimed.

A8 Proof of Proposition 3 and Analysis of Assumptions A6.1 and A6.2

A8.1 Proof of Proposition 3:

We begin by reproducing assumptions A6.1 and A6.2 and restating the proposition.

A6.1:

$$Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}]) + 2Cov(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}]) + Var(\tilde{z}_s^U) \geq 0$$

A6.2:

$$Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}]) + 2Cov(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}]) - Var(\xi_s) \geq 0 \tag{94}$$

Proposition 3: *If assumptions A1-A5 from Proposition 1 and A6.1 hold, then* $Var(\mathbf{Z_{2s}}\mathbf{G_2}) \leq Var(\mathbf{Z_s}\mathbf{\Gamma} + z_s^U)$. *If assumptions A1-A5 and A6.2 hold, then* $Var(\mathbf{Z_{2s}}\mathbf{G_2} + v_s) \leq Var(\mathbf{Z_s}\mathbf{\Gamma} + z_s^U)$.

Proof: By definition, $Var(\mathbf{Z_{2s}G_2})+Var(v_s)$ will understate or equal the true school contribution if

$$Var(\mathbf{Z}_s\mathbf{\Gamma}+z_s^U)\geq Var(\mathbf{Z}_{2s}\mathbf{G}_2)+Var(v_s).$$

Recall the definition $\mathbf{Z}_s\mathbf{\Gamma}\equiv\mathbf{X}_s\mathbf{\Gamma}_1+\mathbf{Z}_{2s}\mathbf{\Gamma}_2$. Also, under the assumptions in Proposition 1, Proposition 2 establishes that $\mathbf{G}_2=\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}$ and $v_s=\tilde{z}_s^U+\xi_s$. Using these three equations, we can rewrite the previous inequality as

$$Var(\mathbf{X}_s\mathbf{\Gamma}_1+\mathbf{Z}_{2s}\mathbf{\Gamma}_2+z_s^U)\geq Var(\mathbf{Z}_{2s}(\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}))+Var(\tilde{z}_s^U+\xi_s). \tag{95}$$

Next, substituting for z_s^U using the projection equation $z_s^U=\mathbf{X}_s\mathbf{\Pi}_{z_s^U X_s}+\mathbf{Z}_{2s}\mathbf{\Pi}_{z_s^U Z_{2s}}+\tilde{z}_s^U$ we obtain

$$Var(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}]+\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}]+\tilde{z}_s^U)\geq Var(\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}])+Var(\tilde{z}_s^U+\xi_s). \tag{96}$$

Using the formula for the variance of a sum and the lack of correlation (by definition) between all components of $\mathbf{Z}_s$ and ξ_s,

$$\begin{aligned}&Var(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}])+2Cov(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}],\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}])\\&+Var(\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}])+Var(\tilde{z}_s^U)\geq Var(\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}])+Var(\tilde{z}_s^U)+Var(\xi_s)\end{aligned} \tag{97}$$

Cancelling common terms from both sides yields A6.2:

$$Var(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}])+2Cov(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}],\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}])-Var(\xi_s)\geq 0 \tag{98}$$

In the case of our more conservative estimator, $Var(\mathbf{Z_{2s}G_2})$, the proof follows the exact same logic, except that the terms $Var(\tilde{z}_s^U)$ and $Var(\xi_s)$ do not appear on the right side in (95), and thus $Var(\tilde{z}_s^U)$ does not cancel and $Var(\xi_s)$ need not be subtracted in (98). This leaves the inequality A6.1:

$$Var(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}])+2Cov(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}],\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}])+Var(\tilde{z}_s^U)\geq 0 \tag{99}$$

This concludes the proof.

A8.2 Analysis of Assumption 6

In this subsection we present theoretical and statistical considerations as well as the empirical evidence specific to our application that all indicate that

$$Var(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}])+2Cov(\mathbf{X}_s[\mathbf{\Gamma}_1+\mathbf{\Pi}_{z_s^U X_s}],\mathbf{Z}_{2s}[\mathbf{\Gamma}_2+\mathbf{\Pi}_{z_s^U Z_{2s}}])\geq 0. \tag{100}$$

This condition is stronger than A6.1 (since it omits $Var(\tilde{z}_s^U)$), and is equivalent to A6.2 in contexts where common shocks either do not exist (such as high school graduation) or are considered part of the group treatment effect component $\tilde{z}_s^U$ (since individuals who choose different schools will receive different common shocks). Because 1) we discuss common shocks elsewhere, 2) we introduced the conservative estimator $\widehat{Var}(\mathbf{Z_{2s}G_2})$ to eliminate their influence, and 3) our sorting model provides no guidance about their size, we focus attention here on the case where common shocks do not exist ($\xi_s = 0 \; \forall \; s$).

From a theoretical standpoint, note that standard sorting patterns would suggest a positive rather than a negative covariance between the observable peer effect term $\mathbf{X_s}\mathbf{\Gamma}_1$ and the observed school input term $\mathbf{Z_{2s}}\mathbf{\Gamma}_2$. Without loss of generality, suppose that each element $\mathbf{X}_i$ has been defined so that higher values increase Y_i (i.e. each element of $\boldsymbol{\beta}$ is positive). Most evidence suggests that concentrations of better prepared students and parents (high values of $\mathbf{X_s}$) are likely to provide stronger peer effects relative to concentrations of less prepared students and parents, suggesting that $\mathbf{X_s}\mathbf{\Gamma}_1$ would be positive when $\mathbf{X_s}$ values are high. And wealthier, more educated parents with more able children (presumed to be positive $\mathbf{X_i}$ inputs) tend to be willing to pay more for neighborhoods featuring schools with better inputs. Thus, standard assumptions about sorting would predict that $\mathbf{X_s}\mathbf{\Gamma}_1$ would display a positive covariance with the direct school inputs captured by $\mathbf{Z_{2s}}(\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}})$. And because z_s^U is likely to partially represent peer effects associated with unobserved school characteristics $\mathbf{X_s^U}\mathbf{\Gamma}_1^U$, and $\mathbf{X_s^U}$ projects fully onto $\mathbf{X_s}$ under Proposition 1, $\mathbf{X_s}\mathbf{\Pi}_{z_s^U \mathbf{X_s}}$ is likely to also capture peer effects associated with concentrations of parents/students with high values of unobserved characteristics. So we would also expect $Cov(\mathbf{X_s}\mathbf{\Pi}_{z_s^U \mathbf{X_s}}, \mathbf{Z_{2s}}(\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}))$ to be positive as well.

The most plausible scenario that could produce a negative covariance is one in which parents mostly value the composition of students at schools, and states seek to compensate for high peer effects at some schools by, for example, allocating less funding to these schools or providing incentives for high quality teachers to move to schools in disadvantaged neighborhoods. If such compensation were sufficiently strong, this could in principle create a negative correlation between peer inputs and direct school inputs. However, in order to produce a violation of Assumption 6, the correlation would need to be quite negative and $Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}])$ would need to be fairly small.

To see this, first note that since $v_s = \tilde{z}_s^U$ is uncorrelated by definition with $\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}]$,

$$Cov(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}]) = Cov(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}] + \tilde{z}_s^U).$$

Next use this result and the definition of correlation to rewrite (100) as

$$\begin{aligned}&Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}]) + \\ &2Corr(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}] + \tilde{z}_s^U)\sqrt{Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}])}\sqrt{Var(\mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}] + \tilde{z}_s^U)} \geq 0.\end{aligned} \tag{101}$$

Define the following scalar parameters:

$$\rho \equiv Corr(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}] + \tilde{z}_s^U)$$
$$\mu \equiv \frac{Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}])}{Var(\mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}] + \tilde{z}_s^U)} \equiv \frac{Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}])}{Var(\mathbf{Z_{2s}}\mathbf{G_2} + v_s)}.$$

ρ captures the correlation between the school inputs that project onto $\mathbf{X_s}$ and the school inputs that either project onto $\mathbf{Z_{2s}}$ or form the residual. The parameter μ captures the ratio of variances of these objects. Then we can rewrite the difference between our "lower bound" estimator and the true variance in school effects $Var(\mathbf{Z_s\Gamma} + z_s^U)$ as a fraction of our estimator (i.e. the size of the potential overstatement of the true variance in percentage terms) in terms of only ρ and μ:

$$\begin{aligned}
&\frac{Var(\mathbf{Z_s\Gamma} + z_s^U) - Var(\mathbf{Z_{2s}G_2} + v_s)}{Var(\mathbf{Z_{2s}G_2} + v_s)} \\
&= \frac{Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}]) + 2Cov(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}], \mathbf{Z_{2s}}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}] + \tilde{z}_s^U)}{Var(\mathbf{Z_{2s}G_2} + v_s)} \\
&= \frac{\mu Var(\mathbf{Z_{2s}G_2} + v_s) + 2\rho\sqrt{\mu} Var(\mathbf{Z_{2s}G_2} + v_s)}{Var(\mathbf{Z_{2s}G_2} + v_s)} \\
&= \mu + 2\rho\sqrt{\mu}
\end{aligned}$$

Generally speaking, this expression is only negative (implying that our estimator overstates the true treatment effect variance) for combinations of highly negative values of ρ and low values of μ. Table A23 summarizes the relationship between ρ, μ, and $\mu + 2\rho\sqrt{\mu}$.

Specifically, the rows in Column 1 display the values of ρ from -0.1 to -1. Column 2 displays the maximum overstatement of treatment effect variance as a fraction of our estimate for each value of ρ (i.e. $\max_\mu \mu + 2\rho\sqrt{\mu}$), while Column 3 displays the size of μ that generates this maximum ($\arg\max_\mu \mu + 2\rho\sqrt{\mu}$). Column 4 provides the threshold value of μ (denoted $\mu_0(\rho)$) beyond which Assumption 6 is satisfied (i.e. the value of μ such that our lower bound estimator actually equals the true school treatment effect variance: $\mu_0(\rho) + 2\rho\sqrt{\mu_0(\rho)} = 0$).

Table A23 shows that when moderate compensation exists (e.g. $\rho = -0.2$), the maximum bias is very small: even our larger lower bound estimator only overstates the true school/neighborhood effect variance by 4 percent. And even this scenario requires that peer effects and other school inputs projecting onto $\mathbf{X_s}$ be quite trivial in magnitude. The bias is maximized when $Var(\mathbf{X_s}[\mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}])$ is only 4 percent as large as $Var(\mathbf{Z_{2s}G_2} + v_s)$. Note that $Var(\mathbf{Z_{2s}G_2} + v_s)$ is typically estimated to be around 2 percent of the variance in the latent index determining our binary outcomes. Furthermore, the overstatement is eliminated when μ is 16 percent of $Var(\mathbf{Z_{2s}G_2} + v_s)$; higher levels of μ lead our estimator to understate the true school effect variance. Indeed, large overstatement of the treatment effect variance can only occur with arguably unrealistically strong compensation by states and schools. Specifically, the scenarios that produce large overstatement of true school effects generally require peer effects to be quite weak compared to $\mathbf{Z_{2s}}(\mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}})$ (low μ). But if compensation using these school inputs was so strong *and* these inputs were so important relative to peer effects (so that schools are dramatically overcompensating for peer effects), it is hard to believe that the parents/students with high individual contributions $\mathbf{X_i}$ would continue to cluster in the schools providing such low value added.

Finally, other aspects of our variance decompositions also suggest that a violation of Assumption 6 is unlikely. Specifically, we report $2Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{X_s G_1}) = 2Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2} + v_s) + Var(\mathbf{X_s G_1})$ for each of our outcomes and specifications in Appendix Tables (A20)-(A22). From (75), we know that $\mathbf{G_1} \neq \mathbf{\Gamma}_1$ due to the presence of $\mathbf{\Pi}_{\tilde{x}_i^U \mathbf{X_s}}$ (our control function absorbs sorting on unobservables!). However, the magnitude of $2Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{X_s G_1})$ is generally at least the half the size of (and often exceeding) that of $Var(\mathbf{Z_{2s} G_2})$, suggesting that the sorting on unobservables component $\mathbf{X_s \Pi}_{\tilde{x}_i^U \mathbf{X_s}}$ would need to be quite substantial for μ to be low enough to be consistent with a violation of Assumption 6.

Indeed, as we pointed out in Section 8 of the paper, under the stronger but standard selection-on-observables-only assumption, $\mathbf{X_s B}$ would fully capture student sorting, $\mathbf{\Pi}_{\tilde{x}_i^U \mathbf{X_s}} = \mathbf{0}$ and $\mathbf{G_1} = \mathbf{\Gamma}_1 + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}$. Thus, under selection-on-observables, $2Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{X_s G_1}) = Var(\mathbf{Z_s \Gamma} + \xi_s) - Var(\mathbf{Z_{2s} G_2} + v_s)$, capturing directly the degree to which our estimator $Var(\mathbf{Z_{2s} G_2} + v_s)$ understates or overstates the true neighborhood/school effect variance. And this sum is positive in every specification and outcome we use, and usually substantially so. Thus, the limited evidence that our empirical estimates provide about Assumption 6 strongly suggest that it holds in our data.

Even when Assumption 6 is violated (but Assumptions 1-5 hold), the quantity $Var(\mathbf{Z_{2s} G_2})$ is still an object of interest. In particular, it still represents a component of variance that purely captures across school/neighborhood differences in external inputs. When interpreting our estimated shifts in outcomes from moving from a 10th to 90th quantile school, we have generally considered the move from the perspective a single family making a school/neighborhood choice. Through that lens, when Assumption 6 is violated the shifts we estimate could overstate the change in the expected outcome for the student from such a family, because the improvement in school inputs and policies captured by $\mathbf{Z_{2s} G_2}$ would be partially offset by a decrease in peer inputs (or other school inputs that project onto $\mathbf{X_s}$).

However, the same 10th-to-90th quantile shifts could also be interpreted as an estimate of the gain in expected outcomes of students at the 10th quantile school that would occur if the non-peer school inputs and policies $\mathbf{Z_{2s}}$ of a school at the 90th quantile of $Var(\mathbf{Z_{2s} G_2})$ were bestowed upon the 10th quantile school. This counterfactual is more relevant for the state policy maker or principal, who wants a broader understanding of how important school inputs and policies are for student outcomes. Importantly, this counterfactual holds the peer inputs at each school fixed when changing school inputs, so that the correlation between peer and school inputs induced by sorting or compensation is irrelevant. Thus, violations of Assumption 6 do not change what we learn about the importance of school inputs in producing student outcomes, and about the potential for outcome gains from applying superior school inputs or successful policies currently observed in some high value-added schools (conditional on peer inputs) to other schools in the nation.

A9 Estimation of Model Parameters

In this section we discuss estimation of the coefficients $\mathbf{B}$, $\mathbf{G_1}$, $\mathbf{G_2}$ and the variances of the error components $Var(v_s)$ and $(v_{si} - v_s)$. The estimation strategy depends on the outcome, so we consider the outcomes in turn. To simplify the notation, let $v'_{si} \equiv v_{si} - v_s$

A9.1 Years of Postsecondary Academic Education

Parameter estimation is most straightforward in the case of years of postsecondary academic education. Recall that $\mathbf{Z_s}$ is comprised of two components: $\mathbf{Z_s} = [\mathbf{X_s}; \mathbf{Z_{2s}}]$. $\mathbf{Z_{2s}}$ consists of school and neighborhood characteristics for which direct measures are available, such as student/teacher ratio, city size, and school type. $\mathbf{X_s}$ consists of school wide averages for each variable in $\mathbf{X_i}$, 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 $\mathbf{X_s}$ differs across specifications that use different $\mathbf{X}$ vectors. $\mathbf{G_1}$ and $\mathbf{G_2}$ are the corresponding subsets of the coefficients in $\mathbf{G}$. We replace $\mathbf{X_s}$ with $\hat{\mathbf{X}}_\mathbf{s}$, where $\hat{\mathbf{X}}_\mathbf{s}$ is the average of $\mathbf{X_i}$ computed over all available students from the school, leading to the equation below.[44] The regression model is

$$Y_{si} = \mathbf{X_i B} + \hat{\mathbf{X}}_\mathbf{s} \mathbf{G_1} + \mathbf{Z_{2s} G_2} + [\mathbf{X_s} - \hat{\mathbf{X}}_\mathbf{s}]\mathbf{G_1} + v_s + v'_{si} \tag{102}$$

using the appropriate panel weights from the surveys. We estimate the model parameters using restricted maximum likelihood (REML). We treat v_s as a random effect and assume the error components are normally distributed, ignoring the error component $[\mathbf{X_s} - \hat{\mathbf{X}}_\mathbf{s}]\mathbf{G_1}$. REML accounts for degrees of freedom in estimating $Var(v_s)$ and $Var(v'_{si})$, while maximum likelihood does not.[45] Computations were performed using the STATA Version 14 procedure *mixed* with the REML option. Because we experienced computational difficulties when using panel weights, the estimates are unweighted.

A9.2 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

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

[44] A substantial number of students who appear in the base year of the surveys can be used to construct $\hat{\mathbf{X}}_\mathbf{s}$ but cannot be used to estimate (102) 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 6, we impute missing values for most of our explanatory variables prior to estimating $\mathbf{B}$ and $\mathbf{G}$, but we do not use the imputed values when constructing the school averages.

[45] See Harvey (1977) for an overview. In the normal regression model without the random effect v_s, the REML estimator of $Var(v'_{si})$ is the usual OLS estimator—the sum of squared residuals divided by the sample size minus 1 plus the number of regressors. The ML estimator of $Var(v'_{si})$ divides the sum of squared residuals by the sample size only, thus ignoring the lost degrees of freedom absorbed by additional regressors.

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. ς_{sit} is a stationary component that evolves as a result of luck in the job search process or within a company, changes in motivation or productivity due to health and other short term factors that may persist for up to 7 years. It also includes measurement error.[46] The determination of Y_{si} is given by (8) which leads to the regression equation (12). After substituting for Y_{si} and replacing $\mathbf{X_s}$ with $\hat{\mathbf{X}}_\mathbf{s}$, the wage equation is

$$Y_{sit} = \mathbf{X_i B} + \hat{\mathbf{X}}_\mathbf{s}\mathbf{G_1} + \mathbf{Z_{2s}G_2} + (\mathbf{X_s} - \hat{\mathbf{X}}_\mathbf{s})\mathbf{G_1} + v_s + v'_{si} + \varsigma_{sit}.$$

We estimate the model by REML under the assumption that the error components are normally distributed and that $Cov(\varsigma_{si1979}, \varsigma_{si1986})$ is 0.[47]

A9.3 High School Graduation and College Enrollment

For binary outcomes such as 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},

$$HSGRAD_{si} = 1(\mathbf{X_i B} + \mathbf{X_s G_1} + \mathbf{Z_{2s}G_2} + v_s + v'_{si} > 0). \tag{103}$$

We replace $\mathbf{X_s}$ with $\hat{\mathbf{X}}_\mathbf{s}$ and estimate the equation

$$HSGRAD_{si} = 1(\mathbf{X_i B} + \hat{\mathbf{X}}_\mathbf{s}\mathbf{G_1} + \mathbf{Z_{2s}G_2} + (\mathbf{X_s} - \hat{\mathbf{X}}_\mathbf{s})\mathbf{G_1} + v_s + v'_{si} > 0) \tag{104}$$

via maximum likelihood random effects probit using STATA xtprobit (version 14). Because of software constraints, student weights are set to the average student-level weights for the students' schools. The procedure for enrollment in a four-year college is analogous to that of high school graduation. In both cases, we adjust the ML estimates of $Var(v_s)$ and $Var(v'_{si})$ to correct for degrees of freedom. Essentially, we treat the equation for the latent variable Y_{si} as a continuous regression

[46] In prior drafts of the paper we used a different estimation procedure based on the method of moments. We were able to include a random walk component e_{sit} as well as ς_{sit}, and we did so because the earnings dynamics literature typically finds evidence of a highly persistent wage component. Some studies fail to reject the hypothesis that e_{sit} is a random walk. Recent examples include Baker and Solon (2003), Haider (2001), and Meghir and Pistaferri (2004). We were unable to modify the method of moments estimator of $Var(v_s)$ to account for the degrees of freedom used in estimating the regression coefficients. The mixed effects estimator we use assumes that $Var(\varsigma_{sit})$ does not depend on t. This rules out a random walk component.

[47] In reality, we also include a vector $\mathbf{T_{it}}$ 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 $\boldsymbol{\Upsilon}$ be the corresponding vector of wage coefficients. The term $\mathbf{T_{it}}\hat{\boldsymbol{\Upsilon}}$ does not contribute to Y_{si} and does not play a role in the variance decompositions. The estimate of $\hat{\boldsymbol{\Upsilon}}$ depends on whether tests, postsecondary education, or both are in $\mathbf{X_i}$. We report results with and without these variables. In our main specification, we exclude postsecondary education from $\mathbf{X_i}$.

model and assume that the small sample bias corrections for the regression model carry over. We provide the necessary detail in the next section.

A9.3.1 Formulae for Estimating $Var(v_i)$ and $Var(v_s)$ in the Case of Binary Outcome Variables

We start with formulae for the unweighted case. In the random effects regression model with a continuous dependent variable, the formula for the unbiased estimator of $Var(v_i)$ is:

$$\widehat{Var}(v_i) = \frac{\sum_{s=1}^{S}\sum_{i=1}^{N_s}\hat{e_{is}}^2}{N-S-K_w-1}, \tag{105}$$

where K_w is the number elements of $\mathbf{X_i}$ (i.e., the number of regressors that vary within schools) and S is the number of schools. In the continuous case, the ML estimator is:

$$\widehat{Var}(v_i)_{ML} = \frac{\sum_{s=1}^{S}\sum_{i=1}^{N_s}\hat{e_{is}}^2}{N-S-1}. \tag{106}$$

In the binary case Y_{si} is latent, so the squared residual $\hat{e_{it}}^2$ is also unobserved. However, since the variance of the latent index is not identified, the probit estimator normalizes scale so that the variance of e_{is} is 1. In the ML case, this means that the scale is chosen so that:

$$\widehat{Var}(v_i)_{ML} = \frac{\sum_{s=1}^{S}\sum_{i=1}^{N_s}\hat{e_{is}}^2}{N-S-1} = 1. \tag{107}$$

This equation and (105) implies that we can estimate $Var(v_i)$ via

$$\widehat{Var}(v_i) = \frac{N-S-1}{N-S-K_w-1}\widehat{Var}(v_i)_{ML} = \frac{N}{N-S-K_w-1}(1). \tag{108}$$

In the continuous case the standard unbiased estimator for the group-level error component, $Var(v_s)$, is:

$$\widehat{Var}(v_s) = \max\{0, \frac{SSR_b}{S-K_b} - \frac{\widehat{Var}(v_i)}{\overline{N}_s}\} \tag{109}$$

where the max function is taken to prevent a negative variance estimate, K_b is the number of variables that vary only across s ($\mathbf{X_s}$ and $\mathbf{Z_{2s}}$), $\overline{N}_s$ is the harmonic mean of the number of observations (students) per school, and SSR_b is the sum of squared residuals from the between-group regression:

$$SSR_b = \sum_{s=1}^{S}(\overline{Y}_s - \mathbf{X_s}\hat{\mathbf{B}} - \mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1} - \mathbf{Z_{2s}}\hat{\mathbf{G}}_\mathbf{2})^2.$$

Let $\widehat{Var}(v_s)_{ML}$ denote the ML estimator of $Var(v_s)$, which does not correct for the degrees of freedom used to estimate $\mathbf{G_1}$ and $\mathbf{G_2}$:

$$\widehat{Var}(v_s)_{ML} = \max\{0, \frac{SSR_b}{S} - \frac{\widehat{Var}(v_i)}{\overline{N}_s}\}. \tag{110}$$

In the binary case the group mean $\overline{Y}_s$ of the latent index Y_i is unobserved for the binary outcomes, so we must approximate SSR_b. In the continuous case one can see from (110) that ML will choose $Var(v_s)$ to be equal to the remaining between-school variance not accounted for by (1) the between school variance contributed by $\mathbf{X_s\hat{B}} + \mathbf{X_s\hat{G}_1} + \mathbf{Z_{2s}\hat{G}_2}$ or (2) variation across schools in the mean of v_i for students chosen for the sample. Reasoning by analogy (and assuming a non-negative variance estimate), in the binary case we approximate $\widehat{Var}(v_s)_{ML}$ as:

$$\widehat{Var}(v_s)_{ML} \approx \frac{SSR_b}{S} - \frac{\widehat{Var}(v_i)}{\overline{N}_s}, \tag{111}$$

where S is the number of schools. Rearranging, we obtain:

$$SSR_b \approx S(\widehat{Var}(v_s)_{ML} + \frac{\widehat{Var}(v_i)}{\overline{N}_s}). \tag{112}$$

Using the above approximation and incorporating the estimator $\widehat{Var}(v_i)$, our estimator of $Var(v_s)$ becomes:

$$\widehat{Var}(v_s) = \max\{0, \frac{S(\widehat{Var}(v_s)_{ML} + \frac{\widehat{Var}(v_i)}{\overline{N}_s})}{S-K_b} - \frac{\widehat{Var}(v_i)}{\overline{N}_s}\} = \max\{0, \frac{S}{S-K_b}\widehat{Var}(v_s)_{ML} + (\frac{S}{S-K_b} - 1)(\frac{\widehat{Var}(v_i)}{\overline{N}_s})\}. \tag{113}$$

where $\widehat{Var}(v_i)$ is given by (108) above.

The bias-corrected estimators must be modified when sampling weights are incorporated into the ML estimator. We replace the sample size N in (108) by Kish's "effective sample size", $N^{Eff} = \frac{(\sum_{i=1}^{N} w_i)^2}{\sum_{i=1}^{N} w_i^2}$, where w_i is the observation weight:

$$\widehat{Var}(v_i) = \frac{N^{eff}}{N^{eff} - S - K_w - 1}(1). \tag{114}$$

We redefine SSR_b and $\bar{N}_s$ to be their school-weighted counterparts and we replace S with Kish's "effective sample size" of schools, $S^{Eff} = \frac{(\sum_{s=1}^{S} w_s)^2}{\sum_{s=1}^{S} w_s^2}$, where w_s is the mean of the individual weights w_i of the sampled students at the school. Let $\widehat{Var}(v_s)_{WML}$ be the weighted maximum likelihood estimator. This yields:

$$\widehat{Var}(v_s) = \max\{0, \frac{S^{Eff}}{S^{Eff} - K_b}\widehat{Var}(v_s)_{WML} + (\frac{S^{Eff}}{S^{Eff} - K_b} - 1)(\frac{\widehat{Var}(v_i)}{\overline{N}_s})\}. \tag{115}$$

where our estimator for $\widehat{Var}(v_i)$ is given by (114) above.

In the probit model $Var(v_i)$ is normed to 1, so the final step is to divide $\widehat{Var}(v_s)$ by $\widehat{Var}(v_i)$ and then set $\widehat{Var}(v_i)$ to 1. We perform analogous scale adjustments to the estimates of the variance and covariances among the regressions indices that enter into the variance decompositions and the $10-90$ treatment effect calculations.

A9.4 Estimating the Variances and Covariances of the Components of the Regression Index

Here we describe how we account for the effects of sampling error in the $\hat{\mathbf{B}}$, $\hat{\mathbf{G}}_1$, and $\hat{\mathbf{G}}_2$ coefficient vectors when estimating $Var(\mathbf{X}_i\mathbf{B})$, $Var(\mathbf{X}_s\mathbf{G}_1)$, $Var(\mathbf{Z}_{2s}\mathbf{G}_2)$, and the covariance terms that enter the variance decompositions reported in the paper. Consider the case of $Var(\mathbf{Z}_s\mathbf{G}_2)$. Recalling that $\mathbf{Z_{2s}}$ has mean $\mathbf{0}$, note first that

$$Var(\mathbf{Z_{2s}}\hat{\mathbf{G}}_\mathbf{2}) = \frac{1}{N}\sum_i(\mathbf{Z_{s(i)}}\hat{\mathbf{G}}_\mathbf{2}\hat{\mathbf{G}}_\mathbf{2}'\mathbf{Z}_\mathbf{s(i)}') \tag{116}$$

$$=\frac{1}{N}\sum_i(\mathbf{Z_{2s(i)}}\mathbf{G_2}\mathbf{G_2}'\mathbf{Z}_\mathbf{2s(i)}') + \frac{1}{N}\sum_i \mathbf{Z_{s(i)}}[\hat{\mathbf{G}}_\mathbf{2}-\mathbf{G_2}][\hat{\mathbf{G}}_\mathbf{2}-\mathbf{G_2}]'\mathbf{Z}_\mathbf{s(i)}'. \tag{117}$$

In the above equation we have made the dependence of s on i explicit. The expectation of the first term on the right is $Var(\mathbf{Z_{2s}G_2})$. The second term is the effect of sampling error in $\hat{\mathbf{G}}_\mathbf{2}$, conditional on $\mathbf{Z_{2s(i)}}$. It has expectation

$$\frac{1}{N}\sum_i \mathbf{Z_{2s(i)}}\mathbf{Var}(\hat{\mathbf{G}}_\mathbf{2})\mathbf{Z}_\mathbf{2s(i)}'. \tag{118}$$

Using (118) we generate a bias-adjusted estimator of $Var(\mathbf{Z_{2s}G_2})$ as

$$\widehat{Var}(\mathbf{Z_{2s}G_2}) = [\frac{1}{N}\sum_i(\mathbf{Z_{2s(i)}}\hat{\mathbf{G}}_\mathbf{2}\hat{\mathbf{G}}_\mathbf{2}'\mathbf{Z}_\mathbf{2s(i)}') - \frac{1}{N}\sum_i \mathbf{Z_{2s(i)}}\widehat{\mathbf{Var}}(\hat{\mathbf{G}}_\mathbf{2})\mathbf{Z}_\mathbf{2s(i)}']$$

In the above formula $\widehat{\mathbf{Var}}(\hat{\mathbf{G}}_\mathbf{2})$ is the estimator based on the formula for the asymptotic variance associated with the estimator $\hat{\mathbf{G}}_\mathbf{1}$, which depends on the outcome. We do not account for the use of imputed data. In practice, we report population weighted variances, so sample weights appear in the two sums. The estimators of $Var(\mathbf{X_iB})$, $Var(\mathbf{X_sB})$ and $Var(\mathbf{Z_{2s}G_2})$ are almost exactly analogous.

To estimate $Cov(\mathbf{X_sG_1}, \mathbf{Z_{2s}G_2})$, we first estimate $Var(\mathbf{Z_sG}) \equiv Var(\mathbf{X_sG_1} + \mathbf{Z_{2s}G_2})$ via:

$$\widehat{Var}(\mathbf{Z_sG}) = [\frac{1}{N}\sum_i(\mathbf{Z_{s(i)}}\hat{\mathbf{G}}\hat{\mathbf{G}}'\mathbf{Z}_\mathbf{s(i)}') - \frac{1}{N}\sum_i \mathbf{Z_{s(i)}}\widehat{\mathbf{Var}}(\hat{\mathbf{G}})\mathbf{Z}_\mathbf{s(i)}'] \tag{119}$$

Then we generate $\widehat{Cov}(\mathbf{X_sG_1}, \mathbf{Z_{2s}G_2})$ via:

$$\widehat{Cov}(\mathbf{X_sG_1}, \mathbf{Z_{2s}G_2}) = \frac{\widehat{Var}(\mathbf{Z_sG}) - \widehat{Var}(\mathbf{X_sG_1}) - \widehat{Var}(\mathbf{Z_{2s}G_2})}{2} \tag{120}$$

We use an analogous procedure for $Cov(\mathbf{X_sB}, \mathbf{X_sG_1})$ and $Cov(\mathbf{X_sB}, \mathbf{Z_{2s}G_2})$.

A10 Decomposing the Variance in Educational Attainment and Permanent Wages

In this section we discuss an analysis of variance based on (23) that can be used to place a lower bound on the importance of factors that are common to students from the same school.[48] As with parameter estimation, the details of our procedure depend upon the outcome. We begin with years of postsecondary education and permanent wages.

A10.1 Years of Postsecondary Education and Wages

We start with years of 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_s is the average outcome for students from s. From (21) we obtain

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

and

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

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

$$Var(Y_{si}) = [Var((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + Var(v_{si} - v_s)] + \tag{121}$$
$$[Var(\mathbf{X_s B}) + 2Cov(\mathbf{X_s B}, \mathbf{X_s G_1}) + 2Cov(\mathbf{X_s B}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{X_s G_1}) + \tag{122}$$
$$2Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{Z_{2s} G_2}) + Var(v_s)]. \tag{123}$$

We replace the population moments on the right hand side with the population weighted estimates discussed in the proceeding section. The sum of the weighted estimates of the components of $Var(Y_{si})$ need not equal the weighted sample variance of Y_{si}, so we use:

$$\widehat{Var}(Y_{si}) \equiv [\widehat{Var}((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + \widehat{Var}(v_{si} - v_s)] + \tag{124}$$
$$[\widehat{Var}(\mathbf{X_s B}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{X_s G_1}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{X_s G_1}) + \tag{125}$$
$$2\widehat{Cov}(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{Z_{2s} G_2}) + \widehat{Var}(v_s)]. \tag{126}$$

as the estimator of $Var(Y_{si})$. We express the variance and covariance estimates as fractions of $Var(Y_{si})$ by dividing the variance and covariance terms by $\widehat{Var}(Y_{si})$. For example, we compute $\widehat{Var}(\mathbf{Z_{2s} G_2})/\widehat{Var}(Y_\mathbf{s})$. The fractions are reported in Table 6, Columns 1 and 2 and Appendix Table

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

[49] The equation below imposes $Cov(\mathbf{X_i}, v_{si} - v_s) = 0$, which is implied by our definition of $\mathbf{B}$ and $v_{si} - v_s$. The equation also imposes $Cov(\mathbf{X_s}, v_s) = 0$ and $Cov(\mathbf{Z_{2s}}, v_s) = 0$, which are implied by our definition of $[\mathbf{G_1}, \mathbf{G_2}]$ and v_s (see Section 4).

A22, Columns 1 and 2.

The procedure for decomposing the variance of the permanent log wage component Y_{si} is essentially the same as the procedure just described for years of postsecondary education. We exclude $\widehat{Var}(\varsigma_{sit})$ from $\widehat{Var}(Y_{si})$ because ς_{sit} is transitory. The fractions of variance for permanent log wages are reported in Table 6, Columns 3-6 and Appendix Table A22, Columns 3-6.

A10.2 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

$$Var(Y_{si}) \equiv Var((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + 1 + \tag{127}$$

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

$$2Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{Z_{2s} G_2}) + Var(v_s)] \tag{128}$$

We thus estimate $Var(Y_{si})$ via:

$$\widehat{Var}(Y_{si}) \equiv \widehat{Var}((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + 1 \tag{129}$$

$$[\widehat{Var}(\mathbf{X_s B}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{X_s G_1}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{X_s G_1}) + \tag{130}$$

$$2\widehat{Cov}(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{Z_{2s} G_2}) + \widehat{Var}(v_s)]. \tag{131}$$

In the tables we report the fractions of $Var(Y_{si})$ contributed by the various components.

A10.3 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 using 500 replications. We bootstrap the entire estimation procedure, including imputation of missing data, estimation of model parameters, variance decompositions, and treatment effects. To preserve the size distribution of the samples of students from particular schools, we divide the sample into five school sample size classes and re-sample schools within class.

A11 Using the North Carolina Data to Assess the Magnitude of Bias from Limited Samples of Students Per School

The monte carlo simulations in Online Appendix A6 suggest that estimation based on subsamples of 20 students per school (similar to those used to construct $\mathbf{X_s}$ in the three panel survey 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 200 samples drawn from each sampling scheme.

Tables A10 and A11 present the results of this exercise for the baseline and full specifications, respectively. For comparison, the first column of Panel A in each table presents the variance decomposition described in Section 5 for the entire 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 4 display the results from recomputing these estimates for subsamples of the North Carolina population featuring the same distributions of school-specific sample sizes as the high schools in NLS72 and ELS2002 and the 8th grade schools in NELS88. In both tables, Columns 2-4 report very similar numbers to one another, and reveal that the use of small student samples at each school may produce relatively small amounts of bias for each of our panel survey datasets. Most of the rows of Panel A exhibit close matches between the true results in Column 1 and the sample-based results in Columns 2-4. Of particular interest are the last two rows of Panel A. In the baseline specification in Table A10, we see that the panel survey sample size distributions lead to an understatement of the true variance fraction for the lower bound without common shocks ($Var(\mathbf{Z_{2s}G_2})$) of around 25% but fairly accurate estimates of the unobserved school component ($Var(v_s)$). These translate to underestimates of the impact of a 10th-90th quantile shift in school quality on the probability of graduation of around one percentage point for both the estimators that include and exclude $Var(v_s)$. The results for the full specification in Table A11 show much smaller understatement of $Var(\mathbf{Z_{2s}G_2})$ (around 10%), but now also display a 10% understatement of $Var(v_s)$. Overall, the effects of 10th-to-90th shifts in school/neighborhood quality are understated by less than half a percentage point for our more conservative estimator based on $Var(\mathbf{Z_{2s}G_2})$ and by closer to a percentage point for our less conservative estimator based on $Var(\mathbf{Z_{2s}G_2} + v_s)$.

Taken together, Tables A10 and A11 show that the use of small samples from each school to construct the school averages $\mathbf{X_s}$ need not generate significant bias in our lower bound estimators of the impact of shifts in school/neighborhood quality. If anything, the slight bias that is generated only serves to make the lower bound estimators more conservative.

A12 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 when possible for a number of reasons. The first is to reduce the influence of choice-based sampling, which is an issue in NELS88. It is also a potential issue for the wage analysis based on NLS72, but we had difficulty computing weighted estimates. 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. For all outcomes, including wages, we employ sample weights when using the regression model parameters to construct estimates of $Var(\mathbf{Z}_{2s}\mathbf{G}_2)$, $Var(\mathbf{X}_s\mathbf{G}_1)$. Note, though, that we do not adjust weights for item non-response associated with the key variables required for inclusion in our sample. As discussed in section 5.1, due to computational difficulties, for our continuous outcomes (years of postsecondary education and log wages) we do not incorporate weights into the REML procedure used to estimate the coefficients $\mathbf{B}$, $\mathbf{G_1}$, and $\mathbf{G_2}$ and the error component variances $Var(v_{si} - v_s)$ and $Var(v_s)$. However, conditional on the estimated coefficient vectors and error variances, panel weights are still used to compute the variances of the various regression indices (such as $Var(\mathbf{Z_{2s}G_2})$), and they are still used to average over the population of students when estimating the impact of 10th-to-90th quantile shifts in school quality.

A13 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 than 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.

A13.1 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 $\mathbf{A_c}$. Some of the amenities are likely to reflect the demographic makeup of the class and thus are endogenous to classroom assignment. Others can be considered exogenous to the principal's student-to-classroom allocation problem. These would 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.

We can then adapt the utility function featured in (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, $\mathbf{X_i}$ is a vector of student characteristics that are observed by the econometrician and are relevant for the outcome Y_{si}, the student's end-of-year standardized test score. Similarly, $\mathbf{X_i^U}$ is a vector of student characteristics that are unobserved by the econometrician but are observed by the principal and are relevant for test score performance, and $\mathbf{Q_i}$ 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 $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}^{\mathbf{U}}$ and $\boldsymbol{\Theta}^{Q}$ matrices might capture a principal's belief about which types of students are most likely to benefit from a better teacher or difficulty level. $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}^{\mathbf{U}}$, and $\boldsymbol{\Theta}^{Q}$ might also reflect the desire to placate parents or students, where students/parents with certain values of $\mathbf{X_i}$, $\mathbf{X_i^U}$, or $\mathbf{Q_i}$ 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.[50]

[50] 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 $\mathbf{X_i}$ and student characteristics that are only available to the principal (part of $\mathbf{X_i^U}$). He discusses bias in VAM models that include $\mathbf{X_i}$ and possibly other controls. He does not consider the potential for $\mathbf{X_c}$ to control for $\mathbf{X_c^U}$.

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). First we consider the special case in which none of the amenities reflect the demographic makeup of the class and thus $\mathbf{A_c}$ can be considered exogenous to the principal's student-to-classroom allocation problem. 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:

$$\begin{aligned}
&\max_{c:\mathscr{I}\rightarrow\mathscr{C}} \sum_{i\in\mathscr{I}} U_{ic(i)} \\
&s.t. \sum_{c'} \mathbb{1}(c(i) = c') = 1 \ \forall\ i \\
&s.t. \sum_{i'} \mathbb{1}(c(i') = c) = \overline{C}_c \ \forall\ c \in \mathscr{C}
\end{aligned} \tag{132}$$

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:

$$\begin{aligned}
&\max_{\mathbf{d}} \ \mathbf{a} * \mathbf{d} \\
&s.t. \ \mathbf{M}_i * \mathbf{d} = 1 \ \forall\ i \in \mathscr{I} \\
&s.t. \ \mathbf{N}_c * \mathbf{d} = \overline{C}_c \ \forall\ c \in \mathscr{C} \\
&s.t. \ \mathbf{d} \in \{0,1\}
\end{aligned} \tag{133}$$

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

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

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):

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

$\mathbf{N}_c$ 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 $\{P_c\}$.[51] This means that the optimal assignment for each individual is also the solution to his/her utility maximization problem:

$$c(i) = \arg\max_c \tilde{U}_{ic} - P_c \equiv U_{ic} \tag{134}$$

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 $\mathbf{\Theta^U} = \mathbf{R}\tilde{\mathbf{\Theta}}$ is satisfied for some matrix $\mathbf{R}$, $\mathbf{X_c}$ will be a linear function of $\mathbf{X_c^U}$.

[51] The case they consider is 1:1, but it easy to recast the classroom assignment problem as assignment of students to seats. Each class room has a fixed number of seats that have exactly the same value of $\mathbf{A_c}$ and the same shadow price. A student's preferred seats will all be in one classroom, and he/she will be indifferent among them. The student lets the principal (who is also indifferent) assign a specific seat.

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. However, in the elementary and middle school contexts, it is likely that some elements of $\mathbf{A_c}$ reflect the student makeup of the class. Anticipated peer effects complicates the principal's problem, 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 $\mathbf{A_c}$, while parents would take $\mathbf{A_s}$ as given. We have not yet extended Proposition 1 to a classroom assignment problem with endogenous amenities.

A13.2 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 $\mathbf{Z_c}\mathbf{\Gamma} + z_c^U + \eta_{ci}$, mirroring (8). As before, partition the vector of observed classroom characteristics into two parts $\mathbf{Z_c} = [\mathbf{X_c}, \mathbf{Z_{2c}}]$, where $\mathbf{X_c}$ captures classroom averages of observed student characteristics and $\mathbf{Z_{2c}}$ represents other observed classroom characteristics. Consider the classroom version of our estimating equation (21):

$$Y_{si} = \mathbf{X_i}\boldsymbol{B} + \mathbf{X_c}\mathbf{G_1} + \mathbf{Z_{2c}}\mathbf{G_2} + v_{ci}, \tag{135}$$

When past test scores are elements of $\mathbf{X_i}$ and a design matrix $\mathbf{D_{c(i)}}$ indicating which classrooms were taught by which teachers is included in $\mathbf{Z_{2c}}$, (135) represents a standard teacher value-added specification.[52]

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 $\mathbf{X_c}$ and $\mathbf{X_c}^U$ are linearly dependent. Suppose in addition that the principal's perception of teacher quality is noisy, so that $\mathbf{D_c}$ is not collinear with $\mathbf{A_c}$ (and therefore not collinear with $\mathbf{X_c}$). Then our analysis in Section 4.3 suggests that $\mathbf{G_2} = \mathbf{\Gamma}_2 + \mathbf{\Pi}_{\mathbf{Z_c}^U\mathbf{Z_{2c}}}$. Since $\mathbf{Z_{2c}}$ includes the teacher design matrix $\mathbf{D_{c(i)}}$, we see that including classroom averages of student characteristics $\mathbf{X_c}$ 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 $\mathbf{\Pi}_{\mathbf{Z_c}^U\mathbf{Z_{2c}}}$ stems from the possible correlation between the assignment of the chosen teacher to the classroom and other aspects of the classroom environment. Note that $\mathbf{G_1}$ should be allowed to differ across schools or districts if the preference parameters $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ are believed to differ.

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 (i.e. are elements of $\mathbf{A_c}$) or are directly included as other controls in $\mathbf{Z_{2c}}$. If in addition the mapping from $\mathbf{A_c}$ to $\mathbf{X_c}$ is linear, then the analysis in Section 4.3.1 reveals that including classroom averages of observed

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

student characteristics will also purge teacher value-added estimates $\mathbf{G_2}$ of any omitted variables bias driven by inequitable access to advantageous classroom environments (the subvector of $\mathbf{\Pi_{Z_c^U Z_{2c}}}$ corresponding to the teacher design matrix $\mathbf{D_c}$ will equal $\mathbf{0}$).

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 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 $\mathbf{X_i}$ 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.	0.052	0.042	0.090	0.083	0.049	0.032
$Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(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	0.181	0.162	0.183	0.175	0.095	0.077
Based on $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(0.025)	(0.015)	(0.015)	(0.019)	(0.009)	(0.009)
No Unobs. Sort.: 10th-50th	0.100	0.089	0.111	0.105	0.056	0.044
Based on $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(0.017)	(0.010)	(0.024)	(0.013)	(0.006)	(0.006)
Sample Mean	0.769	0.769	0.827	0.827	0.897	0.897

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 $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$, which captures the variance in true school/neighborhood contributions under the assumption that sorting is driven only by $\mathbf{X_i}$.

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 Online Appendix Table A20.

Online Appendices A9 and A10 discuss estimation of model parameters and the variance decompositions. Section 5.4 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 $\mathbf{X_i}$ 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.	0.062	0.046	0.076	0.071	0.068	0.043
$Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(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	0.216	0.184	0.261	0.246	0.264	0.204
Based on $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(0.017)	(0.016)	(0.018)	(0.016)	(0.018)	(0.015)
No Unobs. Sort.: 10th-50th	0.097	0.084	0.117	0.112	0.124	0.097
Based on $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)
Sample Mean	.267	.267	.310	.310	.365	.365

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 $\mathbf{X_s}$ 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	0.028	0.017	0.023	0.022	0.024	0.016
$Var(\mathbf{Z_{2s}G_2})$	(0.007)	(0.004)	(0.005)	(0.005)	(0.010)	(0.007)
LB w/ unobs	0.066	0.049	0.075	0.073	0.072	0.050
$Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.015)	(0.008)	(0.009)	(0.009)	(0.023)	(0.016)

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	0.141	0.111	0.138	0.136	0.152	0.124
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.013)	(0.011)	(0.014)	(0.013)	(0.014)	(0.013)
LB w/ unobs: 10th-90th	0.219	0.187	0.253	0.246	0.266	0.218
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.016)	(0.014)	(0.014)	(0.013)	(0.017)	(0.015)
LB no unobs: 10th-50th	0.066	0.052	0.065	0.064	0.073	0.060
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)
LB w/ unobs: 10th-50th	0.098	0.085	0.114	0.111	0.124	0.104
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.007)
Sample Mean	.267	.267	.310	.310	.365	.365

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

The notes to Table 2 apply, except that Table A3 reports results for enrollment in a 4-year college two years after graduation, and the naive OLS specification and estimates are used, as described in Section 7.5

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 $\mathbf{X_s}$ omitted from estimating equation)

Panel A: Fraction of Variance Attributable to School/Neighborhood Quality		
Lower Bound	Yrs. Postsec. Ed.	
	Baseline	Full
	(1)	(2)
LB no unobs	0.006	0.003
$Var(\mathbf{Z_{2s}G_2})$	(0.002)	(0.001)
LB w/ unobs	0.024	0.014
$Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.004)	(0.003)

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.	
	Baseline	Full
	(1)	(2)
LB no unobs: 10th-90th	0.353	0.227
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.052)	(0.045)
LB w/unobs: 10th-90th	0.679	0.526
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.059)	(0.056)
LB no unobs: 10th-50th	0.176	0.114
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.026)	(0.022)
LB w/unobs: 10th-50th	0.339	0.263
Based on $Var(\mathbf{Z_{2s}G_2}+v_s)$	(0.029)	(0.028)
Sample Mean	1.62	1.62

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

Panel A reports lower bound estimates of the fraction of variance of years of postsecondary education that can be directly attributed to school/neighborhood choices in NLS72.

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 $[\widehat{Var}(\mathbf{Z_{2s}G_2}+v_s)]^{0.5}$ or $[\widehat{Var}(\mathbf{Z_{2s}G_2})]^{0.5}$ in the corresponding column of the table.

Table A5: Principal Components Analysis of the Vector of School Average Observable Characteristics $\mathbf{X_s}$

Panel A: Fraction of Total Variance in $\mathbf{X_s}$ 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 $\mathbf{X_s}$	32	34	39	49	40	51
	# Factors Needed to Explain:						
(2)	75% of Total $\mathbf{X_s}$ Var.	7	7	7	9	6	8
		[7,8]	[8,8]	[7,8]	[8,9]	[6,7]	[7,8]
(3)	90% of Total $\mathbf{X_s}$ Var.	12	12	13	16	11	14
		[11,12]	[12,13]	[11,13]	[14,15]	[11,12]	[14,15]
(4)	95% of Total $\mathbf{X_s}$ Var.	15	15	17	20	14	19
		[14,15]	[14,15]	[14,16]	[18,19]	[14,15]	[17,19]
(5)	99% of Total $\mathbf{X_s}$ Var.	20	21	22	26	20	25
		[18,19]	[17,18]	[19,21]	[23,25]	[18,20]	[23,25]
(6)	100% of Total $\mathbf{X_s}$ Var.	24	25	27	32	26	33
		[21,23]	[18,19]	[23,26]	[29,31]	[23,25]	[29,31]

Panel B: Fraction of Variance in the Regression Index $\mathbf{X_s\hat{G}_1}$ 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 $\mathbf{X_s}$	32	34	39	49	40	51
	# Factors Needed to Explain:						
(2)	75% of $Var(\mathbf{X_s G_1})$	3	3	6	5	2	5
		[3,5]	[3,6]	[3,7]	[5,8]	[2,3]	[4,7]
(3)	90% of $Var(\mathbf{X_s G_1})$	8	7	10	10	5	11
		[5,9]	[5,10]	[6,11]	[9,14]	[3,7]	[8,14]
(4)	95% of $Var(\mathbf{X_s G_1})$	10	9	13	13	7	15
		[8,13]	[7,11]	[9,14]	[12,17]	[5,11]	[11,17]
(5)	99% of $Var(\mathbf{X_s G_1})$	14	15	19	20	14	22
		[13,17]	[10,15]	[13,19]	[19,24]	[11,16]	[17,23]
(6)	100% of $Var(\mathbf{X_s G_1})$	24	25	27	32	26	33
		[21,23]	[18,19]	[23,26]	[29,31]	[23,25]	[29,31]

See Online Appendix A3 for details. The numbers in brackets are bootstrapped 90% confidence interval estimates of the number of factors required to explain the variance fraction specified in a given row.

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

		Dataset (Number of Variables in $\mathbf{X_s}$)					
		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 $\mathbf{X_s}$ reveals the number of amenity factors driving sorting. See Online Appendix A3 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 $\mathbf{X_s}$ 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 $\mathbf{X_s}$ 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 ($dim(\mathbf{A_s})$): Kleibergen and Paap (2006) Heteroskedasticity-Robust and Cluster Robust Tests of the Rank of the $\mathbf{X_s}$ Covariance Matrix (Full Specification Results)

		Dataset (Number of Variables in $\mathbf{X_s}$)					
		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 $\mathbf{X_s}$ reveals the number of amenity factors driving sorting. See Online Appendix A3 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 $\mathbf{X_s}$ 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 $\mathbf{X_s}$ 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 ($\boldsymbol{\Theta}^U = \mathbf{R}\boldsymbol{\Theta}$ For Some $\mathbf{R}$)

Row	# Stu.	# Sch.	# Con.	# Ob.	# Un.	# Am.	$\rho_{\boldsymbol{\Theta}}$	$\frac{Var(\mathbf{X_s^U B^U})}{Var(Y)}$	Adj-R-Sq (All)	Resid (All)	Adj-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	0.9979	.0003	.869 .926 .959	.016 .009 .005
(2)	500	50	50	10	10	5	0.25	.123	0.9961	.0005	.869 .926 .960	.016 .009 .005
(3)	2000	50	50	10	10	5	0.25	.122	0.9989	.0001	.869 .925 .959	.016 .009 .005
(4)	1000	100	50	10	10	5	0.25	.122	0.9979	.0002	.868 .924 .959	.017 .009 .006
(5)	1000	50	10	10	10	5	0.25	.100	0.9976	.0001	.835 .908 .950	.016 .009 .005
(6)	1000	50	50	20	20	5	0.25	.122	0.9993	.0002	.897 .944 .971	.014 .007 .003
(7)	1000	50	50	10	10	10	0.25	.136	0.9933	.0009	.872 .923 .952	.018 .010 .006
(8)	1000	50	50	20	20	10	0.25	.135	0.9988	.0002	.909 .951 .973	.014 .007 .003
(9)	1000	50	50	10	10	5	0	.048	0.9941	.0003	.649 .779 .867	.016 .010 .006

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

$\rho_{\boldsymbol{\Theta}}$: Correlation in Θ_{lk} taste parameters across student characteristics for a given amenity and across amenities for a given student characteristic

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

Adj-R-sq (All): Fraction of between-school variance in unobservable student characteristics $\mathbf{X_s^U}\beta^\mathbf{U}$ explained by the control function $\overline{\mathbf{X}}_\mathbf{s}$ (sample averages of both $\mathbf{X_s}$ and $\mathbf{X_s^U}$ 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 $\mathbf{X_s}$, $[(1-Adj-R^2)Var(\mathbf{X_s^U}\beta^\mathbf{U})]/Var(Y_i)$ (sample averages of both $\mathbf{X_s}$ and $\mathbf{X_s^U}$ are computed using all students)

Adj-R-sq (10/20/40): Fraction of between-school variance in unobservable student characteristics $\mathbf{X_s^U}\beta^\mathbf{U}$ explained by the control function $\mathbf{X_s}$ (sample school averages of $\mathbf{X_s}$ are constructed using 10/20/40 students, while school averages of $\mathbf{X_s^U}$ are estimated 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 $\overline{\mathbf{X}}_\mathbf{s}$ (sample averages of $\mathbf{X_s}$ are computed using 10/20/40 students, while school averages of $\mathbf{X_s^U}$ are computed using all students.)

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

Row	$\mathbf{X}/\mathbf{X}^{\mathbf{U}}$ Corr. Structure	WTP for A1-A4 Depends On	WTP for A5 Depends On	Assu. (A5) Satisfied	$\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{A}}\boldsymbol{\Theta}$?	$\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{B}}\boldsymbol{\Pi}^{X^U X}\boldsymbol{\Theta}^U$?	$\frac{Var(\mathbf{X}_i^{\mathbf{U}}\mathbf{B}^{\mathbf{U}})}{Var(Y)}$	Adj-R-Sq (All)	Resid (All)	Adj-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 $\mathbf{X_i}$ and $\mathbf{X_i^U}$	All elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$	Yes	Yes	Yes	.122	0.998	.0003	.869 .926 .959	.016 .009 .005
(2)	Elements of $\mathbf{X^U}$ independent of elements of $\mathbf{X}$	All elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$	All elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$	Yes	No	Yes	.101	0.965	.0035	.492 .607 .717	.052 .040 .029
(3)	Corr = .25 for each pair of (obs. or unobs.) char.	All elements of $\mathbf{X_i}$	All elements of $\mathbf{X_i^U}$	Yes	Yes	No	.049	0.967	.0015	.622 .720 .795	.019 .014 .010
(4)	Elements of $\mathbf{X^U}$ independent of elements of $\mathbf{X}$	All elements of $\mathbf{X_i}$	All elements of $\mathbf{X_i^U}$	No	No	No	.069	0.148	.0591	.114 .128 .139	.062 .060 .059
(5)	Elements of $\mathbf{X^U}$ independent of elements of $\mathbf{X}$	All elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$	All elements of $\mathbf{X_i^U}$	No	No	No	.098	0.524	.0465	.325 .374 .421	.065 .062 .057
(6)	Elements of $\mathbf{X^U}$ independent of elements of $\mathbf{X}$	All elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$	$X_{i,10}^U$ only	No	No	No	.109	0.948	.0051	.589 .682 .760	.045 .035 .026
(7)	Elements of $\mathbf{X^U}$ independent of elements of $\mathbf{X}$	All obs. and unobs. char. except $X_{i,10}^U$	$X_{i,10}^U$ only	No	No	No	.095	0.952	.0050	.580 .677 .756	.040 .030 .023
(8)	Corr = .25 for each pair of obs. or unobs. char. except $X_{i,10}^U$ (independent)	All elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$	$X_{i,10}^U$ only	No	No	No	.117	0.997	.0003	.906 .947 .971	.013 .006 .004
(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	0.997	.0003	.893 .942 .969	.013 .006 .004

All specifications share the following parameter values: # Stu. = 1000, # Sch. = 50, # Con. = 50, # Ob = 10, # Un = 10, # Am. = 5, $\rho_{\boldsymbol{\Theta}} = 0.25$ (See Online Appendix Table A8 for definitions of parameters).

The column labeled "$\mathbf{X}/\mathbf{X}^{\mathbf{U}}$ Corr. Structure" describes the correlation structure among and between the elements of the vectors of observed and unobserved individual characteristics $\mathbf{X_i}$ and $\mathbf{X_i^U}$.

The columns labeled "WTP for A1-A4 Depends On" and "WTP for A5 Depends On" specifies which elements of the observable ($\mathbf{X_i}$) and unobservable ($\mathbf{X_i^U}$) characteristics predict willingness-to-pay for amenity factors 1-4 and amenity factor 5, respectively.

The columns labeled "Assu. A5 Satisfied", "$\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{A}}\boldsymbol{\Theta}$?", and "$\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{B}}\boldsymbol{\Pi}^{X^U X}\boldsymbol{\Theta}^U$?" specify whether the taste matrix $\boldsymbol{\Theta}^U$ can be written as $\boldsymbol{\Theta}^U = \mathbf{R}\tilde{\boldsymbol{\Theta}}$ (i.e. Assumption A5 is satisfied), $\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{A}}\boldsymbol{\Theta}$, and $\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{B}}\boldsymbol{\Pi}^{X^U X}\boldsymbol{\Theta}^U$, for some matrix (matrices) $\mathbf{R}$, $\mathbf{R}^{\mathbf{A}}$, and $\mathbf{R}^{\mathbf{B}}$, respectively. The condition $\boldsymbol{\Theta}^U = \mathbf{R}\tilde{\boldsymbol{\Theta}}$ for some matrix $\mathbf{R}$ (Assumption A5) is a necessary condition for Proposition 1 to hold, while the conditions $\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{A}}\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}^U = \mathbf{R}^{\mathbf{B}}\boldsymbol{\Pi}^{X^U X}\boldsymbol{\Theta}^U$ for some matrices $\mathbf{R}^{\mathbf{A}}$ and $\mathbf{R}^{\mathbf{B}}$ are each sufficient conditions for A5 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 - Baseline Specification

Panel A: Fractions of Total Outcome Variance

Row	Full NC Sample	NLS72	NELSg8	ELS2002
Within School:				
Total $Var(Y_{is} - Y_s)$	0.913	0.916	0.917	0.916
Observable Student-Level (Within): $Var((\mathbf{X_{si}} - \mathbf{X_s})\mathbf{B})$	0.123	0.119	0.119	0.118
Unobservable Student-Level (Within) $Var(v_{si} - v_s)$	0.790	0.797	0.798	0.798
Between School:				
Total $Var(Y_s)$	0.087	0.084	0.083	0.084
Observable Student-Level: $Var(\mathbf{X_s B})$	0.018	0.016	0.016	0.016
Student-Level/ School-Level Covariance $2 * Cov(\mathbf{X_s B}, \mathbf{X_s G_1} + \mathbf{Z_{2s} G_2})$	0.017	0.019	0.018	0.018
School-Avg. Student-Level/ School Char. Covariance $2 * Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2})$	-0.016	-0.007	-0.008	-0.008
School-Avg. Student-Level $Var(\mathbf{X_s G_1})$	0.017	0.009	0.010	0.010
School Char. $Var(\mathbf{Z_{2s} G_2})$	0.018	0.013	0.013	0.014
Unobservable School-Level $Var(v_s)$	0.033	0.033	0.032	0.033

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

Row	Full NC Sample	NLS72	NELSg8	ELS2002
LB no unobs $Var(\mathbf{Z_{2s} G_2})$	0.104	0.088	0.089	0.091
LB w/unobs $Var(\mathbf{Z_{2s} G_2} + v_s)$	0.178	0.167	0.167	0.169

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 Online Appendix Table A20.

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: 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 - Full Specification

Panel A: Fractions of Total Outcome Variance

Row	Full NC Sample	NLS72	NELSg8	ELS2002
Within School:				
Total $Var(Y_{is} - Y_s)$	0.913	0.919	0.919	0.918
Observable Student-Level (Within): $Var((\mathbf{X_{si}} - \mathbf{X_s})\mathbf{B})$	0.229	0.231	0.232	0.230
Unobservable Student-Level (Within) $Var(v_{si} - v_s)$	0.684	0.688	0.687	0.688
Between School:				
Total $Var(Y_s)$	0.087	0.081	0.081	0.082
Observable Student-Level: $Var(\mathbf{X_s B})$	0.033	0.032	0.032	0.033
Student-Level/ School-Level Covariance $2 * Cov(\mathbf{X_s B}, \mathbf{X_s G_1} + \mathbf{Z_{2s} G_2})$	0.012	0.014	0.012	0.012
School-Avg. Student-Level/ School Char. Covariance $2 * Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2})$	-0.007	-0.006	-0.006	-0.006
School-Avg. Student-Level $Var(\mathbf{X_s G_1})$	0.011	0.008	0.008	0.008
School Char. $Var(\mathbf{Z_{2s} G_2})$	0.010	0.010	0.009	0.010
Unobservable School-Level $Var(v_s)$	0.027	0.023	0.024	0.025

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

Row	Full NC Sample	NLS72	NELSg8	ELS2002
LB no unobs $Var(\mathbf{Z_{2s} G_2})$	0.079	0.076	0.075	0.078
LB w/unobs $Var(\mathbf{Z_{2s} G_2} + v_s)$	0.153	0.142	0.144	0.146

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 Online Appendix Table A20.

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 A12: Summary Statistics for Student Characteristics in NLS72

Variable	% Imputed	Mean	Std. Dev.
Student Demographics			
1(Female)	0.00	.505	.500
1(Black)	0.00	.088	.283
1(Hispanic)	0.00	.034	.181
1(Asian)	0.00	.010	.101
Student Ability			
Std. Math Score	0.00	.007	.997
Std. Reading Score	0.00	.005	.989
Student Behavior			
[None]			
Family Background Characteristics			
SES Index	0.00	-.028	1.01
Number of Siblings	2.90	2.81	2.04
1(Both Parents Present)	43.17	.754	.360
1(Mother, Male Guardian)	43.17	.020	.117
1(Mother Only Present)	43.17	.123	.272
1(Father Only Present)	43.17	.040	.162
Father's Years of Educ.	0.74	12.53	2.47
Mother's Years of Educ.	0.00	12.28	2.05
1(Mother's Ed. Missing)	0.00	.003	.057
Log(Family Income)	19.98	10.89	.661
1(Eng. Spoken at Home)	0.46	.920	.271
1(Home Environ. Index)	3.33	.112	1.25
1(No Religion)	0.00	.052	.222
1(Eastern Religion)	0.00	.041	.199
1(Jewish)	0.00	.023	.151
1(Catholic)	0.00	.313	.464
1(Oth. Christian Relig.)	0.00	.181	.385
1(Fath. Occ.: Service)	22.21	.106	.276
1(Fath. Occ.: Security/Military)	22.21	.050	.195
1(Fath. Occ.: Farmer/Laborer)	22.21	.309	.415
1(Fath. Occ.: Craftsman/Technician)	22.21	.214	.362
1(Fath. Occ.: Manager)	22.21	.126	.306
1(Fath. Occ.: Owner)	22.21	.067	.227
1(Fath. Occ.: Professional)	22.21	.125	.313
1(Moth. Occ.: Sales)	18.42	.035	.171
1(Moth. Occ.: Service)	18.42	.060	.216
1(Moth. Occ.: Clerical)	18.42	.147	.328
1(Moth. Occ.: Professional)	18.42	.088	.267
1(Moth. Occ.: Other)	18.42	.095	.267
Parental Beliefs/Desires			
[None]			
Outcomes			
1(Enrolled at a 4-Yr. Coll.)	0.00	.267	.442
Years of Postsec. Education	0.00	1.62	1.72
Log Wage (1979)	0.00	2.78	.451
Log Wage (1986)	0.00	2.98	.479
Log Wage (Pooled)	0.00	2.88	.475

The summary statistics reported above incorporate sample weights. See Appendix A12 for further details about these weights.

Table A13: Summary Statistics for School Characteristics in NLS72

Variable	% Imputed	Mean	Std. Dev.
School Characteristics (Treated as elements of $\mathbf{X_s}$)*			
% Minority Students	1.87	.146	.228
School Characteristics (Treated as elements of $\mathbf{Z_{2s}}$)*			
1(Catholic School)	3.52	.074	.259
1(Private School)	3.52	.004	.060
% of Teachers with Masters' Deg.	1.03	.412	.210
Teacher Turnover Rate	0.27	.082	.087
Total School Enrollment	0.86	1362	864
Student-to-Teacher Ratio	1.51	20.30	4.35
% of Minority Teachers	2.61	.070	.137
1(Tracking System Exists)	17.80	.761	.385
Age of School Building	1.32	20.83	16.84
Neighborhood Characteristics			
Distance to 4-Year College	4.61	18.70	24.99
Distance to Community College	4.64	18.12	25.02
1(South Region)	0.00	.282	.450
1(Midwest Region)	0.00	.296	.457
1(West Region)	0.00	.167	.373
1(Small Town)	0.00	.294	.456
1(Medium-Sized City)	0.00	.087	.282
1(Suburb of Medium-Sized City)	0.00	.054	.225
1(Large City)	0.00	.096	.295
1(Suburb of Large City)	0.00	.113	.316
1(Huge City)	0.00	.074	.262
1(Suburb of Huge City)	0.00	.087	.281

*School characteristics treated as elements of $\mathbf{X_s}$ 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.

The summary statistics reported above incorporate sample weights. See Appendix A12 for further details about these weights.

Table A14: Summary Statistics for Student Characteristics in NELS88

Variable	% Imputed	Mean	Std. Dev.
Student Demographics			
1(Female)	0.00	.503	.500
1(Black)	0.00	.127	.333
1(Hispanic)	0.00	.099	.299
1(Asian)	0.00	.033	.179
1(Immigrant)	6.80	.048	.205
Student Ability			
Std. Math Score (8th grd.)	0.00	.060	1.01
Std. Reading Score (8th grd.)	0.00	.061	1.00
Student Behavior			
Parent checks HW	0.36	.448	.496
# Weekly HW Hours	5.71	5.85	4.93
# Weekly Reading Hours	4.28	2.21	2.58
# Weekly TV Hours	14.15	22.09	10.20
1(Often Missing Pencil)	4.20	.221	.406
1(Fought at School)	1.45	.226	.415
Family Background Characteristics			
SES Index	0.00	.034	1.01
Number of Siblings	0.46	2.31	1.58
1(Both Parents Present)	0.84	.648	.476
1(Mother, Male Guardian)	0.00	.115	.319
1(Mother Only Present)	0.00	.149	.357
1(Father Only Present)	0.00	.053	.225
Father's Years of Educ.	6.38	13.24	2.92
Mother's Years of Educ.	0.00	12.91	2.32
1(Mother's Ed. Missing)	0.00	.024	.152
Log(Family Income)	9.67	10.87	.910
1(Eng. Spoken at Home)	0.87	.902	.295
1(Moth. Is Immigrant)	7.66	.113	.306
1(Fath. Is Immigrant)	8.62	.106	.296
1(Parents Married)	7.70	.776	.403
1(No Religion)	0.00	.023	.148
1(Eastern Religion)	0.00	.039	.193
1(Jewish)	0.00	.019	.138
1(Catholic)	0.00	.286	.452
1(Oth. Christian Relig.)	0.00	.072	.258
1(Home Environ. Index)	6.49	-.010	1.41
1(Fath. Occ.: Service)	24.39	.109	.267
1(Fath. Occ.: Security/Military)	24.39	.047	.183
1(Fath. Occ.: Farmer/Laborer)	24.39	.286	.403
1(Fath. Occ.: Craftsman/Technician)	24.39	.201	.344
1(Fath. Occ.: Dentist/Lawyer/Etc.)	24.39	.040	.207
1(Fath. Occ.: Accountant/Nurse/Etc.)	24.39	.093	.287
1(Fath. Occ.: Manager)	24.39	.120	.313
1(Fath. Occ.: Owner)	24.39	.076	.237
1(Moth. Occ.: Sales)	11.23	.055	.218
1(Moth. Occ.: Service)	11.23	.132	.320
1(Moth. Occ.: Clerical)	11.23	.231	.405
1(Moth. Occ.: Teacher)	11.23	.075	.258
1(Moth. Occ.: Accountant/Nurse/Etc.)	11.23	.090	.278
1(Moth. Occ.: Other)	11.23	.256	.410
Parental Sch. Engage. Index	10.79	-.079	1.46
Parental Beliefs/Desires			
Moth. Desired Educ. for Child	12.63	16.20	1.94
Fath. Desired Educ. for Child	16.09	16.13	1.94
Outcomes			
1(High School Graduate)	0.00	.827	.379
1(Enrolled at a 4-Yr. Coll.)	0.00	.310	.463

The summary statistics reported above incorporate sample weights. See Appendix A12 for further details about these weights.

Table A15: Summary Statistics for School Characteristics in NELS88

Variable	% Imputed	Mean	Std. Dev.
School Characteristics (Treated as elements of $\mathbf{X_s}$)*			
% Minority Students	1.51	.232	.300
% Limited English Proficient	1.31	.071	.090
% Receiving Free/Reduced Price Lunch	1.49	.243	.234
% in Special Ed.	1.31	.068	.058
% in Remedial Reading	1.19	.104	.127
% in Remedial Math	1.19	.081	.112
Admin's Perceived Sch. Problems Index	1.16	3.07	.671
School Characteristics (Treated as elements of $\mathbf{Z_{2s}}$)*			
1(Catholic School)	0.00	.076	.267
1(Private School)	0.00	.038	.190
% of Teachers with Masters' Deg.	3.75	.473	.246
Total School Enrollment	1.05	675.2	368.7
Student-to-Teacher Ratio	1.05	17.87	4.82
% of Minority Teachers	2.92	.118	.192
Log(Minimum Teacher Salary)	2.51	9.76	.188
1(Collectively Bargained Contracts)	1.49	.590	.491
1(Gifted Program Exists)	1.05	.693	.461
Admin.'s Reported Security. Policies Index (1)	1.36	.219	1.05
Admin.'s Reported Security. Policies Index (2)	1.36	-.046	1.03
Neighborhood Characteristics			
1(Urban Neighborhood)	0.00	.248	.432
1(Suburban Neighborhood)	0.00	.437	.496
1(South Region)	0.00	.358	.479
1(Midwest Region)	0.00	.260	.439
1(West Region)	0.00	.189	.391

*School characteristics treated as elements of $\mathbf{X_s}$ 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.

The summary statistics reported above incorporate sample weights. See Appendix A12 for further details about these weights.

Table A16: Summary Statistics for Student Characteristics in ELS2002

Variable	% Imputed	Mean	Std. Dev.
Student Demographics			
1(Female)	0.00	.506	.500
1(Black)	0.00	.137	.344
1(Hispanic)	0.00	.152	.359
1(Asian)	0.00	.037	.189
1(Immigrant)	10.78	.082	.256
Student Ability			
Std. Math Score	0.00	.038	1.01
Std. Reading Score	0.00	.036	1.01
Student Behavior			
Parent checks HW	14.43	.345	.440
# Weekly HW Hours	3.72	10.49	8.80
# Weekly Reading Hours	4.06	2.81	4.10
# Weekly Computer Hours	3.92	2.19	1.69
# Weekly TV Hours	4.01	23.21	11.98
1(Often Missing Pencil)	1.71	.172	.374
1(Fought at School)	0.85	.137	.342
Family Background Characteristics			
SES Index	0.00	.014	.997
Number of Siblings	17.22	2.34	1.39
1(Both Parents Present)	10.44	.571	.471
1(Mother, Male Guardian)	10.44	.131	.320
1(Mother Only Present)	10.44	.185	.367
1(Father Only Present)	10.44	.071	.235
Father's Years of Educ.	9.24	13.61	2.53
Mother's Years of Educ.	0.00	13.47	2.26
1(Mother's Ed. Missing)	0.00	.034	.182
Avg. Grandparents' Educ.	23.77	12.15	1.64
Log(Family Income)	21.01	10.87	.894
1(Eng. Spoken at Home)	13.32	.895	.286
1(Moth. Is Immigrant)	11.38	.176	.363
1(Fath. Is Immigrant)	12.33	.176	.363
1(Parents Married)	10.85	.723	.423
1(No Religion)	18.55	.033	.161
1(Eastern Religion)	18.55	.064	.215
1(Jewish)	18.55	.011	.091
1(Catholic)	18.55	.334	.437
1(Oth. Christian Relig.)	18.55	.199	.363
1(Home Environ. Index)	13.35	-.095	1.38
1(Fath. Occ.: Service)	30.74	.116	.259
1(Fath. Occ.: Security/Military)	30.74	.050	.177
1(Fath. Occ.: Farmer/Laborer)	30.74	.285	.377
1(Fath. Occ.: Craftsman/Technician)	30.74	.202	.328
1(Fath. Occ.: Dentist/Lawyer/Etc.)	30.74	.038	.197
1(Fath. Occ.: Accountant/Nurse/Etc.)	30.74	.103	.289
1(Fath. Occ.: Manager)	30.74	.149	.306
1(Fath. Occ.: Owner)	30.74	.051	.192
1(Fath. Occ.: Other)	30.74	.004	.047
1(Moth. Occ.: Sales)	21.10	.047	.187
1(Moth. Occ.: Service)	21.10	.158	.318
1(Moth. Occ.: Clerical)	21.10	.181	.342
1(Moth. Occ.: Teacher)	21.10	.070	.239
1(Moth. Occ.: Accountant etc.)	21.10	.148	.333
1(Moth. Occ.: Other)	21.10	.248	.378
Parental Sch. Engage. Index	20.71	-.141	1.34
Parental Beliefs/Desires			
Moth. Desired Educ. for Child	15.90	16.55	2.21
Fath. Desired Educ. for Child	23.04	16.48	2.20
Outcomes			
1(High School Graduate)	0.00	.897	.305
1(Enrolled at a 4-Yr. Coll.)	0.00	.365	.481

The summary statistics reported above incorporate sample weights. See Appendix A12 for further details about these weights.

Table A17: Summary Statistics for School Characteristics in ELS2002

Variable	% Imputed	Mean	Std. Dev.
School Characteristics (Treated as elements of $\mathbf{X_s}$)*			
% Minority Students	1.53	.338	.303
% Limited English Proficient	4.71	.047	.085
% Receiving Free/Reduced Price Lunch	7.68	.255	.234
% in Special Ed.	5.98	.104	.074
% in Remedial Reading	17.81	.049	.073
% in Remedial Math	19.24	.065	.089
Admin's Perceived Sch. Problems Index	15.74	3.46	.768
School Characteristics (Treated as elements of $\mathbf{Z_{2s}}$)*			
1(Catholic School)	1.23	.044	.205
1(Private School)	1.84	.032	.175
% of Teachers with Masters' Deg.	33.72	.450	.182
Teacher Turnover Rate	28.01	.056	.049
Total School Enrollment	0.34	1408	830
Student-to-Teacher Ratio	2.67	17.1	3.99
% of Minority Teachers	37.99	.137	.174
Log(Minimum Teacher Salary)	20.01	10.26	.155
% of Teachers with Certification	3.35	95.37	12.82
Teacher Evaluation Policy Index	14.42	-.141	.941
Teacher Incentive Pay Index (1)	13.25	.023	1.34
Teacher Incentive Pay Index (2)	13.25	-.086	1.06
Teaching Technology Index	16.29	.190	1.47
1(High Stakes Competency Exam)	0.00	.994	.077
Observed Sch. Cleanliness/Disorder Index (1)	29.85	.021	1.78
Observed Sch. Cleanliness/Disorder Index (2)	29.85	.030	1.18
Security Policy Implementation Index (1)	8.56	.073	1.34
Security Policy Implementation Index (2)	8.56	-.152	.934
Admin.'s Reported Security. Policies Index (1)	15.78	.157	1.48
Admin.'s Reported Security. Policies Index (2)	15.78	-.257	1.09
Admin.'s Impression of Fac. Quality Index (1)	19.31	.187	2.20
Admin.'s Impression of Fac. Quality Index (2)	19.31	.025	1.03
Neighborhood Characteristics			
1(Rural within MSA)	0.24	.108	.310
1(Small Town)	0.24	.103	.304
1(Large Town)	0.24	.014	.118
1(Suburb of Medium City)	0.24	.091	.288
1(Suburb of Large City)	0.24	.286	.452
1(Medium City)	0.24	.163	.369
1(Large City)	0.24	.133	.340
1(South Region)	0.00	.345	.476
1(Midwest Region)	0.00	.252	.434
1(West Region)	0.00	.220	.414
Admin. Perception of N-Hood Crime	12.24	2.93	.595

*School characteristics treated as elements of $\mathbf{X_s}$ 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.
The summary statistics reported above incorporate sample weights. See Appendix A12 for further details about these weights.

Table A18: 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 A19: Summary Statistics for School Characteristics in North Carolina Administrative Data

Variable	% Imputed	Mean	Std. Dev.
School Characteristics (Treated as elements of $\mathbf{Z_{2s}}$)			
# 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 A20: 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	0.913	0.913	0.812	0.812	0.891	0.904
$Var(Y_i - Y_s)$	(0.021)	(0.014)	(0.019)	(0.019)	(0.014)	(0.015)
Observable Student-Level (Within):	0.123	0.229	0.145	0.201	0.125	0.213
$Var((\mathbf{X_i} - \mathbf{X_s})\mathbf{B})$	(0.004)	(0.005)	(0.013)	(0.014)	(0.012)	(0.015)
Unobservable Student-Level (Within)	0.790	0.684	0.667	0.611	0.767	0.690
$Var(v_{si} - v_s)$	(0.018)	(0.010)	(0.020)	(0.019)	(0.016)	(0.017)
Between School:						
Total	0.087	0.087	0.188	0.188	0.109	0.097
$Var(Y_s)$	(0.021)	(0.014)	(0.019)	(0.019)	(0.014)	(0.015)
Observable Student-Level:	0.018	0.033	0.057	0.069	0.031	0.054
$Var(\mathbf{X_s B})$	(0.001)	(0.003)	(0.012)	(0.012)	(0.005)	(0.007)
Student-Level/ School-Level Covariance	0.017	0.012	0.042	0.035	0.029	0.010
$2 * Cov(\mathbf{X_s B}, \mathbf{X_s G_1} + \mathbf{Z_{2s} G_2})$	(0.004)	(0.005)	(0.018)	(0.019)	(0.010)	(0.013)
School-Avg. Student-Level/ School Char. Covariance	-0.016	-0.007	0.016	0.015	0.007	0.006
$2 * Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2})$	(0.011)	(0.005)	(0.007)	(0.007)	(0.011)	(0.011)
School-Avg. Student-Level	0.017	0.011	0.025	0.024	0.010	0.006
$Var(\mathbf{X_s G_1})$	(0.010)	(0.005)	(0.010)	(0.007)	(0.012)	(0.011)
School Char.	0.018	0.010	0.011	0.006	0.012	0.009
$Var(\mathbf{Z_{2s} G_2})$	(0.008)	(0.004)	(0.008)	(0.007)	(0.010)	(0.009)
Unobservable School-Level	0.033	0.027	0.038	0.038	0.023	0.012
$Var(v_s)$	(0.013)	(0.008)	(0.008)	(0.008)	(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.

Online Appendices A9 and A10 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 Table 1

Table A21: 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	0.861	0.861	0.775	0.777	0.786	0.793
$Var(Y_{is} - Y_s)$	(0.012)	(0.012)	(0.017)	(0.017)	(0.069)	(0.066)
Observable Student-Level (Within):	0.170	0.354	0.170	0.261	0.180	0.327
$Var((\mathbf{X_{si}} - \mathbf{X_s})\mathbf{B})$	(0.010)	(0.012)	(0.012)	(0.013)	(0.257)	(0.201)
Unobservable Student-Level (Within)	0.691	0.507	0.606	0.517	0.606	0.466
$Var(v_{si} - v_s)$	(0.014)	(0.013)	(0.018)	(0.016)	(0.189)	(0.137)
Between School:						
Total	0.139	0.139	0.225	0.223	0.214	0.207
$Var(Y_s)$	(0.012)	(0.012)	(0.017)	(0.017)	(0.069)	(0.066)
Observable Student-Level:	0.041	0.062	0.071	0.104	0.073	0.122
$Var(\mathbf{X_s B})$	(0.005)	(0.006)	(0.015)	(0.015)	(0.025)	(0.040)
Student-Level/ School-Level Covariance	0.036	0.032	0.078	0.047	0.072	0.042
$2 * Cov(\mathbf{X_s B}, \mathbf{X_s G_1} + \mathbf{Z_{2s} G_2})$	(0.007)	(0.009)	(0.014)	(0.016)	(0.024)	(0.016)
School-Avg. Student-Level/ School Char. Covariance	-0.003	-0.004	0.006	0.008	-0.001	-0.001
$2 * Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2})$	(0.006)	(0.005)	(0.006)	(0.005)	(0.008)	(0.006)
School-Avg. Student-Level	0.018	0.012	0.021	0.025	0.017	0.007
$Var(\mathbf{X_s G_1})$	(0.006)	(0.005)	(0.006)	(0.006)	(0.011)	(0.007)
School Char.	0.027	0.018	0.017	0.015	0.019	0.014
$Var(\mathbf{Z_{2s} G_2})$	(0.006)	(0.005)	(0.006)	(0.005)	(0.010)	(0.008)
Unobservable School-Level	0.021	0.020	0.032	0.024	0.033	0.023
$Var(v_s)$	(0.006)	(0.006)	(0.006)	(0.005)	(0.008)	(0.005)

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 Online Appendix Table A20 for additional details.

Table A22: 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	0.899	0.901	0.831	0.826	0.844	0.839
$Var(Y_{is} - Y_s)$	(0.008)	(0.008)	(0.021)	(0.021)	(0.024)	(0.012)
Observable Student-Level (Within):	0.146	0.271	0.138	0.184	0.113	0.137
$Var((\mathbf{X_{si}} - \mathbf{X_s})\mathbf{B})$	(0.006)	(0.007)	(0.012)	(0.014)	(0.012)	(0.012)
Unobservable Student-Level (Within)	0.753	0.630	0.694	0.642	0.731	0.703
$Var(v_{si} - v_s)$	(0.009)	(0.008)	(0.023)	(0.024)	(0.026)	(0.013)
Between School:						
Total	0.101	0.099	0.169	0.174	0.156	0.161
$Var(Y_s)$	(0.008)	(0.008)	(0.021)	(0.021)	(0.024)	(0.012)
Observable Student-Level:	0.038	0.056	0.044	0.056	0.029	0.036
$Var(\mathbf{X_s B})$	(0.003)	(0.004)	(0.006)	(0.007)	(0.005)	(0.006)
Student-Level/ School-Level Covariance	0.036	0.027	0.031	0.029	0.020	0.021
$2 * Cov(\mathbf{X_s B}, \mathbf{X_s G_1} + \mathbf{Z_{2s} G_2})$	(0.005)	(0.006)	(0.009)	(0.010)	(0.009)	(0.009)
School-Avg. Student-Level/ School Char. Covariance	0.001	0.001	0.005	0.007	0.003	0.007
$2 * Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2})$	(0.002)	(0.002)	(0.010)	(0.009)	(0.013)	(0.011)
School-Avg. Student-Level	0.013	0.007	0.019	0.013	0.018	0.015
$Var(\mathbf{X_s G_1})$	(0.004)	(0.002)	(0.011)	(0.009)	(0.013)	(0.005)
School Char.	0.002	0.000	0.025	0.028	0.032	0.033
$Var(\mathbf{Z_{2s} G_2})$	(0.002)	(0.001)	(0.010)	(0.011)	(0.012)	(0.005)
Unobservable School-Level	0.011	0.009	0.045	0.042	0.053	0.048
$Var(v_s)$	(0.003)	(0.002)	(0.016)	(0.015)	(0.019)	(0.006)

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 Online Appendix Table A20 for additional details.

Table A23: Potential Bias from Violations of Assumption 6.2

ρ	Maximum Bias $\max_\mu \mu + 2\rho\sqrt{\mu}$	μ:Max. Bias $\arg\max_\mu \mu + 2\rho\sqrt{\mu}$	μ:Zero Bias $\mu_0(\rho)$
-0.1	-0.01	0.01	0.04
-0.2	-0.04	0.04	0.16
-0.3	-0.09	0.09	0.36
-0.4	-0.16	0.16	0.64
-0.5	-0.25	0.25	1
-0.6	-0.36	0.36	1.45
-0.7	-0.49	0.49	1.96
-0.8	-0.64	0.64	2.56
-0.9	-0.81	0.81	3.23
-1	-1	1	4