Quantifying Family, School, and Location Effects in the Presence of Complementarities and Sorting

Mohit Agrawal
Yale University

Joseph G. Altonji
Yale University and NBER

Richard K. Mansfield
University of Colorado-Boulder and NBER *

January 20, 2018

Abstract

We extend the control function approach of Altonji and Mansfield (2016) for bounding the variance of group treatment effects to allow for multiple group levels and complementarities between individual and group-level variables. Our analysis provides a foundation for causal interpretation of multilevel mixed effects models in the presence of sorting. We apply the methods using the NELS88 and ELS02 and obtain lower bound estimates of the importance of school and commuting zone inputs during adolescence for determining long-run educational attainment and wages. Experiencing a school/location combination at the 90th versus 10th percentile of the school/location treatment effect distribution increases the high school graduation probability and college enrollment probability for a randomly selected student by at least .05 and .16, respectively, with dropout rates for disadvantaged students and college attendance for average and advantaged students exhibiting particular sensitivity to school/location inputs. Heterogeneity in effects is primarily due to nonlinearity in the educational attainment model rather than interactions between student characteristics and school and location characteristics.

*Agrawal: Department of Economics, Yale University, PO Box 208264, New Haven CT 06520-8264, mohit.agrawal@yale.edu. Altonji: Department of Economics, Yale University, joseph.altonji@yale.edu. Mansfield: Department of Economics, University of Colorado, Boulder. richard.mansfield@colorado.edu. We thank David Card, Mel Stevens, participants in the NBER pre-conference and conference and participants in seminars at EIEF, Oxford, Princeton, the University of Chicago Harris School and Yale for many valuable comments. The shortcomings of the paper are our responsibility.

1 Introduction

What drives differences in educational attainment and labor market success? Research since Coleman et al. (1966) and Jencks and Brown (1975) indicates that the most of the variation is attributable to family-level influences associated with parents or to idiosyncratic person level influences. However, research and introspection suggest that broader environmental factors associated with neighborhoods, schools, and regions also play a role. Given the prevalence of sorting, neighborhood-, school-, and regional-level determinants of success could potentially be an important source of disparities in adult outcomes along the lines of race, ethnicity, family structure, parental education and parental income.

In this paper, we assess the relative contributions of combined student/family-level influences, neighborhood influences, school influences, and regional influences to students' educational attainment and early career wages, building on a vast literature. We consider both observed and unobserved factors. And we allow the effects of neighborhood, school, and regional influences to depend upon student characteristics.

Designers of policy interventions often must choose whether to target neighborhoods, schools, or broader local or regional areas. Policies aimed at reducing the availability of drugs, poverty, crime and violence, and at improving the physical environment and providing recreational opportunities are frequently targeted to neighborhoods. The Harlem Children's Zone, evaluated by Dobbie et al. (2011) is an example of a community-focused approach aimed at providing a positive social environment outside of school.

Policies promoting access to good schools take two main forms: policies designed to improve schools directly, such as curriculum reform, increased spending, accountability systems, and changes in personnel practices, or policies that seek to improve access to good schools by broadening school choice. These include school busing programs, private school vouchers, open enrollment, and charter schools.

Similarly, regional policy influences economic development and access to higher education. The availability of colleges and universities and efforts to stimulate local labor demand may have important effects on educational attainment and wages.

Policy makers wish to know which policies with which target areas or locations are most effective, but research has been very slow in providing answers Katz (2015). We focus on the more modest goal of measuring the overall importance of neighborhoods, schools, and local areas for adult outcomes. This provides information about where inequality arises, and therefore guidance about where the most potential progress can be made, even if it does not provide a specific plan for how to reduce it.

The efficacy of particular policies and the relative importance of neighborhood vs. school. vs. broader local area inputs more generally is likely to vary across the student population. For example, school-level dropout prevention policies may improve outcomes for struggling students but negligibly affect more mature or better-supported students, while opening a satellite campus of a flagship research university may improve attainment primarily for high achieving but low income students. Such considerations motivate us to work with a model that permits the relative sensitivity of particular subpopulations to external inputs to vary across external environments (i.e. neighborhoods/schools/local areas).

We are far from the first to consider a model in which individual characteristics interact with observed and unobserved school and area characteristics. Indeed, the spread of multilevel mixed effects modeling (aka., hierarchical linear models and random coefficients models) was greatly stimulated by interest in the interplay among the student, classroom, and school factors that determine educational outcomes.[1] We estimate a multilevel mixed effects specification of outcomes in this paper.

Perhaps the single most difficult challenge in assessing the importance of school and location for the success of children is that families sort into them. As Durlauf (2004) and Graham (2016) emphasize, differences in outcomes across neighborhoods, for example, combine the causal effects of neighborhood with differences across neighborhoods in the observed and unobserved attributes of children that matter for their outcomes. The same issue arises in assessing schools. The multilevel modeling literature is mindful of the potential for bias posed by sorting across schools and location, but has not addressed it.

Our main methodological contribution is to show that, even in the presence of endogenous sorting of individuals to groups, multilevel mixed effects estimates still contain sufficient information to distinguish among four sources of the variation in outcomes. The sources are individual contributions that are common across groups, group contributions that are common across individuals, contributions that consist of interactions between student and group inputs, and a set of ambiguous contributions that either reflect common group inputs or group-averages of individual inputs. We then show that this decomposition is sufficient to generate meaningful (albeit conservative) answers to two key questions. How much do the schools and areas we choose for our children matter for their outcomes? And how much does the importance of schools and locations depend on the students' and parents' own inputs?

To address the sorting problem, we build upon Altonji and Mansfield (2016) (hereafter, AM). AM work with a fairly standard model of choice of school/neighborhood combinations based upon willingness to pay for the attributes associated with each choice (air quality, noise, access to transportation, crime, school quality, etc.). They show that the choice model implies a relationship between school/neighborhood averages of observed student characteristics and averages of unobserved student characteristics. Consequently, school/neighborhood averages of observed student characteristics can serve as controls for the averages of unobserved student characteristics that would otherwise bias estimates of the importance of school and neighborhood effects. Our methodological contribution is to extend AM's identification results to allow for interactions between student-level

[1] Garner and Raudenbush (1991) is a good example. The early editions of Goldstein (2011) and Raudenbush and Bryk (2002) became key references for empirical researchers.

characteristics and both observed and unobserved group-level (i.e. neighborhood/school/regional) characteristics.

A precise statement of our theoretical results and the assumptions they require is best postponed until later in the paper, but they may be summarized as follows. Roughly speaking, we study the coefficient vector $\mathbf{r_2}$ on the interaction between a student characteristic, say mother's education, and a set of observed "group" characteristics, where "group" may refer to the neighborhood, the school, and/or the local area. We maintain the assumptions justifying AM's control function approach (reviewed below) and add additional assumptions.[2] Under these assumptions, we show that $\mathbf{r_2}$ is the sum of two components. The first is the causal effect of the interaction between mother's education and the observed group characteristics. The second is the effect of the interaction between mother's education and the part of the contribution of *unobserved* group characteristics that covaries with the observed group characteristic, holding student characteristics fixed. The key point is that the estimates of $\mathbf{r_2}$ reflect interactions between the student and the environment and not sorting.

We also study the coefficient $\mathbf{r_1}$ on the interaction between a student's characteristic and both observed and unobserved peer characteristics that matter for the outcome. We show that $\mathbf{r_1}$ is the sum of the causal effects of (1) interactions between the student's characteristic and observed and unobserved peer characteristics and (2) interactions between the student's characteristic and unobserved school and location characteristics, such as the quality of a principal.

What about the main effects of school and location on student outcomes? We extend AM's results and show that even in the presence of endogenous sorting and interactions involving unobserved school and area characteristics, the coefficient on observed school and area characteristics (referred to as ($\mathbf{G_2}$ below) captures causal effects and is not contaminated by sorting. It picks up the causal effect of the school and area characteristics plus a second component that reflects their association with unobserved school and area characteristics that affect outcomes.

Finally, we consider the error components in the model. We show that AM's result that school and neighborhood error components are not contaminated by the effects of sorting carries over to the model with interaction effects, but only under somewhat stronger assumptions. We also justify a causal interpretation for a random slope coefficient that captures the interaction of an observed student characteristic with an index of unobserved school and area characteristics.

As in AM's analysis, the use of school/location averages of observed student characteristics to control for school/location averages of unobserved characteristics is likely to lead to an understatement of the importance of school and neighborhood. The main reason is that by treating the averages as controls that absorb sorting bias, we are discarding the main effects of peers on outcomes. A second reason is that the group averages will absorb a portion of the unobserved school or area components that are uncorrelated with the observed components and are correlated with the amenities that families sort on. Consequently, our analysis only places a lower bound on the overall

[2] We delay discussion of the assumptions, but the most important additional one is that the within group covariance between student characteristics that influence treatment effect heterogeneity (mother's education in the example) and other outcome relevant student characteristics do not vary with outcome relevant school and location characteristics.

importance of school and area factors for student outcomes.

Using the above analysis as the foundation, we present and interpret empirical mixed effects estimates from two cohort-specific panel surveys, the National Educational Longitudinal Survey of 1988 (NELS88) and the Educational Longitudinal Survey of 2002 (ELS2002). These data sets provide a rich set of student characteristics for samples of students from each of a large sample of schools. Students are followed for several years, which permits investigation of longer run outcomes such as high school graduation, attendance at a four year college, and attainment of a college degree. In addition, they contain location of residence and school identifiers. Specifically, NELS88 identifies the ZIP code of residence for students attending the same 8th grade. The ELS2002 data contain the block group of residence when the student is in 10th grade. Using the ELS data we experiment with both ZIP code and block group as the definition of "neighborhood".

We restrict attention primarily to interactions between regression indices of student level and school, neighborhood or commuting zone level variables, but also include a small number of interactions involving particular student characteristics, such as underrepresented minority status. This is due to a need for parsimony, but also reflects an absence of prior consensus from the literature about which school and location variables matter most, let alone matter differentially by student type. Even with a small set of interactions, estimating a non-linear mixed effects model with four levels places heavy demands on our data, particularly given limited samples of students per block group and school. Consequently, some of our estimates are imprecise. However, the empirical approach we demonstrate in this paper could be implemented more richly with the type of linked administrative data that is currently being developed.

Our estimates of the effects on education outcomes of interactions between observed student characteristics and either school and area composition ($\mathbf{r_1}$ above) or non-compositional direct school and area characteristics ($\mathbf{r_2}$) are typically small and imprecise. They do not provide strong evidence that the latent variable governing educational attainment depends on the interaction between observed student characteristics and characteristics of the school and commuting zone. We do find that important differences in treatment effects exist across students, but they are primarily due to the fact that the probit function that we use to map the latent index into the outcome probability for binary outcomes naturally features treatment effect heterogeneity. The binary outcomes of students with characteristics suggesting they are close to the decision margin are more sensitive to the school and location environment they experience.

We use a version of the model without interaction terms to separately quantify the degree of sorting at the immediate neighborhood, the school and associated broader neighborhood, and the region (commuting zone) levels and perform variance decompositions of alternative educational attainment measures as well as log wages at age 25. Our four-level commuting zone, school, neighborhood, and individual variance decompositions are consistent with prior work based on two level models featuring individual and school (Coleman et al. (1966), Jencks and Brown (1975), Betts (1995), Altonji and Mansfield (2011) and Alexander and Morgan (2016)) or individuals and neighborhood (Solon et al. (2000)) in that they indicate that the individual-level factors dominate. However, our decomposition provides further insights about the relative importance of the four levels. For example, in ELS for college attendance the within-neighborhood (block group) share is 73 percent, while the neighborhood share is 2.7 percent, the school share is 14.8 percent, and the commuting zone share is 9.8 percent. Attaining a 4 year college degree and high school graduation both features a somewhat larger within-neighborhood component and smaller neighborhood, school and commuting zone shares. Commuting zone accounts for about 7 percent of the variance in wages, and block group within a school accounts for about 3%. Within a school there is very little clustering at the zipcode level. We also show that segregation by student quality (defined by an index of characteristics that promote education and wages) is primarily across schools and commuting zones rather then within zipcode or block group among schoolmates.

The key product of our empirical analysis uses the multilevel model estimates to measure the consequences of exposure to a low-quality school versus a high-quality school, and a low-quality commuting zone versus a high-quality commuting zone. Here we define "quality" narrowly as attributes that contribute to the outcomes we consider (educational attainment or wages). Our main results concern the "treatment effects" of 10th-to-90th percentile shifts in school quality, in commuting zone quality, and in combined school and commuting zone quality. First, we consider average effects. We take the average over the student population of the effect of a 10th-to-90th percentile shift in the combined school/commuting zone environment. Such a shift increases the high school graduation rate by 8.2 percentage points in NELS88 and by about 5.2 percent in ELS2002 (which has a higher baseline graduation rate–92%). The value for the school treatment is 6.1 points in NELS and about 3.2 points in ELS. The values for the 10th-to-90th quantile shift in commuting zone quality are between 3.8 and 5.5 percentage points.

For college enrollment, the combined school and commuting zone treatment effect is 17.9 percentage points for NELS and about 15 for ELS. The values for the school treatment effect are about 13 points, while the values for the commuting zone treatment are 11.4 points for NELS and about 8.5 points for ELS. The effects of the combined treatment, the school treatment and the commuting zone treatment on attainment of a BA degree are also large in NELS: 11.9, 10.1, and 9.2 percentage points, respectively. They are smaller in ELS. For wage rates at age 25, the combined school and commuting zone treatment ranges between 9.3 and 12.9 log points. The values for the school treatment range from 5 to 8.9 points. The values for the commuting zone only treatment range from 6.8 to 11.3 points. Overall, the results suggest that large changes in school and commuting zone inputs can make a substantial difference for students' educational attainment and wage rates.

We also present estimates of school and commuting zone "treatment" effects for particular subpopulations: blacks, Hispanics, low income families, whites with two resident parents with college degrees, and students at each percentile of an index of student level variables that predict the outcomes. As noted above, the dropout rates of subpopulations that tend to be disadvantaged are particularly sensitive to the external environment, while few students from advantaged subpopulations are near the margin. For college enrollment and college graduation, superior school and commuting zone inputs are important for all, but particularly important for students near or above the middle of the distribution of student and family background.

Our paper builds on and contributes to several literatures. The model of school and location choice that we use to address sorting on unobserved student characteristics is essentially that of AM, although we are more specific about the distinction between neighborhood choice, school choice, and region.[3] AM in turn draw on the rich theoretical and empirical literature on equilibrium sorting and matching across several fields, including marriage (Browning et al. (2014)), firms and workers (Lise et al. (2013), Lindenlaub (2017), Rosen (1974) and Ekeland et al. (2004)), and papers on sorting across neighborhoods and schools following Tiebout (1956). Particularly noteworthy are Epple and Platt (1998) and Epple and Sieg (1999). Epple and Sieg (1999) study a model with one dimension of neighborhood quality and two dimensions of heterogeneity across households–income and tastes for a public good. In the equilibrium of their model 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. This is the same type of mapping that exploited here.

Control function approaches appear in number of settings.[4] We extend AM's use of the control function to outcome models that include interactions of student characteristics with both observed and unobserved school and area characteristics.

On the empirical side, the literatures on the importance of families, neighborhoods, schools, and region are each too vast to discuss meaningfully here. We have already mentioned that a number of papers, like ours, descend from the Coleman Report's (Coleman et al. (1966)) examination of the importance of family background, peer characteristics, and school inputs using data with a multi-level structure similar to the NELS88 and ELS2002.[5] They found, in keeping with most subsequent research, that family background is by far the most important determinant of education success. Jencks and Brown (1975)), Betts (1995), Altonji and Mansfield (2011) are contributions to the large literature in economics, sociology, and education that performs variance decompositions separating the contribution of school- or neighborhood-level factors versus student-level factors for test scores, educational attainments, and in a few cases wage rates. Focusing on neighborhood, 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), Durlauf and Ioannides (2010), Harding et al. (2011), Sharkey and Faber (2014), and Graham (2016).[6] Many of the papers emphasize that estimates of the impacts of particular characteristics, such as percent minority, segregation measures, poverty rates, and income per capita on socioeco-

[3] The multinomial choice formulation that we use to characterize the school/location choice problem with heterogenous preferences is standard in the consumer choice literature, drawing on McFadden (1984), Berry (1994) and many subsequent papers, including Bayer et al. (2007)'s study of the demand for housing and location.

[4] Examples include the estimation of firm production functions (e.g., Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015)), 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)).

[5] See Alexander and Morgan (2016) and Lucas (2016) for a recent discussions of the legacy of the Coleman report and its influence on subsequent research.

[6] Kline and Moretti (2014) provide a recent analysis of placed based policies.

nomic outcomes of children, often by racial group and poverty status, are subject to the problem of sorting bias. A good example is Card and Rothstein (2007), who carefully address the issue in their analysis of the effects of segregation on the test score gap between blacks and whites.

Duncan and Murnane (2011) contains several recent papers on school and/or neighborhood effects, with references to the literature. Meghir et al. (2011) discuss alternative approaches to estimating school fixed effects and the effects of particular school inputs, and highlight the problem of endogenous selection of schools and neighborhoods, among other econometric issues.

There is also a small but growing experimental or quasi-experimental literature that attempts to measure neighborhood or school effects. 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. They obtain small estimates, in contrast to Chyn (2016). A number of authors, including Kling et al. (2007), use the Moving to Opportunity experiment to study effects of growing up in a low poverty neighborhood. Most of the work finds small effects. However, Chetty et al. (2016) revisit the MTO experiment using Internal Revenue Service data on young adult outcomes, including earnings, college attendance, and single parenthood. They find substantial effects for children who move before age 13 but not for older children. 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.

Using a sibling differences approach that also exploits high quality data from tax records, Chetty and Hendren (2015) identify county-level and commuting zone 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.[7]

Bergman (2016) finds that a lottery based opportunity to transfer from a predominantly minority school district to a high income predominantly white district increases college enrollment by 10 percentage points. Deming et al. (2014) exploit randomized lottery outcomes from the school choice plan in the Charlotte-Mecklenburg district and find large effects of attending a chosen public school on high school graduation, college enrollment, and college completion for students coming from low quality urban schools. Angrist et al. (2016) also use admissions lotteries. They find positive effects of attending a Boston charter high school on performance on both high and low stakes and on 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 do not find an effect on the high school graduation probability. Dobbie et al. (2011) is a rare quasi-experimental attempt to distinguish neighborhood effects from school effects. They find that the Harlem Children's Zone has little effect on test scores, while Promise Academy Charter schools have large effects that do not vary with residence in the zone.

[7]Using a sibling difference design, Altonji and Dunn (1996) find that teacher salaries and teacher/student ratios (considered separately) increase wage rates.

The paper continues in Section 2 which presents the model of location and school choice and AM's control function result. Section 3 presents the model of outcomes used in the paper. Section 4 specifies the estimating equation used in the paper and establishes key identification results relating slope parameters and error terms of the estimating equation to the parameters of the underlying production function. Section 5 discusses the data and variables used in the study. Section 6 presents the estimation methodology. Section 7 presents the model estimates and the variance decompositions that are derived from them. Section 8 presents lower bound estimates of treatment effects of shifts in school and commuting zone quality, both on average and for particular subpopulations. Section 9 concludes.

2 A Multinomial Model of School and Location Choice and Sorting

In this section we present AM's model of how families choose school systems and associated neighborhoods, recasted slightly to more explicitly consider neighborhood, school, and labor market area.[8] The presentation draws very heavily on AM, with small sections verbatim. We repeat some of the discussion here so that the assumptions required for identification of the main effects of school and location characteristics as well as their interactions with student characteristics (not considered in AM) will be clear. Throughout the paper, matrices, vectors, and matrix or vector valued functions are in bold. The "prime" symbol denotes matrix or vector transposes.

Parents choose a neighborhood n from the set $\mathscr{N}$ of neighborhoods. Due to attendance boundaries and travel costs, the choice of n then restricts the choice of school s to the subset $\mathscr{S}(n)$ of the full set $\mathscr{S}$. Since neighborhoods are embedded in labor market areas, the choice of neighborhood also implicitly involves the choice of a commuting zone c. In our main specification we assume parents choose among all available (n,s) pairs.

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

$$\mathscr{U}_i(ns) = \mathbf{W_i A_{ns}} + \varepsilon_{nsi} - P_{ns}. \tag{1}$$

In the above equation $\mathbf{A_{ns}} \equiv [A_{1ns}, \dots, A_{Kns}]'$ is a $K \times 1$ vector of underlying latent amenities that characterize the neighborhood n, the school s and the commuting zone c that (n,s) is associated with. We exclude the commuting zone subscript c because commuting zone is fully determined by n. $\mathbf{W_i} \equiv [W_{1i}, \dots, W_{Ki}]$ is a $1 \times K$ vector of weights that captures the increases in family i's willingness to pay for a neighborhood and school per unit increase in each of its K amenity factors $A_{1ns}, \dots, A_{Kns}$, respectively. P_{ns} is the price of living in n plus a fixed utility cost associated with the logistics of attending school s from neighborhood n. The component ε_{nsi} is an idiosyncratic taste of the parent/student i for the particular location and school (n,s).

[8] AM model the choice of school attendance area, defining the choices of school and neighborhood to be synonymous.

Next we specify a linear relationship between willingness to pay (hereafter denoted WTP) for particular amenities across parent/student combinations and the families' observable ($\mathbf{X_i}$) and unobservable ($\mathbf{X_i^U}$) characteristics that affect the school outcome (say college attendance) and an additional set of variables $\mathbf{Q}_i$:

$$\mathbf{W_i} = \mathbf{X_i}\boldsymbol{\Theta} + \mathbf{X_i^U}\boldsymbol{\Theta}^\mathbf{U} + \mathbf{Q_i}\boldsymbol{\Theta}^\mathbf{Q} \ . \tag{2}$$

Here $\mathbf{X_i}$ has L elements and $\mathbf{X_i^U}$ has L^U elements. The coefficient matrix $\boldsymbol{\Theta}$ ($\boldsymbol{\Theta}^\mathbf{U}$) 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 $\mathbf{A_{ns}}$ varies with the ℓ-th element of $\mathbf{X_i}$ ($\mathbf{X_i^U}$). We sometimes refer to the elements of $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}^\mathbf{U}$, and $\boldsymbol{\Theta}^\mathbf{Q}$ as WTP coefficients. The $1 \times L^Q$ vector $\mathbf{Q_i}$ captures the components of i's taste for the K amenities in $\mathbf{A_{ns}}$ that are uncorrelated with $[\mathbf{X_i}, \mathbf{X_i^U}]$. Below we define $\mathbf{X_i^U}$ so that $[\mathbf{X_i}, \mathbf{X_i^U}]$ represents the complete set of student attributes that determine Y_{si}, so the elements of $\mathbf{Q_i}$ influence school choice but have no direct effect on student outcomes.

Substituting equation (2) into equation (1), we obtain:

$$\mathscr{U}_i(ns) = (\mathbf{X_i}\boldsymbol{\Theta} + \mathbf{X_i^U}\boldsymbol{\Theta}^\mathbf{U} + \mathbf{Q_i}\boldsymbol{\Theta}^\mathbf{Q})\mathbf{A_{ns}} + \varepsilon_{nsi} - P_{ns} \tag{3}$$

As AM discuss, this formulation of utility allows for a fairly 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 (1).

Expected utility is taken with respect to the information available when n and s is chosen. The information set includes the price and the amenity vector in each school and neighborhood as well as student/parent characteristics $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$ and the values of ε_{nsi}, where $n \in \mathscr{N}$ and $s \in \mathscr{S}(n)$. The information set excludes any local shocks that are determined after the start of secondary school. It also excludes components of neighborhood, school quality and commuting zone quality that are not observable to families when a location is chosen. Some of the elements of $\mathbf{A_{ns}}$ may depend on school and neighborhood characteristics that influence educational attainment and labor market outcomes (denoted $\mathbf{Z_n^{N*}}$ and $\mathbf{Z_s^{S*}}$ in the production function introduced later). The amenities may also include or depend on aspects of the demographic composition of the neighborhood and school. Some determinants of amenities (such as spending per pupil) may be affected by demographic composition. Thus, some of the amenities are influenced by the sorting equilibrium.

The parents of i choose (n,s) if net utility $\mathscr{U}_i(ns)$ is the highest among the options. That is,

$$(n(i), s(i)) = \arg \max_{n \in \mathscr{N}, s \in \mathscr{S}(n)} \mathscr{U}_i(ns)$$

Parents behave competitively in the sense that prices and $\mathbf{A_{ns}}$ are taken as given, and choice is unrestricted. In equilibrium the values of some elements of $\mathbf{A_{ns}}$ may in fact depend on the averages of $\mathbf{X_i}$ and $\mathbf{X_i^U}$ for the parents who choose n and the parents who choose s (not all of whom come from n) but parents ignore the externalities they are imposing on others.

Proposition 1 of AM establishes that the expectation $\mathbf{X_{ns}^U} \equiv \mathbf{E}[\mathbf{X_i^U} | (n(i), s(i)) = (n,s)]$ is linearly dependent on the expectation $\mathbf{X_{ns}} \equiv \mathbf{E}[\mathbf{X_i}|(n(i),s(i)) = (n,s)]$ if five assumptions hold. Decompose $\mathbf{X_i^U}$ into its linear prediction given $\mathbf{X_i}$ and an uncorrelated residual vector $\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}$:

$$\mathbf{X_i^U} = \mathbf{X_i}\mathbf{\Pi_{X^UX}} + \tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}. \tag{4}$$

Use (4) to rewrite (2) 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 $\tilde{\mathbf{\Theta}} = [\mathbf{\Theta} + \mathbf{\Pi_{X^UX}}\mathbf{\Theta^U}]$. In the rewritten form, the three components of $\mathbf{W_i}$ are mutually orthogonal.

Proposition 1: (AM (2016)) *Assume the following assumptions hold:*

A1: *Preferences are given by (3).*

A2: *Parents take P_{ns} and $\mathbf{A_{ns}}$ as given when choosing location, and face a common choice set.*

A3: *The idiosyncratic preference components ε_{nsi} have a mean of* 0 *and are independent of* $\mathbf{X_i}$, $\mathbf{X_i^U}$, $\mathbf{Q_i}$, and $\mathbf{A_{ns}}$ *for all* (n,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,

$$\mathbf{\Theta^U} = \mathbf{R}\tilde{\mathbf{\Theta}} \tag{5}$$

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

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

$$\mathbf{X_{ns}^U} = \mathbf{X_{ns}}[\mathbf{\Pi_{X^UX}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})] \tag{6}$$

The proof is in AM. Proposition 1 states that the sorting model introduced in this section implies that the vector $\mathbf{X_{ns}}$ can serve as a set of controls for $\mathbf{X_{ns}^U}$. This is key to distinguishing the causal effect of school and location on outcomes from sorting on unobservable student characteristics that affect outcomes. AM provide a detailed discussion of the assumptions, and devote particular attention to the spanning assumption A5. We will not repeat that discussion here. Note, though, that the proposition characterizes the relationship between *expected values* of observable and unobservable student characteristics conditional on choice of school and neighborhood; the exact linear dependence need not hold when samples of students are taken at each school. However, AM provide a monte carlo analysis suggesting that (6) is a good approximation even with samples of 20 students at each school (around the number observed per school used to construct $\mathbf{X}_s$ in the samples we employ below).

3 The Model of Education Attainment and Wage Rates

In this section we present the model of outcomes. The model is similar to its counterpart from AM, but 1) is explicit about the separate roles played by neighborhood, school, and local area inputs, and 2) introduces potential interactions or complementarities between individual inputs and each category of group-level inputs.

In our application the outcomes are high school graduation, attendance at a four-year college, graduation from a four-year college, and the wage rate. The outcome Y_i of student i whose family has chosen the neighborhood $n(i)$, school $s(i)$ and the associated commuting zone $c(i)$ is determined according to

$$\begin{aligned} Y_i \quad = \quad & \mathbf{X_i^*}\boldsymbol{\beta}^* + \mathbf{Z_n^{N*}}\mathbf{\Gamma^{N*}} + \mathbf{Z_s^{S*}}\mathbf{\Gamma^{N*}} + \mathbf{Z_c^{C*}}\mathbf{\Gamma^{C*}} \\ & + M_i\mathbf{Z_n^{N*}}\boldsymbol{\rho}^*_{\mathbf{N}} + M_i\mathbf{Z_s^{S*}}\boldsymbol{\rho}^*_{\mathbf{S}} + M_i\mathbf{Z_c^{C*}}\boldsymbol{\rho}^*_{\mathbf{C}} \\ & + \eta_{nsci} + \xi_{nsci} \end{aligned} \tag{7}$$

where we have dropped the dependence of the neighborhood, school, and commuting zone subscripts on the student subscript i to simplify the notation. To simplify the presentation and the proofs of identification results in Propositions 2 and 3 below, we focus on the case in which M_i is a scalar. In the empirical specification, M_i is replaced with a vector.

For the education outcomes, Y_i is the latent variable that determines the binary outcome. The row vectors $\mathbf{X_i^*}$, $\mathbf{Z_n^{N*}}$, $\mathbf{Z_s^{S*}}$, and $\mathbf{Z_c^{C*}}$ respectively denote the exhaustive set of child and family characteristics, neighborhood characteristics, school characteristics, and commuting zone characteristics that determine outcomes. Many of them are not observed by the econometrician. All are normalized to have a population mean of 0.

The parameters $\boldsymbol{\beta}^*$, $\mathbf{\Gamma^{N*}}$, $\mathbf{\Gamma^{S*}}$, and $\mathbf{\Gamma^{C*}}$ are the corresponding slope coefficients or input productivities. By virtue of our normalization, $\mathbf{\Gamma^{N*}}$, $\mathbf{\Gamma^{S*}}$, and $\mathbf{\Gamma^{C*}}$ capture the effects of school and location variables at the mean of M_i.

M_i is a known scalar-valued linear function of $\mathbf{X_i}$. The parameters $\boldsymbol{\rho}^{\mathbf{N}*}$, $\boldsymbol{\rho}^{\mathbf{S}*}$, and $\boldsymbol{\rho}^{\mathbf{C}*}$ represent unknown parameter vectors capturing the strength of interactions or complementarities between student- and group-level inputs. In the empirical work, we will place restrictions on the $\boldsymbol{\rho}$ vectors and will often use estimated parameter vectors to construct the M_i values from $\mathbf{X_i}$. Our current analysis excludes interactions between $\mathbf{X_i^U}$ and $\mathbf{Z}^*$. We leave the possibility of relaxing this restriction to future work.

The unobserved scalar index η_{nsci} captures variation among students within a neighborhood/school combination in the neighborhood and school inputs they experience, such as the characteristics of immediate neighbors and characteristics of nearby children, distinct course tracks and the luck of the draw in teacher quality at the school. Importantly, η_{nsci} captures the extent to which different students receive *different* school and location "treatments", while the $\boldsymbol{\rho}$ parameters capture the extent to which different students respond differently to the *same* school and location treatments (i.e. treatment effect heterogeneity). Some of the factors that determine η_{nsci} may represent within-neighborhood components of $\mathbf{Z_n^{N*}}$ and within-school components of $\mathbf{Z_s^{S*}}$.

The component ξ_{nsci} is the sum of influences at the n, s, and c level that are determined after secondary school and are unrelated to the other variables in the model, both with and across groups. We use the nsc subscript to allow for shocks at all three levels, but we primarily have in mind shocks at the commuting zone level. These might include the opening of a local college or local labor demand shocks that occur after high school is completed. It will prove useful to write ξ_{snci} as $\xi_{snc} + \xi_i$, where ξ_{snc} is common to all students in nsc and ξ_i is idiosyncratic. ξ_{nsc} is taken to be 0 for the high school graduation outcome.

Note that the characteristics/inputs at each level of observation can represent non-linear functions of other inputs from that level (e.g. an element of $\mathbf{X}_i^*$ can represent the square of another element of $\mathbf{X}_i^*$). Thus, the linear-in-parameters specification for Y_{is} is more general than it first appears.

The productivity parameters $\boldsymbol{\beta}^*$, $\mathbf{\Gamma}^{N*}$, $\mathbf{\Gamma}^{S*}$, and $\mathbf{\Gamma}^{C*}$, the complementarity parameters $\boldsymbol{\rho}_N^*$, $\boldsymbol{\rho}_S^*$, $\boldsymbol{\rho}_C^*$, and the error components η_{nsci} and ξ_{nsci} depend implicitly upon the specific outcome under consideration as well as the time period in the case of wages. To simply the subscripts, we often use g to refer to the combined "group" index nsc. Furthermore, in this section and in our empirical work we restrict neighborhoods to be nested within schools, so that the set $\mathscr{S}(n)$ from the previous section is restricted to be a singleton for each n (each school is still associated with several neighborhoods).[9] The schools are then nested within commuting zones. We use the terms "area" or "location" to refer to the joint influence of neighborhood and commuting zone inputs.

This production function yields a clear definition of the "neighborhood" effect associated with a given neighborhood, as experienced by a particular student type (indexed by M_i). Specifically, compared to a neighborhood with population-average inputs ($\mathbf{Z_n^{N*}} = \mathbf{0}$), a randomly selected student with a particular value $M_i = M_i^a$ who grows up in a neighborhood n^1 featuring neighborhood inputs $\mathbf{Z_{n^1}^{N*}}$ can expect an increase (or decrease) in outcome Y_i of $[\mathbf{Z_{n^1}^{N*}\Gamma^{N*}} + M_i^a \mathbf{Z_{n^1}^{N*}}\boldsymbol{\rho}^{\mathbf{N}}]$, holding the distribution of school quality and commuting zone quality constant. A student with the population mean value $M_i = 0$ can expect an increase of $\mathbf{Z}_{n^1}^{N*}\mathbf{\Gamma}^{N*}$. The corresponding expression for the effect of attending school s^1 rather then a school with average inputs (holding location inputs constant) is

$$\mathbf{Z_{s^1}^{S*}\Gamma^{S*}} + M_i^a \mathbf{Z_{s^1}^{S*}}\boldsymbol{\rho}^{\mathbf{S}}.$$

The corresponding effect of community zone c^1 versus a commuting zone with average inputs (holding school and neighborhood quality constant) is

$$\mathbf{Z_{c^1}^{C*}\Gamma^{C*}} + M_i^a \mathbf{Z_{c^1}^{C*}}\boldsymbol{\rho}^{\mathbf{C}}.$$

The impact at M_i^a of growing up in a particular (n,s,c) combination (n^1,s^1,c^1) relative to one

[9] We discussion the issue in Section 6.3 below.

featuring average inputs at each level is:

$$[\mathbf{Z}_{\mathbf{n}^1}^{\mathbf{N}*}\mathbf{\Gamma}^{\mathbf{N}*} + \mathbf{Z}_{\mathbf{s}^1}^{\mathbf{S}*}\mathbf{\Gamma}^{\mathbf{S}*} + \mathbf{Z}_{\mathbf{c}^1}^{\mathbf{C}*}\mathbf{\Gamma}^{\mathbf{C}*} + M_i^a(\mathbf{Z}_{\mathbf{n}^1}^{\mathbf{N}*}\boldsymbol{\rho}^{\mathbf{N}} + \mathbf{Z}_{\mathbf{s}^1}^{\mathbf{S}*}\boldsymbol{\rho}^{\mathbf{S}} + \mathbf{Z}_{\mathbf{c}^1}^{\mathbf{C}*}\boldsymbol{\rho}^{\mathbf{C}})]$$

A natural extension would be to consider interactions between neighborhood, school, and commuting zone characteristics, but we leave that to future research.

We wish to quantify the contributions of differences in neighborhood factors, school factors, and commuting zone factors to education and labor market outcomes. In the case of college attendance, college graduation, and wage rates, the expected outcome from growing up in a particular school/location combination will also reflect ξ_{nsc^1}, which is common to all individuals in commuting zone c but is determined after high school.[10]

The productivity coefficients $\boldsymbol{\beta}^*$ do not have a straightforward causal interpretation. Some components of $\mathbf{X}_i^*$ associated with student inputs (for example, student aptitude) will have been determined by past parental inputs such as family income (Todd and Wolpin (2003) and Cunha et al. (2006)). Such links make it difficult to interpret the productivity associated with a given component of $\mathbf{X}_i^*$, once we have conditioned on the other components. Consequently, we do not attempt to tease apart the distinct influences of child characteristics, family characteristics, and early childhood schooling and location inputs, respectively. We introduce the $\boldsymbol{\beta}^*$ notation in order to clearly demonstrate the impact of student sorting on our ability to identify the causal effects associated with group-level inputs as well as the degree to which these causal effects vary across students with different individual characteristics.

Nor do we attempt to estimate the causal effects of particular neighborhood, school, or commuting zone inputs (and so will not aim to separately identify particular elements of $\mathbf{\Gamma}^*$). This is because the control function variables only addresses sorting bias. It does not eliminate omitted variable bias that arises because observed neighborhood, school, or region inputs may be correlated with unobserved inputs. Instead, we aim to distinguish the combined outcome effects of neighborhood factors, school factors, and commuting zone factors, respectively, from the effects of student, family, and prior school and location factors.

4 Identification Results

In this section we present our estimating equation for the outcome Y_i and discuss the relationship between the slope parameters and error components recovered from estimation and the parameters of the outcome production function (7) presented in Section 3. In particular, we show that the regression coefficients and error components estimated via OLS allow us to divide the contribution of student inputs and school/location inputs to Y_i into four components. The first component consists of the main effect of student inputs on the outcome. Its effect is common across groups

[10]The outcomes of a specific student i will also differ across neighborhoods, schools and commuting zones because the values of the idiosyncratic terms η_{nsci} differ.

(school/location combinations). The second component consists of the main effect of group-level inputs. These effects are common to all students. The third consists of interactions between student inputs and group inputs. The fourth component is associated with the control function $\mathbf{X_{ns}}$. It captures a combination of group inputs (mostly peer influences) and group averages of student inputs that affect outcomes regardless of the school and location (reflecting sorting). In Section 8 we show how to use the identification results to provide conservative estimates of the average impact associated with "treatments" where group-level inputs are shifted. We also show how to estimate the impact of group treatments on particular student types (allowing a characterization of the degree of treatment effect heterogeneity). As we explain below, we are not able to fully distinguish among the roles of unobserved neighborhood characteristics, unobserved school characteristics, and unobserved commuting zone characteristics.

4.1 Toward an Estimating Equation

In this subsection we introduce some additional notation. We use it to write both the production function (7) and our estimating equation in a way that facilitates the analysis of identification in subsections 4.2 and 4.3.

To this end, we first decompose the vector $\mathbf{X_i^*}$ into its observable and unobservable components: $[\mathbf{X_i}, \mathbf{X_i}^{\mathbf{U}}]$. Note that the superscript "U" denotes "unobserved" throughout the paper. We similarly decompose $\boldsymbol{\beta}^* \equiv [\boldsymbol{\beta}, \boldsymbol{\beta}^{\mathbf{U}}]$ so that the index $\mathbf{X_i^*}\boldsymbol{\beta}^*$ is equal to $\mathbf{X_i}\boldsymbol{\beta} + \mathbf{X_i}^{\mathbf{U}}\boldsymbol{\beta}^{\mathbf{U}}$.

Similarly, the neighborhood influence term $\mathbf{Z_n^{N*}}\boldsymbol{\Gamma}^{\mathbf{N*}}$ may be written as

$$\mathbf{Z_n^{N*}}\boldsymbol{\Gamma}^{\mathbf{N*}} = \mathbf{X_n}\boldsymbol{\Gamma}_\mathbf{1}^\mathbf{N} + \mathbf{Z_{2n}^N}\boldsymbol{\Gamma}_\mathbf{2}^\mathbf{N} + z_n^{NU}$$

where $\mathbf{Z_n^{N*}} = [\mathbf{X_n}, \mathbf{Z_{2n}^N}, \mathbf{Z_n^{NU}}]$. To better utilize the theoretical insights from the choice model above, we distinguish $\mathbf{X_n}$, which is the vector of neighborhood means of $\mathbf{X_i}$, from $\mathbf{Z_{2n}^N}$, which is a set of other observed neighborhood characteristics that are not mechanically related to peer characteristics. Both of these components are observed. The variable $z_n^{NU} \equiv \mathbf{Z_n^{NU}}\boldsymbol{\Gamma}^{\mathbf{NU}}$ is an index of the unobserved neighborhood characteristics (possibly including the neighborhood mean $\mathbf{X_n^U}$ of $\mathbf{X_i^U}$). $\boldsymbol{\Gamma}_\mathbf{1}^\mathbf{N}$, $\boldsymbol{\Gamma}_\mathbf{2}^\mathbf{N}$, and $\boldsymbol{\Gamma}^{\mathbf{NU}}$ are the subvectors that make up $\boldsymbol{\Gamma}^{\mathbf{N*}}$. Thus, $\mathbf{X_n}\boldsymbol{\Gamma}_\mathbf{1}^\mathbf{N}$ captures peer influences generated by $\mathbf{X_n}$, and z_n^{NU} will partly reflect peer influences generated by $\mathbf{X_n^U}$.

Along the same lines, we can write the indices summarizing the components of the impact of school and commuting zone inputs that are common to all students as

$$\mathbf{Z_s^{S*}}\boldsymbol{\Gamma}^{\mathbf{S*}} \equiv \mathbf{X_s}\boldsymbol{\Gamma}_\mathbf{1}^\mathbf{S} + \mathbf{Z_{2s}^S}\boldsymbol{\Gamma}_\mathbf{2}^\mathbf{S} + z_s^{SU}$$

and

$$\mathbf{Z_c^{C*}}\boldsymbol{\Gamma}^{\mathbf{C*}} \equiv \mathbf{X_c}\boldsymbol{\Gamma}_\mathbf{1}^\mathbf{C} + \mathbf{Z_{2c}^C}\boldsymbol{\Gamma}_\mathbf{2}^\mathbf{C} + z_c^{CU}.$$

Furthermore, to conserve notation, we introduce the subscript $g \equiv (n,s,c)$ to denote the combined neighborhood-school-commuting zone combination or "group" experienced by a given child. Thus, we may define the following vector of observed group-level inputs $\mathbf{Z_g} \equiv [\mathbf{Z_n^N}, \mathbf{Z_s^S}, \mathbf{Z_c^C}] \equiv [\mathbf{X_n^N}, \mathbf{Z_{2n}^N}, \mathbf{X_s^S}, \mathbf{Z_{2s}^S}, \mathbf{X_c^C}, \mathbf{Z_{2c}^C}]$. We define the vector $\mathbf{Z_g^U}$ analogously. The index of group-level unobserved impacts, z_g^U, is defined as $z_n^{NU} + z_s^{SU} + z_c^{CU}$.

Similarly, we define $\mathbf{\Gamma} \equiv [\mathbf{\Gamma^{N\prime}}, \mathbf{\Gamma^{S\prime}}, \mathbf{\Gamma^{C\prime}}]' \equiv [\mathbf{\Gamma_1^{N\prime}}, \mathbf{\Gamma_2^{N\prime}}, \mathbf{\Gamma_1^{S\prime}}, \mathbf{\Gamma_2^{S\prime}}, \mathbf{\Gamma_1^{C\prime}}, \mathbf{\Gamma_2^{C\prime}}]'$ and we define $\boldsymbol{\rho} \equiv [\boldsymbol{\rho}^{\mathbf{N}\prime}, \boldsymbol{\rho}^{\mathbf{S}\prime}, \boldsymbol{\rho}^{\mathbf{C}\prime}]' \equiv [\boldsymbol{\rho}_\mathbf{1}^{\mathbf{N}\prime}, \boldsymbol{\rho}_\mathbf{2}^{\mathbf{N}\prime}, \boldsymbol{\rho}_\mathbf{1}^{\mathbf{S}\prime}, \boldsymbol{\rho}_\mathbf{2}^{\mathbf{S}\prime}, \boldsymbol{\rho}_\mathbf{1}^{\mathbf{C}\prime}, \boldsymbol{\rho}_\mathbf{2}^{\mathbf{C}\prime}]'$. $\mathbf{\Gamma^U}$ and $\boldsymbol{\rho}^\mathbf{U}$ are defined analogously.

Using this notation, the production function for outcomes may be written in compact form as

$$Y_i = \mathbf{X_i}\boldsymbol{\beta} + \mathbf{Z_g}\mathbf{\Gamma} + M_i\mathbf{Z_g}\boldsymbol{\rho} + x_i^U + z_g^U + M_i\mathbf{Z_g^U}\boldsymbol{\rho}^\mathbf{U} + \eta_{gi} + \xi_{gi} \, . \tag{8}$$

For purposes of comparison with our estimating equation, however, we usually refer to the version that distinguishes peer inputs from other group-level inputs:

$$\begin{aligned} Y_i \quad = \quad & \mathbf{X_i}\boldsymbol{\beta} + \mathbf{X_g}\mathbf{\Gamma_1} + \mathbf{Z_g}\mathbf{\Gamma_2} + M_i\mathbf{X_g}\boldsymbol{\rho}_\mathbf{1} + M_i\mathbf{Z_{2g}}\boldsymbol{\rho}_\mathbf{2} \\ & + M_i\mathbf{X_g^U}\boldsymbol{\rho}_\mathbf{1}^U + M_i\mathbf{Z_g^U}\boldsymbol{\rho}_\mathbf{2}^U + x_i^U + z_g^U + \eta_{gi} + \xi_{gi} \, . \end{aligned} \tag{9}$$

AM provide an extensive discussion of which variables should be included in $\mathbf{X_g}$ and $\mathbf{Z}_{2g}$, respectively, which we summarize in Section 5.3.

Using the same group notation, our estimating equation takes form

$$Y_{gi} = \mathbf{X_i}\mathbf{B} + \mathbf{X_g}\mathbf{G_1} + \mathbf{Z_{2g}}\mathbf{G_2} + M_i\mathbf{X_g}\mathbf{r_1} + M_i\mathbf{Z_{2g}}\mathbf{r_2} + M_g\mathbf{X_g}\mathbf{G_3} + M_g\mathbf{Z_{2g}}\mathbf{G_4} + v_g + (v_{gi} - v_g). \tag{10}$$

The parameters refer to the coefficient vectors from the linear least-squares projection of Y_{gi} on the right hand side variables. In practice, we will dramatically restrict the estimating equation rather than freely introduce all of the interaction terms.

The inclusion of the row vectors $M_g\mathbf{X_g}$ and $M_g\mathbf{Z_g}$ would ensure that identification of $\mathbf{r_1}$ and $\mathbf{r_2}$ is obtained exclusively from within-group variation in $M_i\mathbf{X_g}$ and $M_i\mathbf{Z_{2g}}$. We show below that Assumptions 1-5 of the sorting model combined with an additional assumption imply that $\mathbf{G}_3$ and $\mathbf{G}_4$ are 0, and so in the end we exclude $M_g\mathbf{X_g}$ and $M_g\mathbf{Z_g}$ from the models we estimate.

We now consider how the parameters of the estimating equation (10) relate to the production function parameters in (9). We first consider $\mathbf{B}$, $\mathbf{r}_1$ and $\mathbf{r}_2$ and then turn to $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$, and $\mathbf{G}_4$.

4.2 Interpretation of B, $\mathbf{r_1}$, and $\mathbf{r_2}$

Let D represent the within-group linear operator, so that $DY_{gi} \equiv Y_{gi} - E[Y_{gi}|g(i) = g] \equiv Y_{gi} - Y_g$, where Y_g is the expected value of Y_{gi} conditional on student i's choice of g. The within-group counterpart of our estimating equation (10) can be written as:

$$DY_{gi} = \mathbf{DX_i}\mathbf{B} + [(DM_i)\mathbf{X_g}]\mathbf{r_1} + [(DM_i)\mathbf{Z_{2g}}]\mathbf{r_2} + Dv_{gi} \tag{11}$$

Given the presence in (10) of the group means $\mathbf{X}_g$, $M_g\mathbf{X}_g$ and $M_g\mathbf{Z}_{2g}$, OLS estimation of $\mathbf{B}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ using (10) is equivalent to OLS estimation using the within-group equation (11). Thus, we can start by restricting attention to (11).

Next, we make the following strong but powerful assumption:

A6: $\mathbf{Cov}[DM_i\mathbf{DX}_{\mathbf{i}}^{*\prime}, \mathbf{Z}_{\mathbf{g}}^*] = 0$

Assumption 6 states that the cross products of the within group deviations of M_i and the student variables $\mathbf{X_i}$ and $\mathbf{X_i^U}$ do not covary with any of the observed or unobserved neighborhood/school variables. Because $\mathbf{Z}_g^*$ is normalized to have a mean of 0, assumption A6 is equivalent to the assumption $\mathbf{E}[DM_i\mathbf{DX}_{\mathbf{i}}^{*\prime}, \mathbf{Z}_{\mathbf{g}}^*] = 0$. A6 addresses a potentially important identification problem. The variation that will potentially identify interaction effects in our estimating equation consists of differences in the strength of the relationship between individual characteristics and outcomes across groups with different characteristics (e.g. student-teacher ratio). If the covariance between parental income (observed) and parental motivation (unobserved) is larger at schools with smaller student-teacher ratios, then parental income may predict greater increases in outcomes at schools with smaller student-teacher ratios because parental income better predicts parental motivation at such schools, even if students of richer parents are not differentially sensitive to low student-teacher ratios. Such a mechanism would yield spurious "interaction" effects. Assumption A6 rules out this possibility, so that any differential impact of M_i across groups with different observed inputs can be interpreted as evidence of true student-group interactions rather than evidence of differential sorting-induced joint distributions of $[\mathbf{X_i}, \mathbf{X_i^U}]$ across groups.

How plausible is Assumption A6? We fully allow the possibility that the means of M_i and $\mathbf{X_i^*}$ vary across g and even systematically with $\mathbf{Z_g^*}$. To the extent that choice of g is heavily influenced by $\mathbf{Q_i}$ and idiosyncratic factors, then the within group second moments may not vary much with $\mathbf{Z}_g^*$. Below we use the index $\mathbf{X_i B}$ as an M_i variable. One could test A6 for observed components of $\mathbf{DX}_i^{*\prime}$ and $\mathbf{Z}_g^*$, although we have not yet done so. The assumption will fail when M_i is set to binary variables such as the indicators for minority status and low income status that vary substantially across g. The quantitative significance of failures of A6 requires further investigation.

We also require a second additional assumption:

A7 $\mathbf{E}[DM_i^2|g(i) = g]$ is a constant.

One could test this assumption for the case $M_i = \mathbf{X_i B}$.It does not hold when M_i is the minority status indicator, given that the mean and thus the within group variance of M_i varies substantially across groups. Finally, we assume:

A8. $E[D\tilde{\eta}_{gi}DM_i[\mathbf{X_g}, \mathbf{Z_{2g}}]] = 0$

To interpret A8, recall that η_{gi} reflects differential treatment of i in g that is due to random factors. The interaction term $DM_i\mathbf{Z}_g^*\rho$ in the outcome equation (8) captures differential impacts of a common treatment. A8 concerns the residual component of the projection (13) of $D\eta_{gi}$ on $D\mathbf{X_i}$. A8 is perfectly consistent with the possibility that $D\eta_{gi}$ is related to DM_i through the relationship between $D\eta_{gi}$ and $D\mathbf{X_i}$. If DM_i is a linear function of $\mathbf{DX}_i$, $D\tilde{\eta}_{gi}$ and DM_i are orthogonal in the population. A8 will hold if they do not covary within a group.

The statement of Proposition 2 below uses the projection matrices from the following projection equations:

$$\mathbf{DX_i^U} = \mathbf{DX}_i\mathbf{\Pi}_{\mathbf{DX_iDX_i^U}} + \widetilde{\mathbf{DX_i^U}} \tag{12}$$

$$D\eta_{gi} = \mathbf{DX_i}\mathbf{\Pi}_{D\eta_{gi}D\mathbf{X_i}} + D\tilde{\eta}_{gi}. \tag{13}$$

$$\mathbf{X_g^U} = \mathbf{X_g}\mathbf{\Pi}_{\mathbf{X_g^UX_g}} \tag{14}$$

$$\mathbf{Z_{2g}^U} = \mathbf{X_g}\mathbf{\Pi}_{\mathbf{Z_{2g}^UX_g}} + \mathbf{Z_{2g}}\mathbf{\Pi}_{\mathbf{Z_{2g}^UZ_{2g}}} + \widetilde{\mathbf{Z_{2g}^U}} \tag{15}$$

Note that $\mathbf{X_g^U}$ is perfectly predicted by $\mathbf{X_g}$ under the assumptions of Proposition 1 of Altonji and Mansfield (2016), with $\mathbf{\Pi}_{\mathbf{X_g^UX_g}} \equiv [\mathbf{\Pi}_{\mathbf{X^UX}} + \mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})]$. This is why (14) does not have an error term.

We are now ready to present the relationship between $\mathbf{B}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ in (11) (and therefore in (10) as well) and the production function parameters in (9).

Proposition 2:

Suppose assumptions A1-A8 hold. Then:

$$\mathbf{B} = \boldsymbol{\beta} + \mathbf{\Pi}_{\mathbf{DX_iDX_i^U}}\boldsymbol{\beta}^U + \mathbf{\Pi}_{D\eta_{gi}D\mathbf{X_i}} \tag{16}$$

$$\mathbf{r_1} = \boldsymbol{\rho}_\mathbf{1} + \mathbf{\Pi}_{\mathbf{X_g^UX_g}}\boldsymbol{\rho}_\mathbf{1}^\mathbf{U} + \mathbf{\Pi}_{\mathbf{Z_{2g}^UX_g}}\boldsymbol{\rho}_2^U \tag{17}$$

$$\mathbf{r_2} = \boldsymbol{\rho}_2 + \mathbf{\Pi}_{\mathbf{Z_{2g}^UZ_{2g}}}\boldsymbol{\rho}_\mathbf{2}^\mathbf{U}. \tag{18}$$

The proof is in Appendix A1. Note that the coefficient vector $\mathbf{B}$ on $\mathbf{X_i}$ is the same as the coefficient vector in the model without interactions considered by AM. The coefficient $\mathbf{r_1}$ consists of $\boldsymbol{\rho}_\mathbf{1}$, which is the vector of interactions between M_i and the group averages of the observable student characteristics $\mathbf{X_g}$, plus two other terms. The second term $\mathbf{\Pi}_{\mathbf{X_g^UX_g}}\boldsymbol{\rho}_\mathbf{1}$ captures interactions between M_i and the group average of the unobservables $\mathbf{X_g^U}$ (Recall that under proposition 1, $\mathbf{X_g^U} = \mathbf{X_g}\mathbf{\Pi}_{\mathbf{X_g^UX_g}}$). Thus, the first two terms of $\mathbf{r_1}$ capture the fact that the effect of $\mathbf{X_i}$ on Y_i depends upon the characteristics of the population in g. The third term $\mathbf{\Pi}_{\mathbf{Z_{2g}^UX_g}}\boldsymbol{\rho}_\mathbf{2}^\mathbf{U}$ is present because the unobserved school and area characteristics $\mathbf{Z_{2g}^U}$ may vary with $\mathbf{X_g}$ conditional on $\mathbf{Z_{2g}}$. The coefficient vector $\mathbf{r_2}$ is the sum of $\boldsymbol{\rho}_\mathbf{2}$, the effect of the interaction between M_i and $\mathbf{Z_{2g}}$, along with the effect of the interaction between M_i and the portion of $\mathbf{Z_{2g}^U}$ that is predictable by $\mathbf{Z_{2g}}$ holding $\mathbf{X_g}$ constant. Importantly, both $\mathbf{r_1}$ and

$\mathbf{r_2}$ exclusively reflect interactions between $\mathbf{X_i}$ and group-level characteristics, rather than individual contributions that are common across groups or group-level contributions that are common across individuals.

4.3 Interpretation of $\mathbf{G_1}, \mathbf{G_2}, \mathbf{G_3}$, and $\mathbf{G_4}$

This subsection presents Proposition 3, which establishes the relationship between the production parameters in (8) and the coefficients $\mathbf{G_1}, \mathbf{G_2}, \mathbf{G_3}$, and $\mathbf{G_4}$ identified by OLS. Note that just as $\mathbf{B}$, $\mathbf{r_1}$, and $\mathbf{r_2}$ are identified exclusively from within-group variation, $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$, and $\mathbf{G_4}$ are identified exclusively from between-group variation. This fact means that the OLS coefficients $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$, and $\mathbf{G_4}$ are numerically identical to the coefficients of the projection of the adjusted group g mean of Y_{gi}, $Y_g - [\mathbf{X_g B} + M_g \mathbf{X_g r_1} + M_g \mathbf{Z_{2g} r_2}]$, onto $\mathbf{X_g}$, $\mathbf{Z_{2g}}$, $M_g \mathbf{X}_g$, and $M_g \mathbf{Z_{2g}}$.

First we need to define some projection coefficients that appear in Proposition 3. Let the projections of the unobserved production function index z^U_{2g} onto the space of group-level observables $[\mathbf{X_g}, \mathbf{Z_{2g}}, M_g \mathbf{X_g}, M_g \mathbf{Z_{2g}}]$ be given by

$$z^U_{2g} = \mathbf{X_g} \mathbf{\Pi}_{z^U_{2g} \mathbf{X_g}} + \mathbf{Z_{2g}} \mathbf{\Pi}_{z^U_{2g} \mathbf{Z_{2g}}} + M_g \mathbf{X_g} \mathbf{\Pi}_{z^U_{2g}, M_g \mathbf{X_g}} + M_g \mathbf{Z_{2g}} \mathbf{\Pi}_{z^U_{2g}, M_g \mathbf{Z_{2g}}} + \widetilde{z}^U_{2g}$$

We also introduce an additional assumption:

A9 $\widetilde{\mathbf{Z^U_{2g}}}$ *has mean 0 and is independent of* $\mathbf{X_g}$ *and* $\mathbf{Z_{2g}}$.

By definition, the error component $\widetilde{\mathbf{Z^U_{2g}}}$ in the projection equation (15) is uncorrelated with $\mathbf{X_g}$ and $\mathbf{Z_{2g}}$. A9 strengthens zero correlation to independence. Doing so rules out the possibility that $\mathbf{X_g}$ or $\mathbf{Z_{2g}}$ might be predictive of the value of the average interaction $M_g \mathbf{Z^U_{2g}}$, even conditional on $M_g \mathbf{X_g}$ and $M_g \mathbf{Z_{2g}}$.

We are now ready to state the proposition.

Proposition 3:

Suppose assumptions A1-A9 hold. Then:

$$\begin{aligned}
\mathbf{G_1} &= [(\boldsymbol{\beta} - \mathbf{B}) + \mathbf{\Pi}_{x^U_g \mathbf{X_g}}] + [\mathbf{\Gamma_1} + \mathbf{\Pi}_{\mathbf{X^U_g} \mathbf{X_g}} \mathbf{\Gamma^U_1} + \mathbf{\Pi}_{z^U_{2g} \mathbf{X_g}}] \\
\mathbf{G_2} &= \mathbf{\Gamma_2} + \mathbf{\Pi}_{z^U_{2g} \mathbf{Z_{2g}}} \\
\mathbf{G_3} &= 0 \\
\mathbf{G_4} &= 0
\end{aligned}$$

The proof is in Appendix A2. Proposition 3 states that under A1-A9 the coefficients $\mathbf{G_1}$ and $\mathbf{G_2}$ on $\mathbf{X_g}$ and $\mathbf{Z_{2g}}$ in the model with interactions are the same as the corresponding coefficients in the model without interactions considered by AM. $\mathbf{G_1}$ consists of the causal peer effects of $\mathbf{X_g}$ and $\mathbf{X_g^U}$ plus the part of the effect of unobserved group inputs $\mathbf{Z_{2g}^U}$ predicted by $\mathbf{X_g}$. However, it also picks up the bias term $(\boldsymbol{\beta} - \mathbf{B})$ and the term $\mathbf{\Pi}_{x_g^U \mathbf{X_g}}$, which captures differences across school and location in x_g^U. Since these last two components represent student contributions to Y rather than group-level contributions, we exclude $\mathbf{X_g G_1}$ when constructing lower bound estimates for school and location treatment effects.

$\mathbf{G_2}$ captures the causal effects of the observed group-level inputs $\mathbf{Z_{2g}}$ along with associated variation in unobserved group-level inputs z_g^U on the outcome of a student at the population mean of $\mathbf{M_i}$. Thus, the index $\mathbf{Z_{2g}G_2}$ only reflects the contributions of group-level inputs whose impacts are common across all students. It does not include student inputs or interactions between student- and group-level inputs. This result is key to our ability to characterize both average treatment effects associated with shifts in neighborhood quality as well as the degree of treatment effect heterogeneity across types of students.

Finally, $\mathbf{G_3}$ and $\mathbf{G_4}$ both enter with zero coefficients. Consequently, we can exclude both $M_g\mathbf{X_g}$ and $M_g\mathbf{Z_{2g}}$ from the estimating equation (10).

4.3.1 Interpretation of the Error Components v_g and $(v_{gi} - v_g)$

The final proposition establishes the relationship between the individual-level and group-level error components in (8) and (10).

Proposition 4:

Suppose assumptions A1-A9 hold. Then:

$$v_g = \tilde{z}_{2g}^U + M_g \tilde{\mathbf{Z}}_{2g}^{\mathbf{U}} \boldsymbol{\rho}_2^{\mathbf{U}} + \xi_g \tag{19}$$

$$v_{gi} - v_g = \widetilde{Dx_i^U} + D\eta_{gi} + D\xi_{gi}. \tag{20}$$

We relegate the proof to a footnote[11]. Equation (19) reveals that v_g is the sum of group-level effects that are common across students ($\tilde{z}_{2g}^U$ and ξ_g) and group-averages of interactions between student inputs and group inputs ($M_g \widetilde{\mathbf{Z}}_{2g}^{\mathbf{U}} \boldsymbol{\rho}_2^{\mathbf{U}}$). In principal, one could estimate the distribution of $M_g \widetilde{\mathbf{Z}}_{2g}^{\mathbf{U}} \boldsymbol{\rho}_2^{\mathbf{U}}$ by incorporating random neighborhood-, school-, and commuting zone-specific slopes

[11] Proposition 4 can be proved simply by (1) substituting into the estimating equation (10) the expressions from Propositions 2 and 3 for $\mathbf{B}$, $\mathbf{r_1}$, $\mathbf{r_1}$, $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$, and $\mathbf{G_4}$, (2) subtracting off all the observed regression indices from the production function for Y_i given by (9), and (3) taking group means and within-group deviations of the production function components that remain. To obtain $\tilde{z}_{2g}^U$ in (19) we also use the fact that by definition z_g^U, which appears in (9), is equal to $\mathbf{X_g^U}\mathbf{\Gamma_1^U} + z_{2g}^U$, where $z_{2g}^U = \mathbf{Z_{2g}^U}\mathbf{\Gamma_2^U}$.

with respect to M_i into a mixed effects procedure. This would complete the full separation of the contributions of interaction components from the common impacts of group-level inputs that we discuss in the introduction. However, in practice we have had trouble identifying such random slopes in our application.

Finally, (20) shows that $v_{gi} - v_g$ reflects only student level inputs and perhaps (via $D\eta_{gi}$) differential treatments within groups (e.g. the sequence of assigned teachers), rather than interactions with group-level inputs that would reflect heterogeneity in the impact of the same group-level treatments.

5 Data, Variable Selection, and Specification of Interactions

5.1 Overview of Data Sources

We use two panel data sets, the National Educational Longitudinal Survey of 1988 (NELS88) and the Educational Longitudinal Survey of 2002 (ELS2002).[12] 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 8th graders in 1988 for NELS88, and 10th graders in 2002 for ELS2002. Second, each source provides a representative sample of American 8th grades and high schools, respectively, and samples of students are selected within each school. Public, private, and parochial schools are represented.[13] Enough students are sampled from each school to permit construction of estimates of the school means of a large set of student-specific variables and to provide sufficient within-school variation to support the variance decomposition described below. Third, each survey administered questionnaires to school administrators in addition to sampling individuals at each school. This provides us with a rich set of both individual-level and school-level variables to examine, allowing a meaningful decomposition of observable versus unobservable variation at both levels of observation. Fourth, each survey contains information on the student's location of residence. In the case of NELS, we observe the student's ZIP code in grade 8 and at age 25. In ELS, we observe the student's census block of residence in grade 10, as well as each student's ZIP code of residence in each of three follow up surveys. Observing residential location is critical for characterizing the relative contributions of neighborhood versus school inputs as well as the relative importance of neighborhood versus school amenities in driving student sorting. Finally, each survey collects follow-up information from each student past high school graduation, facilitating analysis of the impact of high school environment on outcomes economists and policymakers particularly care about: educational attainment and wage rates at age

[12] This section draws heavily on Section 6 of AM.

[13] We include private schools because they are an important part of the education landscape. The model in Section 2 says that we should include group averages for the neighborhood/school pairs to control for sorting on unobservables, but that is not practical, because sample sizes are too small. We include $\mathbf{X_s}$ for the school and census based demographic variables at the zip code or block group level, which we denote $\mathbf{X_n}$ even though they include characteristics of residents who are not students or parents.

25.[14]

While there is considerable overlap in the survey questionnaires associated with NELS88 and ELS2002, we chose not to restrict attention to the set of variables that are available and measured consistently across the two datasets. This is because the efficacy of the control function approach depends on the richness and diversity of our student-level measures, and using the intersection of the two datasets limits the diversity of student characteristics. In practice, the variables lists are similar, as one can see from the variable list in Appendix Table 1. Section 5.3 describes the process by which we chose what to include in $\mathbf{X_i}$, $\mathbf{X_n}$, $\mathbf{X_s}$, $\mathbf{X_c}$, $\mathbf{Z_{2s}^S}$, and $\mathbf{Z_{2c}^C}$. Unfortunately, the panel surveys do not contain neighborhood level variables, though we do merge in Census-based and LODES based neighborhood-level demographic averages to bolster our control function.

Coding of most of the variables is straightforward, but in some cases variables are the first principle components constructed from the responses to batteries of questions about topics such as the home environment, school policies for reducing dropout rates, and quality of school facilities, among others.

As AM discuss, a drawback of the two panel surveys is that only around 20 students per school are generally sampled. Simulation results in AM indicate that samples of this size may reduce to some degree the ability of sample school averages of observable characteristics to serve as an effective control function for variation in average unobservable student contributions across schools, but they perform fairly well[15]

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. We then impute values for the other explanatory variables to preserve the sample size, since no other single variable is critical to our analysis.[16]

5.2 Outcome Measures and Weighting

HSGRAD is an indicator for whether a student has a high school diploma (not including a GED) as of two years after the high school graduation year of his/her cohort. ENROLL is an indicator for whether the student is enrolled in a four year college in October of the second year beyond the high school graduation year of his/her cohort. COLLBA is an indicator for whether the student has a

[14] AM (2011) and AM (2016) use the National Longitudinal Survey of the Class of 1972 as well as NELS88 and ELS2002. The NLS72 is less rich in terms of student and school covariates, does not identify ZIP code of residence separately from high school, and does not start until 12th grade. However, it contains panel data on wages and follows students up to 13 years after 12th grade. Thus, it is better suited for the study of permanent wage rates.

[15] AM's simulations are for an additive model. Using an additive model, AM also study high school graduation using administrative data from the student record system of North Carolina. They report simulation experiments that compare estimates based upon the full North Carolina sample with estimates based upon samples of students that match the distributions in NELS88 and ELS2002 of students per school. The results suggest that the modest samples per school in NELS88 and ELS2002 is not a major problem.

[16] 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.

four-year degree at age 25. Log wages are hourly and in 2009 dollars. They are reported at about age 25.

Use of COLLBA and log wage results in a loss of sample, because it requires use of the NELS 4th follow-up and the ELS 3rd follow-up. High school graduation rates in the previous follow-up surveys are somewhat higher among respondents who are also observed in the final followups. Summary statistics for the outcome measures are in Table 1.

One might want to use weights for three reasons. The first is to account for the stratified sampling regime used by the sample designers. The second is differential attrition associated with the explanatory variables. The third is because follow-up probabilities are based on outcomes. We do not use panel weights because we experienced difficulties in estimating the variance components of the mixed level models when using weights. The slope parameter estimates are not very sensitive to weighting, but we do not know whether weighting would affect the estimates of the error component variances.

5.3 Selection of $\mathbf{X_i}$, $\mathbf{X_n}$, $\mathbf{X_s}$, $\mathbf{Z_{2n}^N}$, $\mathbf{Z_{2s}^S}$, and $\mathbf{Z_{2c}^C}$

AM discuss the principles governing variable selection. $\mathbf{X_i}$ should include variables that directly affect the outcome and/or are correlated with unobserved student level characteristics that affect the outcome. We focus attention on a "full" specification which includes in $\mathbf{X_i}$ measures of student behavior, parental expectations, and student academic ability (standardized test scores). Such measures may be influenced directly by school and location inputs, so including them could cause an underestimate of the contribution of school and location inputs. As a result, our lower bound estimates will be too conservative. On the other hand, excluding such measures could instead cause an overestimate of the contribution of location- and school-level inputs if the more limited set of student observables no longer satisfies the spanning condition $A5$ stated in Proposition 1. In that 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. We also discuss results for a "basic" specification that only includes student-level characteristics that are unlikely to be affected by the neighborhood and high school the child attends.

For purposes of the control function, $\mathbf{X_g}$ should contain aggregates of $\mathbf{X_i}$ at the (n,s) school/neighborhood level. In practice, due to sample size limitations we include $\mathbf{X_s}$ as well as $\mathbf{X_n}$ variables, but not averages over (n,s) pairs. The $\mathbf{X_n}$ variables are census-based measures of the demographic makeup of the neighborhood.[17]

$\mathbf{Z_{2n}^N}$, $\mathbf{Z_{2s}^S}$, and $\mathbf{Z_{2c}^C}$ should include observed neighborhood and school characteristics that could plausibly influence the socioeconomic outcome of interest, including school policies that may be partially affected by student composition. $\mathbf{Z_{2n}^N}$, $\mathbf{Z_{2s}^S}$, and $\mathbf{Z_{2c}^C}$ should exclude variables that are sim-

[17]As AM discuss, in principle the control function variables can be augmented with aggregates of outcome-irrelevant characteristics $\mathbf{Q_i}$ or even directly observed amenities in $\mathbf{A_g}$, since its purpose is span the space of amenities that drive sorting on $\mathbf{X_i^U}$.

ple aggregates of parent/student traits that might also affect willingness to pay for neighborhood characteristics and thus lead to sorting. These are $\mathbf{X_n}$, $\mathbf{X_s}$, or $\mathbf{X_c}$ variables regardless of whether the source is aggregates of the student micro data, Census data, or administrative data from the schools. Unfortunately, we do not observe any neighborhood-level variables that are not population aggregates, so we do not include any $\mathbf{Z_{2n}^N}$ variables in our empirical work. As a result, below we primarily focus attention on the impact of shifts in school- and commuting zone-level treatments.

The sample sizes, number of neighborhoods (zip code or block group), number of schools, and number of commuting zones are reported in Appendix Table 2. While the number of commuting zones and schools is substantial, the demands of estimating our model are also substantial given the lack of prior information about which school and commuting zone variable are likely to have large effects. The precision of our estimates is reduced by the need to work with a fairly large number of $\mathbf{Z_{2s}^S}$ and $\mathbf{Z_{2c}^C}$ variables, along with large numbers variables in the control functions $\mathbf{X_n}$ and $\mathbf{X_s}$.

As AM discuss, group-level variables such as a school's frequency of fights or average 10th grade test scores that capture earlier outcomes that were jointly produced by both individual- and group-level variables fall in a grey area. We exclude such variables entirely from the baseline specification (on the basis that they are determined by other observed and unobserved variables in the model), and include them in the control function $\mathbf{X_g}$ in the full specification. To the extent that school policy and the skill of teachers and the administration have a large effect on fighting and/or test scores, assigning these measures to the control function leads to conservative estimates of group effects.

Appendix Table 1 lists the final choices of individual-level and school-level explanatory measures used in each dataset.[18] The table also provides the mean, standard deviation, and percent of observations imputed for each individual-level and school-level characteristics for each of our data sets. Appendix Table 2 provides the number of neighborhoods, schools, and commuting zones for each combination of dataset and outcome.

6 Estimation Methods

6.1 Restricting $\mathbf{M_i Z_s^* \boldsymbol{\rho}}$

As we mentioned earlier, we have to reduce the dimensionality of the interactions between student characteristics and peer characteristics as well as between student characteristics and neighborhood, school, and commuting zone characteristics. In our main specification with interactions, we set $\mathbf{M_i}$ to consist of the index $\mathbf{X_i B}$ and indicators for whether the student is female, a member of

[18] In preliminary work we experimented with a grouped backward stepwise regression procedure to pare down the variable sets at each unit level. One could also consider other procedures, such as group Lasso (Meier et al. (2008) and Yuan and Lin (2006)). Ultimately, we chose not to use these procedures because of concerns about how to do statistical inference, and about the computational feasibility and statistical properties of a bootstrap procedure that accounts for both variable selection and sampling error given variable choice. The three step estimation method we use is computationally demanding.

an underrepresented minority group (denoted URM below and coded as Hispanic or non-Hispanic black), and whether the student's family is in the bottom quartile in our sample of the ratio of family income to average rent in the commuting zone (LOWINC). Including $\mathbf{X_i B}$ in $\mathbf{M_i}$ imposes the restriction that for white males from high income families, the strength of interactions depends upon components of $\mathbf{X_i}$ in proportion to their direct effects on the outcome. Including the three additional indicator variables in $\mathbf{M_i}$ allows a somewhat more general pattern of interactions for particular subpopulations for whom differential interactions have often been posited in the literature.

In addition, to keep the main specification parsimonious, we further restrict the interactions involving $\mathbf{X_i B}$ and $\mathbf{Z_s^*}$ to occur only through separate interactions between $\mathbf{X_i B}$ and the indices $\mathbf{X_n G_1^N}$, $\mathbf{X_s G_s^S}$, $\mathbf{Z_{2s}^S G_2^S}$, and $\mathbf{Z_{2c}^C G_2^C}$. The three subpopulation indicators are restricted to interact with only the indices $\mathbf{Z_{2s}^S G_2^S}$, and $\mathbf{Z_{2c}^C G_2^C}$.

6.2 Estimating the Parameters of the Model

We estimate the model in three steps:

Step 1: Estimate $\mathbf{B}$, $\mathbf{G_1^N}$, $\mathbf{G_1^S}$, $\mathbf{G_2^S}$, and $\mathbf{G_2^C}$

We estimate $\mathbf{B}$, $\mathbf{G_1^N}$, $\mathbf{G_1^S}$, $\mathbf{G_2^S}$, and $\mathbf{G_2^C}$ using models that include the interaction terms but ignore the multilevel random effects error structure. That is, we estimate the following specification:

$$\begin{aligned} Y_{gi} = {} & \mathbf{X_i B} + \mathbf{X_n G_1^N} + \mathbf{X_s G_1^S} + \mathbf{Z_{2s}^S G_2^S} + \mathbf{Z_{2c}^C G_2^C} \\ & + \mathbf{X_i B X_n G_1^N} r_1^n + \mathbf{X_i B X_s G_1^S} r_1^s + \mathbf{M_i} \otimes \mathbf{Z_{2s}^S G_2^S r_2^s} + \mathbf{M_i} \otimes \mathbf{Z_{2c}^C G_2^C r_2^{\mathit{c}}} + v_{gi}, \end{aligned} \tag{21}$$

where Y_{gi} is the latent index and $\mathbf{M_i} \equiv [\mathbf{X_i B}, 1(Female), 1(URM), 1(Low\_Income)]$ and as a reminder, the Kronecker product $\mathbf{M_i} \otimes \mathbf{Z_{2s}^S G_2^S}$ is the row vector $[\mathbf{X_i B}\, \mathbf{Z_{2s}^S G_2^S}, 1(Female)\, \mathbf{Z_{2s}^S G_2^S}, 1(URM) \mathbf{Z_{2s}^S G_2^S},$ $1(Low\_Income) \mathbf{Z_{2s}^S G_2^S}]$. [19]

For binary education outcomes, Y_{gi} is the latent index of a nonlinear probit model, and the model is estimated via maximum likelihood with parameter restrictions on the regressor indices imposed. For log wages, (21) we estimate using nonlinear least squares.[20]

[19] To improve the allocation of school level factors to estimates of the school treatment effects and the allocation of community level factors to the estimates of commuting level effects, when estimating (21) we also include school and commuting zone averages of the index $\mathbf{X_n \hat{G}_1^N}$ as well as the commuting zone averages of the indices $\mathbf{X_s \hat{G}_1^S}$ and $\mathbf{Z_{2s}^S \hat{G}_2^S}$ with separate coefficients. The school and commuting zone averages of $\mathbf{X_n \hat{G}_1^N}$ and the commuting zone average of $\mathbf{X_s \hat{G}_1^S}$ are treated as control function variables and are contained in $\mathbf{X_s \hat{G}_1^S}$ when estimating (22) below. The commuting zone average of the index $\mathbf{Z_{2s}^S \hat{G}_2^S}$ is treated as a commuting zone characteristic that is contained in $\mathbf{Z_{2c}^C \hat{G}_2^C}$ in (22) below. It contributes to our estimates of the variance in commuting zone treatment effects described below. Since its coefficient should not be affected by sorting bias under assumptions A1-A9, and is identified purely from between-commuting zone variation, it is likely to capture otherwise unobserved commuting zone inputs).

[20] To impose the restrictions that the interactions operate through the same regressor indices as the main group effects, we choose initial values $\mathbf{B^0}$, $\mathbf{G_1^{N,0}}$, $\mathbf{G_1^{S,0}}$, $\mathbf{G_2^{S,0}}$, and $\mathbf{G_2^{C,0}}$. Then, letting k denote the iteration number, we implement an iterative estimation procedure in which (temporary) main effect parameters $\mathbf{B^k}$ $\mathbf{G_1^{N,k}}$, $\mathbf{G_1^{S,k}}$, $\mathbf{G_2^{S,k}}$, and $\mathbf{G_2^{C,k}}$ and the interaction coefficients r_{1n}^k, r_{1s}^k, $\mathbf{r_{2s}^k}$ and $\mathbf{r_{2c}^k}$ are estimated while holding fixed the regressor indices entering the interaction

Note that Propositions 2 and 3 imply that the coefficients $\mathbf{B}$, $\mathbf{G_1^N}$, $\mathbf{G_1^S}$, $\mathbf{G_2^S}$, and $\mathbf{G_2^C}$ in (21) match those in (10), so there is no inconsistency in using the same symbols to label the coefficients in the two equations.

Step 2: Estimate the restricted mixed level model

In the second step, we estimate the following model:

$$
\begin{aligned}
Y_i = {} & \alpha_0 + \alpha_1 \frac{\mathbf{X_i\hat{B}}}{sd(\mathbf{X_i\hat{B}})} + \alpha_2 \frac{\mathbf{X_n\hat{G}_1^N}}{sd(\mathbf{X_n\hat{G}_1^N})} + \alpha_3 \frac{\mathbf{X_s\hat{G}_1^S}}{sd(\mathbf{X_s\hat{G}_1^S})} + \alpha_4 \frac{\mathbf{Z_{2s}^S\hat{G}_2^S}}{sd(\mathbf{Z_{2s}^S\hat{G}_2^S})} + \alpha_5 \frac{\mathbf{Z_{2c}^C\hat{G}_2^C}}{sd(\mathbf{Z_{2c}^C\hat{G}_2^C})} + \\
& + r_1^N \frac{\mathbf{X_i\hat{B}}}{sd(\mathbf{X_i\hat{B}})} \frac{\mathbf{X_n\hat{G}_1^N}}{sd(\mathbf{X_n\hat{G}_1^N})} + r_1^S \frac{\mathbf{X_i\hat{B}}}{sd(\mathbf{X_i\hat{B}})} \frac{\mathbf{X_s\hat{G}_1^S}}{sd(\mathbf{X_s\hat{G}_1^S})} + [\mathbf{\hat{M}_i} \otimes \frac{\mathbf{Z_{2s}^S\hat{G}_2^S}}{sd(\mathbf{Z_{2s}^S\hat{G}_2^S})}]\mathbf{r}_2^S + [\mathbf{\hat{M}_i} \otimes \frac{\mathbf{Z_{2c}^C\hat{G}_2^C}}{sd(\mathbf{Z_{2c}^C\hat{G}_2^C})}]\mathbf{r}_2^C \\
& + v_c + (v_s - v_c) + (v_n - v_s) + (v_i - v_n) \qquad (22)
\end{aligned}
$$

If we excluded the interaction terms and used a simple probit model, which norms the variance of the composite error term to 1, then by construction the parameter α_1 would equal $sd(\mathbf{X_i\hat{B}})$, α_2 would equal $sd(\mathbf{X_n\hat{G}_1^N})$, α_3 would equal $sd(\mathbf{X_s\hat{G}_2^S})$, and so on. We allow the estimates $\{\hat{\alpha}_k\}$ to differ from 1 because the mixed level probit model norms $Var(v_i - v_n)$ to 1 instead. In practice, estimates are very close to the implied values. Recall that all variables, including the indices, are entered as deviations from samplewide means, so the main effects can be interpreted as effects at the mean.

We experienced computational difficulties when attempting to estimate (22) via maximum likelihood. Instead, we adopt a Bayesian approach to estimating the slope parameters and the variances $Var(v_c)$, $Var(v_s - v_c)$, $Var(v_n - v_s)$, and $Var(v_i - v_n)$. Specifically, we assign prior distributions from which our slope parameters and random effect variances are drawn, and use Markov-chain Monte Carlo methods to estimate the means of the posterior distributions governing these parameters. We use these posterior means in place of the fixed parameters defined in (22) when reporting results and computing estimates of the impact of alternative group-level treatments below. The estimates are not very sensitive to modest changes in the priors for $Var(v_c)$, $Var(v_s - v_c)$, $Var(v_n - v_s)$. We treat the Bayesian approach as a computational device.

Step 3: Adjust estimates of α_1-α_5 and error component variances for sampling error in $\mathbf{\hat{B}}$, $\mathbf{\hat{G}_1^N}$, $\mathbf{\hat{G}_1^S}$, $\mathbf{\hat{G}_2^S}$, and $\mathbf{\hat{G}_2^C}$.

The parameter α_3 captures $sd(\mathbf{X_sG_1^S})$, where the standard deviation is taken over the student-weighted distribution of $\mathbf{X_sG_1^S}$. $\hat{\alpha}_3$ corresponds to $sd(\mathbf{X_s\hat{G}_1^S})$. It will be biased upward by sampling error in $\mathbf{\hat{G}_1^S}$. The same issue arises for the other α parameters. Because the hierarchical nature of the model and the control function strategy we adopt requires the use of a large number of variables at multiple group levels, the bias from sampling error is not negligible given our sample sizes. We

terms at their values from the previous iteration ($\mathbf{X_iB^{k-1}}$, $\mathbf{X_nG_1^{N,k-1}}$, $\mathbf{X_sG_1^{S,k-1}}$, $\mathbf{Z_{2s}^SG_2^{S,k-1}}$, and $\mathbf{Z_{2c}^CG_2^{C,k-1}}$). The iterative routine ends when successive iterations produce sufficiently similar parameter estimates.

separate the signal and noise subcomponents of $\hat{\alpha}_3$ via:

$$[\hat{\alpha}_3]^2 = Var(\mathbf{X_{s(i)}}\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S}) = \frac{1}{N}\sum_i(\mathbf{X_{s(i)}}\mathbf{G_1^S}\mathbf{G_1^{S\prime}}\mathbf{X}'_{\mathbf{s(i)}}) + \frac{1}{N}\sum_i \mathbf{X_{s(i)}}[\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S} - \mathbf{G_1^S}][\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S} - \mathbf{G_1^S}]'\mathbf{X}'_{\mathbf{s(i)}}. \quad (23)$$

The expectation of the second (sampling variance) term, conditional on $\mathbf{X_{s(i)}}$, is

$$\frac{1}{N}\sum_i \mathbf{X_{s(i)}}\mathbf{Var}(\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S})\mathbf{X}'_{\mathbf{s(i)}} \quad (24)$$

Using (24) we generate a bias-adjusted estimate of α_3 via $[(\hat{\alpha}_3)^2 - \frac{1}{N}\sum_i \mathbf{X_{s(i)}}\widehat{\mathbf{Var}}(\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S})\mathbf{X}'_{\mathbf{s(i)}}]^{.5}$, where $\hat{\alpha}_3$ is the estimate from the multilevel mixed effects model in Step 2. We use estimates of $\widehat{\mathbf{Var}}(\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S})$ based on the formula for the asymptotic variance of our Step 1 probit estimates. We account for clustering at the commuting zone level but not for the use of imputed data. We perform similar adjustments to the Step 2 estimates of the other α parameters, but not to the coefficients on the interactions (where we conjecture that the sampling error in the indices might bias estimates of interaction effects toward zero). Note, however, that the sampling error in the regression indices reflects true contributions of the error components, so that such sampling error implies downward bias in estimates of the variances of the error components $Var(v_c)$, $Var(v_s - v_c)$, $Var(v_n - v_s)$, and $Var(v_i - v_n)$. We discuss how we allocate the bias adjustments across these error components to remove this downward bias in Web Appendix A3.[21]

In the case of the wages, we construct an estimate of the variance of the permanent component of v_i under the assumption that the permanent component of the wage makes up the same 45.7% share of cross-sectional wage variance at age 25 in the ELS and NELS samples as it did for the high school class of 1972 cohort (NLS72) examined in AM (2016).

The standard errors in the paper are based on a bootstrap approach encompassing the entire estimation procedure, including the construction of $\mathbf{X_s}$, imputation of missing data, and the bias corrections to the α parameters. In particular, they account for the fact that the coefficients that define the index variables used in the second step are estimated.[22] Across outcome variables, data sets, and specifications, we find that the mean of the bootstrap replications of the bias corrected slope coefficients on $\frac{\mathbf{X_n}\hat{\mathbf{G}}_\mathbf{1}^\mathbf{N}}{sd(\mathbf{X_n}\hat{\mathbf{G}}_\mathbf{1}^\mathbf{N})}$, $\frac{\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S}}{sd(\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1}^\mathbf{S})}$, $\frac{\mathbf{Z_{2s}^S}\hat{\mathbf{G}}_\mathbf{2}^\mathbf{S}}{sd(\mathbf{Z_{2s}^S}\hat{\mathbf{G}}_\mathbf{2}^\mathbf{S})}$, and $\frac{\mathbf{Z_{2c}^C}\hat{\mathbf{G}}_\mathbf{2}^\mathbf{C}}{sd(\mathbf{Z_{2c}^C}\hat{\mathbf{G}}_\mathbf{2}^\mathbf{C})}$ are above the point estimates. Not

[21] If the bias adjusted estimate of a variance is negative, we set it to 0. The bias adjusted estimate of the covariance between two terms is set to 0 if the estimate of the variance of one of the terms is 0. If the correlation between two terms implied by bias adjusted estimates exceeds 1 in absolute value, we adjust the covariance to make the correlation 1 in absolute value.

[22] The standard error estimates are based on re-sampling commuting zones with replacement. To preserve the size distribution of the samples of students from particular commuting zones, we divide the sample into ten CZ sample size classes and resample CZs within class. For CZs in the largest size class, we break the CZs into two groups, each containing half the schools. We sample these half CZs instead to prevent any one bootstrap cluster from accounting for too large a share of the sample. Due to the considerable computational burden of the model estimation and the simulations relative to the computer resources available for use with the restricted-use versions of ELS and NELS, we use 100 replications to form the bootstrap estimates for this submission. Furthermore, we use fewer (4) MCMC chains of length 300 rather than the 20 chains of length 500 that we use to compute the point estimates. This might lead to an overstatement of the standard errors.

surprisingly, the disparity is even greater for the bias corrected variance component estimates based on them, which underlie the variance decompositions in Table 3-1 to 3-4. The distribution of the bootstrap estimates display a right skew as well as a mean shift. For this reason, we report 5th and 95th percentile values of the bootstrap distribution of our estimated variance components. The distribution of the estimates of the treatment effects of a 10th-90th percentile shift in school and/or commuting zone quality are wider than we would like, but are less sensitive to the issue. More work on the best way to implement bias corrections and do statistical inference is needed.

6.3 Variance Decomposition Methodology

Here we describe the simple variance decomposition procedure that we use to 1) provide an initial, descriptive sense of the relative importance of inputs at each level (individual, neighborhood, school, commuting zone) in determining the outcomes of interest, and 2) assess the degree to which amenities at the neighborhood, school, and commuting zone level are driving student sorting.

To simplify the empirical analysis we define neighborhoods to be nested within schools, which are themselves nested within commuting zones.[23] We consider a linear version of our estimating equation in which the vector of interaction coefficients $[r_1^N, r_1^S, \mathbf{r_2^S}, \mathbf{r_2}]$ is restricted to equal $\mathbf{0}$.[24]

$$Y_i = \mathbf{X_i B} + \mathbf{Z_n^N G^N} + \mathbf{Z_s^S G^S} + \mathbf{Z_c^C G^C} + (v_i - v_n) + (v_n - v_s) + (v_s - v_c) + v_c. \tag{25}$$

One can then write the outcome as the sum of orthogonal components:

$$Y_i = (Y_i - Y_n) + (Y_n - Y_s) + (Y_s - Y_c) + Y_c. \tag{26}$$

Because the components in (26) are mutually orthogonal, $Var(Y_i)$ is:

$$Var(Y_i) = Var(Y_i - Y_n) + Var(Y_n - Y_s) + Var(Y_s - Y_c) + Var(Y_c) \tag{27}$$

While characterizing sorting on unobservable student characteristics is daunting (even our control function approach absorbs group-level sorting at the expense of absorbing part of the group treatment effect), we can use an analogous four-component decomposition to analyze the structure

[23] In the empirical work our narrowest definition of neighborhood is the census block group. In a few cases, students from the same block group choose different schools. We treat students from the same block group who attend different schools as if they live in different block groups by using school-block group combinations as our definition of a neighborhood. Nesting neighborhood within school greatly simplifies the empirical analysis and is required by the routine we use to estimate the variances of the error components. But it may diminish the precision of our estimates of neighborhood and school effects on outcomes. The same issue arises when zip code is the neighborhood definition. Note that the choice model in Section 2 allows families who live in a given neighborhood to choose different schools.

[24] When interactions are included in the model, the within-neighborhood variance will differ across neighborhoods based on the degree to which students sensitive to environment are located in neighborhoods/schools offering more supportive environments, undermining the validity of a simple decomposition.

of sorting on the index of observable characteristics $\mathbf{X_i B}$ that best predicts the outcome:

$$Var(X_iB) = Var((X_i - X_n)B) + Var((X_n - X_s)B) + Var((X_s - X_c)B) + Var(X_cB) \tag{28}$$

This is likely to be a reasonable approximation of the overall structure of sorting if either observable characteristics drive the bulk of sorting or if the relative weights in the taste matrix $\mathbf{\Theta}^U$ placed on amenities at the neighborhood, school, and commuting zone levels by unobservable characteristics mirror the weights implicit in $\mathbf{\Theta}$.

6.4 Measuring the Effect of Shifts in School and Commuting Zone Quality

While decompositions of variance provide an overall sense of the relative importance of individual, neighborhood, school, and commuting zone inputs in determining later educational outcomes, they do not permit one to easily gauge the impact that a substantial improvement in external environment can have on a student's expected educational attainment. Consequently, in this section we extend to the non-linear context AM's methodology for converting variance components into expected impacts on outcomes of particular "treatments", or shifts in school and/or community input quality. Here we describe the three distinct treatments we consider.

6.4.1 The Combined School and Commuting Zone Treatment

First, we evaluate the expected change in outcomes associated with moving a randomly chosen student from a school/commuting zone combination at the 10th percentile of the combined school/commuting zone quality distribution to the 90th percentile. From our production function (7), the true distribution of combined school/commuting zone quality may be defined as the distribution of $\mathbf{Z_s^{S*}}\mathbf{\Gamma}^{S*} + \mathbf{Z_c^{C*}}\mathbf{\Gamma}^{S*}$. We approximate this combined school/commuting zone quality distribution using the distribution of $T = \mathbf{Z_{2s}^S G_2^S} + \mathbf{Z_{2c}^C G_2^C} + (v_s - v_c) + v_c$, where we use T to denote the particular "treatment" chosen. Because we exclude the control function $\mathbf{X_n G_1^N} + \mathbf{X_s G_1^S}$, which may capture peer effects and other unobserved school and commuting zone inputs in $[\mathbf{Z^{S*}}, \mathbf{Z^{C*}}]$ in addition to student sorting, our estimated impacts of 10th-to-90th percentile shifts in school/commuting zone quality will likely understate the impact of the corresponding 10th-to-90th percentile shifts in $\mathbf{Z_s^{S*}}\mathbf{\Gamma}^{S*} + \mathbf{Z_c^{C*}}\mathbf{\Gamma}^{S*}$. However, as the discussion in Section 6.3 made clear, to the extent that unobserved neighborhood inputs are clustered in particular schools and commuting zones, such inputs could contribute to $\mathbf{G_2^S}$, $\mathbf{G_2^C}$, $v_s - v_c$, and v_c. To the extent that such clustering is significant, we could interpret our estimates instead as lower bound estimates of the impact of a shift in the combined neighborhood/school/commuting zone quality index $\mathbf{Z_n^{N*}}\mathbf{\Gamma}^{N*} + \mathbf{Z_s^{S*}}\mathbf{\Gamma}^{S*} + \mathbf{Z_c^{C*}}\mathbf{\Gamma}^{S*}$.

Building an estimator of the impact of these quantile shift "treatments" is complicated by the interaction terms in (10). First, we assume that the treatment distribution $T \equiv \mathbf{Z_{2s}^S G_2^S} + \mathbf{Z_{2c}^C G_2^C} + (v_s - v_c) + v_c$ is normally distributed, so that the q-th treatment quantile (denoted T^q) is given by $T^q = \hat{Var}(T)^{.5}\Phi^{-1}(q)$ where $\Phi(*)$ is the CDF of the standard normal distribution and $\Phi^{-1}(*)$ is its inverse. Next, note that the interaction terms $\mathbf{M_i} \otimes \mathbf{Z^S_{2s}G^S_2r^S_2}$ and $\mathbf{M_i} \otimes \mathbf{Z^C_{2c}G^C_2r^C_2}$ depend separately on the subcomponents $\mathbf{Z^S_{2s}G^S_2}$ and $\mathbf{Z^C_{2c}G^C_2}$ of the full treatment T. We handle this by effectively integrating over the joint conditional distribution $f(\mathbf{Z^S_{2s}G^S_2}, \mathbf{Z^C_{2c}G^C_2}|T = T^q)$. We do this by taking P draws of the vector $[\mathbf{Z^S_{2s}G^S_2}, \mathbf{Z^C_{2c}G^C_2}]$ from the appropriate joint multivariate normal conditional distribution and averaging our predicted outcomes over these P draws. The parameters of that distribution are based on the bias-corrected estimates of the variances and covariances of the components of the vector. In our main results we also integrate over the distribution of student and neighborhood inputs. We use the empirical joint distribution of observed student and neighborhood inputs from our sample by averaging over the observed $[\mathbf{X_iB}, \mathbf{X^N_n}, \mathbf{X^S_s}]$ vectors of all I students in the sample.

Thus, our estimator of the expected outcome at a chosen quantile q of the "treatment effect" distribution is:

$$\begin{aligned} E[\hat{Y}^q] = \frac{1}{P}\sum_p \frac{1}{I}\sum_i \Phi(&\mathbf{X_i\hat{B}} + \mathbf{X^N_n\hat{G}^N_1} + \mathbf{X^S_s\hat{G}^S_1} + T^q \\ &+ (\mathbf{X_i\hat{B}})(\mathbf{X^N_n\hat{G}^N_1})\hat{r}^N_1 + (\mathbf{X_i\hat{B}})(\mathbf{X^S_s\hat{G}^S_1})\hat{r}^S_1 + \mathbf{M_i} \otimes (\mathbf{Z^S_{2s}\hat{G}^S_2})_p \mathbf{\hat{r}^S_2} \\ &+ \mathbf{M_i} \otimes (\mathbf{Z^C_{2c}\hat{G}^C_2})_p \mathbf{\hat{r}^C_2})/(1 + Var(v_n - v_s)) \end{aligned} \tag{29}$$

where $(\mathbf{Z^S_{2s}G^S_2})_p$ and $(\mathbf{Z^C_{2c}G^C_2})_p$ represent the p-th draws of these regression indices from the conditional joint distribution $f(\mathbf{Z^S_{2s}G^S_2}, \mathbf{Z^C_{2c}G^C_2}|T = T^q)$. Note that $Var(v_i - v_n)$ has been normalized to 1 in the denominator, since the scale of the latent index for binary outcomes is not identified.

We then compute the difference $E[\hat{Y}^{90}] - E[\hat{Y}^{10}]$ to estimate the change in expected outcome (e.g. the increase in the probability of high school graduation) for a randomly chosen high school student from a 10th-to-90th percentile shift in school/commuting zone quality. Alternatively, this quantity can be thought of as the increase in the population average outcome if we placed every student in a 10th percentile school/commuting zone, and then moved them each to a 90th percentile school/commuting zone, but held the distribution of peer effects fixed as it was in our sample. We refer to this counterfactual as the "School and CZ" counterfactual in our tables and discussion.

6.4.2 The School Treatment

The second counterfactual thought experiment or "treatment" we consider consists of only replacing each student's school inputs with those of the school at the 10th percentile versus 90th percentile of the school quality distribution (defined by $\mathbf{Z^{S*}_s\Gamma}^{S*}$), holding neighborhood and commuting zone inputs fixed. We approximate the distribution of true school quality $\mathbf{Z^{S*}_s\Gamma}^{S*}$ with the distribution of $T \equiv \mathbf{Z^S_{2s}G^S_2} + (v_s - v_c)$. Our analysis in Section 4 suggests that both $\mathbf{Z^S_{2s}G^S_2}$ and $(v_s - v_c)$ should be purged of any observed or unobserved student inputs by virtue of including $\mathbf{X_s}$ in the control function. However, just as in our first treatment, $\mathbf{Z^S_{2s}G^S_2} + v_s$ will not include any peer effects or unobserved inputs that are predicted by $\mathbf{Z^S_{1s}}$, and it may include unobserved neighborhood inputs, to the extent that they cluster at the school (but not CZ) level. As we discussed in section 6.3, if the full set of commuting zone averages of $\mathbf{Z^S_{2s}G^S_2}$ were included in $\mathbf{Z^C_{2c}}$, then we could rule out the possibility that any unobserved commuting zone inputs could project onto $\mathbf{Z_{2s}^{S}}$ and be captured by $\mathbf{G_2^S}$. However, to guard against overfitting we only include the commuting zone average of the single index $\mathbf{Z_{2s}^{S}G_2^S}$ in $\mathbf{Z_{2c}^{C}}$, so that our "school only" treatment could in principle be slightly influenced by unobserved commuting zone inputs as well. Thus, the counterfactual impacts we estimate are simply the best approximation of the true counterfactual impacts $E[Y|T^{90}]-E[Y|T^{10}]$ that our data and model allow. We refer to this counterfactual as the "School only" counterfactual in our tables and discussion.

Specifically, our estimator of the expected outcome for a randomly chosen student who is assigned a school at the q-th percentile of quality is:

$$E[\hat{Y}^q] = \frac{1}{P}\sum_p \frac{1}{I}\sum_i \Phi(\mathbf{X_i\hat{B}} + \mathbf{X_n^N\hat{G}_1^N} + \mathbf{X_s^S\hat{G}_1^S} + T^q + (\mathbf{Z_{2c}^C G_2^C})_p + (v_c)_p \tag{30}$$

$$+(\mathbf{X_i\hat{B}})(\mathbf{X_n^N\hat{G}_1^N})\hat{r}_1^N + (\mathbf{X_i\hat{B}})(\mathbf{X_s^S\hat{G}_1^S})\hat{r}_1^S + \mathbf{M_i}\otimes(\mathbf{Z_{2s}^S\hat{G}_2^S})_p\mathbf{\hat{r}_2^S} \tag{31}$$

$$+\mathbf{M_i}\otimes(\mathbf{Z_{2c}^C\hat{G}_2^C})_p\mathbf{\hat{r}_2^C})/(1+Var(v_n-v_s)) \tag{32}$$

where $(\mathbf{Z_{2s}^S G_2^S})_p$ represents the p-th draw from the conditional distribution $f(\mathbf{Z_{2s}^S G_2^S}|T=T^q)$ and $(\mathbf{Z_{2c}^C G_2^C})_p$ and $(v_c)_p$ represent the p-th draws from the unconditional joint distribution $f(\mathbf{Z_{2c}^C G_2^C}, v_c)$. This counterfactual attempts to isolate the sensitivity of student outcomes to the quality of the school as distinct from the inputs of the surrounding commuting zone.

6.4.3 The Commuting Zone Treatment

The third counterfactual "treatment" replaces each student's commuting zone inputs with those of the commuting zone at the 10th percentile versus 90th percentile of the CZ quality distribution defined by $\mathbf{Z_c^{C*}}\mathbf{\Gamma}^{C*}$, holding neighborhood and school inputs fixed. Importantly, this treatment does not include the subcomponent of $\mathbf{Z_s^{S*}}\mathbf{\Gamma}_2^{S*}$ that varies across commuting zones. Thus, this counterfactual is designed to gauge the importance of commuting zone-level inputs in their own right, rather than the importance of the choice of commuting zone in which to live (which combines commuting zone inputs with differences in the distributions of school inputs across commuting zones). We approximate the distribution of true commuting zone quality $\mathbf{Z_c^{C*}}\mathbf{\Gamma}^{C*}$ with the distribution of $T\equiv \mathbf{Z_{2c}^C G_2^C}+v_c$ (and we do not include the between-CZ component of $\mathbf{Z_{2s}^S G_2^S}$, since G_2^S is identified using between-school/within-CZ variation and is designed to capture school effects).[25] This is again an approximation due to the possibility that $\mathbf{Z_{2c}^C G_2^C}$ and v_c could partly reflect unobserved school or neighborhood inputs, to the extent that good schools or neighborhoods tend to cluster in particular commuting zones. We refer to this counterfactual as the "CZ only" counterfactual in our tables and discussion.

Our estimator of the expected outcome for a randomly chosen student who is assigned a com-

[25]Recall from footnote 19 in section 6.2 that the commuting zone average of $Z_{2s}^S G_2^S$ enters our estimating equation as a separate commuting zone variable with its own coefficient. Since the coefficient on this index is identified purely from between-commuting zone variation, we include this index as part of our commuting zone treatment.

muting zone at the q-th percentile of quality is:

$$E[\hat{Y}^q] = \frac{1}{P}\sum_p \frac{1}{I}\sum_i \Phi(\mathbf{X_i}\hat{\mathbf{B}} + \mathbf{X^N_{1n}}\hat{\mathbf{G}}^{\mathbf{N}}_{\mathbf{1}} + \mathbf{X^S_{1s}}\hat{\mathbf{G}}^{\mathbf{S}}_{\mathbf{1}} + (\mathbf{Z^S_{2s}}\mathbf{G^S_2})_p + (v_s - v_c)_p + T^q$$
$$+ (\mathbf{X_i}\hat{\mathbf{B}})(\mathbf{X^N_n}\hat{\mathbf{G}}^{\mathbf{N}}_{\mathbf{1}})\hat{r}^N_1 + (\mathbf{X_i}\hat{\mathbf{B}})(\mathbf{X^S_s}\hat{\mathbf{G}}^{\mathbf{S}}_{\mathbf{1}})\hat{r}^S_1 + \mathbf{M_i} \otimes (\mathbf{Z^S_{2s}}\hat{\mathbf{G}}^{\mathbf{S}}_{\mathbf{2}})_p \hat{\mathbf{r}}^{\mathbf{S}}_{\mathbf{2}}$$
$$+ \mathbf{M_i} \otimes (\mathbf{Z^C_{2c}}\hat{\mathbf{G}}^{\mathbf{C}}_{\mathbf{2}})_p \hat{\mathbf{r}}^{\mathbf{C}}_{\mathbf{2}})/(1 + Var(v_n - v_s)) \tag{33}$$

where $(\mathbf{Z^S_{2s}G^S_2})_p$ and $(v_s - v_c)_p$ are the p-th draws from the unconditional joint distribution of $(\mathbf{Z^S_{2s}G^S_2})$ and $(v_s - v_c)$ and $(\mathbf{Z^C_{2c}G^C_2})_p$ and $(v_c)_p$ are the p-th draws from the conditional joint distribution $f(\mathbf{Z^C_{2c}G^C_2}, v_c | T \equiv \mathbf{Z^C_{2c}G^C_2} + v_c = T^q)$.

6.4.4 Estimating Impacts of Shifts in School and Commuting Zone Quality for Particular Subpopulations

In contrast to AM (2016), the introduction of interaction terms in the estimated production function (10) in this paper allows us to characterize the degree to which the outcomes of specific subpopulations are particularly sensitive to the quality of external inputs at the neighborhood, school, or commuting zone levels.[26] Thus, in this subsection we briefly describe how to extend the methodology introduced in the last subsection to capture treatment heterogeneity across particular subpopulations.

The most straightforward approach is simply to restrict the sample used for the counterfactual treatments to members of a particular subpopulation. We report results from the following subpopulations: Hispanic students, non-Hispanic black students, students in a single-mother household where the mother has a high-school education or less, and students in a two-parent-college-educated household. We use the empirical distribution of individual and neighborhood inputs $\mathbf{X_i}\hat{\mathbf{B}} + \mathbf{X^N_n}\hat{\mathbf{G}}^{\mathbf{N}}_{\mathbf{1}} + \mathbf{X^S_s}\hat{\mathbf{G}}^{\mathbf{S}}_{\mathbf{1}}$, so restricting the sample naturally imposes the chosen sample's joint distribution of observed individual and neighborhood inputs. Furthermore, recall that the unobserved components $v_i - v_n$ and $v_n - v_s$ are defined to be uncorrelated with all of the observable characteristics used to define the subpopulation. Thus, the formulas (29) - (33) are still valid, with i and I now indexing the particular individual and number of individuals among the chosen subpopulation. All elements of $\mathbf{M_i}$ take on the values for i, so that the results for Hispanic students, for example, reflect not only the interaction terms involving the minority (non-Hispanic black or Hispanic) indicator but also differences across groups in the distribution of the other elements of $\mathbf{M}_i$, such as low income status, weighted by the corresponding elements of the interaction coefficients $\hat{\mathbf{r}}^{\mathbf{S}}_{\mathbf{2}}$ and $\hat{\mathbf{r}}^{\mathbf{C}}_{\mathbf{2}}$.

However, our rich set of observed individual characteristics also allow us to investigate the degree to which school and CZ treatment effect heterogeneity is related to the individual's own contribution to the outcome. To do this, we fix $\mathbf{X_i B}$ at each quintile dividing point $[.05, \ldots, .95]$ in its empirical distribution in the sample, and compute the change in expected outcome for each

[26]The only non-linearity in AM (2016) came from the probit function for binary outcomes.

of our three counterfactual quality shifts ("School and CZ", "School only", and "CZ only", described above) for randomly chosen individuals at the chosen quintile of $\mathbf{X_i B}$. We integrate over the joint distribution of $v_i - v_n$, $v_n - v_s$, $\mathbf{X_n^N G_1^N}$ and $\mathbf{X_s^S G_1^S}$. This means that we are not holding fixed the kind of neighborhood such students tend to experience, but are instead randomly assigning a neighborhood from the full population distribution for both the low ($E[Y|T^{10}]$) and high ($E[Y|T^{90}]$) school/commuting zone treatments. Specifically, the expected outcome of a randomly chosen student at a particular $\mathbf{X_i B}$ percentile q' (denoted $(\mathbf{X_i \hat{B}})^{q'}$ below) who is assigned a school-commuting zone combination at the q-th percentile in the "School and CZ" counterfactual is estimated via:

$$\begin{aligned} E[\hat{Y}^q] = \frac{1}{P}\sum_p \frac{1}{I}\sum_i \Phi((\mathbf{X_i \hat{B}})^{q'} + \mathbf{X_n^N \hat{G}_1^N} + \mathbf{X_s^S \hat{G}_1^S} + T^q \\ + (\mathbf{X_i \hat{B}})^{q'}(\mathbf{X_n^N \hat{G}_1^N})\hat{r}_1^N + (\mathbf{X_i \hat{B}})^{q'}(\mathbf{X_s^S \hat{G}_1^S})\hat{r}_1^S + \mathbf{M_i}^{q'} \otimes (\mathbf{Z_{2s}^S \hat{G}_2^S})_p \mathbf{\hat{r}_2^S} \\ + \mathbf{M_i}^{q'} \otimes (\mathbf{Z_{2c}^C \hat{G}_2^C})_p \mathbf{\hat{r}_2^C})/(1 + Var(v_n - v_s)), \end{aligned} \tag{34}$$

where $\mathbf{M}_i^{q'} = [(\mathbf{X_i B})^{q'}, 1(Female), 1(URM), 1(Low\_Income)]$. We can then examine how sensitivity to school and commuting zone inputs systematically varies as one moves through the distribution of observed student contributions.

Note that, just as in AM (2016), part of the differential sensitivity will arise from the nonlinearity of the probit function. In other words, even if each of the interaction parameters $\{r_1^N, r_1^S, \mathbf{r_2^S}, \mathbf{r_2^C}\}$ were equal to zero, we should still expect heterogeneity in outcome sensitivity even in the absence of heterogeneity in sensitivity of the latent probit index to the school and location treatments. This is because some subpopulations have a larger mass of students near the decision margin (as evidenced by different mean outcome values). Below we sometimes compare the predicted effects we obtain from subpopulation-specific 10th-to-90th quantile shifts based on the non-linear probit index model to the corresponding predicted effects based on the linear version of our estimating equation (25) in which the vector of interaction coefficients is restricted to $\mathbf{0}$. We do this to better gauge the degree to which the treatment effect heterogeneity generated by our full model is well-approximated by the more parsimonious model that removes interaction terms.

7 Results: Main Estimates and Variance Decompositions

In section 7.1 we discuss the estimates of (22). In section 7.2 we present the decompositions of variance discussed in (6.3) for each of our outcomes from NELS and ELS. These are based on the linear version of our estimating equation (25) in which the interaction parameters $\{r_1^N, r_1^S, \mathbf{r_2^S}, \mathbf{r_2^C}\}$ have been restricted to zero. We focus on the estimates for the "full" set of $\mathbf{X_i}$ variables. Model estimates and variance decompositions for the "basic" set are in the supplemental appendix. As expected, they usually imply a more important role for school and commuting zone factors in education and wages. We discuss treatment effect estimates based on them in Section 8

7.1 Model Estimates

Since we do not seek to interpret any of the particular elements of the coefficient vectors $\mathbf{B}$, $\mathbf{G}_1$, or $\mathbf{G}_2$, we focus here on characterizing the relative importance of the regression indices at each school and location level, as well as the strength of the interaction coefficients. To this end, the columns of Table 2-1 present bias-corrected estimates of (22) for the educational attainment outcomes.[27] Keep in mind that the indices are normed to have a standard deviation of one (they were already mean zero by construction), so that the coefficients α_1-α_5 capture the degree to which a one standard deviation change in a given regression index increases the latent probit indices that determine the outcomes we consider. Table 2-2 presents a corresponding set of estimates for $\ln(wage)$ at about age 25. Zero values for point estimates of the α parameters indicate that the bias correction led to a negative value.

First, note that the individual-level observable index $\mathbf{X_i B}$ is an extremely powerful predictor of educational attainment outcomes, particularly ENROLL and COLLBA. The estimate is similar when we exclude interaction terms (See Web Appendix Table 7-1). As we will demonstrate in Figure 3a below, one can predict educational attainment quite well using observed student and family characteristics even in 8th grade (NELS results). In the case of $\ln(wage)$, a one standard deviation increase in $\mathbf{X_i B}$ raises the permanent wage by 0.165 in the case of ELS and 0.134 in the case of NELS. These values are substantial relative to the standard deviation of permanent wages, which we estimate to be 0.338 for ELS and 0.361 for NELS. The mid-20's wages used here do not fully capture the divergence that will occur later in life.

The second notable pattern is that the index of observed neighborhood (zip code or block group) characteristics taken from Census Bureau and LODES data demonstrates very little predictive power across outcomes and datasets. Indeed, once the bias correction has been applied, in 5 of 12 cases we find that there is no more variance in the observed neighborhood index than we would expect due to sampling variance (so that the bias corrected coefficient is zeroed out).

Third, the group-level unobserved components are small to modest in size for each outcome, with the neighborhood component being the most important followed by the school and then the commuting zone. The standard deviation of $v_n - v_s$ is close to 0.15 at the block group level in ELS for all 3 education outcomes. It is around 0.11 at the zipcode level across the three education outcomes for both ELS and NELS. Not surprisingly, these values are small relative to $\mathbf{X_i B}$ or the idiosyncratic component $v_i - v_n$, which has a standard deviation of 1. The estimates of the standard deviation of $v_s - v_c$ are tightly concentrated around 0.07 across data sets and specifications for HSGRAD, around 0.11 for ENROLL, and around 0.06 for COLLBA. We do not have strong priors as to whether to expect a larger value for the lower-level education outcomes than for COLLBA.

[27] As noted above, we do not use weights in the estimation because using them led to computational difficulties in estimating the error component variances. We have also experimented with estimating $\alpha_1 - \alpha_5$ and the interaction coefficients treating the error term as a composite error. The uncorrected estimates are very close to the uncorrected estimates that underlie the bias corrected estimates reported in the tables. Not surprisingly, discrepancies do arise for some of the coefficients on the interactions, which have large estimated standard errors. We do not know if the error variance estimates are sensitive to weighting.

Commuting zone error components v_c typically feature standard deviations about half to two thirds as large as the school level component $v_s - v_c$.

In the case of wages, the standard deviation of the neighborhood random component $v_n - v_s$ is .043 (.008) log points in the ELS block group case. It is around .027 at the zip code level for both ELS and NELS. School effects and commuting zone effects are small, on the order of .01-.02 log points.

Next, we turn to the indices of observed group-level inputs that are not averages of individual inputs, $\mathbf{Z_{2s}^S}$ and $\mathbf{Z_{2c}^C}$. The ability of these indices to predict outcomes, along the standard deviations of $v_s - v_c$ and v_c, form the core of our counterfactual shifts in school- and commuting zone-level inputs presented later in this section.

The estimated (bias-corrected) standard deviations of $\mathbf{Z_{2c}^C}$ ($\hat{\alpha}_5$) are relatively robust across alternative neighborhood definitions, linear/nonlinear specifications, and datasets for both high school graduation and four-year college enrollment. The values range between .10 and .15, with slightly smaller NELS88 values for high school graduation and slightly larger values for college enrollment. Standard deviations for college graduation are somewhat smaller, between .06 and .12. As we will see in Section 8, these commuting zone inputs predict enough outcome variation to imply relatively large impacts from shifting from low to high commuting zone environments. Standard errors are typically around .035 and tend to be larger in models with interaction terms.

The estimated (bias-corrected) standard deviations of $\mathbf{Z_{2s}^S}$ ($\hat{\alpha}_4$) are mostly of similar size to (or slightly smaller than) $\mathbf{Z_{2c}^C}$, suggesting similar importance for school and commuting-zone level inputs. Recall that commuting zone averages of $\mathbf{Z_{2s}^S}\hat{\mathbf{G}}_\mathbf{2}$ are included in $\mathbf{Z_{2c}^C}$, so that $\hat{\alpha}_4$ is based only on within-CZ variation across schools, preventing CZ-level inputs from being captured by $\hat{\alpha}_4$. These estimates are also generally robust to excluding interactions and changing the neighborhood definition, with the notable exception of high school graduation in ELS. In the ELS case, $\hat{\alpha}_4$ is much smaller with a much larger standard error ($\hat{\alpha}_4 =$.009 (.073)) than when interactions are restricted to 0 (0.124 (.051)), particularly when block group is used to designate a neighborhood. While we do not have a convincing explanation for the sensitivity to interactions in this one context, we should note that dropping out is a rare outcome in the unweighted ELS sample (only 8%), which may lead these estimates to be less robust. Indeed, the corresponding ELS high school graduation results for the basic specification (displayed in Web Appendix Table 1-1) display a much smaller drop in $\hat{\alpha}_4$.

Neither $\mathbf{Z_{2s}^S}$ nor $\mathbf{Z_{2c}^C}$ predicts log wages particularly well, with $\hat{\alpha}_4$ generally between .01 and .03 log points and $\hat{\alpha}_5$ between .02 and .04 log points. $\hat{\alpha}_4$ is not statistically significant in either the ELS block group or zipcode specifications. This may reflect the early age at which log wages are measured in both NELS and ELS.

Finally, the estimates of the interaction coefficients are somewhat noisy. In the case of the interactions between $\mathbf{X_iB}$ and the neighborhood composition index $\mathbf{X}_n^N\mathbf{G_1^N}$, the estimates of r_1^N are negative and statistically significant in the equations for *ENROLL* and *COLLBA*, suggesting that disadvantaged students disproportionately benefit from living in a stronger neighborhood. However, we do not make too much of this result because the main effect of $\mathbf{X_n^N G_1^N}$ is weak and the result does not carry over to high school graduation or wages. The estimates of r_1^S also tend to be small and statistically insignificant (partly due to large standard errors), and the signs vary across outcomes and data sets. The sensitivity of the estimates partly reflects the relatively low predictive power of $\mathbf{Z_{2s}G_2}$ (particularly for ELS high school graduation), so that there is limited variation with which to identify interactions with $\mathbf{Z_{2s}G_2}$. Alternatively, for the binary outcomes in particular, this may suggest that the non-linear probit function that maps the probit index into the outcome probability does a fairly good job of capturing the differential sensitivity of students with higher and lower observed inputs to group-level inputs. We investigate this further in section 8.

7.2 Decompositions of Variance

7.2.1 The Extent of Neighborhood-, School-, and Commuting Zone-Level Clustering in Outcomes

Columns 1-3 of Table 3-1 report the decomposition of the variance of the latent probit index Y_i that determines high school graduation into components at the individual, neighborhood, school, and commuting zone levels based on the linear specification (25) that excludes the interaction terms. We report the 5th and 95th percentiles of the bootstrap replications in brackets and give a few examples in the text. The range is often substantial relative to the point estimates. Figure 1 displays the decomposition for each dataset and outcome graphically.

The first row of Column 1 suggests that about 82.3% of the variance in the latent index determining graduation exists among students living within the same grade 8 neighborhoods in NELS. Row 2 shows that just 1.2% [0.5%, 1.6%] consists of different neighborhood averages among students attending the same school, while Row 3 indicates that another 10.1% [8.0%, 13.5%] consists of different school averages among students residing in the same commuting zone. Finally, row 4 shows that the remaining 6.4% [5.5%, 9.0%] consists of differences in commuting zone averages. While the small neighborhood component is surprising, note that this does not necessarily mean that neighborhood effects are unimportant. Rather, it could be that high and low quality neighborhoods tend to be sufficiently grouped within school attendance zones so that most of the variation in neighborhood inputs exists across schools. As discussed in section 6.3, our variance decompositions are crafted to prevent higher-level inputs (eg., commuting zone) from being reflected in lower-level variance components, but between-school variation in neighborhood quality will still be captured by the between-school variance components. Alternatively, ZIP code may be too coarse a measure of neighborhood, with most ZIP codes containing a mix of both good and bad neighborhoods (so that the bulk of neighborhood effect variation is assigned to the "within-neighborhood" category in Table 3-1). The larger neighborhood-level variance component in the ELS block group results suggest that this is part of the story.

Column 3 presents a variance decomposition of the determinants of HSGRAD for ELS using zip code. A greater share of the variance is now within-neighborhood (84.2%), with slightly smaller shares at the neighborhood, school and commuting zone levels (1.2%, 9.1%, and 5.4% respectively). The ELS results using block group as the neighborhood definition (column 2) are similar, but the contribution of neighborhood is higher (2.1% versus 1.2%).

Columns 4-6 present the same decomposition for the latent index that determines enrollment in a four-year college. The NELS grade 8 results (Col. 4) display considerably larger school and commuting zone components (14.7% and 9.8%, respectively) at the expense of within-neighborhood variance (74.2%), with the neighborhood component only increasing to 1.4%. The ELS zip code results mirror the 10th grade NELS results except for a slight decrease in the within-neighborhood component (71.9%) and slightly larger school and commuting zone components (16.7 and 10.6%). The ELS block group results are similar, but the neighborhood component rises from 0.8% to 2.7% [2.0%,4.9%].[28]

Columns 7-9 show the results for COLLBA. The neighborhood, school, and commuting zone shares of variance are between those for high school graduation and college enrollment, but closer to the high school graduation shares.

The results for log wages (columns 10-12) show a much larger role for the variance of student level factors within neighborhood for NELS, with only 1.8% of the variance existing across schools/within commuting zones, and 6.9% existing across commuting zones. ELS results display a similarly-sized commuting zone component, but a much larger school component ($\sim$ 8.5%).

7.3 The Extent of Student Sorting at the Neighborhood, School, and Commuting Zone Levels

The decompositions in Table 3-1 convey the degree of outcome clustering at each level of aggregation, but they combine group-level differences that are attributable to student sorting with true group-level inputs. In this subsection we focus attention on the degree to which neighborhood, school, and commuting zone amenities and job opportunities segregate the population on the observable individual characteristics that best predict educational outcomes and wages. Specifically, the last row of Table 3-2 provides the fraction of the total variance in the latent index for each outcome that is attributable to the regression index of individual-level observables $\mathbf{X_i B}$. Rows 1 through 4 decompose this variance component into within-neighborhood, between-neighborhood/within-school, between-school/within-commuting zone, and between-commuting zone components to illustrate the nature of student sorting on outcome-relevant observables.

Column 1 displays the results for grade 8 neighborhoods, schools, and commuting zones in NELS for the index that determines high school graduation. Row 5 shows that individual observables have a fair amount of predictive power. They account for 22.9% of the total variance in the latent index Y_i. Rows 1-4 show that 73.4% of this variance is within neighborhood, while the

[28] It bears repeating that we are not using sampling weights due to computational difficulties and so are not accounting for differences in designs of the ELS and NELS samples. It is possible that this contributes to differences in the estimates across ELS and NELS.

neighborhood-level, school-level, and commuting zone level inputs each account for 1.4%, 14.4%, and 10.8% of $Var(\mathbf{X_i B})$, respectively. Consistent with the results in the previous section, there is virtually no evidence of outcome-relevant sorting across ZIP codes within school attendance zones. Perhaps this is simply an indication that ZIP codes are poor proxies for neighborhoods. The table shows that there is substantial student sorting on observables both across commuting zones and across schools within commuting zones. Such sorting underscores the need for econometric techniques to distill the contribution of school and commuting zone inputs, since a substantial portion of differences in raw outcome means across schools and commuting zones is directly attributable to differences in the kinds of students these schools and areas attract.

The ELS results using zipcode (col. 3) are similar but with a substantially larger value for schools (18.7%) and only 0.5 percent for zipcode. The ELS block group estimates (col. 3) are similar to the ELS zipcode results, but show a slightly larger value for neighborhood (2%) within school. Nonetheless, this suggests that school attendance is the critical level at which sorting occurs, rather than neighborhoods within attendance zones

Columns 4-6 display the corresponding decomposition of $Var(\mathbf{X_i B})$ for the latent index determining college enrollment. Interestingly, the bottom row indicates that the individual-level observable characteristics are much more powerful predictors of ENROLL than high school graduation; for ENROLL they account for around 34% of the total outcome variance in NELS and around 40% in the two ELS specifications. The relative importance of neighborhood, school, and commuting zone in driving sorting is essentially unchanged across outcomes.

The decompositions for the latent index determining college graduation (col. 7-9) are roughly similar to those for high school graduation and college enrollment; $Var(\mathbf{X_i B})$ accounts for between 32% and 39% of the total variance, and the relative importance of the different groups closely mirrors the high school graduation results.

In the case of permanent wages (col. 10-12), between 66% and 72% of the variance in $\mathbf{X_i B}$ is within neighborhood across datasets, with 16-19% of the variance existing between schools/within-commuting zones, and around 10-11% between commuting zones. The between-neighborhood/within school share is again much larger (4.7%) for the ELS block group specification than for ELS or NELS zip code specifications (around 1%), underscoring the importance of better finer neighborhood designations.

8 Results from Counterfactual Shifts in School and Commuting Zone Quality

This section reports results from the three types of counterfactual shifts in school and commuting zone quality described in section 6.4, both for the full population and for particular subpopulations. We organize the discussion by outcome. We focus on the treatment effects estimates for the full set of $\mathbf{X_i}$ variables but briefly discuss estimates using the basic set, which are typically larger.

8.1 High School Graduation

Columns 1-3 of Table 4 report the expected difference in the probability of graduating high school for a student randomly chosen from the full population between "treatments" in which the student grows up in a 10th versus 90th percentile school and/or commuting zone environment. Standard errors are in parentheses.

The entries in the first row consider the "School and CZ" treatment (labeled Sch+CZ). Here we set the combined school and commuting zone inputs that the student receives to the 10th vs. 90th percentile of the distribution of combined inputs $\mathbf{Z^{S*}}\mathbf{\Gamma}^{S*}+\mathbf{Z^{C*}}\mathbf{\Gamma}^{C*}$, which we approximate with the estimated distribution of $\mathbf{Z_{2s}^{S}}\mathbf{G_{2}^{S}}+\mathbf{Z_{2c}^{C}}\mathbf{G_{2}^{C}}+(v_s-v_c)+v_c$. The entries in the second row consider the "School only" specification (labeled Sch Only). Here we aim to set school inputs only at the 10th vs. 90th percentile of the true school input distribution $\mathbf{Z^{S*}}\mathbf{\Gamma}^{S*}$. We approximate these quantities using the distribution of $\mathbf{Z_{2s}^{S}}\mathbf{G_{2}^{S}}+(v_s-v_c)$. Finally, the third row entries display results from the "CZ only" treatment. Here we isolate the importance of commuting-zone level inputs by using quantiles of the $\mathbf{Z_{2c}^{C}}\mathbf{G_{2}^{C}}+v_c$ distribution to approximate the distribution of true CZ inputs $\mathbf{Z^{C*}}\mathbf{\Gamma}^{C*}$.

Focusing first on the NELS grade 8 results in column 1, we see that growing up in a 90th percentile school/commuting zone environment rather than a 10th percentile environment increases a randomly chosen student's probability of graduation from .821 to .903, an increase of 8.2 (1.7) percentage points. It is very large relative to the dropout rate. A corresponding shift in school environment only (while receiving a random draw from the distribution of commuting zone inputs) changes the graduation probability from .832 to .893, an increase of 6.1 (1.9) percentage points. A 10th-to-90th percentile shift in commuting zone inputs generates a predicted increase of 5.5 (2.0) percentage points (.835 to .890).

The estimates for ELS are considerably smaller. Column 3 reports ELS results for block group. The "School and CZ", "School Only", and "CZ Only" treatments raise the graduation probability by 5.4 (2.2) percentage points (.901 to .955), 3.0 (2.0) percentage points, and 4.5 (1.8) percentage points, respectively. The values are slightly smaller when we use zipcode as the neighborhood measure (column 2). One reason for the smaller ELS estimates is that the unweighted high school graduation probability in the sample is very high: 0.92. Thus, there is not much room for increases in the graduation probability. However, as discussed earlier, the estimated standard deviation of $\mathbf{Z_{2s}^{S}}\mathbf{\hat{G}_{2}}$ is anomalously small in the nonlinear full specification for high school graduation in ELS, relative to the NELS estimates, the linear version (no interaction terms) of the full specification in ELS (not reported), or the nonlinear specification in ELS that uses only the basic $\mathbf{X_i}$ variables. Indeed, the same 10th-90th probability shifts for the "School and CZ", "School Only", and "CZ Only" treatments using the linear specification (25) with block group as the neighborhood designation are 6.6 percentage points, 4.9 percentage points, and 4.3 percentage points, respectively. Estimates are similar for the basic non-linear specification (Web Appendix Table 1-1) and the full non-linear specification using zip code to define the neighborhood (Column 3).

While the estimates are somewhat noisy (standard errors are usually nearly 2 percentage points), taken together the results for HSGRAD indicate that large shifts in school and commuting zone inputs could generate socially significant impacts on graduation probabilities. Note that this is true even though the components $(\mathbf{Z^S_{2s}G^S_2}, \mathbf{Z^C_{2c}G^C_2}, (v_s - v_c), v_c)$ that we ascribe to school and commuting zone inputs comprise relatively small fractions of the outcome variance. Appendix Table 3 displays the variance components that underlie the three counterfactual treatments.[29] The variance of the distribution of combined school and CZ inputs in the "School and CZ" specification comprises just 2.0 percent of the full latent index variance in NELS, and only 2.4 percent in the ELS block group specification. Similarly, the variance of the inputs that make up the school quality and CZ quality distributions in the "School Only" and "CZ Only" specifications each account for only about 1 percent of the latent index variance in NELS grade 8. How do 10th-to-90th quantile shifts in distributions featuring such small variances generate such substantial outcome changes? One reason is that variance components, by virtue of squaring deviations, tend to exacerbate differences in the relative importance of various inputs; standard deviations of these components are considerably closer in size to those corresponding to individual inputs. However, perhaps more importantly, for binary outcomes even a small shift in underlying propensity to graduate can have a significant impact on graduation outcomes if many students are near the decision margin.

The ELS estimates using the basic set of $\mathbf{X_i}$ variables in Web Appendix Table 5 are typically about 1 percentage point larger for School and CZ and for School, but not for CZ. The NELS estimates change very little. As we have discussed, the estimates based on the full set are more conservative. The inclusion in the full set of test scores and other variables that are likely to be influenced by the school and location will generally lead to an understatement of treatment effects. Including them in $\mathbf{X_s}$ also potentially absorbs more of the true variation school and location quality, leading to a further understatement. The advantage of the full set is that the expanded $\mathbf{X_s}$ reduces the risk of sorting bias.

8.1.1 Subpopulation Results for High School Graduation

We now examine the degree of heterogeneity in sensitivity to our counterfactual treatments across subpopulations.

Treatment effect heterogeneity across groups arises through three pathways. The first is that the model includes interactions involving female, underrepresented minority, and low income indicator variables. The second and third both arise from differences in the distributions of $\mathbf{X_iB}$ across subpopulations. The second pathway arises because the model contains interactions between $\mathbf{X_iB}$ and $\mathbf{Z^S_{2s}G^S_2}$ and $\mathbf{Z^C_{2c}G^C_2}$, respectively. The third arises because differences in the locations of the $\mathbf{X_iB}$ distribution across subgroups produces effect heterogeneity for binary outcomes even in the absence

[29]Note that these variance components are computed based on the specification in Columns 1-3 of Table 2-1. In computing variance fractions, we ignore the variance contribution of the interaction terms because a simple variance decomposition does not exist for the model with interactions. The interactions are accounted for in the 10-90 estimates. The corresponding table for the linear specification is available upon request. The values are close for all outcomes and samples, with the exception of HSGRAD for the ELS block group sample.

of explicit interaction terms. Due to the nonlinearity in the probit function that links Y to the binary outcome indicators for education, the sensitivity to school quality is higher for subpopulations with values of $\mathbf{X_i\hat{B}}$ that place them closer to an outcome probability of 0.5. Consequently, HSGRAD is more sensitive to school or commuting zone quality for disadvantaged subpopulations and less sensitive for advantaged subpopulations. The opposite tends to be true for ENROLL and COLLBA. In the extreme case, if every member of the subpopulation had a sufficiently high $\mathbf{X_iB}$ value, they would all be presumed to attain the positive outcome regardless of their school/commuting zone environment.

Table 2-1, column 1 shows that r^S_{21}, the coefficient on the interaction between $\mathbf{X_iB}$ and $\mathbf{Z^S_{2s}G^S_2}$, is only .008 (.033) for NELS 8th graders. The corresponding values for ELS block group and ELS zipcode are -.011 (.035) and -.011 (.034). None of these values are statistically nor socially significant. The NELS coefficient implies that experiencing a school with a one standard deviation above average value of the index $\mathbf{Z^S_{2s}G^S_2}$ would increase the probit index by .008 more for a student with a one standard deviation above average value of the index $\mathbf{X_iB}$ than for a student at the mean of $\mathbf{X_iB}$.

The coefficient r^C_{21} on $(\mathbf{X_iB})(\mathbf{Z^C_{2c}G^C_2})$ is small and positive for NELS and ELS blockgroup and -.017 (.029) for ELS zip code. Again, none are statistically significant. Columns 2, 4, and 6 also report coefficients on the other interactions. They are mostly small and inconsistent in sign. The estimates are noisy, so we cannot rule out the possibility of important interaction effects in the probit index. But the interaction terms are not the primary source of the substantial differences across groups in treatment effects, to which we now turn.

Table 5 displays the impacts of the 10th-to-90th quantile shifts in quality for each of our three counterfactuals for randomly chosen individuals from specific subpopulations. We consider Hispanic students, non-Hispanic black students, white students, non-Hispanic white students with a single mother with a high school education or less, and non-Hispanic white students living with two biological parents with four-year college degrees or above.

White students dominate the sample, and so estimates are similar to but a bit smaller than those for full sample. Black students are moderately more sensitive to each of the three treatments than the population at large, while Hispanic students and non-Hispanic white students with a single mother with a high school education or less, who have the lowest mean $\mathbf{X_iB}$ values, are the most sensitive. Specifically, for the "School and CZ" 10th-to-90th quality shift, the predicted increase in high school graduation probability is 8.7 (2.1) percentage points (81.4 to 90.1) for a random member of the black subpopulation and 13.6 (2.7) percentage points (65.6 to 79.3) for a member of the subpopulation with less-educated white single mothers. In contrast, non-Hispanic white students with two college-graduate parents can only expect a 4.2 (0.9) percentage point increase from the same shift in school/commuting zone inputs. The ELS results with either block group or zip code as the neighborhood are smaller: 7.6 (3.3) and 6.7 (2.9) percentage points, respectively, for Hispanic and non-Hispanic black students, and only about 1.8 (0.9) percentage points for a randomly chosen white student with two college-graduate parents. The smaller ELS results are primarily due to the high sample mean graduation rate of 92% in the ELS sample: high income students are sufficiently well-supported that they are likely to graduate high school regardless of the school and commuting zone environment they experience in the ELS cohort.

Similar patterns emerge for the "School only" and "CZ only" counterfactuals, for which the subpopulations feature variation in their 10th-90th probability differences of a couple of percentage points around the full sample means of 3.0 and 4.5 percentage points for ELS block group, and 6.1 and 5.5 percentage points for NELS zip code. Interestingly, the "CZ only" counterfactual continues to have a slightly larger impact than the "School only" counterfactual for all subgroups, despite the fact that the estimated interaction coefficients imply that disadvantaged subgroups might be less sensitive to CZ inputs than the full population and more sensitive to school inputs. This implies that the contribution of the estimated interaction effects to treatment effect heterogeneity is swamped by the effect heterogeneity generated by the highly nonlinear probit function, which for high school graduation makes all disadvantaged populations more sensitive to external inputs (a greater share are near the decision margin). Consist with this, we obtain similar estimates of 10th-90th treatment effects using the models without interactions (not shown).

Finally, Figures 3a and 3b and Figure 4 illustrate more generally how the predicted outcome changes as the percentile of the $\mathbf{X_i B}$ index increases (based on (34)). The solid dark line in Figure 3a (3b) shows how in the NELS (ELS block group) case the probability of high school graduation (vertical axis) varies with the percentile of $\mathbf{X_i B}$ for a student who receives the median value of the School and CZ treatment. It provides a yard stick for thinking about the magnitude of treatment effects. Figure 4 graphs 10th-to-90th treatment effects against the $\mathbf{X_i B}$ percentile. We exclude confidence intervals to avoid cluttering the graph, but Web Appendix Table 4 reports estimates as well as standard errors of treatment effects for student at the 10th, 50th and 90th quantiles of $\mathbf{X_i B}$. The impact of each 10th-to-90th treatment decreases monotonically with the $\mathbf{X_i B}$ quantile. Specifically, students in NELS (light grey line) at the 5th quantile of the $\mathbf{X_i B}$ distribution move from a 51.8% chance of graduating to a 68.9% chance when the quality of the combined school/commuting zone environment shifts from the 10th to the 90th percentile, whereas students at the 95th quantile only move from a 98.0% graduation rate to a 99.3% graduation rate. So 8th grade school and commuting zone environment seems to be critically important for high school graduation for particularly disadvantaged populations, and essentially irrelevant for particularly advantaged populations. The ELS results for both neighborhood definitions display a similar pattern of sensitivity to school and CZ inputs, but with proportionately smaller impacts for all subpopulations.

8.2 Enrollment in a Four-Year College

Columns 4-6 of Table 4 reports the impacts of the corresponding counterfactual shifts in school and/or commuting zone environment for whether the student enrolls in a four-year college within two years of expected high school graduation.

A 10th-to-90th percentile shift in combined school/commuting zone quality predicts increases in enrollment probability of 17.9 (2.3) percentage points (23.4 to .41.3) for 8th graders in NELS, while the corresponding shifts in school inputs exclusively and commuting zone inputs exclusively generate predicted increases of 13.0 (25.7 to 38.9) and 11.4 (26.5 to 37.9) percentage points, respectively. The effect of the "School and CZ" treatment in the ELS zip code and ELS block group specifications are 16.2% (2.7%) and 15.8% (3.6%), respectively, with a reduced importance of CZ compared to school.

Note that the impacts of school and commuting zone on this outcome are considerably larger than those for high school graduation. For NELS, this partly reflects a larger variance in the component of Y_i (the latent index) that is attributed to school and commuting zone inputs (2 vs. 3.5 percent of $Var(Y_i)$, from Appendix Table 3). However, the variance components underlying the ELS block group and ELS zipcode counterfactuals are very similar in magnitude to those used for the high school graduation counterfactuals ($\sim$ 2.3 percent of $Var(Y_i)$ in each case). Instead, the larger estimated impacts for college enrollment primarily reflects the fact that the population outcome mean for college enrollment is much closer to .5 (.327 in NELS and .422 in ELS, from Table 1) than for high school graduation (.853 in NELS and .919 in ELS). The probit model assumes a normal distribution of unobserved inputs, so that an outcome mean of .5 implies that many students are near the decision margin. Thus, the same change in the latent probit index translates to a much greater shift in outcome probability for college enrollment.

Treatment effects for the basic specification of $\mathbf{X_i}$ are about 4 percentage points larger in ELS and about 0.5 percentage points larger in NELS (Web Appendix Table 5, columns 4-6).

8.2.1 Subpopulation Results for College Enrollment

Table 5 reports the impacts of our counterfactual shifts for particular subgroups of NELS 8th graders. Because the population mean for college enrollment is below .5, the probit model implies that more disadvantaged subpopulations have fewer students near the decision margin, so that the largest impacts occur for the most advantaged subgroup (in this case, white students with two college-graduate parents) and the lowest value is for the most disadvantaged subgroup (children of single white mothers with education less than or equal to high school). The impacts of the "School and CZ" treatment range from 12.2 to 21.7 percentage points. The "School only" treatment displays slightly higher effects than the "CZ only" treatments with similar degrees of heterogeneity: Impacts range from 6.9-14.1 percentage points for "School only" and 9.2-15.7 percentage points for the "CZ only" treatment.

The dashed line (zip code) and dotted line (block group) in Figures 3a and 3b display for NELS and ELS block group the same strong link between $\mathbf{X_iB}$ and college attendance as was found for high school graduation.

Figure 5 shows the relationship between the treatment effects on college attendance probability and the quantiles of $\mathbf{X_iB}$. The impact of the counterfactual treatments is nonmonotonic in $\mathbf{X_iB}$, increasing and then decreasing as one moves through the distribution of $\mathbf{X_iB}$. For example, for the "School and CZ" counterfactual in NELS, the estimated impact for the 10th, 70th (near the maximum impact), and 95th quantiles of $\mathbf{X_i B}$ are 7.3 percentage points, 25.2 percentage points, and 18.3 percentage points, respectively.

The pattern occurs despite the fact that in the NELS case the coefficient on the interaction between $\mathbf{X_i B}$ and $\mathbf{Z^S_{2s} G^S_2}$ is substantial: -.087 (0.038). Part of the decline at the top of the $\mathbf{X_i B}$ distribution may be due to this. However, the pattern is primarily due to a striking fact about the variation in predicted outcome probabilities across $\mathbf{X_i B}$ quantiles that is established in Figure 1a. In NELS (solid light line) for a school/CZ combination of median quality, the predicted enrollment rates for a student at the 10th, 30th, 70th and 95th quantiles of $\mathbf{X_i B}$ are 6.1%, 21.6%, 60.7%, and 89.3%, respectively. The figure demonstrates just how strongly observed family background and student aptitude measures (as of 8th grade) predict enrollment at a four-year postsecondary institution.

The estimates of the treatment effects for the ELS zip code and block group specifications follow the same pattern as in NELS, although the estimates are typically about 1 point smaller in size and fall in a narrower range. The variance components attributed to school inputs ($Var(\mathbf{Z^S_{2s} G^S_2})$ and $Var(v_s - v_c)$) are substantially larger in ELS than those attributed to CZ inputs ($Var(\mathbf{Z^C_{2c} G^C_2})$ and $Var(v_c)$), so that counterfactual impacts are substantially larger for the "School Only" specification than the "CZ only" specification, particularly for subgroups featuring low values of $\mathbf{X_i B}$. For example, the impact of a 10th-to-90th percentile shift in the quality of school inputs increases the predicted enrollment probability by 11 percentage points for a random member of the Hispanic subpopulation, while the corresponding shift in the quality of commuting zone inputs only produces a 7 percentage point increase. The estimated impacts associated with 10th-90th shifts in combined school and CZ inputs display similar magnitudes and patterns of heterogeneity to the NELS sample.

The ELS zip code specification provides a particularly useful opportunity to assess the importance of the interaction terms in the model. First, the bias-corrected standard deviations in $\mathbf{Z^S_{2s} \hat{G}_2}$ and $\mathbf{Z^C_{2c} G^C_2}$ are both substantial, so that there is more signal contained in these regression indices with which to identify true interactions. Second, a number of the interaction terms are seemingly non-negligible in size (albeit still noisily estimated). Specifically, the interaction coefficients on $\mathbf{1}(\mathbf{LowInc}) \times \mathbf{Z^S_{2s} G^S_2}$ and $\mathbf{1}(\mathbf{LowInc}) \times \mathbf{Z^C_{2c} G^C_2}$ are 0.079 and 0.053 respectively, suggesting greater sensitivity of students from modest backgrounds to both school and CZ inputs. The interaction coefficients on $(\mathbf{1}(\mathbf{URM})\mathbf{Z^S_{2s} G^S_2})$ and $(\mathbf{1}(\mathbf{URM})\mathbf{Z^C_{2c} G^C_2})$ are also substantial, but are conflicting in sign (-0.044 and 0.071 respectively). Third, the linear and nonlinear specifications feature very similar values of the variance components we use for our "School and CZ" treatment ($Var(\mathbf{Z^S_{2s} \hat{G}_2} + \mathbf{Z^C_{2c} G^C_2} + (v_s - v_c) + v_c)$ makes up 2.37% of $Var(Y_i)$ in both cases), so that we can directly compare treatment effects when interaction terms are both excluded and included, effectively holding fixed the importance of the main effects. Indeed, the mean impact of the 10th-90th "School and CZ" treatment is 16.1 percentage points in the linear specification (not shown) and 16.2 (Table 4) percentage points in the nonlinear specification.

However, none of the subgroups we consider displays more than a 0.3 percentage point difference in treatment effects between the linear and non-linear specification. While the imprecision of our interaction estimates cautions against overinterpreting the results, our limited evidence suggests that the probit function seems to be doing an effective job of capturing heterogeneity in sensitivity to higher quality school and commuting zone inputs.

8.3 Graduation from a Four-Year College

Columns 7-9 of Table 4 report the estimated impacts of the same counterfactual treatments on a randomly selected student's probability of graduating college by age 25. The NELS results (col. 7) indicate that replacing a combined school/commuting zone at the 10th percentile of the quality distribution with one at the 90th percentile increases the college graduation probability by 11.9 (2.1) percentage points (38.6% to 50.5%), while shifting only school inputs or only CZ inputs from their respective 10th to 90th percentiles increases the graduation probability by 10.1 (2.8) and 9.2 (2.3) percentage points, respectively. The small increase in impact from shifting school and CZ quality together relative to separately is partly due to an estimated negative covariance between $\mathbf{Z^S_{2s}G^S_2}$ and $\mathbf{Z^C_{2c}G^C_2}$, but is also due to the fact that doubling a variance only increases the standard deviation by a factor of $\sqrt{2}$.

However, the ELS block group estimates (col. 8) are only 7.3 (2.2), 5.6 (2.4), and 5.7 (2.4) percentages points, respectively, with the ELS zip code estimates (col. 9) displaying similar magnitudes. The discrepancy is much larger in percentage terms than the discrepancy in the results for ENROLL. COLLBA has a mean of 0.376 in ELS and 0.443 in NELS (Table 1), so one cannot attribute more than a small share of the larger treatment effects in NELS to nonlinearity of the probit model. Instead, the share of variance in the latent index attributed to school and CZ inputs is only half as large in ELS as in NELS (Appendix Table 3). While this could be a true cohort effect, the fact that college graduation outcome is measured four years later in a separate survey than our enrollment and high school graduation outcomes may be a contributing factor. The change in sample is larger in NELS (the sample size drops by about a quarter), and may have had a different composition relative to ELS. Notably, the school effect and the commuting zone effect are similar in size to each other in both data sets.

Switching from the full specification to the basic specification (displayed in Web Appendix Table 5) increases the estimates in ELS by about 2.6% and in NELS by about 1.2%, narrowing the gap between the data sets.

8.3.1 Subpopulation Results for College Graduation

Table 5 reveals modest heterogeneity in estimated impacts from the counterfactual treatments across subgroups for NELS 8th graders. For the "School and CZ" counterfactual, a 10th-to-90th shift in combined school/CZ input quality increases the probability of college graduation by between 9.6% and 12.4%. The range across subgroups of the estimated impacts of the "School only" and "CZ only" counterfactuals are 8.2-10.5 and 7.2-9.7 percentage points. However, this is partly because the

$\mathbf{X_iB}$ distributions for these subgroups feature similar fractions of students for whom the treatment effects are largest (those at the 40th to 90th percentiles of the distribution), in contrast to high school graduation, so that the degree of heterogeneity across student types is somewhat obscured by these subgroup choices.

Once again, examining the variation in treatment impacts across $\mathbf{X_iB}$ quantiles reveals a greater degree of heterogeneity (Figure 4). For NELS, the "School and CZ" impacts for the 5th, 70th, and 95th quantiles of $\mathbf{X_iB}$ are 5.8, 14.8, and 8.4 percentage points, respectively, demonstrating the same non-monotonicity of treatment impacts across the $\mathbf{X_iB}$ as was observed for the college enrollment outcome. Figure 1a shows that in NELS the probability of college graduation is 12.8% for a student at the 10th quantile of $\mathbf{X_iB}$ and 79.5% at the 90th (evaluated at the median value of the "School and CZ" treatment).

8.4 Log Wage Rates

Columns 10-12 of Table 4 report the estimated impacts of the same counterfactual treatments on a randomly selected student's log wage as of about age 25. The NELS results (col. 10) indicate that the effect of the 10th-90th percentile combined school/commuting zone treatment is 0.129 (0.031) log points (which corresponds to a 13.8% wage increase). A one standard deviation improvement in the treatment would raise the wage by 0.051 (a 5.2% wage increase). In comparison, the standard deviation of the permanent component of the wage is 0.361. The values for ELS are smaller. The estimate of the combined school/community 10th-90th treatment effect is only 0.093 (0.023) log points in the block group specification (a 10% wage increase). Estimates using the base specification in Web Appendix Table 3 are similar.

The effect of the 10th-90th commuting zone treatment is larger than the school treatment in both NELS and in ELS, in contrast to what we found for education. One would expect commuting zone characteristics to be particularly important for wages, where opportunities in the local labor market are paramount.

Since the wage model is linear, the only source of treatment effect heterogeneity consists of the interaction terms between $\mathbf{M_i}$ and $\mathbf{Z^S_{2s}G^S_2}$ and $\mathbf{Z^C_{2c}G^C_2}$. The coefficients on the interactions terms are small, so we do not find important subgroup differences.

8.5 Summary of Results from Counterfactual Shifts

There are several broad takeaways from the estimated impacts of counterfactual 10th-to-90th quantile shifts in school and/or commuting zone inputs. First, large changes in school and commuting zone inputs can make a substantial difference in students' educational attainment.

Second, the impact of these inputs is quite heterogeneous across disadvantaged vs. advantaged students. The dropout rates of disadvantaged students are particularly sensitive to the external environment, while few advantaged students are near the margin. In contrast, for college enrollment and graduation, superior school and commuting zone inputs are important for all, but particularly consequential for students near (or above) the middle of the distribution of student and family background.

Third, the more general model featuring interactions between student and school/commuting zone inputs produces predicted impacts of shifts in school and commuting zone environment that differ little from the results using the simpler probit specification without interactions. This suggests that the nonlinearity inherent in the probit function does a fairly good job of capturing differential sensitivity to school and region-level inputs, and that a linear specification for the probit index may suffice in similar contexts. But this need not be true in other applications and the estimates of the interaction coefficients are imprecise.[30]

Finally, both school and commuting zone inputs seem to be important, and with relatively comparable magnitudes. However, this conclusion is more tentative, since our ability to distinguish school from commuting zone inputs is somewhat limited.

9 Conclusion

Educational attainment and wages, like many adult outcomes, are influenced by factors that are specific to the individual as well as aspects of the broader social environment. In this paper, we build on the rich literature on sorting, school and neighborhood effects, and multilevel modeling to assess the relative importance of neighborhood, school, and broader local area factors in shaping student's educational attainment and early career wages, and the degree to which this relative importance differs across students from different backgrounds. We extend the control function result of Altonji and Mansfield (2016) that group averages of individual-level observables can fully control for sorting bias from group averages of unobservables to allow for multiple group levels and interactions between individual and group level factors.

Our theoretical results demonstrate the existence a structural decomposition of variation in educational and labor market outcomes of interest into four components: (1) individual contributions that are common across groups, (2) group contributions that are common across individuals, (3) contributions that consist of interactions between student and group inputs, and a (4) set of ambiguous contributions, absorbed by our control function of group-level averages, that reflect a combination of common group inputs and group-averages of individual inputs.

We implement this structural decomposition using a multilevel mixed effects model, and use the results to generate lower bound estimates of the average impact among the student population of "treatments" consisting of shifts in school, commuting zone, and combined school/CZ quality from the 10th-to-90th quantile of their respective distributions. We also produce estimates customized for particular student subpopulations that exploit the treatment effect heterogeneity accommodated

[30]In retrospect, it may have been better to put indicators for the subgroups we consider in M_i in place of LOWINC and URM.

by our model.

Despite accounting for a small share of the total variance in the latent indices that generate our binary outcomes, our population average estimates suggest that school and commuting zone inputs play an important role in determining educational attainment. Moving from a combined school/CZ combination at the 10th quality quantile to a 90th quantile combination is estimated to increase the probability of high school graduation by at least 5-8 percentage points across datasets and specifications, the probability of enrollment at a four-year college by at least 16-18 percentage points, and the probability of college graduation by at least 7-12 percentage points. School and commuting zone inputs seem to play a roughly equal role in producing educational attainment, though the model's ability to distinguish between the two is more limited. Estimates for log wages of the impact of 10th-90th quantile shifts in school/CZ quality are smaller, but still meaningful (10-14 percent increases), though the age 25 wage data we use likely understates effects on permanent income.

Our subpopulation estimates show that the high school graduation rates of disadvantaged populations are considerably more sensitive to group-level inputs, while college graduation rates are more sensitive to group-level inputs for advantaged populations. However, such heterogeneity in treatment effects is driven by the predicted share of the subpopulation near the decision margin in our probit models, rather than by fundamental differences in input sensitivity: our estimated interaction effects are small (though imprecise), suggesting that the inherent nonlinearity in the probit model well-approximates the heterogeneity in school and commuting zone treatment effects.

Our analysis of a regression index of student-level observables suggests that most outcome-relevant sorting takes place at the level of school attendance zones, with a modest role for CZ-level sorting and a very small role for neighborhood sorting, though the neighborhood component may be understated by the use of block group or zipcode to define neighborhood boundaries.

While the panel surveys we use offer many advantages, a multilevel mixed effects model featuring interactions among multiple combinations of levels places strong demands on the data. This leads some of our estimates (particularly interactions) to be imprecise, and highlights the value of increasingly available linked administrative data that offer both many groups at each hierarchical level and large numbers of individuals per group, along with rich sets of observable characteristics at each level. A particular data limitation is the lack of observed neighborhood characteristics that are plausible candidates for sources of neighborhood-level causal effects.

Going forward, natural extensions to the model include incorporating interactions with unobservable individual characteristics and examining the sensitivity of results to departures from the assumption (A6) that the across-group variation in the within-group covariance matrix of individual-level characteristics across groups is unrelated to group-level characteristics.

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

Alexander, Karl, and Stephen L Morgan (2016) 'The coleman report at fifty: Its legacy and implications for future research on equality of opportunity.' *RSF*

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

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

Altonji, Joseph G, and Richard Mansfield (2016) 'Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects.' *Working Paper*

Altonji, Joseph G, and Thomas A Dunn (1996) 'Using siblings to estimate the effect of school quality on wages.' *The Review of Economics and Statistics* pp. 665–671

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

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

Bergman, Peter Leopold S (2016) 'The effects of school integration: Evidence from a randomized desegregation program.' Technical Report, CESifo Group Munich

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

Betts, Julian R (1995) 'Does school quality matter? evidence from the national longitudinal survey of youth.' *The Review of Economics and Statistics* pp. 231–250

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

Card, David, and Jesse Rothstein (2007) 'Racial segregation and the black–white test score gap.' *Journal of Public Economics* 91(11), 2158–2184

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

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

Chyn, Eric (2016) 'Moved to opportunity: The long-run effect of public housing demolition on labor market outcomes of children.' *Unpublished paper. University of Michigan, Ann Arbor*

Coleman, James S, Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederick Weinfeld, and Robert York (1966) 'The coleman report.' *Equality of Educational Opportunity*

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

Dobbie, Will, Roland G Fryer, and G Fryer Jr (2011) 'Are high-quality schools enough to increase achievement among the poor? evidence from the harlem children's zone.' *American Economic Journal: Applied Economics* 3(3), 158–187

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

Durlauf, Steven N, and Yannis M Ioannides (2010) 'Social interactions.' *Annu. Rev. Econ.* 2(1), 451–478

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

Garner, Catherine L, and Stephen W Raudenbush (1991) 'Neighborhood effects on educational attainment: A multilevel analysis.' *Sociology of education* pp. 251–262

Goldstein, Harvey (2011) *Multilevel statistical models,* vol. 922 (John Wiley & Sons)

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

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

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

Katz, Lawrence F (2015) ‘Reducing inequality: Neighborhood and school interventions.’ *Focus* 31(2), 12–17

Kline, Patrick, and Enrico Moretti (2014) ‘People, places, and public policy: Some simple welfare economics of local economic development programs.’ *Annual Review of Economics* 6(1), 629–662

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

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

Lucas, Samuel R (2016) ‘First-and second-order methodological developments from the coleman report.’ *RSF*

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

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

Meier, Lukas, Sara Van De Geer, and Peter Bühlmann (2008) ‘The group lasso for logistic regression.’ *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(1), 53–71

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

Raudenbush, Stephen W, and Anthony S Bryk (2002) *Hierarchical linear models: Applications and data analysis methods,* vol. 1 (Sage)

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

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

Sharkey, Patrick, and Jacob W Faber (2014) 'Where, when, why, and for whom do residential contexts matter? moving away from the dichotomous understanding of neighborhood effects.' *Annual Review of Sociology* 40, 559–579

Solon, Gary, Marianne E Page, and Greg J Duncan (2000) 'Correlations between neighboring children in their subsequent educational attainment.' *Review of Economics and Statistics* 82(3), 383–392

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

Yuan, Ming, and Yi Lin (2006) 'Model selection and estimation in regression with grouped variables.' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(1), 49–67

Appendices

A1 Proof of Proposition 2:

In deviation from mean form, the model is

$$\begin{aligned} DY_i &= \mathbf{DX}_i\boldsymbol{\beta} + \boldsymbol{DM}_i\mathrm{X}_g\boldsymbol{\rho}_1 + \mathbf{DM}_i\mathbf{Z}_{2g}\boldsymbol{\rho}_2 \\ &\quad + \boldsymbol{DM}_i\mathbf{X}_g^U\boldsymbol{\rho}_1^U + \boldsymbol{DM}_i\mathbf{Z}_{2g}^U\boldsymbol{\rho}_2^U + Dx_i^U + D\eta_{gi} + D\xi_{gi}\ . \end{aligned}$$

Using $\mathbf{X_i^*}\boldsymbol{\beta}^* = \mathrm{X}_i\boldsymbol{\beta} + x_i^U$, $\boldsymbol{DM}_i\mathbf{Z}_g^*\boldsymbol{\rho}^* = \boldsymbol{DM}_i\mathrm{X}_g\boldsymbol{\rho}_1 + \boldsymbol{DM}_i\mathbf{Z}_{2g}\boldsymbol{\rho}_2 + \boldsymbol{DM}_i\mathbf{X}_g^U\boldsymbol{\rho}_1^U + \boldsymbol{DM}_i\mathbf{Z}_{2g}^U\boldsymbol{\rho}_2^U$, and $D\eta_{gi} = \mathbf{DX_i}\boldsymbol{\Pi}_{\eta_{gi}\mathbf{X_i}} + D\tilde{\eta}_{gi}$, we can rewrite the above equation as

$$DY_i = \mathbf{DX}_i^*\boldsymbol{\beta}^* + \boldsymbol{DM}_i\mathbf{Z}_g^*\boldsymbol{\rho}^* + \boldsymbol{DX}_i\boldsymbol{\Pi}_{\eta_{gi}D\mathbf{X_i}} + D\tilde{\eta}_{gi} + D\xi_{gi} \tag{35}$$

$D\xi_{gi}$ is uncorrelated with all variables in the model. $D\tilde{\eta}_{gi}$ in uncorrelated with DX_i by construction, and A8 with the interaction terms involving DM_i are uncorrelated with all variables in the model. Consequently, these components do not contribute to $\mathbf{B}$, $\mathbf{r_1}$, or $\mathbf{r_2}$ and can be ignored for the rest of the proof.

Write the projection of the within g deviation in individual i's composite individual contribution $D\mathbf{X_i^*}\boldsymbol{\beta}^*$ on the observable space $[\mathbf{DX_i}, \mathbf{DM}_i\mathbf{X_g}, \mathbf{DM_iZ_{2g}}]$ as

$$\boldsymbol{DX}_i^*\boldsymbol{\beta}^* = \boldsymbol{DX}_i\boldsymbol{\Pi}_1 + \boldsymbol{DM}_i\mathrm{X}_g\boldsymbol{\Pi}_2 + \boldsymbol{DM}_i\mathbf{Z}_{2g}\boldsymbol{\Pi}_3 + \widetilde{D\mathbf{X_i^*}\boldsymbol{\beta}^*} \tag{36}$$

We now establish that $\boldsymbol{\Pi}_2$ and $\boldsymbol{\Pi}_3$ are 0. From basic regression theory, $\boldsymbol{\Pi}_2$ and $\boldsymbol{\Pi}_3$ are 0 if $\mathbf{DM}_i\mathbf{X_g}$ and $\mathbf{DM_iZ_g}$ are both uncorrelated with $D\mathbf{X_i^*}\boldsymbol{\beta}^*$ and with $\mathbf{DX_i}$. Using the law of iterated expectations it is easy to show that $E[(D\mathbf{M}_i)\mathbf{X_g}]E[D\mathbf{X_i^*}\boldsymbol{\beta}^*] = \mathbf{0}$, $E[(DM_i)\mathbf{Z}_{2g}]E[D\mathbf{X_i^*}\boldsymbol{\beta}^*] = \mathbf{0}$, $E[(DM_\mathbf{i})\mathbf{X_g}]'E[\mathbf{X_i}] = \mathbf{0}$ and $E[(DM_i)\mathbf{Z_{2g}}]'E[\mathbf{X_i}] = \mathbf{0}$.[31] These results and A6 imply that $\mathbf{Cov}(((DM_i)\mathbf{X_g})', D\mathbf{X_i^*}\boldsymbol{\beta}^*) = \mathbf{0}$, $\mathbf{Cov}(((DM_i)\mathbf{Z_{2g}})', D\mathbf{X_i^*}\boldsymbol{\beta}^*) = \mathbf{0}$, $\mathbf{Cov}((DM_i\mathbf{X_g})', D\mathbf{X_i}) = \mathbf{0}$ and $\mathbf{Cov}((DM_i\mathbf{Z_{2g}})', D\mathbf{X_i}) = \mathbf{0}$.[32] We

[31] For example, $E[(DM_i\mathbf{X_g}]E[\mathbf{X_i}] = E[E[(DM_i)\mathbf{X_g}|\mathbf{g}]]E[\mathbf{X_i}] = E[E[(DM_\mathbf{i}|\mathbf{g})\mathbf{X_g}]]E[\mathbf{X_i}] = 0$ because $DM_\mathbf{i}$ is the deviation from the mean for g.

[32] To see this for $\mathbf{Cov}((D\mathbf{M_i}\mathbf{X_g})', D\mathbf{X_i^*}\boldsymbol{\beta}^*)$, note that

$$\begin{aligned} &\mathbf{Cov}((\boldsymbol{D}\mathbf{M}_i\mathbf{X_g})', \boldsymbol{DX}_i^*\boldsymbol{\beta}^*) = \mathbf{Cov}(\mathbf{X_g'DM_i}, D\mathbf{X_i^*}\boldsymbol{\beta}^*) \\ &= \boldsymbol{E}[\boldsymbol{X}_g'\boldsymbol{DM}_i\boldsymbol{DX}_i^*]\boldsymbol{\beta}^* - E[\mathbf{X_g'DM_i}]\mathbf{E}[\boldsymbol{DX}_i^*]\boldsymbol{\beta}^*. \end{aligned} \tag{37}$$

Since X_g is a subvector of Z_g^*, A6 implies directly that $\boldsymbol{E}[\mathbf{X_g'}DM_i\mathbf{DX}_i^*]\boldsymbol{\beta}^*$ is 0. Since $\mathbf{E}[\mathbf{X_g'}DM_i]$ and $\mathbf{E}[\mathbf{DX}_i^*]$ both involve deviations from the mean of g, both are 0 and thus $\mathbf{E}[\mathbf{X_g'DM_i}]\mathbf{E}[\mathbf{DX}_i^*]\boldsymbol{\beta}^* = 0$. Thus both terms of (37) are 0, which establishes the result. Replacing $\mathbf{X}_g$ with $\mathbf{Z_{2g}}$ in this derivation demonstrates that $\mathbf{Cov}((DM_i\mathbf{Z_{2g}})', D\mathbf{X_i^*}\boldsymbol{\beta}^*) = \mathbf{0}$, while replacing $D\mathbf{X_i^*}\boldsymbol{\beta}^*$ with $\mathbf{DX}_i$ demonstrates that $\mathbf{Cov}((DM_i\mathbf{X_g})', \mathbf{DX_i}) = \mathbf{0}$ as well. Using both replacements yields $\mathbf{Cov}((DM_i\mathbf{Z_{2g}})', \mathbf{DX_i}) = \mathbf{0}$.

conclude that $DM_i\mathbf{X_g}$ and $DM_i\mathbf{Z_{2g}}$ are both uncorrelated with $D\mathbf{X_i^*}\boldsymbol{\beta}^*$ and with $\mathbf{DX_i}$. It follows that $\mathbf{\Pi}_2$ and $\mathbf{\Pi}_3$ are each $\mathbf{0}$, and more generally that $Proj(\mathbf{DX_i^*}\boldsymbol{\beta}^*|\mathbf{DX_i},\mathbf{DX_iX_g},\mathbf{DX_iZ_{2g}}) = Proj(\mathbf{DX_i^*}\boldsymbol{\beta}^*|\mathbf{DX_i})$.

Since projection is a linear transformation,

$$\Pr oj(D\mathbf{X_i^*}\boldsymbol{\beta}^*|D\mathbf{X_i}) = \Pr oj(D\mathbf{X_i}\boldsymbol{\beta}|\mathbf{DX_i}) + \Pr oj(D\mathbf{X_i^U}\boldsymbol{\beta}^U|\mathbf{DX_i}) = D\mathbf{X_i}\boldsymbol{\beta} + D\mathbf{X_i}\mathbf{\Pi}_{DX^UDX}\boldsymbol{\beta}^U) = D\mathbf{X}_i(\boldsymbol{\beta} + \mathbf{\Pi}_{DX^UDX}\boldsymbol{\beta}^U),$$

where $\mathbf{\Pi}_{DX^UDX}$ is the projection coefficient from (12). Thus, $\mathbf{\Pi}_1 = \boldsymbol{\beta} + \mathbf{\Pi}_{DX^UDX}\boldsymbol{\beta}^U$.

Next, project $DM_i\mathbf{Z_g^*}\boldsymbol{\rho}^*$ onto the space of observed regressors:

$$DM_i\mathbf{Z_g^*}\boldsymbol{\rho}^* = D\mathbf{X_i}\mathbf{\Pi}_4 + DM_i\mathbf{X_g}\mathbf{\Pi}_5 + DM_i\mathbf{Z_{2g}}\mathbf{\Pi}_6 + \widetilde{DM_i\mathbf{Z_g^*}\boldsymbol{\rho}} \tag{38}$$

First, notice that A6 also implies that $\mathbf{Cov}(D\mathbf{X_i'}, DM_i\mathbf{Z_g^*}\boldsymbol{\rho}^*) = \mathbf{0}$, $\mathbf{Cov}(D\mathbf{X_i'}, DM_i\mathbf{X_g}) = \mathbf{0}$, and $\mathbf{Cov}(D\mathbf{X_i'}, DM_i\mathbf{Z_{2g}}) = \mathbf{0}$ (the argument is essentially the same as in footnote 32). This implies that $\Pr \boldsymbol{oj(DM_i}\mathrm{Z}_g^*\boldsymbol{\rho}^*\boldsymbol{|DX_i, DM_i}\mathrm{X}_g\boldsymbol{, DM_i}\mathrm{Z}_{2g}\boldsymbol{)} = \Pr \boldsymbol{oj(DM_i}\mathrm{Z}_g^*\boldsymbol{\rho}^*\boldsymbol{|DM_i}\mathrm{X}_g\boldsymbol{, DM_i}\mathrm{Z}_{2g}\boldsymbol{)}$, and that $\mathbf{\Pi}_4 = \mathbf{0}$. Thus,

$$\mathbf{B} = \mathbf{\Pi}_1 + \mathbf{\Pi}_4 + \mathbf{\Pi}_{D\eta_{gi}D\mathbf{X_i}} = \boldsymbol{\beta} + \mathbf{\Pi}_{X^UX}\boldsymbol{\beta}^U + \mathbf{\Pi}_{D\eta_{gi}D\mathbf{X_i}}$$

as stated in Proposition 2.

To determine $\mathbf{\Pi}_5$ and $\mathbf{\Pi}_6$, we break $DM_i\mathbf{Z_g^*}\boldsymbol{\rho}^*$ into components. We first demonstrate that $\Pr oj(DM_i\mathbf{Z_g^*}|DM_i[\mathbf{X_g},\mathbf{Z_{2g}}]) = \Pr oj(\mathbf{Z_g^*}|[\mathbf{X_g},\mathbf{Z_{2g}}])$. To see this, note that the projection coefficient vector on $DM_i[\mathbf{X_g},\mathbf{Z_{2g}}]$ in (38) satisfies:

$$\begin{aligned}
&[\mathbf{\Pi}_5', \mathbf{\Pi}_6']' = \mathbf{Var}(D\mathbf{M}_i[\mathbf{X_g},\mathbf{Z_{2g}}])^{-1}\mathbf{Cov}((D\mathbf{M}_i[\mathbf{X_g},\mathbf{Z_{2g}}])', D\mathbf{M}_i\mathbf{Z_g^*}) \\
&= E[DM_iDM_i[\mathbf{X_g},\mathbf{Z_{2g}}]'[\mathbf{X_g}\mathbf{Z_{2g}}]]^{-1}E[DM_iDM_i[\mathbf{X_g},\mathbf{Z_{2g}}]'\mathbf{Z_g^*}] \\
&= E[E[(DM_i)^2[\mathbf{X_g},\mathbf{Z_{2g}}]'[\mathbf{X_g}\mathbf{Z_{2g}}]|g(i)=g]]]^{-1}E[E[(DM_i)^2[\mathbf{X_g},\mathbf{Z_{2g}}]'\mathbf{Z_g^*}|g(i)=g]] \\
&= E[E[(DM_i)^2|g(i)=g][\mathbf{X_g},\mathbf{Z_{2g}}]'[\mathbf{X_g}\mathbf{Z_{2g}}]]]^{-1}E[E[(DM_i)^2|g(i)=g][\mathbf{X_g},\mathbf{Z}_g]'\mathbf{Z_g^*}] \\
&= E[(Var(DM_i|g(i)=g)[\mathbf{X_g},\mathbf{Z_{2g}}]'[\mathbf{X_g}\mathbf{Z_{2g}}]]]^{-1}E[Var(DM_i|g(i)=g)[\mathbf{X_g},\mathbf{Z_{2g}}]'\mathbf{Z_g^*}] \\
&= [(Var(DM_i)E[[\mathbf{X_g},\mathbf{Z_{2g}}]'[\mathbf{X_g},\mathbf{Z_{2g}}]]]^{-1}[Var(DM_i)E[\mathbf{X_g},\mathbf{Z_{2g}}]'\mathbf{Z_g^*}] && (39) \\
&= \mathbf{Var}([\mathbf{X_g},\mathbf{Z_{2g}}])^{-1}\mathbf{Cov}([\mathbf{X_g},\mathbf{Z_{2g}}],\mathbf{Z_g^*}) && (40)
\end{aligned}$$

The first equality uses basic regression theory. The second used the definition of a variance and covariance and the fact that $DM_i[\mathbf{X_g},\mathbf{Z_{2g}}]$ has mean 0. It has a mean of 0 because DM_i has mean 0 within g (by definition of deviation from means) as well as across g. We have also used the fact that DM_i is a scalar to rearrange terms. The third equality imposes the law of iterated expectation, and the fourth uses the fact that $\mathbf{Z_g^*}$ does not vary within a school. The fifth equality uses the fact that $E(DM_i|g) = 0$ by construction and the definition of a variance to replace $E[(DM_i)^2|g(i)=g]$ with $(Var(DM_i|g(i)=g)$. The 6th equality follows from A7, while the seventh collects terms. The last equality illustrates that the projection coefficient on $DM_i[\mathbf{X_g}, \mathbf{Z_{2g}}]$ in (38) equals what one would recover from $Proj(\mathbf{Z_g^*}|[\mathbf{X_g}, \mathbf{Z_{2g}}])$, the simple projection of all school and area inputs on observed school and area inputs.

Next, using the fact that projection is a linear transformation, we can write:

$$
\begin{aligned}
&Proj(\mathbf{Z_g^*}|[\mathbf{X_g}, \mathbf{Z_{2g}}])\boldsymbol{\rho}^* = Proj(\mathbf{X_g}|[\mathbf{X_g}, \mathbf{Z_{2g}}])\boldsymbol{\rho}_1 + Proj(\mathbf{Z_{2g}}|[\mathbf{X_g}, \mathbf{Z_{2g}}])\boldsymbol{\rho}_2 + \\
&Proj(\mathbf{X_g^U}|[\mathbf{X_g}, \mathbf{Z_{2g}}])\boldsymbol{\rho}_1^U + Proj(\mathbf{Z_{2g}^U}|[\mathbf{X_g}, \mathbf{Z_{2g}}])\boldsymbol{\rho}_2^U
\end{aligned} \tag{41}
$$

Considering the four projections on the right hand side in turn, we find:

$$\mathbf{X_g}\boldsymbol{\rho}_1 = \mathbf{X_g}\boldsymbol{\rho}_1 + \mathbf{Z_{2g}}[\mathbf{0}] \tag{42}$$

$$\mathbf{Z_{2g}}\boldsymbol{\rho}_2 = \mathbf{X_g}[\mathbf{0}] + \mathbf{Z_{2g}}\boldsymbol{\rho}_2 \tag{43}$$

$$\mathbf{X_g^U}\boldsymbol{\rho}_3 = \mathbf{X_g}\boldsymbol{\Pi}_{X_g^U X_g}\boldsymbol{\rho}_1^U + \mathbf{Z_{2g}}[\mathbf{0}] \tag{44}$$

$$\mathbf{Z_{2g}^U}\boldsymbol{\rho}_4 = \mathbf{X_g}\boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U + \mathbf{Z_{2g}}\boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U \tag{45}$$

The first and second lines use the fact that $\mathbf{X_g}$ and $\mathbf{Z_{2g}}$ are of course fully determined by $\mathbf{X_g}$ and by $\mathbf{Z_{2g}}$, respectively. The third line uses Proposition 1, which states that $\mathbf{X_g^U}$ is an exact linear function of $\mathbf{X_g}$ (equation (6) provides an explicit expression for $\boldsymbol{\Pi}_{X_g^U X_g}$). The last line echoes (15) defined above Proposition 2.

Collecting terms from equations (42)-(45), we see that:

$$\mathbf{r_1} = \boldsymbol{\rho}_1 + \boldsymbol{\Pi}_{X_g^U X_g}\boldsymbol{\rho}_1^U + \boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U \tag{46}$$

$$\mathbf{r_2} = \boldsymbol{\rho}_2 + \boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U \tag{47}$$

This concludes the proof of Proposition 2.

A2 Proof of Proposition 3

Proposition 3:

Suppose assumptions A1-A9 hold. Then:

$$
\begin{aligned}
\mathbf{G}_1 &= [(\boldsymbol{\beta}-\mathbf{B})+\mathbf{\Pi}_{x_g^U X_g}]+[\mathbf{\Gamma}_1+\mathbf{\Pi}_{X_g^U X_g}\mathbf{\Gamma}_1^U+\mathbf{\Pi}_{z_{2g}^U X_g}] \\
\mathbf{G}_2 &= \mathbf{\Gamma}_2+\mathbf{\Pi}_{z_{2g}^U Z_{2g}} \\
\mathbf{G}_3 &= 0 \\
\mathbf{G}_4 &= 0
\end{aligned}
$$

Proof:

As in the main text, we begin by noting that since none of $\mathbf{X}_g$, $\mathbf{Z}_{2g}$, $M_g\mathbf{X}_g$, and $M_g\mathbf{Z}_{2g}$ vary within groups, $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$, and $\mathbf{G}_4$ are identified exclusively from between-group variation. Thus, we can simplify our analysis by noting that the OLS coefficients $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$, and $\mathbf{G}_4$ are numerically identical to the coefficients of the projection of the adjusted group g mean of Y_{gi}, $Y_g-[\mathbf{X_g B}+M_g\mathbf{X_g r_1}+M_g\mathbf{Z_{2g} r_2}]$, onto $\mathbf{X_g}$, $\mathbf{Z_{2g}}$, $M_g\mathbf{X}_g$, and $M_g\mathbf{Z_{2g}}$.

Using (10), we obtain

$$
\begin{aligned}
Y_g-[\mathbf{X}_g\mathbf{B}+M_g\mathbf{X}_g\mathbf{r}_1+M_g\mathbf{Z}_{2g}\mathbf{r}_2] \quad = \quad & \mathbf{X}_g[\boldsymbol{\beta}-\mathbf{B}+\mathbf{\Gamma}_1]+\mathbf{Z}_{2g}\mathbf{\Gamma}_2+M_g\mathbf{X}_g[\boldsymbol{\rho}_1-\mathbf{r_1}]+M_g\mathbf{Z}_{2g}[\boldsymbol{\rho}_2-\mathbf{r}_2] \\
& +M_g\mathbf{X}_g^U\boldsymbol{\rho}_1^U+M_g\mathbf{Z}_{2g}^U\boldsymbol{\rho}_2^U+x_g^U+z_g^U \qquad (48)
\end{aligned}
$$

Recall that under assumptions A1-A5 $\mathbf{X}_g^U=\mathbf{X}_g\mathbf{\Pi}_{X_g^U X_g}$, so that $x_g^U=\mathbf{X}_g\mathbf{\Pi}_{X^U X}\boldsymbol{\beta}^U\equiv\mathbf{X}_g\mathbf{\Pi}_{x_s^U X_s}$. Furthermore, recall that $z_g^U=\mathbf{X}_g^U\mathbf{\Gamma}_1^U+\mathbf{Z}_{2g}^U\mathbf{\Gamma}_2^U$ and define $z_{2g}^U=\mathbf{Z}_{2g}^U\mathbf{\Gamma}_2^U$. Then we can simplify (48) as follows:

$$
\begin{aligned}
Y_g-[\mathbf{X_g}\mathbf{B}+M_g\mathbf{X}_g\mathbf{r}_1+M_g\mathbf{Z}_{2g}\mathbf{r}_2] \quad = \quad & \mathbf{X}_g[\boldsymbol{\beta}-\mathbf{B}+\mathbf{\Gamma}_1+\mathbf{\Pi}_{x_g^U X_g}+\mathbf{\Pi}_{X_g^U X_g}\mathbf{\Gamma}_1^U] \qquad (49) \\
& +\mathbf{Z}_{2g}\mathbf{\Gamma}_2+M_g\mathbf{X}_g[\boldsymbol{\rho}_1-\mathbf{r}_1+\mathbf{\Pi}_{X_g^U X_g}\boldsymbol{\rho}_1^U]+M_g\mathbf{Z}_{2g}[\boldsymbol{\rho}_2-\mathbf{r}_2] \qquad (50) \\
& +M_g\mathbf{Z}_{2g}^U\boldsymbol{\rho}_2^U+z_{2g}^U. \qquad (51)
\end{aligned}
$$

Let projections of the error components z_{2g}^U and $M_g\mathbf{Z}_{2g}^U\boldsymbol{\rho}_2^U$ in the above equation onto $\mathbf{X}_g$, $\mathbf{Z}_{2g}$, $M_g\mathbf{X}_g$, and $M_g\mathbf{Z}_{2g}$ be given by:

$$
z_{2g}^U=\mathbf{X}_g\mathbf{\Pi}_{z_{2g}^U X_g}+\mathbf{Z}_{2g}\mathbf{\Pi}_{z_{2g}^U Z_{2g}}+M_g\mathbf{X}_g\mathbf{\Pi}_{z_{2g}^U,M_g X_g}+M_g\mathbf{Z}_{2g}\mathbf{\Pi}_{z_{2g}^U,M_g Z_{2g}}+\tilde{z}_{2g}^U \qquad (52)
$$

$$
\begin{aligned}
& M_g\mathbf{Z_{2g}^U}\boldsymbol{\rho}_2^U=\mathbf{X}_g(0)+\mathbf{Z}_{2g}(0)+ \\
& M_g\mathbf{X}_g\mathbf{\Pi}_{M_g Z_{2g}^U,M_g X_g}\boldsymbol{\rho}_2^U+M_g\mathbf{Z}_{2g}\mathbf{\Pi}_{M_g Z_{2g}^U,M_g Z_{2g}}\boldsymbol{\rho}_2^U+\widetilde{M_g\mathbf{Z_{2g}^U}\boldsymbol{\rho}_2^U}. \qquad (53)
\end{aligned}
$$

Note that we have imposed assumption A9 in order to set to zero the projection coefficients on X_g and Z_{2g} in (53). Recall that A9 implies that (15) gives the conditional expectation $\mathbf{E}[\mathbf{Z}_{2g}^U|\mathbf{X}_g,\mathbf{Z}_{2g}]$ and that $\widetilde{\mathbf{Z}_{2g}^U}$ is independent of $\mathbf{X}_g$ and $\mathbf{Z}_{2g}$ and not simply uncorrelated with them.

Collecting terms from equations (51)-(53), we conclude that

$$
\begin{aligned}
\mathbf{G}_1 &= [(\boldsymbol{\beta}-\mathbf{B})+\boldsymbol{\Pi}_{X_g^U X_g}]+[\boldsymbol{\Gamma}_1+\boldsymbol{\Pi}_{X_g^U X_g}\boldsymbol{\Gamma}_1^U+\boldsymbol{\Pi}_{z_{2g}^U X_g}] && (54)\\
\mathbf{G}_2 &= \boldsymbol{\Gamma}_2+\boldsymbol{\Pi}_{z_g^U Z_{2g}} && (55)\\
\mathbf{G}_3 &= [\boldsymbol{\rho}_1-\mathbf{r}_1+\boldsymbol{\Pi}_{X_g^U X_g}\boldsymbol{\rho}_1^U]+\boldsymbol{\Pi}_{z_g^U,M_g X_g}+\boldsymbol{\Pi}_{M_g Z_{2g}^U,M_g X_g}\boldsymbol{\rho}_2^U && (56)\\
\mathbf{G}_4 &= [\boldsymbol{\rho}_2-\mathbf{r}_2]+\boldsymbol{\Pi}_{z_g^U,M_g Z_{2g}}+\boldsymbol{\Pi}_{M_g Z_{2g}^U,M_g Z_{2g}}\boldsymbol{\rho}_2^U && (57)
\end{aligned}
$$

However, the expressions for $\mathbf{G}_3$ and $\mathbf{G}_4$ can be simplified further.

First note that A9 and the fact that M_g is a linear function of $\mathbf{X}_g$ implies that $\tilde{\mathbf{Z}}_{2g}^U$ and $\tilde{z}_{2g}^U$ are independent of and thus do not co-vary with functions of $\mathbf{X}_g$ and $\mathbf{Z}_{2g}$. Thus, the terms $\boldsymbol{\Pi}_{z_{2g}^U,M_g X_g}$ and $\boldsymbol{\Pi}_{z_{2g}^U,M_g X_g}$ in (52) are 0 and therefore drop out of equations (56) and (57), respectively.

Second, note that from (15) we can write $M_g\mathbf{Z}_{\mathbf{2g}}^{\mathbf{U}}\boldsymbol{\rho}_2^U$ as $M_g[\mathbf{X}_g\boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U+\mathbf{Z}_{2g}\boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U]+M_g\widetilde{\mathbf{Z}_{2g}^U}\boldsymbol{\rho}_2^U$ and separately consider the projections of the two terms. The first term is exactly equal to $M_g\mathbf{X}_g\boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U+M_g\mathbf{Z}_{2g}\boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U$, so $\mathbf{X}_g$ and $\mathbf{Z}_{2g}$ do not play a role. Furthermore, $E(M_g\widetilde{\mathbf{Z}_{2g}^U}\boldsymbol{\rho}_2^U|\mathbf{X}_g,\mathbf{Z}_{2g},M_g\mathbf{X}_g,M_g\mathbf{Z}_{2g})=0$ by **A9**, the fact that M_g is function of $\mathbf{X}_g$ and is thus also independent of $\widetilde{\mathbf{Z}_{2g}^U}$, and the fact that the expectation of the product of two independent random variables is the product of the expectations. Consequently, the right hand side of (53) simplifies to $M_g\mathbf{X}_g\boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U+M_g\mathbf{Z}_{2g}\boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U$, and the terms $\boldsymbol{\Pi}_{M_g Z_{2g}^U,M_g X_g}\boldsymbol{\rho}_2^U$ and $\boldsymbol{\Pi}_{M_g Z_{2g}^U,M_g Z_{2g}}\boldsymbol{\rho}_2^U$ become $\boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U$ and $\boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U$ in equations (56) and (57), respectively.

Thus, $\mathbf{G}_3$ and $\mathbf{G}_4$ simplify to:

$$
\mathbf{G}_3=\boldsymbol{\rho}_1-\mathbf{r}_1+\boldsymbol{\Pi}_{X_g^U X_g}\boldsymbol{\rho}_1^U+\boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U \qquad \mathbf{G}_4=\boldsymbol{\rho}_2-\mathbf{r}_2+\boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U \tag{58}
$$

But recall the results of Proposition 2:

$$
\begin{aligned}
\mathbf{r}_1 &= \boldsymbol{\rho}_1+\boldsymbol{\Pi}_{X_g^U X_g}\boldsymbol{\rho}_1^U+\boldsymbol{\Pi}_{Z_{2g}^U X_g}\boldsymbol{\rho}_2^U && (59)\\
\mathbf{r}_2 &= \boldsymbol{\rho}_2+\boldsymbol{\Pi}_{Z_{2g}^U Z_{2g}}\boldsymbol{\rho}_2^U && (60)
\end{aligned}
$$

This implies that both $\mathbf{G_3}$ and $\mathbf{G_4}$ are zero.

Combining these insights we obtain:

$$
\begin{aligned}
\mathbf{G}_1 &= [(\boldsymbol{\beta}-\mathbf{B})+\boldsymbol{\Pi}_{X_g^U X_g}]+[\boldsymbol{\Gamma}_1+\boldsymbol{\Pi}_{X_g^U X_g}\boldsymbol{\Gamma}_1^U+\boldsymbol{\Pi}_{z_{2g}^U X_g}]\\
\mathbf{G}_2 &= \boldsymbol{\Gamma}_2+\boldsymbol{\Pi}_{z_{2g}^U Z_{2g}}\\
\mathbf{G}_3 &= 0\\
\mathbf{G}_4 &= 0
\end{aligned}
$$

This completes the proof.

A3 Allocating Bias Correction Terms Across Error Components

Step 3 of Section 6.2 describes how we implement the bias correction to remove sampling variance from our estimates of the variances and covariances of our observed regression indices. However, consider the bias correction term $\frac{1}{N}\sum_i \mathbf{X_{s(i)}}\mathbf{Var}(\mathbf{\hat{G}_1^S}-\mathbf{G_1^S})\mathbf{X'_{s(i)}}$ that is subtracted from $Var(\mathbf{X_s}\mathbf{\hat{G}_1^S})$ to estimate $Var(\mathbf{X_s}\mathbf{G_1})$. Assuming that the outcome is measured without error, the expected sampling variance captured by this correction term reflects true inputs into Y_i that should have been allocated to the unobserved error components $v_i - v_n$, $v_n - v_s$, or $v_s - v_c$.

To determine the share of the bias correction to allocate to each error component, we ignore the small amount of heterogeneity in the number of sampled students per neighborhood, the number of sampled neighborhoods per school, and the number of sampled schools per commuting zone, and treat these as fixed scalar values $\frac{I}{N}$, $\frac{N}{S}$, and $\frac{S}{C}$, respectively (where I, N, S, and C are the number of sampled individuals, neighborhoods, schools, and commuting zones). We also treat the population number of students per neighborhood, neighborhoods per school, and schools per commuting zone as large, so that such sampling variance would disappear if we observed the full population of high school students in the United States. Then the variance in the sampling error among school averages Y_s within the same commuting zone (for schools each featuring $\frac{I}{S}$ sample members) is given by:

$$
\begin{aligned}
&Var(\frac{1}{I/S}\sum_{i\in s}[(v_i - v_{n(i)}) + (v_{n(i)} - v_{s(i)}) + v_s] \\
&= Var(\frac{1}{I/S}\sum_{i\in s}(v_i - v_n)) + Var(\frac{1}{N/S}\sum_{n'=1}^{\frac{N}{S}}(v_{n'} - v_s)) + Var(v_s) \\
&= \frac{1}{(I/S)^2}Var(\sum_{i\in s}(v_i - v_n)) + \frac{1}{(N/S)^2}Var(\sum_{n'=1}^{\frac{N}{S}}(v_{n'} - v_s)) + Var(v_s) \\
&= \frac{1}{(I/S)^2}(I/S)Var(v_i - v_n) + \frac{1}{(N/S)^2}(N/S)Var(v_n - v_s) + Var(v_s) \\
&= \frac{Var(v_i - v_n)}{I/S} + \frac{Var(v_n - v_s)}{N/S} + Var(v_s), \qquad (61)
\end{aligned}
$$

where we have assumed independence in the draws of $v_i - v_{n(i)}$, $v_n - v_s$, and v_s across individuals, neighborhoods and schools.

Thus, the individual, neighborhood, and school shares of the variance in the sampling error among school averages Y_s is given by:

$$Share_S^I = \frac{\frac{Var(v_i - v_n)}{I/S}}{\frac{Var(v_i - v_n)}{I/S} + \frac{Var(v_n - v_s)}{N/S} + Var(v_s)} \tag{62}$$

$$Share_S^N = \frac{\frac{Var(v_n - v_s)}{N/S}}{\frac{Var(v_i - v_n)}{I/S} + \frac{Var(v_n - v_s)}{N/S} + Var(v_s)} \tag{63}$$

$$Share_S^S = \frac{\frac{Var(v_s)}{N/S}}{\frac{Var(v_i - v_n)}{I/S} + \frac{Var(v_n - v_s)}{N/S} + Var(v_s)} \tag{64}$$

We assume that the sampling variance component of the estimated variance of each school-level regression index (or the estimated covariance among each pair of school-level regression indices) contains individual, neighborhood, and school subcomponents in the same proportions as the overall variance in sampling error among school averages Y_s. Thus, we allocate the estimated sampling variance $\frac{1}{N}\sum_i \mathbf{X_{s(i)}}\mathbf{Var}(\mathbf{\hat{G}_1^S} - \mathbf{G_1^S})\mathbf{X'_{s(i)}}$ associated with $Var(\mathbf{X_s}\mathbf{\hat{G}_1})$, for example, to the individual-level, neighborhood-level, and school-level error variances $Var(v_i - v_n)$, $Var(v_n - v_s)$ and $Var(v_s - v_c)$ according to the shares given in (62) - (64). We use analogous formulae to derive the individual and neighborhood shares used to allocate neighborhood-level sampling variance terms and to derive the individual, neighborhood, school, and commuting zone shares used to allocate commuting zone-level sampling variance terms.