QUANTIFYING FAMILY, SCHOOL, AND LOCATION EFFECTS IN THE PRESENCE
OF COMPLEMENTARITIES AND SORTING

Mohit Agrawal
Joseph G. Altonji
Richard K. Mansfield

Quantifying Family, School, and Location Effects in the Presence of Complementarities and Sorting
Mohit Agrawal, Joseph G. Altonji, and Richard K. Mansfield
NBER Working Paper No. 25167
October 2018
JEL No. C1,C31,I20,I24,I26,R23

ABSTRACT

We extend the control function approach of Altonji and Mansfield (2018) to allow for multiple group levels and complementarities. Our analysis provides a foundation for causal interpretation of multilevel mixed effects models in the presence of sorting. In our empirical application, we obtain lower bound estimates of the importance of school and commuting zone inputs for education and wages. A school/location combination at the 90th versus 10th percentile of the school/location quality distribution increases the high school graduation probability and college enrollment probability by at least .06 and .17, respectively. Treatment effects are heterogeneous across subgroups, primarily due to nonlinearity in the educational attainment model.

Mohit Agrawal
Yale University
PO Box 208268
New Haven, CT 06520-8268
mohitii@gmail.com

Joseph G. Altonji
Department of Economics
Yale University
Box 208264
New Haven, CT 06520-8264
and NBER
joseph.altonji@yale.edu

Richard K. Mansfield
Economics Building Room 04B
University of Colorado-Boulder
Boulder, CO 80309
and NBER
richard.mansfield@colorado.edu

1 Introduction

Designers of policy interventions often must choose whether to target neighborhoods, schools, or broader local or regional areas. They would like to know which policies are most effective for each area, but research has been very slow in providing answers (Katz 2015). In this paper, we focus on the more modest goal of measuring the overall importance of neighborhoods, schools, and local areas for adult outcomes. Information about where inequality arises provides guidance about where the most potential progress can be made, even if it does not provide a specific plan of action.

The efficacy of particular policies and the relative importance of neighborhood vs. school, vs. broader local area factors is likely to vary across students. For example, school-level dropout prevention policies may improve outcomes for struggling students but not stronger or better-supported students, while opening a satellite campus of a flagship research university might benefit high achieving low income students the most. Such considerations motivate us to work with a multilevel mixed effects (MME) model in which individual characteristics interact with observed and unobserved school and area characteristics.

The use of MME models (also known as hierarchical linear models and random coefficients models) was stimulated by interest in the interplay among the student, classroom, and school factors that determine educational outcomes.[1] However, the MME literature is mindful of but does not address sorting bias arising from school and neighborhood choice. As Durlauf (2004) and Graham (2018) emphasize, differences in outcomes across neighborhoods combine the causal effects of neighborhood with differences across neighborhoods in the observed and unobserved attributes of children that matter for their outcomes. Sorting poses similar problems for assessing schools.

Our main methodological contribution is to address the sorting problem present in MME models. We do this by using a flexible production function specification and a rich but tractable model of school and neighborhood choice to derive exact formulas relating the slope parameters and error component variances of a standard MME specification to the corresponding production function parameters. We then use these formulas combined with MME estimates to characterize the importance of school and location choices for students' long-run education and labor market outcomes.

We postpone precise statements of our theoretical results and the assumptions, but the questions about MME models that we address and the answers that we give may be summarized as follows. Consider the following MME level regression model:

$$Y_{gi} = \boldsymbol{X}_i \boldsymbol{B} + \boldsymbol{X}_g \boldsymbol{G}_1 + \boldsymbol{Z}_{2g} \boldsymbol{G}_2 + M_i \boldsymbol{X}_g \boldsymbol{r}_1 + M_i \boldsymbol{Z}_{2g} \boldsymbol{r}_2 + v_g + (v_{gi} - v_g). \tag{1}$$

Here i denotes the student and g denotes the neighborhood, school, and associated local area the student is exposed to. Y_{gi} is a long-run student outcome, such as wages at age 25. $\boldsymbol{X}_i$ is a set of observed student characteristics such as mother's education. $\boldsymbol{X}_g$ is a set of observed peer

---

[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.

1

characteristics—averages of X_i among the students in g. Z_{2g} is a set of observed characteristics of the school and location that influence wages, such as student/teacher ratio. M_i is a student characteristic that interacts with X_g and Z_{2g} in influencing outcomes, such as minority status. v_g is a group specific error and $(v_{gi} - v_g)$ is an idiosyncratic error. The error components are treated as random effects and are defined to be uncorrelated with the explanatory variables. The slope coefficients and error variances are defined to be what one recovers using a random effects estimator with a very large sample.

The production function that actually determines the outcome Y_{gi} is assumed to take the same linear form with interaction terms as the above estimating equation. However, it also includes unobserved (by the econometrician) counterparts to X_i, X_g, Z_{2g}, M_iX_g and M_iZ_{2g}. In general, sorting-driven differences across school/locations in unobserved student characteristics as well as other forms of omitted variables bias lead the parameters of (1) to differ from the production function parameters.

We then build upon and extend to an interactive setting the control function approach of Altonji and Mansfield (2018) (hereafter, AM). In particular, they show that even a flexible version of a standard multinomial 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 (X_g) can serve as controls for the averages of unobserved student characteristics that would otherwise bias estimates of the importance of school and neighborhood effects.

Under somewhat stronger assumptions (discussed below) we show that the coefficient vector r_2 on the interaction between a student characteristic M_i such as mother's education and a set of observed "group" characteristics Z_{2g} 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 (e.g. the quality of a principal) that covary with the observed group characteristics, holding observed and unobserved student characteristics fixed. Analogously, r_1 combines two components. The first is the causal effect of interactions between M_i and both observed and unobserved peer characteristics (such as average mother's education and average unobserved parental motivation). The second is the causal effect of part of the interaction between M_i and unobserved school and location characteristics. The key point is that under the assumptions of the model the estimates of both r_1 and r_2 exclusively reflect interactions between the student and the environment and not sorting.

We also show that AM's expression for G_2, the coefficient vector on the main effects of observed school and area characteristics, remains unchanged when observed and unobserved interactions are introduced to the production function. Specifically, it picks up the causal effect of the observed school and area characteristics plus a second component that reflects their association with unobserved school and area characteristics that affect outcomes. It is not contaminated by sorting bias or unobserved interactions. AM's result that school and neighborhood error components v_g are simi-

2

larly uncontaminated by the effects of sorting also carries over to the model with interaction effects. Finally, we also justify a causal interpretation for a random slope coefficient on $\mathbf{M}_i$ that captures the interaction of an observed student characteristic with an index of unobserved school and area characteristics.

Taken together, these identification results reveal that even in the presence of endogenous sorting of individuals to groups, multilevel mixed effects estimates can 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 (associated with $\boldsymbol{X}_g\boldsymbol{G}_1$) 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 locations 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?

As in AM's analysis, the use of school/location averages of observed student characteristics $\boldsymbol{X}_g$ 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 importance of school and area factors for student outcomes.

We use the above analysis as the theoretical foundation to interpret MME model 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 (ZIP code or block group) and school identifiers.

We use 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 collections of attributes that contribute to the outcomes we consider (educational attainment or log 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 probability by 0.088 in NELS88 and by about 0.061 in ELS2002 (which has a higher baseline graduation rate–0.92). The effect of the school treatment is 0.06 in NELS and about 0.044 in ELS. The values for

3

the 10th-to-90th quantile shift in commuting zone quality are between 0.039 and 0.067.

For the college enrollment probability, the combined school and commuting zone treatment effect is between 0.17 and 0.18 for both NELS and ELS. The values for the school treatment are about 0.125, while the values for the commuting zone treatment are about 0.115 in each dataset. The effects of the combined treatment, the school treatment and the commuting zone treatment on the probability of completing a BA degree are somewhat smaller in both datasets (0.085, 0.077, and 0.046 in NELS and around 0.085, 0.070, and 0.065 in ELS). For wage rates at age 25, the combined school and commuting zone treatment ranges between 11 and 12 percent. The values for the school treatment range from 6.2 to 7.1 percent. The commuting zone effects range from 7.8 to 10.7 percent. 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 population groups: blacks, Hispanics, whites with single mothers who did not attend college, whites with two resident parents with college degrees, as well as for students at each percentile of an index of student level variables that predict the outcomes. We do not find strong evidence that wages or the latent indices governing the binary educational attainment outcomes depend on the interaction between observed student characteristics and characteristics of the school and commuting zone, although we cannot rule out modest interactions. Nevertheless, we do find important differences in treatment effects across students. The heterogeneity is primarily due to the fact that the probit function that we use for the education outcomes naturally imposes greater sensitivity to school and location environment for students predicted to be close to the decision margin. Specifically, high school graduation 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.

Finally, we provide a richer picture of sorting than is possible with two level models featuring individual and school (Coleman et al. 1966; Jencks and Brown 1975; Betts 1995; Altonji and Mansfield 2011; Alexander and Morgan 2016) or individuals and neighborhood (Solon et al. 2000). We find 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 ZIP code or block group among schoolmates. Our variance decompositions of educational attainment and wages consistently show that individual level factors account for the lion's share of the variance, with school and commuting zone level variation making modest contributions, and neighborhood level factors playing a smaller role. For example, in ELS2002 for college attendance, the within-neighborhood (block group) share is 71.4 percent, while the neighborhood share is 3.1 percent, the school share is 15.9 percent, and the commuting zone share is 9.7 percent. School and commuting zone account for about 9 and 6 percent of the variance in log wages respectively, and block group within a school accounts for about 3%. Within a school there is very little clustering in any of our

outcomes at the ZIP code level.

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.[2] AM in turn draw on the rich theoretical and empirical literature on equilibrium sorting and matching other contexts.[3] Control function approaches are widely used in econometric work.

On the empirical side, the literatures on the importance of families, neighborhoods, schools, and region are each too vast to discuss meaningfully here. The most directly relevant are the large number of studies that 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 multilevel structure similar to the NELS88 and ELS2002.[4] 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), and 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 attainment, 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 (2018).[5] Many of the papers emphasize that estimates of the impacts of particular characteristics, such as percent minority, segregation measures, and poverty rates on socioeconomic 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.

Meghir and Rivkin (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 (2018).

---

[2] 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.

[3] These include marriage (eg., Browning et al. 2014), the labor market (eg., Rosen 1974; Ekeland et al. 2004; Lise 2016; Lindenlaub 2017), and neighborhoods and schools (eg., Epple and Platt 1998; Epple and Sieg 1999).

[4] 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.

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

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 benefits for children who move before age 13 but not for older children.

Using a sibling differences approach that also exploits high quality tax data, Chetty and Hendren (2018a; 2018b) 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.[6]

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 exams 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 and Fryer (2011) and Laliberte (2018) are rare quasi-experimental attempts to distinguish neighborhood effects from school effects. Dobbie and Fryer (2011) finds 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. Laliberte (2018) also finds that school effects are much more important than neighborhoods in determining postsecondary enrollment and completion, at least in Montreal.

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 MME 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 on various educational outcomes and on log wages generated by shifts in school and commuting zone quality, both on average and for particular subpopulations. Section 9 concludes.

---

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

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, recast slightly to more explicitly consider neighborhood, school, and labor market area.[7] 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. We use the terms "area" or "location" to refer to the joint influence of neighborhood and commuting zone inputs. 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 $\mathcal{N}$ of neighborhoods. Due to attendance boundaries and travel costs, the choice of n then restricts the choice of school s to the subset $\mathcal{S}(n)$ of the full set $\mathcal{S}$. 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. 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.

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

$$\mathcal{U}_i(g) = \boldsymbol{W}_i \boldsymbol{A}_g + \varepsilon_{gi} - P_g. \tag{2}$$

$\boldsymbol{A}_g \equiv [A_{1g}, \ldots, A_{Kg}]'$ is a $K \times 1$ vector of underlying latent amenities that characterize the neighborhood, the school and the commuting zone combination to which g refers. $\boldsymbol{W}_i \equiv [W_{1i}, \ldots, W_{Ki}]$ is a $1 \times K$ vector of weights that captures the increases in family i's willingness to pay for a neighborhood and school per unit increase in each of its K amenity factors $A_{1g}, \ldots, A_{Kg}$, respectively. P_g 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_{gi} is an idiosyncratic taste of the parent/student i for the particular location and school.

Next we specify an equation for willingness to pay (WTP) for particular amenities:

$$\boldsymbol{W}_i = \boldsymbol{X}_i \boldsymbol{\Theta} + \boldsymbol{X}_i^U \boldsymbol{\Theta}^U + \boldsymbol{Q}_i \boldsymbol{\Theta}^Q. \tag{3}$$

where the row vectors $\boldsymbol{X}_i$, $\boldsymbol{X}_i^U$ and $\boldsymbol{Q}_i$ have L, L^U, and L^Q elements, respectively. The coefficient matrix $\boldsymbol{\Theta}$ is an $L \times K$ matrix whose ℓk-th entry captures the extent to which the willingness to pay for the k-th element of the amenity vector $\boldsymbol{A}_g$ varies with the ℓ-th element of $\boldsymbol{X}_i$. The $L^U \times K$ matrix $\boldsymbol{\Theta}^U$ and the $L^Q \times K$ matrix $\boldsymbol{\Theta}^Q$ are the corresponding coefficient matrices for $\boldsymbol{X}_i^U$ and $\boldsymbol{Q}_i$. We define $\boldsymbol{X}_i^U$ so that $[\boldsymbol{X}_i, \boldsymbol{X}_i^U]$ represents the *complete* set of student attributes that determine Y_{si}. $\boldsymbol{Q}_i$ is a set

---

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

of variables that influence WTP for amenities but have no direct effect on student outcomes. It is convenient to define $\boldsymbol{Q}_i$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}^U$ so that $\boldsymbol{Q}_i$ is orthogonal to $[\boldsymbol{X}_i, \boldsymbol{X}_i^U]$.

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

$$\mathscr{U}_i(g) = (\boldsymbol{X}_i\boldsymbol{\Theta} + \boldsymbol{X}_i^U\boldsymbol{\Theta}^U + \boldsymbol{Q}_i\boldsymbol{\Theta}^Q)\boldsymbol{A}_g + \varepsilon_{gi} - P_g \tag{4}$$

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 (2).

Expected utility is taken with respect to the information available when n and s are chosen. The information set includes the price and the amenity vector in each school and neighborhood as well as student/parent characteristics $[\boldsymbol{X}_i, \boldsymbol{X}_i^U, \boldsymbol{Q}_i]$ and the values of ε_{gi} in each school and neighborhood. 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 elements of $\boldsymbol{A}_g$ may depend on neighborhood and school characteristics that influence educational attainment and labor market outcomes (denoted $\boldsymbol{X}_g^*$ or $\boldsymbol{Z}_{2g}^*$ in the production function introduced later). The amenities may also include or depend on aspects of the demographic composition of the neighborhood and school. Thus, some of the amenities are influenced by the sorting equilibrium.

The parents of i choose $g(i) = g$ if net utility $\mathscr{U}_i(g)$ is the highest among the options. Parents behave competitively in the sense that prices P_g and amenities $\boldsymbol{A}_g$ are taken as given, and choice is unrestricted. In equilibrium the values of some elements of $\boldsymbol{A}_g$ may in fact depend on the averages of $\boldsymbol{X}_i$ and $\boldsymbol{X}_i^U$ for the parents who choose g. Parents ignore the externalities they are imposing on others.

Proposition 1 of AM establishes that the expectation $\boldsymbol{X}_g^U \equiv \boldsymbol{E}[\boldsymbol{X}_i^U|g(i) = g]$ is linearly dependent on the expectation $\boldsymbol{X}_g \equiv \boldsymbol{E}[\boldsymbol{X}_i|g(i) = g]$ if five assumptions hold. Decompose $\boldsymbol{X}_i^U$ into its linear prediction given $\boldsymbol{X}_i$ and an uncorrelated residual vector $\tilde{\boldsymbol{X}}_i^U$:

$$\boldsymbol{X}_i^U = \boldsymbol{X}_i\boldsymbol{\Pi}_{\boldsymbol{X}^U\boldsymbol{X}} + \tilde{\boldsymbol{X}}_i^U. \tag{5}$$

Use (5) to rewrite (3) as $\boldsymbol{W}_i = \boldsymbol{X}_i\tilde{\boldsymbol{\Theta}} + \tilde{\boldsymbol{X}}_i^U\boldsymbol{\Theta}^U + \boldsymbol{Q}_i\boldsymbol{\Theta}^Q$, where $\tilde{\boldsymbol{\Theta}} = [\boldsymbol{\Theta} + \boldsymbol{\Pi}_{\boldsymbol{X}^U\boldsymbol{X}}\boldsymbol{\Theta}^U]$.[8] In the rewritten form, the three components of $\boldsymbol{W}_i$ are mutually orthogonal.

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

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

A2: *Parents take P_g and $\boldsymbol{A}_g$ as given when choosing location, and face a common choice set.*

A3: *The idiosyncratic preference components ε_{gi} have a mean of 0 and are independent of $\boldsymbol{X}_i$.*

---

[8]Throughout the paper we use the symbol $\boldsymbol{\Pi}_{\boldsymbol{HV}}$ to denote the vector or matrix of partial regression coefficients relating the scalar or vector $\boldsymbol{H}$ to the explanatory variables $\boldsymbol{V}$. In the case of $\boldsymbol{\Pi}_{\boldsymbol{X}^U\boldsymbol{X}}$, $\boldsymbol{H} = \boldsymbol{X}_i^U$ and $\boldsymbol{V} = \boldsymbol{X}_i$.

X_i^U, Q_i, and A_g for all g.

A4: $E(X_i|W_i)$ and $E(X_i^U|W_i)$ are linear in W_i.

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

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

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

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

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

The proof is in AM. Proposition 1 states that the sorting model introduced in this section implies that the vector X_g can serve as a set of controls for X_g^U. This is key to distinguishing the causal effects of school and location on outcomes from sorting on unobservable student characteristics that affect these outcomes. The intuition behind the proposition is that X_g^U and X_g are related across g because both X_i and X_i^U influence tastes for A_g. As a result the means X_g^U and X_g are both functions of A_g, and under A1-A5, the maps from A_g to X_g and from A_g to X_g^U determines a map from X_g to X_g^U that takes form (7).[9]

AM provide a detailed discussion of the assumptions. We will not repeat that discussion here, but a few words about A4 and A5 are in order. A sufficient condition for A4 is that the joint distribution of $[X_i, X_i^U, Q_i]$ is a member of the continuous elliptical class. This class includes the normal, the Laplace, and the multivariate exponential power family. In our application X_i includes several discrete variables, so the sufficient condition will not be satisfied. AM show that if the assumption fails, an approximation error will appear in (7) that is likely to contribute to the error component ν_g of (1) and lead to some upward bias in our estimator of the importance of school/location effects. But AM also provide reasons to believe the approximation error will be small.

We consider the spanning condition A5 to be the key assumption for Proposition 1; it ensures that X_g fully determines X_g^U. Roughly speaking, two conditions must hold for X_g^U to be a function of X_g. The first is that the number of elements of X_i must be greater than or equal to the number of amenities that X_i and/or X_i^U affect tastes for. Otherwise, A_g and thus X_g^U will vary conditional on X_g.[10]

The second condition is that there cannot exist any elements of X_i^U that are both uncorrelated

---

[9]Note 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 (7) is a good approximation even with samples of 20 students at each school (around the number observed per school used to construct X_g in the samples we employ below).

[10]The plausibility of the first condition depends on the number and breadth of coverage of the variables in X_i. The condition implies that the number of factors that determine X_g is less than or equal to L. AM formally test this restriction and find that the number of factors is considerably less then L.

with X_i and that shift preferences for amenity factors that neither X_i nor those elements of X_i^U that are correlated with X_i shift preferences for. Otherwise such amenity factors will shift the distribution of these remaining X_i^U elements across schools in a way that X_g cannot predict. While this condition is untestable without further structure, we have struggled to find an example of an unobserved variable that would cause its violation given the rich set of X_i variables that we have.[11]

3   The Production Function for Education and Wages

In this section we present the model of outcomes. This is the "production function", as distinct from the MME estimating equation presented in the introduction. The model is similar to its counterpart from AM but 1) introduces potential interactions or complementarities between individual inputs and each category of group-level inputs, and 2) allows for separate roles played by neighborhood, school, and local area inputs.

Our outcomes are high school graduation, attendance at a four-year college, graduation from a four-year college, and log wages. The outcome Y_i of student i whose family has chosen the neighborhood/school/commuting zone combination $g(i)$ is determined according to

$$Y_i \equiv X_i^* \boldsymbol{\beta}^* + Z_g^* \boldsymbol{\Gamma}^* + M_i Z_g^* \boldsymbol{\rho}^* + \eta_{gi} + \xi_{gi} \qquad (8)$$

where we have dropped the dependence of the neighborhood, school, and commuting zone subscript on the student subscript i to simplify the notation. The parameters of equation (8) also implicitly depend on the outcome. For the education outcomes, Y_i is the latent variable that determines the binary outcome.

The row vector $X_i^* \equiv [X_i, X_i^U]$ denotes the exhaustive set of child and family characteristics. As a reminder, the superscript U denotes "unobserved" throughout the paper. It indicates variables that are unobserved to the econometrician and their associated parameters. The row vector $Z_g^* \equiv [X_g, X_g^U, Z_{2g}, Z_{2g}^U]$ contains four components. The vectors X_g and X_g^U consist of neighborhood, school, and commuting zone averages of $[X_i, X_i^U]$ that capture peer influences. The vectors Z_{2g} and Z_{2g}^U consist of the set of other observed and unobserved location and school characteristics that affect outcomes and are not mechanically related to peer characteristics. All are normalized to have a population mean of 0. To simplify the derivation and presentation of our identification results relating to interactions between individual-level and group-level inputs, we delay explicitly distinguishing between the components of Z_g^* that represent neighborhood, school and commuting zone inputs, respectively, until Section 4.5.

The parameter vectors $\boldsymbol{\beta}^* = [\boldsymbol{\beta}', \boldsymbol{\beta}^{U\prime}]'$ and $\boldsymbol{\Gamma}^* \equiv (\boldsymbol{\Gamma}_{1g}', \boldsymbol{\Gamma}_{1g}^{U\prime}, \boldsymbol{\Gamma}_{2g}', \boldsymbol{\Gamma}_{2g}^{U\prime})'$ are the corresponding slope coefficients or input productivities, some of which may be zero. By virtue of our normalization $\boldsymbol{\Gamma}^*$ captures the effects of school and location variables at the mean of M_i.

---

[11]See AM for a more complete discussion, including a Monte Carlo analysis suggesting that (7) is quite robust to violations of the spanning condition in which just a few outcome-relevant unobservables in X_i^U affect WTP for just a few additional amenities that are not weighted by any elements of X_i.

M_i is a scalar-valued linear function of X_i. We focus on the scalar case to simplify the presentation and the proofs of identification results in Propositions 2 and 3 below. In the empirical specification, M_i is replaced with a vector.

The parameter $\boldsymbol{\rho}^*$ captures the extent to which students with different M_i values respond differently to the *same* school and location inputs or policies. But it also captures how students with different M_i values tend to receive different inputs within a school/location. For example, one might expect a school variable such as tracking to influence the relationship between X_i and the quality of the teachers a student receives at the school. Thus, the relevant elements of $\boldsymbol{\rho}^*$ are defined to also capture the degree to which school and location variables affect which group g resources get to allocated to students with higher M_i values relative to their group g peers.

The vector $\boldsymbol{\rho}^* = [\boldsymbol{\rho}_1', \boldsymbol{\rho}_1^{U\prime}, \boldsymbol{\rho}_2', \boldsymbol{\rho}_2^{U\prime}]'$ has four components. The vectors $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_1^U$ capture the strength of interactions or complementarities between student characteristics and observed and unobserved peer characteristics such as average family income. The vectors $\boldsymbol{\rho}_2$ and $\boldsymbol{\rho}_2^U$ capture interactions between student characteristics and observed and unobserved nonpeer school and location characteristics, such the quality of law enforcement, average class size, and availability of a nearby four-year college.[12] Our current identification results and empirical analysis exclude interaction variables that are functions of the unobservables X_i^U. We leave the task of relaxing this restriction to future work.

The unobserved scalar index η_{gi} captures the remaining random 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, and the luck of the draw in teacher quality at the school. η_{gi} is defined to have a mean of zero within each g, and given the definition of $\boldsymbol{\rho}^*$, it is uncorrelated with the interaction terms in the production function.

The component ξ_{gi} 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 within and across groups. We use the g 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_{gi} as $\xi_g + \xi_i$, where ξ_g is common to all students in g and ξ_i is idiosyncratic. ξ_g is taken to be 0 for the high school graduation outcome. Akin to η_{gi}, any systematic relationships between conditions after high school and observed and unobserved student, location and school characteristics are not deemed shocks and thus are properly reflected in the production function parameters, so that the "shock" component ξ_{gi} should only contain an unpredictable residual uncorrelated with other inputs.

Note that the characteristics/inputs at each level of observation can represent non-linear functions of other inputs from that level. For example, an element of X_i^* can represent the square of another element of X_i^* or the product of two elements. Thus, the linear-in-parameters specification for Y_i is more general than it first appears.

---

[12]In the empirical work, we will place restrictions on the $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ vectors and will often use estimated parameter vectors to construct the M_i values from X_i.

This production function yields a clear definition of the causal effect associated with group membership for a particular student type (indexed by M_i). Compared to a group featuring population-average values of all school and location inputs ($\mathbf{Z}_g^* = \mathbf{0}$), a randomly selected student with a particular value $M_i = M_i^1$ who grows up in group g^1 featuring external inputs $\mathbf{Z}_{g^1}^*$ can expect an increase (or decrease) in outcome Y_i of $[\mathbf{Z}_{g^1}^*,\boldsymbol{\Gamma}^* + M_i^a\mathbf{Z}_{g^1}^*,\boldsymbol{\rho}^*]$. A student with the population mean value $M_i = 0$ can expect an increase of $\mathbf{Z}_{g^1}^*\boldsymbol{\Gamma}^*$. In the case of college attendance, college graduation, and wage rates, the expected outcome from growing up in a particular school/location combination g^1 will also reflect ξ_{g^1}, which is common to all individuals in group g but is determined after high school.[13] Expressions for school effects, commuting zone effects, and a particular neighborhood-school-commuting zone combination are similar, and will be introduced later.

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) are determined by past parental inputs such as family income (Todd and Wolpin 2003; Cunha et al. 2006). 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 do not attempt to estimate $\boldsymbol{\beta}$.

Nor do we attempt to estimate the causal effects of particular neighborhood, school, or commuting zone inputs (and so we will not aim to separately identify particular elements of $\boldsymbol{\Gamma}_2$ and $\boldsymbol{\rho}_2$). This is because the control function variables only address sorting bias. They do 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, of school factors, and of commuting zone factors, respectively, from the effects of student, family, and prior school and location factors.

4   Identification Results for the MME Model

We now return to the MME model (1) for the outcome Y_i that we started with in the introduction, versions of which serve as our estimating equation. We discuss the relationship between the slope parameters and error components recovered from estimation and the parameters of the outcome production function (8) presented in Section 3. In particular, we show that the regression coefficients and error components estimated via a random effects procedure 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 (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 $\mathbf{X}_g$. It captures a combination of group inputs (mostly peer influences) and group averages of student in-

---

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

12

puts 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 Distinguishing Observables and Unobservables

We start by rewriting the production function (8) to distinguish between observables and unobservables. This facilitates the analysis in subsections 4.2 and 4.3 of what estimation of the MMF model identifies. Let $x_i^U \equiv \boldsymbol{X}_i^U \boldsymbol{\beta}^U$ denote the scalar index of unobserved student characteristics and let $z_{2g}^U \equiv \boldsymbol{Z}_{2g}^U \boldsymbol{\Gamma}_2^U$ denote the scalar index of the unobserved location and school inputs. Using these definitions and the subcomponents of the coefficient vectors $\boldsymbol{\beta}^*$, $\boldsymbol{\Gamma}^*$, and $\boldsymbol{\rho}^*$ and the variables $\boldsymbol{X}_g^*$ and $\boldsymbol{Z}_g^*$, one may rewrite (8) as

$$
\begin{aligned}
Y_i \;=\; & \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{X}_g\boldsymbol{\Gamma}_1 + \boldsymbol{Z}_{2g}\boldsymbol{\Gamma}_2 + M_i\boldsymbol{X}_g\boldsymbol{\rho}_1 + M_i\boldsymbol{Z}_{2g}\boldsymbol{\rho}_2 \\
& + M_i\boldsymbol{X}_g^U\boldsymbol{\rho}_1^U + M_i\boldsymbol{Z}_{2g}^U\boldsymbol{\rho}_2^U + x_i^U + z_g^U + \eta_{gi} + \xi_{gi}.
\end{aligned} \tag{9}
$$

Recall that the superscript "U" denotes "unobserved" variables and parameters associated with them.

Using the same notation, our estimating equation takes the form

$$
Y_{gi} = \boldsymbol{X}_i\boldsymbol{B} + \boldsymbol{X}_g\boldsymbol{G}_1 + \boldsymbol{Z}_{2g}\boldsymbol{G}_2 + M_i\boldsymbol{X}_g\boldsymbol{r}_1 + M_i\boldsymbol{Z}_{2g}\boldsymbol{r}_2 + M_g\boldsymbol{X}_g\boldsymbol{G}_3 + M_g\boldsymbol{Z}_{2g}\boldsymbol{G}_4 + v_g + (v_{gi} - v_g). \tag{10}
$$

The parameters refer to the coefficient vectors from the linear least-squares projection of Y_{gi} on the right hand side variables.

Including $M_g\boldsymbol{X}_g$ and $M_g\boldsymbol{Z}_{2g}$ ensures that identification of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ is obtained exclusively from within-group variation in $M_i\boldsymbol{X}_g$ and $M_i\boldsymbol{Z}_{2g}$. We show below that Assumptions A1-A5 of the sorting model plus assumptions A6 and A7 below imply that $\boldsymbol{G}_3$ and $\boldsymbol{G}_4$ are 0. So in the end we exclude $M_g\boldsymbol{X}_g$ and $M_g\boldsymbol{Z}_{2g}$ 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 the individual-level coefficients $\boldsymbol{B}$ and the interaction coefficients $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Next we turn to the group-level coefficients $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, $\boldsymbol{G}_3$, and $\boldsymbol{G}_4$ and then to the error components $v_g + (v_{gi} - v_g)$.

13

4.2 Interpretation of Student-Level Coefficients and Interaction Coefficients

First we consider $\boldsymbol{B}$, $\boldsymbol{r}_1$, and $\boldsymbol{r}_2$, which are identified from variation within g. Let D represent an operator that subtracts group means, so that $DY_{gi} \equiv Y_{gi} - E[Y_{gi}|g(i) = g] \equiv Y_{gi} - Y_g$. The within-group counterpart of our estimating equation (10) can be written as:

$$DY_{gi} = \boldsymbol{DX}_i\boldsymbol{B} + [(DM_i)\boldsymbol{X}_g]\boldsymbol{r}_1 + [(DM_i)\boldsymbol{Z}_{2g}]\boldsymbol{r}_2 + Dv_{gi}. \tag{11}$$

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

The statement of Proposition 2 below refers to the coefficients from the following projection equations:

$$\boldsymbol{DX}_i^U = \boldsymbol{DX}_i\Pi_{\boldsymbol{DX}_i^U\boldsymbol{DX}_i} + \widetilde{\boldsymbol{DX}}_i^U \tag{12}$$

$$D\eta_{gi} = \boldsymbol{DX}_i\Pi_{D\eta_{gi}\boldsymbol{DX}_i} + D\tilde{\eta}_{gi}. \tag{13}$$

$$\boldsymbol{X}_g^U = \boldsymbol{X}_g\Pi_{\boldsymbol{X}_g^U\boldsymbol{X}_g} \tag{14}$$

$$\boldsymbol{Z}_{2g}^U = \boldsymbol{X}_g\Pi_{\boldsymbol{Z}_{2g}^U\boldsymbol{X}_g} + \boldsymbol{Z}_{2g}\Pi_{\boldsymbol{Z}_{2g}^U\boldsymbol{Z}_{2g}} + \widetilde{\boldsymbol{Z}}_{2g}^U. \tag{15}$$

Note that $\boldsymbol{X}_g^U$ is perfectly predicted by $\boldsymbol{X}_g$ under the assumptions A1-A5, with $\Pi_{\boldsymbol{X}_g^U\boldsymbol{X}_g} \equiv [\Pi_{\boldsymbol{X}^U\boldsymbol{X}} + Var(\boldsymbol{X}_i)^{-1}R'Var(\tilde{\boldsymbol{X}}_i^U)]$. This is why (14) does not have an error term. Let $D\tilde{x}_i^U \equiv \widetilde{\boldsymbol{DX}}_i^U\boldsymbol{\beta}$ where $\widetilde{\boldsymbol{DX}}_i^U$ is defined in (12).

Two additional assumptions are needed when the interaction terms are present.

Assumption A6: $Cov(D\tilde{x}_i^U DM_i, \boldsymbol{X}_g) = 0$ and $Cov(D\tilde{x}_i^U DM_i, \boldsymbol{Z}_{2g}) = 0$
Assumption A7: $\tilde{\boldsymbol{Z}}_{2g}^U \perp (\boldsymbol{X}_g, \boldsymbol{Z}_{2g}, \boldsymbol{DX}_i)$

Assumption A6 says that $D\tilde{x}_i^U DM_i$ is uncorrelated with $[\boldsymbol{X}_g, \boldsymbol{Z}_{2g}]$. It addresses a potentially important identification problem. The variation that identifies interaction effects in our estimating equation consists of differences in the strength of the relationship between individual characteristics and outcomes across groups with different values of group characteristics (e.g. student-teacher ratio). Suppose that the covariance between parental income (observed) and parental motivation (unobserved) is larger in schools with smaller student-teacher ratios. In this case, 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. If one added the interaction term $Dincome_i * (students/teachers_g)$ to (11), it would enter with a negative coefficient. One can see that if M_i was parental income, and student-teacher ratio was part of $\boldsymbol{Z}_{2g}$, then the interaction coefficient $\boldsymbol{r}_2$ would be affected. Such a mechanism would yield spurious "interaction" effects.

14

Assumption A6 rules out this possibility. As a result, any differential impact of M_i across groups with different values of $\boldsymbol{X}_g$ or $\boldsymbol{Z}_{2g}$ can be interpreted as evidence of true student-group interactions rather than evidence of differential sorting-induced joint distributions of $[\boldsymbol{DX}_i, \boldsymbol{DX}_i^U]$ across groups.

A6 is a statement about the residual component $D\tilde{x}_i^U$ of the unobservable Dx_i^U, and so is not directly testable. How plausible is it? Note first that since (12) is a projection equation, $D\tilde{x}_i^U$ is orthogonal to $\boldsymbol{DX}_i$. Thus the population mean of $D\tilde{x}_i^U DM_i$ is 0 because DM_i is a linear function of $\boldsymbol{DX}_i$. It is possible, however that the relationship between $\boldsymbol{DX}_i^U$ and $\boldsymbol{DX}_i$ and thus DM_i varies across g. If in addition, $E(D\tilde{x}_i^U DM_i | g)$ varies systematically with $\boldsymbol{X}_g$ or $\boldsymbol{Z}_{2g}$, A6 will fail.

The model of school/location choice implies that the distribution of $\boldsymbol{X}_i^U$ and $\boldsymbol{X}_i$ varies with the amenities, which may be related to $\boldsymbol{Z}_{2g}$. While Proposition 1 directly implies that the mean $\boldsymbol{X}_g^U$ varies linearly with $\boldsymbol{X}_g$, this need not imply a violation of A6, since A6 fully allows the means of M_i, $\boldsymbol{X}_i$ and $\boldsymbol{X}_i^U$ to vary across g and even systematically with $\boldsymbol{Z}_{2g}$. A6 can also accommodate variation in the within-group variances of M_i and X_i^U across groups, as long as DM_i and $\boldsymbol{DX}_i^U$ move together in the same way in each group. Then any differential covariance between DM_i and $\boldsymbol{DX}_i^U$ across groups is stripped out by $\boldsymbol{DX}_i \boldsymbol{\Pi}_{\boldsymbol{DX}_i^U \boldsymbol{DX}}$, leaving no correlation between DM_i and the residual index $D\tilde{x}_i^U$.

We have not found sufficient conditions concerning the distribution of A_g and $[\boldsymbol{X}_i, \boldsymbol{X}_i^U, \boldsymbol{Q}_i]$ such that A6 is exactly satisfied. In Online Appendix B4 we describe and present results from monte carlo simulations of the sorting equilibrium for a case in which $[\boldsymbol{X}_i, \boldsymbol{X}_i^U, \boldsymbol{Q}_i]$ is jointly normally distributed and $\boldsymbol{A}_g$ is also jointly normally distributed. People choose from 100 groups, with 250 individuals per group. We show that departures from A6 are truly negligible in this example: the mean correlation between $DM_i D\tilde{\boldsymbol{X}}_i^U$ and any element of $\boldsymbol{X}_g$ is only .008.

A7 strengthens a zero correlation property to statistical independence. To see this, note that the error component $\widetilde{\boldsymbol{Z}}_{2g}^U$ in the projection equation (15) is uncorrelated with $\boldsymbol{X}_g$ and $\boldsymbol{Z}_{2g}$ by definition and $\boldsymbol{DX}_i$ is uncorrelated with $\widetilde{\boldsymbol{Z}}_{2g}^U$ by construction. Strengthening zero correlation to independence rules out the possibility that $\boldsymbol{X}_g$ or $\boldsymbol{Z}_{2g}$ might be predictive of the value of the average interaction $M_g \boldsymbol{Z}_{2g}^U$, even conditional on $M_g \boldsymbol{X}_g$ and $M_g \boldsymbol{Z}_{2g}$. It also rules out the possibility that higher moments of the distribution of $\boldsymbol{DX}_i$ for a given g predict the mean of the residual component $\widetilde{\boldsymbol{Z}}_{2g}^U$. Consider the race indicator variable $BLACK_i$ and the group mean $BLACK_g$. These variables are included in $\boldsymbol{X}_i$ and $\boldsymbol{X}_g$, respectively. A7 would fail if $\widetilde{\boldsymbol{Z}}_{2g}^U$ is not *independent* of $BLACK_g$ and/or $DBLACK_i$ even though it is uncorrelated with both. A7 seems likely to be a reasonable approximation in most contexts. Furthermore, we show in Online Appendix B4 that A7 nearly holds in the sorting process used in our monte carlo simulations, and the minor violations produce no discernable bias in the estimates of the key coefficients.

We are now ready to present the relationship between $\boldsymbol{B}$, $\boldsymbol{r}_1$, and $\boldsymbol{r}_2$ in the estimating equation (11) (or equivalently (10)) and the production function parameters in (9).

Proposition 2:

15

Suppose assumptions A1-A7 hold. Then:

$$B = \beta + \Pi_{DX_i^U DX_i} \beta^U + \Pi_{D\eta_{gi} DX_i} \tag{16}$$

$$r_1 = \rho_1 + \Pi_{X_g^U X_g} \rho_1^U + \Pi_{Z_{2g}^U X_g} \rho_2^U \tag{17}$$

$$r_2 = \rho_2 + \Pi_{Z_{2g}^U Z_{2g}} \rho_2^U. \tag{18}$$

The proof is in Online Appendix B1. Note that the coefficient vector B on X_i is the same as the coefficient vector in the model without interactions considered by AM. It captures a blend of effects of observable and unobservable student characteristics that are common across groups as well as differential treatments received by students that can be predicted based on X_i. It does not capture any true interaction effects nor any effects of group characteristics that are common to all members. The coefficient r_1 consists of the true coefficients ρ_1 on the interaction between M_i and the group averages of the observable student characteristics X_g, plus two other terms. The second term $\Pi_{X_g^U X_g} \rho_1^U$ captures interactions between M_i and the group average of the unobservables X_g^U. (Recall that under Proposition 1, $X_g^U = X_g \Pi_{X_g^U X_g}$). Thus, the first two terms of r_1 capture the fact that the effect of X_i on Y_i depends upon the characteristics of the population in g. The third term $\Pi_{Z_{2g}^U X_g} \rho_2^U$ is present because the unobserved school and area characteristics Z_{2g}^U that interact with M_i may vary with X_g conditional on Z_{2g}. The coefficient vector r_2 is the sum of ρ_2, the effects of the interactions between M_i and Z_{2g}, along with the effects of the interactions between M_i and the portion of Z_{2g}^U that is predictable by Z_{2g} holding X_g constant. Importantly, both r_1 and r_2 exclusively reflect interactions between X_i and group-level characteristics, rather than individual contributions that are common across groups or group-level contributions that are common across individuals. We have assumed away interactions between components of X_i^U and Z_{2g} and Z_{2g}. We speculate that such interactions will be partially picked up by r_2 to the extent that M_i predicts X_i^U.

4.3 Interpretation of Group-Level Coefficients

This subsection presents Proposition 3, which establishes the relationship between the production parameters in (9) and the group-level coefficients G_1, G_2, G_3, and G_4 identified by OLS estimation of the MME model (10). Note that just as B, r_1, and r_2 are identified exclusively from within-group variation, G_1, G_2, G_3, and G_4 are identified exclusively from between-group variation. This fact means that the OLS coefficients G_1, G_2, G_3, and G_4 are numerically identical to the coefficients of the projection of the adjusted group g mean of Y_{gi}, $Y_g - [X_g B + M_g X_g r_1 + M_g Z_{2g} r_2]$, onto X_g, Z_{2g}, $M_g X_g$, and $M_g Z_{2g}$.

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

16

$$z_{2g}^{U} = X_g \Pi_{z_{2g}^{U}, X_g} + Z_{2g} \Pi_{z_{2g}^{U}, Z_{2g}} + M_g X_g \Pi_{z_{2g}^{U}, M_g X_g} + M_g Z_{2g} \Pi_{z_{2g}^{U}, M_g Z_{2g}} + \tilde{z}_{2g}^{U}$$

Let $x_g^{U} \equiv X_g^{U} \beta^{U}$ and note that the projection matrix $\Pi_{x_g^{U} X_g} = \Pi_{X_g^{U} X_g} \beta^{U}$.

Proposition 3:

Suppose assumptions A1-A7 hold. Then:

$$
\begin{aligned}
G_1 &= \left[(\beta - B) + \Pi_{x_g^{U} X_g} \right] + \left[\Gamma_1 + \Pi_{X_g^{U} X_g} \Gamma_1^{U} + \Pi_{z_{2g}^{U} X_g} \right] \\
G_2 &= \Gamma_2 + \Pi_{z_{2g}^{U} Z_{2g}} \\
G_3 &= 0 \\
G_4 &= 0
\end{aligned}
$$

The proof is in Online Appendix B2. Proposition 3 states that under A1-A7 the coefficients G_1 and G_2 on X_g and Z_{2g} in the model with interactions are exactly the same as the corresponding coefficients in the model without interactions considered by AM (See Proposition 2 in AM). G_1 consists of the causal peer effects of X_g and X_g^{U} plus the part of the effect of unobserved group inputs Z_{2g}^{U} predicted by X_g conditional on Z_{2g}. However, it also picks up the bias term $(\beta - B)$ and the term $\Pi_{x_g^{U} X_g}$, which captures differences across school and location in x_g^{U}. Since these last two components represent student contributions to Y_i rather than group-level contributions, we exclude $X_g G_1$ when constructing lower bound estimates for school and location treatment effects. We are unable to assign the variation in $X_g G_1$ exclusively to groups or individuals because X_g also serves as a control function for X_g^{U}. This is the price of preserving meaningful assignment of other components of variance in the absence of random group assignment.

G_2 captures the causal effects of the observed group-level inputs Z_{2g} along with the causal effects of associated variation in unobserved group-level inputs z_g^{U} on the outcome of a student at the population mean of M_i. Thus, the index $Z_{2g} G_2$ 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, G_3 and G_4 are both zero. Consequently, we can drop both $M_g X_g$ and $M_g Z_{2g}$ from the estimating equation (10).

4.4 Interpretation of the Error Components

The final proposition establishes the relationship between the individual-level and group-level error components in (9) and (10).

Proposition 4:

Suppose assumptions A1-A7 hold. Then:

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

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

We relegate the proof to a footnote.[14] Equation (19) reveals that v_g is the sum of group-level effects that are common across students ($\bar{z}_{2g}^U$ and ξ_g) and group-averages of interactions between student inputs and group inputs ($M_g \tilde{Z}_{2g}^U \boldsymbol{\rho}_2^U$). In principle, one could estimate the distribution of $M_g \tilde{Z}_{2g}^U \boldsymbol{\rho}_2^U$ by incorporating random neighborhood-, school-, and commuting zone-specific slopes with respect to M_i into an MME 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, 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.

4.5 Explicitly Incorporating Multiple Group Levels

To this point we have lumped neighborhood-level, school-level, and CZ-level variables together to simplify the notation, but we wish to assess the relative importance of these different categories of external inputs in the empirical work. Thus, we now break up X_g, X_g^U, Z_{2g}, Z_{2g}^U, the interaction terms, and the associated parameters into their n level, s level, and c level subcomponents and rewrite the production function and estimating equation in the expanded notation. Let $X_g^* \equiv [X_n^*, X_s^*, X_c^*]$, $Z_g^* \equiv [Z_n^{N*}, Z_s^{S*}, Z_c^{C*}]$, with corresponding notation for the level-specific subvectors of $\boldsymbol{\Gamma}^*$ and $\boldsymbol{\rho}^*$. Then define the subcomponents of these level-specific subvectors analogously to those defined in Section 3 (e.g. $\boldsymbol{\rho}^N \equiv [\boldsymbol{\rho}_1^{Nr}, \boldsymbol{\rho}_1^{NUr}, \boldsymbol{\rho}_2^{Nr}, \boldsymbol{\rho}_2^{NUr}]'$). Also, replace the scalar M_i with the vector $\boldsymbol{M}_i$. Then

---

[14]Proposition 4 can be proved simply by (1) substituting into the estimating equation (10) the expressions from Propositions 2 and 3 for $\boldsymbol{B}$, $\boldsymbol{r}_1$, $\boldsymbol{r}_1$, $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, $\boldsymbol{G}_3$, and $\boldsymbol{G}_4$, (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 $\bar{z}_{2g}^U$ in (19) we also use the fact that by definition z_g^U, which appears in (9), is equal to $X_g^U \boldsymbol{\Gamma}_1^U + z_{2g}^U$, where $z_{2g}^U = Z_{2g}^U \boldsymbol{\Gamma}_2^U$.

the production function (9) becomes

$$
\begin{aligned}
Y_i \;=\; & X_i\boldsymbol{\beta} + X_n\boldsymbol{\Gamma}_1^N + X_s\boldsymbol{\Gamma}_1^S + X_c\boldsymbol{\Gamma}_1^C + Z_{2n}^N\boldsymbol{\Gamma}_2^N + Z_{2s}^S\boldsymbol{\Gamma}_2^S + Z_{2c}^C\boldsymbol{\Gamma}_{2c}^C \\
& + M_i\otimes X_n\boldsymbol{\rho}_1^N + M_i\otimes X_s\boldsymbol{\rho}_1^S + M_i\otimes X_c\boldsymbol{\rho}_1^C + M_i\otimes Z_{2n}^N\boldsymbol{\rho}_2^N + M_i\otimes Z_{2s}^S\boldsymbol{\rho}_2^S + M_i\otimes Z_{2c}^C\boldsymbol{\rho}_2^C \\
& + X_n^U\boldsymbol{\Gamma}_1^{NU} + X_s^U\boldsymbol{\Gamma}_1^{SU} + X_c^U\boldsymbol{\Gamma}_1^{CU} + Z_{2N}^{NU}\boldsymbol{\Gamma}_2^{NU} + Z_{2s}^{SU}\boldsymbol{\Gamma}_2^{SU} + Z_{2c}^{CU}\boldsymbol{\Gamma}_2^{GU} \\
& M_i\otimes X_n^U\boldsymbol{\rho}_1^{NU} + M_i\otimes X_s^U\boldsymbol{\rho}_1^{SU} + M_i\otimes X_c^U\boldsymbol{\rho}_1^{CU} + M_i\otimes Z_{2n}^{CU}\boldsymbol{\rho}_2^{NU} + M_i\otimes Z_{2s}^{SU}\boldsymbol{\rho}_2^{SU} + M_i\otimes Z_{2c}^{CU}\boldsymbol{\rho}_2^{GU} \\
& + x_i^U + \eta_{nsci} + \xi_{nsci},
\end{aligned}
\tag{21}
$$

where $\otimes$ is the Kronecker product. Note that by decomposing $\boldsymbol{Z}_g^*$ into distinct neighborhood, school, and commuting zone subcomponents, we have implicitly assumed away cross-level interactions of neighborhood, school and commuting zone characteristics, such as average teacher salary with the availability of local four year colleges.[15]

The corresponding MME model (10), becomes

$$
\begin{aligned}
Y_{gi} \;=\; & X_i\boldsymbol{B} + X_n\boldsymbol{G}_1^N + X_s\boldsymbol{G}_1^S + X_c\boldsymbol{G}_1^C + Z_{2n}^N\boldsymbol{G}_2^N + Z_{2s}^S\boldsymbol{G}_2^S + Z_{2c}^C\boldsymbol{G}_2^C \\
& + [M_i\otimes X_n]\boldsymbol{r}_1^N + [M_i\otimes X_s]\boldsymbol{r}_1^S + [M_i\otimes X_c]\boldsymbol{r}_1^C \\
& + [M_i\otimes Z_{2n}^N]\boldsymbol{r}_2^N + [M_i\otimes Z_{2s}^S]\boldsymbol{r}_2^S + [M_i\otimes Z_{2c}^C]\boldsymbol{r}_2^C \\
& + (v_n - v_s) + (v_s - v_c) + v_c + (v_i - v_n),
\end{aligned}
\tag{22}
$$

where we have imposed the fact that Proposition 3 implies that $\boldsymbol{G}_3$ and $\boldsymbol{G}_4$ are zero.

4.5.1 The Challenge of Separately Identifying Unobserved Neighborhood, School and Commuting Zone Influences

The expanded notation facilitates discussion of an important limitation on our ability to fully identify the relative contributions of neighborhood, school, and commuting zone inputs. The problem stems from the unobservables $\boldsymbol{Z}_{2g}^U$. While the school component $\boldsymbol{Z}_{2s}^S\,\boldsymbol{G}_2^S$ will not include any peer effects or individual-level contributions, Proposition 3 implies that it may include unobserved neighborhood inputs if they cluster at the school (but not CZ) level. To see this, note first that $\boldsymbol{G}_2 = \boldsymbol{\Gamma}_2 + \boldsymbol{\Pi}_{z_{2g}^U\boldsymbol{z}_{2g}}$ from Proposition 3. Next, note that $z_{2g}^U \equiv Z_{2n}^{NU}\boldsymbol{\Gamma}_2^{NU} + Z_{2s}^{SU}\boldsymbol{\Gamma}_2^{SU} + Z_{2c}^{CU}\boldsymbol{\Gamma}_2^{CU}$, and that $\boldsymbol{Z}_{2s}^S$ is the school subvector of $\boldsymbol{Z}_{2g}$. If $Z_{2n}^{NU}\boldsymbol{\Gamma}_2^{NU}$ is clustered at the school level and predictable by $\boldsymbol{Z}_{2s}^S$, this will be reflected in the columns of $\boldsymbol{\Pi}_{z_{2g}^U\boldsymbol{z}_{2g}}$ corresponding to $\boldsymbol{Z}_{2s}^S$. As a result, Proposition 3 says that the school component $\boldsymbol{G}_2^S$ of $\boldsymbol{G}_2$ will pick up part of $Z_{2N}^{NU}\boldsymbol{\Gamma}_2^{NU}$. Similarly, if $Z_{2n}^{NU}\boldsymbol{\Gamma}_2^{NU}$ or $Z_{2s}^{SU}\boldsymbol{\Gamma}_2^{SU}$ is clustered at the commuting zone level, it may be reflected in the CZ component $\boldsymbol{G}_2^C$. For example, if the commuting zones that feature many colleges have schools with better principals on average, and this advantage is not predictable based on the observed school characteristics $\boldsymbol{X}_s$ and

---

[15]One could define a school effect for a particular type of commuting zone, or an effect of the marginal distribution of commuting zone characteristics. We lack the statistical power to identify such effects, especially given the large number of possibilities and the lack of guidance from prior literature, so we do not pursue this here.

19

Z_{2s}^S, then it will be reflected in G_2^C if Z_{2c} includes a variable capturing the number of colleges in the CZ.

Furthermore, $v_s - v_c$ and v_c will include the part of $Z_{2n}^{NU}\Gamma_2^{NU}$ that varies across schools or across CZs (respectively) that is not predictable by X_g and Z_{2g}. Analogous problems occur with unobserved school inputs that cluster at the CZ level. For example, if the commuting zone containing Boston has schools with better average principal quality than would be predicted by X_g and Z_{2g}, then v_c will be higher for the Boston CZ.

Similar issues also affect interpretation of the interaction coefficients r_2^S and r_2^C. Proposition 3 states that $r_2 = \rho_2 + \Pi_{Z_{2g}^U Z_{2g}} \rho_2^U$, where $\Pi_{Z_{2g}^U Z_{2g}}$ is defined in (15). The CZ subvector r_2^C of r_2 will capture not just interactions involving Z_{2c}^{CU}, but also part of interactions involving Z_{2n}^{NU} and Z_{2s}^{SU} to the extent that Z_{2c} predicts Z_{2n}^{NU} and Z_{2s}^{SU} conditional on X_g, Z_{2n}^N, and Z_{2s}^S.

What about CZ-level unobservables represented by $Z_{2c}^{CU}\Gamma_{2c}^{CU}$? If the full set of commuting zone averages of Z_{2n}^N and Z_{2s}^S were included in Z_{2c}^C, then we could rule out the possibility that any unobserved commuting zone inputs could project onto either Z_{2n}^S or Z_{2s}^S and be captured by G_2^N or G_2^S. This is because neighborhoods and schools are nested within commuting zones. More generally, with nested levels, one can prevent unobservable inputs from being captured by regressors (including interactions) associated with lower levels of aggregation by including the appropriate higher level averages of lower-level characteristics. Thus, with nested levels one could in principle place a lower bound on the relative importance of inputs at the lowest level (neighborhoods in our empirical work) and an upper bound on the relative importance of inputs at the highest level (CZs). We seek to approximate this approach, though we cannot fully implement it due to concerns about overfitting given limited numbers of units at each level.

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).[16] 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.[17] 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. In ELS, we observe the student's census block and ZIP code of residence in grade 10. 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.

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 in Appendix A1. Section 5.3 describes the process by which we chose which variables to include in X_i, X_n, X_s, X_c, Z_{2s}^S, and 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 the control function.

Coding of most of the variables is straightforward, but in some cases variables are the first principal 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 surveys is that the number of students sampled per school is only about 24 for NELS and 20 for ELS.[18] Simulation results in AM indicate that samples of this size may reduce to a small 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. In Online Appendix B4 we present additional simulations featuring estimates of G_2 and r_2 from specifications in which X_g is formed using only 10, 20, or 40 individuals per group. We show that the estimates are close to the corresponding true values unless the correlation between the amenity factors and the elements of Z_{2g} is quite high. In this case the use of small samples causes slight underestimates of the magnitude of the interaction coefficients r_2 and overestimates of the magnitude of the main effects G_2.

We restrict our samples to those individuals whose school administrator filled out a school sur-

_____

[18] School averages are computing using all students in the base-year surveys of NELS and ELS, respectively, not just those in the estimation samples.

vey, 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.[19]

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 four-year degree at around 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 and four year enrollment rates in the previous follow-up surveys are somewhat higher among respondents who are also observed in the final followups.[20] Summary statistics for the outcome measures are in Table 1-1.

We do not use panel weights because we experienced convergence problems when we attempted to estimate the variance components of the mixed level models using weights. 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. The slope parameter estimates are not sensitive to weighting, but we do not know whether weighting would affect the estimates of the error component variances.

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

AM discuss the principles governing variable selection. 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 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

---

[19]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.

[20]This partly contributes to the higher mean of COLLBA than ENROLL in NELS: the mean of ENROLL among sample members of the 3rd follow-up survey in NELS is 0.353 vs. 0.327 among those in the 2nd follow-up (when the outcome is determined), suggesting some differential attribution is taking place. However, transfers from two-year colleges and other delayed entrants to four year college could also account for a higher graduation than enrollment rate, given the way these outcomes are defined.

22

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, $\boldsymbol{X}_g$ should contain aggregates of $\boldsymbol{X}_i$ at the (n, s) school/neighborhood level. In practice, due to sample size limitations we include $\boldsymbol{X}_s$ as well as $\boldsymbol{X}_n$ variables, but not averages over (n, s) pairs. The $\boldsymbol{X}_n$ variables are census-based measures of the demographic makeup of the neighborhood.[21]

$\boldsymbol{Z}_{2n}^N$, $\boldsymbol{Z}_{2s}^S$, and $\boldsymbol{Z}_{2c}^C$ should include observed neighborhood, school and commuting zone characteristics that could plausibly influence the socioeconomic outcome of interest, including school and location policies that may be partially affected by student composition. $\boldsymbol{Z}_{2n}^N$ and $\boldsymbol{Z}_{2s}^S$ should exclude variables that are simple aggregates of parent/student traits that might also affect willingness to pay for neighborhood characteristics and thus lead to sorting. These are $\boldsymbol{X}_n$ or $\boldsymbol{X}_s$ 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 $\boldsymbol{Z}_{2n}^N$ 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. If the composition variables included in $\boldsymbol{X}_n$ and $\boldsymbol{X}_s$ are sufficiently rich to plausibly satisfy the spanning assumption A5, then the coefficients on commuting-zone demographic averages that would otherwise belong in $\boldsymbol{X}_c$ should not capture any student sorting on unobservables under the assumptions of the model. In this case such averages may be considered part of $\boldsymbol{Z}_{2c}^C$, since they are likely to only absorb the impacts of unobserved causal commuting zone characteristics.

As AM discuss, group-level variables such as a school's frequency of fights or average 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 grounds that they are determined by other observed and unobserved variables in the model. We include them in the control function $\boldsymbol{X}_g$ 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 A1 lists the final choices of individual-level and school-level explanatory measures used in each dataset.[22] Online Appendix table B2 provides the mean, standard deviation, and percent

---

[21] As AM discuss, in principle the control function variables can be augmented with aggregates of outcome-irrelevant characteristics $\boldsymbol{O}_i$ or even directly observed amenities in $\boldsymbol{A}_g$, since its purpose is to span the space of amenities that drive sorting on $\boldsymbol{X}_i^U$.

[22] 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; Yuan and Lin 2006). Ultimately, we chose not to use these procedures because of uncertainty about how to do statistical inference, about the computational feasibility, about statistical properties of a bootstrap procedure that accounts for both variable selection and sampling error given variable choice, and about how to correct for finite sample bias in the variance estimates. The three step estimation method we use is computationally demanding.

of observations imputed for each individual-level and school-level characteristics for each of our data sets.

Table 1-2 provides the number of neighborhoods, schools, and commuting zones for each combination of dataset and outcome. 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 Z_{2s}^S and Z_{2c}^C variables, along with large numbers of variables in the control functions X_n and X_s.

6   Estimation Methods

6.1   Restricting the Estimating Equation

We estimate the following restricted version of the MME estimating equation (22):

$$Y_{gi} = X_i B + X_n G_1^N + X_s G_1^S + Z_{2s}^S G_2^S + Z_{2c}^C G_2^C$$
$$+ (X_i B)(X_n G_1^N) r_1^N + (X_i B)(X_s G_1^S) r_1^S + M_i \otimes (Z_{2s}^S G_2^S) r_2^S + M_i \otimes (Z_{2c}^C G_2^C) r_2^C$$
$$+ (v_i - v_n) + (v_n - v_s) + (v_s - v_c) + v_c, \tag{23}$$

where Y_{gi} is a latent index for the binary education outcomes. Several restrictions merit mention. First, we have excluded $Z_{2n}^N G_2^N$ and $(M_i \otimes Z_{2n}^N) r_2^N$ because our only neighborhood-level variables are demographic means that belong in X_n. Second, we have also excluded $X_c G_1^C$ because average student characteristics at the neighborhood and school levels should fully control for sorting bias. This is because CZ is implicitly a part of the choice of n and s. Thus, to prevent overly conservative treatment effect estimates, CZ composition variables are placed in Z_{2c}^C. We also exclude the interaction term $(M_i \otimes X_c) r_1^C$.

Third, to retain sufficient precision, 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. We set M_i to consist of the index $X_i B$ and indicators for whether the student is female, a member of an underrepresented minority group (denoted 1(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 (1(LOWINC)). Thus, using the definition of the Kronecker product $\otimes$,

$$M_i \otimes Z_{2s}^S G_2^S = [(X_i B) Z_{2s}^S G_2^S, \ 1(Female) Z_{2s}^S G_2^S, \ 1(URM) Z_{2s}^S G_2^S, \ 1(Low\_Income) Z_{2s}^S G_2^S],$$

and $M_i \otimes (Z_{2c}^C G_2^C)$ has the same form. Including $X_i B$ in M_i imposes the restriction that for white males from high income families, the strength of interactions depends upon components of X_i in proportion to their direct effects on the outcome. Including the three additional indicator variables in M_i allows a somewhat more general pattern of interactions for particular subpopulations for whom

24

differential interactions have often been posited in the literature.

In addition, to keep the main specification parsimonious, we only allow student characteristics to interact with the indices $X_n G_1^N$, $X_s G_s^S$, $Z_{2s}^S G_2^S$, and $Z_{2c}^C G_2^C$. 1(Female), 1(URM), and 1(LOWINC) are further restricted to interact only with $Z_{2s}^S G_2^S$ and $Z_{2c}^C G_2^C$ and not with $X_n G_1^N$ or $X_s G_s^S$. We experimented with additional interactions, but did not find sufficiently compelling evidence for them given the loss of precision they required.[23]

Finally, as discussed in section 4.5.1, 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 CZ-level effects, we also include school and commuting zone averages of the index $X_n \hat{G}_1^N$ as well as the commuting zone averages of the indices $X_s \hat{G}_1^S$ and $Z_{2s}^S \hat{G}_2^S$ with separate coefficients. The school and commuting zone averages of $X_n \hat{G}_1^N$ and the commuting zone average of $X_s \hat{G}_1^S$ are treated as control function variables and are excluded from the estimates of school and commuting zone treatment effects described below. The commuting zone average of the index $Z_{2s}^S \hat{G}_2^S$ is treated as a commuting zone characteristic that is contained in Z_{2c}^C in (23). 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-A7, and is identified purely from between-commuting zone variation, it is likely to capture unobserved commuting zone inputs.

6.2   Estimating the Model Parameters

For estimation of (23), we assume that the random effects are all normally distributed. Our analysis of the importance of neighborhood, school and commuting zone depends heavily on estimating the student-weighted standard deviations and covariances of the indices $X_n G_1^N$, $X_s G_2^S$, $Z_{2s}^S G_2^S$, and $Z_{2c}^C G_2^C$ as well as the standard deviations of v_c, $v_s - v_c$, $v_n - v_s$, and $v_i - v_n$. In principle, one could estimate (23) in one step by maximum likelihood with index restrictions imposed. Due to the computational burden, we estimate the slope parameters B, G_1^N, G_1^S, G_2^S, G_2^C, r_1^N, r_1^S, r_2^S, and r_2^C and obtain initial estimates of $sd(X_i B)$, $sd(X_n G_1^N)$, $sd(X_s G_s^S)$, $sd(Z_{2s}^S G_2^S)$, $sd(Z_{2c}^C G_2^C)$ and $sd(v_c)$, $sd(v_s - v_c)$, $sd(v_n - v_s)$, and $sd(v_i - v_n)$ in two steps. First, we estimate the slope parameters with nonlinear restrictions imposed, but treat the error term as a composite. We use nonlinear least squares for log wages and MLE probit (given the normality assumption) for the binary education outcomes. In both cases we ignore the error correlation structure. In the second step we estimate $sd(v_c)$, $sd(v_s - v_c)$, $sd(v_n - v_s)$, and $sd(v_i - v_n)$ and update r_1^N, r_1^S, r_2^S, and r_2^C. We also implicitly allow the elements of $\hat{B}$ to update by a common factor of proportionately, and we do the same for $\hat{G}_1^N$, $\hat{G}_1^S$, $\hat{G}_2^S$, $\hat{G}_2^C$ (in practice, this made little difference). We experienced computational difficulties when using maximum likelihood in the second step. Instead, we adopt a Bayesian approach, which we treat as a

---

[23]We considered specifications in which 1) all of the variables in M_i interact with $X_n G_1^N$ and $X_s G_s^S$, 2) 1($Hispanic$) and 1($Non-Hispanic Black$) entered with separate interactions with the indices of external inputs, and 3) interactions between 1($Female$) $*$ 1($Black$) and $Z_{2s}^S G_2^S$ and $Z_{2c}^C G_2^C$ were introduced to allow race-by-gender specific sensitivity to external inputs. Unfortunately, such interactions were so noisy that even large effects were statistically insignificant.

computational device that we use to compute MLE estimates.[24] Appendix B3.1 provides a more detailed description of the two-step procedure.

The third step of the estimation procedure is to adjust the updated estimates of $sd(\boldsymbol{X}_i\boldsymbol{B})$, $sd(\boldsymbol{X}_n\boldsymbol{G}_1^N)$, $sd(\boldsymbol{X}_s\boldsymbol{G}_s^S)$, $sd(\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S)$, $sd(\boldsymbol{Z}_{2c}^C\boldsymbol{G}_2^C)$ and the estimates of $sd(v_c)$, $sd(v_s-v_c)$, $sd(v_n-v_s)$, and $sd(v_i-v_n)$ from step 2 for finite sample bias. This would be necessary even if we had estimated the model via MLE in one step. Consider $sd(\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S)$. The initial estimate corresponds to $sd(\boldsymbol{Z}_{2s}^S\hat{\boldsymbol{G}}_2^S)$, the standard deviation of the index based on the estimate $\hat{\boldsymbol{G}}_2^S$. It will be biased upward by sampling error in $\hat{\boldsymbol{G}}_2^S$. The same problem affects $sd(\boldsymbol{X}_n\hat{\boldsymbol{G}}_1^N)$, $sd(\boldsymbol{X}_s\hat{\boldsymbol{G}}_s^S)$, and $sd(\boldsymbol{Z}_{2c}^C\hat{\boldsymbol{G}}_2^C)$. The bias arises for the same reason R^2 overstates the explanatory power of a linear regression in finite samples. Because the hierarchical nature of the model and the control function strategy we adopt requires the use of a large number of variables at four different levels, finite sample bias from sampling error is not negligible given our sample sizes. Keeping in mind that all variables are mean zero, we separate the signal and noise subcomponents of the initial estimate $sd(\boldsymbol{Z}_{2s}^S\hat{\boldsymbol{G}}_{2s}^S)$ using the equation

$$[sd(\boldsymbol{Z}_{2s}\hat{\boldsymbol{G}}_2^S)]^2 = \frac{1}{N}\sum_i(\boldsymbol{Z}_{2s(i)}^S\boldsymbol{G}_2^S\boldsymbol{G}_2^{S\prime}\boldsymbol{Z}_{2s(i)}^{S\prime}) + \frac{1}{N}\sum_i\boldsymbol{Z}_{2s(i)}^S[\hat{\boldsymbol{G}}_2^S-\boldsymbol{G}_2^S][\hat{\boldsymbol{G}}_2^S-\boldsymbol{G}_2^S]'\boldsymbol{Z}_{2s(i)}^S, \quad (24)$$

where N is the sample size. The expectation of the second (sampling variance) term, conditional on $\boldsymbol{Z}_{2s(i)}^S$, is

$$\frac{1}{N}\sum_i\boldsymbol{Z}_{2s(i)}^S var(\hat{\boldsymbol{G}}_2^S)\boldsymbol{Z}_{2s(i)}^S \quad (25)$$

Using (25) the bias-adjusted estimator of $sd(\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S)$ is

$$\widehat{sd(\boldsymbol{Z}_{2s}\boldsymbol{G}_2^S)} = [sd(\boldsymbol{Z}_{2s}\hat{\boldsymbol{G}}_2^S)^{.5} - [\frac{1}{N}\sum_i\boldsymbol{Z}_{2s(i)}^S\widehat{Var(\hat{\boldsymbol{G}}_2^S)}\boldsymbol{Z}_{2s(i)}^S]^{.5}.$$

However, we do not adjust the r coefficients on the interaction terms (where we conjecture that the sampling error in the indices might bias estimates of interaction effects toward zero).[25]

Since the sampling error in the regression indices reflects true contributions of the error components to Y_i, the flip side of upward bias in $Var(\boldsymbol{X}_n\hat{\boldsymbol{G}}_1^N)$, $Var(\boldsymbol{X}_s\hat{\boldsymbol{G}}_s^S)$, $Var(\boldsymbol{Z}_{2s}^S\hat{\boldsymbol{G}}_2^S)$, and $Var(\boldsymbol{Z}_{2c}^C\hat{\boldsymbol{G}}_2^C)$ is downward bias in the estimates of $Var(v_c)$, $Var(v_s-v_c)$, $Var(v_n-v_s)$, and $Var(v_i-v_n)$, which do not account for degrees of freedom used in estimation.[26] We discuss how we allocate the bias ad-

---

[24]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 (23) 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 $Var(v_c)$, $Var(v_s-v_c)$, $Var(v_n-v_s)$.

[25]We estimate $\widehat{Var(\hat{\boldsymbol{G}}_2^S)}$ using the formula for the asymptotic variance of the Step 1 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 index standard deviations.

[26]The same issue arises in linear regression with one normally distributed error component. In this case, the MLE estimator of the error variance is the sum of squared residuals divided by the sample size. The unbiased estimator adjusts for degrees of freedom by dividing by the sample size minus 1 plus the number of regressors. This is a restricted maximum likelihood estimator (REML). Unfortunately, the REML estimator does not exist for multilevel probit models. AM use

26

justments across these error components to remove this downward bias in Online Appendix B3.2.[27]

In the case of the log wage outcome, v_i contains the age 25 or 26 error component and thus includes both permanent and transitory variation. We lack the wage panel data necessary to isolate the permanent component. Consequently, we estimate the variance of the permanent component of v_i under the assumption that the transitory component of the wage makes up the same 52.3% 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 (sampled in NLS72) examined in Altonji and Mansfield (2011).

The standard errors in the paper are based on a bootstrap approach encompassing the entire estimation procedure, including the construction of X_s, imputation of missing data, and the bias corrections to the standard deviations of the regression indices. In particular, they account for the fact that the coefficients that define the index variables used in the second step are estimated.[28] Across outcome variables, data sets, and specifications, we find that the means of the bootstrap replications of $\widehat{sd}(X_n G_1^N)$, $\widehat{sd}(X_s G_s^S)$, $\widehat{sd}(Z_{2s}^S G_2^S)$, $\widehat{sd}(Z_{2c}^C G_2^C)$ are above the point estimates. Not surprisingly, the disparity is even greater for the variance component estimates based on them, which underlie the variance decompositions in Tables 2 and 3. 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 less sensitive to the issue. More work on the best way to implement bias corrections and perform related 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.

---

the ML estimator for the two level random effects probit and then correct for bias based on the formulas relating the unbiased estimators and the ML estimators in the case of a continuous outcome. We explored the possibility of extending AM's approach to the four level case, but were not able to find (or derive) the necessary formulas for four-level ML and REML variance component estimators.

[27]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.

[28]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 200 replications to form the bootstrap estimates. Furthermore, we use fewer (8) MCMC chains of length 500 rather than the 30 chains of length 1000 that we use to compute the point estimates after first verifying that estimates are not very sensitive to this choice. This might lead to an overstatement of the standard errors.

To simplify the empirical analysis we define neighborhoods to be nested within schools, which are themselves nested within commuting zones.[29] We use a version of our estimating equation with interaction terms excluded.[30]

$$Y_i = X_i B + X_n G_1^N + X_s G_1^S + Z_{2s}^S G_2^S + Z_{2c}^C G_2^C + (v_i - v_n) + (v_n - v_s) + (v_s - v_c) + v_c. \qquad (26)$$

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

$$Y_i = (Y_i - Y_n) + (Y_n - Y_s) + (Y_s - Y_c) + Y_c. \qquad (27)$$

Because the components in (27) are mutually orthogonal, $Var(Y_i)$ is:

$$Var(Y_i) = Var(Y_i - Y_n) + Var(Y_n - Y_s) + Var(Y_s - Y_c) + Var(Y_c) \qquad (28)$$

We are not able to fully characterize sorting on both observed and unobserved student level characteristics because group averages of X_i absorb not only the unobserved group averages of X_i^U but also part of the group treatment effect. Instead, we use an analogous four-component decomposition to analyze the structure of sorting on the index of observable characteristics $X_i B$ that best predicts the outcome:

$$Var(X_i B) = Var((X_i - X_n)B) + Var((X_n - X_s)B) + Var((X_s - X_c)B) + Var(X_c B) \qquad (29)$$

To our knowledge, we are the first to perform such a decomposition based on commuting zone, school and census block group combinations.

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

Decompositions of variance provide an overall sense of the relative importance of individual, neighborhood, school, and commuting zone inputs in determining later educational outcomes. However, 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, we extend to models with student-specific slopes AM's methodology for converting variance components into expected impacts on outcomes of particular shifts in school and/or community input quality. Here we describe the three distinct treatments we consider.

---

[29] 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 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.

[30] 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, preventing a simple decomposition

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 (8), the true distribution of combined school/commuting zone quality may be defined as the distribution of $\mathbf{Z}_s^{S*}\mathbf{\Gamma}^{S*} + \mathbf{Z}_c^{C*}\mathbf{\Gamma}^{C*}$. We approximate this combined school/commuting zone quality distribution using the distribution of $T = \mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S} + \mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C} + (v_s - v_c) + v_c$, where we use T to denote the particular "treatment" chosen. Because we exclude the control function $\mathbf{X}_n\mathbf{G}_1^{N} + \mathbf{X}_s\mathbf{G}_1^{S}$, which may capture peer effects and other unobserved school and commuting zone inputs 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 $\mathbf{Z}_s^{S*}\mathbf{\Gamma}^{S*} + \mathbf{Z}_c^{C*}\mathbf{\Gamma}^{C*}$. On the other hand, as the discussion in subsection 4.5.1 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.[31]

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 $T \equiv \mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S} + \mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C} + (v_s - v_c) + v_c$ 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 $\Phi^{-1}(*)$ is the inverse CDF of the standard normal distribution. Next, note that the interaction terms $\mathbf{M}_i \otimes (\mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S})r_2^{S}$ and $\mathbf{M}_i \otimes (\mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C})r_2^{C}$ depend separately on the sub-components $\mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S}$ and $\mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C}$ of the full treatment T. We handle this by integrating over the joint conditional distribution $f(\mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S}, \mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C}|T = T^q)$. We do this by taking P draws of the vector $[\mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S}, \mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C}]$ 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 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 $[\mathbf{X}_i\hat{\mathbf{B}}, \mathbf{X}_n\hat{\mathbf{G}}_1^{N}, \mathbf{X}_s\hat{\mathbf{G}}_1^{S}]$ 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:

$$E[\hat{Y}^q] \equiv \frac{1}{P}\sum_p \frac{1}{I}\sum_i \Phi([\mathbf{X}_i\hat{\mathbf{B}} + \mathbf{X}_n\hat{\mathbf{G}}_1^{N} + \mathbf{X}_s\hat{\mathbf{G}}_1^{S} + T^q$$
$$+ (\mathbf{X}_i\hat{\mathbf{B}})(\mathbf{X}_n\hat{\mathbf{G}}_1^{N})\hat{r}_1^{N} + (\mathbf{X}_i\hat{\mathbf{B}})(\mathbf{X}_s\hat{\mathbf{G}}_1^{S})\hat{r}_1^{S} + \mathbf{M}_i \otimes (\mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S})_p\hat{r}_2^{S}$$
$$+ \mathbf{M}_i \otimes (\mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C})_p\hat{r}_2^{C}]/[1 + Var(v_n - v_s)]) \qquad (30)$$

where $(\mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S})_p$ and $(\mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C})_p$ represent the p-th draws of these regression indices from the conditional joint distribution $f(\mathbf{Z}_{2s}^{S}\mathbf{G}_2^{S}, \mathbf{Z}_{2c}^{C}\mathbf{G}_2^{C}|T = T^q)$. Note that $Var(v_i - v_n)$ has been normalized to 1

---

[31] Even if such clustering is significant, one could interpret our estimates instead as lower bound estimates of the impact of a shift in the combined neighborhood/school/commuting zone quality index $\mathbf{Z}_n^{N*}\mathbf{\Gamma}^{N*} + \mathbf{Z}_s^{S*}\mathbf{\Gamma}^{S*} + \mathbf{Z}_c^{C*}\mathbf{\Gamma}^{C*}$.

in the denominator, because 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 each of them to a 90th percentile school/commuting zone, but held the distribution of peer effects fixed. We refer to this counterfactual as the "School and CZ" counterfactual in our tables and discussion.

6.4.2 The School Treatment and the Commuting Zone Treatment

The school counterfactual thought experiment or "treatment" consists of 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 $\boldsymbol{Z}_s^S \boldsymbol{\Gamma}^{S*}$), holding neighborhood and commuting zone inputs fixed. We approximate the distribution of true school quality with the distribution of $T \equiv \boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S + (v_s - v_c)$. This variable excludes the peer effects components $\boldsymbol{X}_s \boldsymbol{\Gamma}_1^S$ and $\boldsymbol{X}_s^U \boldsymbol{\Gamma}_1^{SU}$ and also part of $\boldsymbol{Z}_{2s}^{SU} \boldsymbol{\Gamma}_{2s}^{SU}$ and so is likely to understate school effects. But as we discussed in 4.5.1, it may capture unobserved neighborhood inputs that are clustered in particular schools.

The commuting zone 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 $\boldsymbol{Z}_c^C \boldsymbol{\Gamma}^{C*}$, holding neighborhood and school inputs fixed. Importantly, this treatment does not include the subcomponent of school inputs that varies across commuting zones (given by the commuting zone average of $\boldsymbol{Z}_s^{S*} \boldsymbol{\Gamma}^{S*}$). 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 neighborhood and school inputs across commuting zones). We approximate the distribution of true commuting zone quality $\boldsymbol{Z}_c^C \boldsymbol{\Gamma}^{C*}$ with the distribution of $T \equiv \boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C + v_c$ (and we do not include the between-CZ component of $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$, since $\boldsymbol{G}_2^S$ is identified using between-school/within-CZ variation and is designed to capture school effects).[32] This is again an approximation due to the possibility that $\boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C$ and v_c could partly reflect unobserved school or neighborhood inputs. We refer to this counterfactual as the "CZ only" counterfactual in our tables and discussion.

Computation of the school and CZ treatments are similar to the combined school and CZ treatment. The formulas are in Online Appendix B3.3.

---

[32]Recall that the commuting zone average of $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$ enters our estimating equation as a separate commuting zone variable with its own coefficient. Because the coefficient on this index is identified purely from between-commuting zone variation, we include this index as part of our commuting zone treatment.

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

In contrast to AM, 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.[33]

The most straightforward approach is simply to restrict the sample used for the counterfactual treatments to members of a particular subpopulation. We report results for the following subpopulations: Hispanic students, non-Hispanic black students, white students in a single-mother household where the mother has a high-school education or less, and white students in a two-parent-college-educated household. We use the empirical distribution of individual and neighborhood inputs $X_i \hat{B} + X_n \hat{G}_1^N + X_s \hat{G}_1^S$, so restricting the sample naturally imposes the chosen sample's joint distribution of observed individual and neighborhood inputs. All elements of 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 the subgroup-specific distribution of the other elements of M_i, such as low income status, weighted by the corresponding elements of the interaction coefficients $\hat{r}_2^S$ and $\hat{r}_2^C$.

Our rich set of observed individual characteristics also allow us to investigate the degree to which school and CZ treatment effects vary with the individual's own contribution to the outcome, as summarized by $X_i B$. To do this, we fix $X_i B$ at each ventile dividing point $[.05, \ldots, .95]$ in its empirical distribution in the sample, and compute the change in expected outcome for each of our three counterfactual quality shifts ("School and CZ", "School only", and "CZ only", described above) for randomly chosen individuals at the chosen ventile of $X_i B$. We do not hold 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 (T^{10}) and high (T^{90}) treatment values.

See Online Appendix B3.3.2.

As we discuss in detail in Section 8.1.1, part of the differential sensitivity for binary outcomes arises from the nonlinearity of the probit function. Even if each of the interaction parameters $\{r_1^N, r_1^S, r_2^S, r_2^C\}$ were equal to zero, we should expect heterogeneity in outcome sensitivity. This is because some subpopulations have more students near the decision margin (as evidenced by different mean outcome values).

7 Results

We begin by presenting in Section 7.1 the decompositions of variance described in Section 6.3. Section 7.2 then presents estimates of the interaction coefficients and the standard deviations of the regression indices and random effects from the interactive specification (23). We focus on the

---

[33]The only non-linearity in AM comes from the probit function for binary outcomes.

estimates for the "full" set of X_i variables. Model estimates and variance decompositions for the "basic" set are in the online appendix. As expected, they usually imply a more important role for school and commuting zone factors in education and wages. We present the treatment effect estimates based on them in Section 8.

7.1 Decompositions of Variance

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

Table 2 decomposes the variance of log wages and the latent probit indices Y_i that determine the three education outcomes into components at the individual, neighborhood, school, and commuting zone levels. These are based on results for the linear version of (26) in which the interaction parameters $\{r_1^N, r_1^S, r_2^S, r_2^C\}$ have been restricted to zero. We report 5th and 95th quantile confidence interval estimates in brackets in the text in a few instances. The NELS, ELS block group and ELS ZIP code results are in the columns labeled NELS, ELSbg and ELSz. We sometimes report pooled estimates that place half of the weight on the NELS estimate and half on the ELS ZIP code and block group estimates. These are not reported in Table 2.

The main finding across data sets and outcomes is that individual-level factors account for the lion's share of the variance, with school and commuting zone level variation making modest contributions, and neighborhood level factors playing a smaller role. In the case of HSGRAD, the individual-level component accounts for 81.9% [77.5, 85.0] of the outcome variance for NELS, 81.6% [75.2, 85.0] for the ELS block group specification, and 85.6% [77.2, 87.1] for the ELS ZIP code specification. The percentage of variance at the individual level is similar for wages and slightly lower for college graduation. Interestingly, individual level factors are somewhat lower for enrollment in a four-year college, averaging 72.3% across NELS and the two ELS specifications. We do not have a good explanation for this.[34]

The school level accounts for a substantial fraction of the variance in education outcomes. The average of the NELS and the ELS estimates is 9.7% for high school graduation, 15.9% for 4 year enrollment, and 10.8% for COLLBA. For wages, the estimates are about 9.0% for ELS but only 3.1% for NELS.

Variation at the commuting zone level accounts for between 3.9% and 6.6 of the latent variable for high school graduation, about 10% for college enrollment, between 6.2% and 7.9% for college graduation and between 6.2% and 7.6% for wages.

We were surprised to find that differences across neighborhoods among students attending the same schools do not account for much variation, even at the census block level. In the ZIP code case, the percentage varies between 0.6% and 1.3% across education outcomes. For wages the estimates

---

[34]It bears repeating that we are not using sampling weights due to computational difficulties in the second step and so are not accounting for differences in designs of the ELS and NELS samples. This could contribute to differences in the estimates across ELS and NELS, although we noted earlier that the first step estimates of parameters involving the observables are not sensitive to weighting.

are 0.7% for NELS and 1.1% for ELS. The ELS block group estimates are larger: 2.2% for HS-GRAD, 3.1% for ENROLL, and 2.4% for COLLBA and 3.2% for wages. While the relatively small neighborhood component is surprising, this does not necessarily mean that neighborhood effects and neighborhood-level sorting are unimportant. Alternatively, the high and low quality neighborhoods might be sufficiently grouped within school attendance zones that most of the variation in neighborhood inputs is across schools. As discussed in sections 4.5.1 and 6.3, the 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. This may be reinforced by the fact that the ELS and NELS samples are clustered at the school level rather than at the neighborhood level. In addition, ZIP code may be too coarse a measure of neighborhood, with many ZIP codes containing a mix of both good and bad neighborhoods (so that the bulk of neighborhood variation is assigned to the "within-neighborhood" category in Table 2). But the estimates are still relatively small when we use block group. Clustering is clearly much stronger across schools.

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

The decompositions in Table 2 convey the degree to which outcomes are clustered at each level of aggregation, but they combine group-level differences that are attributable to student sorting with true group-level inputs. Table 3 examines 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. Note that the individual observables X_i have a fair amount of predictive power for Y_i, especially for postsecondary education (bottom row). Across NELS and the two ELS specifications, the average of the estimates of the share of outcome variance attributable to $X_i B$ is 23.9% for HSGRAD and considerably larger for ENROLL (38.6%) and COLLBA (35.6%). For wages, the share is 13.9% for NELS and about 23.3% for ELS.

The row labeled "Individual Share of $Var(XB)$" reports the fraction of $var(X_i B)$ that is within neighborhood. The estimates average 70.5% for both HSGRAD and for the wages, but are somewhat smaller for ENROLL (65.0%) and COLLBA (63.1%). The individual shares are about 3% lower when the neighborhood definition is block group rather than ZIP code.

We also find substantial sorting across schools both within and across commuting zones. The school level share estimates average 17.2% for HSGRAD, 20.8% for ENROLL, 22.3% for COLLBA and 17.6% for wages. The corresponding commuting zone shares are 10.8% for HSGRAD, 12.6% for ENROLL, 13.2% for COLLBA and 10.2% for wages. The substantial amount of sorting on $X_i B$ underscores the need for econometric techniques to distinguish the contribution of school and commuting zone inputs from sorting. A large 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.

Consistent with the results in the previous section, there is not much evidence of outcome-relevant sorting across ZIP codes within school attendance zones. In the case of HSGRAD, the neighborhood-level shares of $Var(\boldsymbol{X}_i\boldsymbol{B})$ are 1.5% for NELS and 0.5% for ELS ZIP code. They average less than 1% for each of the other three outcomes. This might simply indicate that ZIP codes are poor proxies for neighborhoods. The ELS block group estimates are substantially larger: 2.3% for HSGRAD, about 3.5% for ENROLL, 3.6% for COLLBA, and 4.6% for wages, underscoring the importance of finer neighborhood designations.[35]

7.2 Model Estimates

7.2.1 Regression Indices and Random Effects

Because we do not seek to interpret the particular elements of the coefficient vectors $\boldsymbol{B}$, $\boldsymbol{G}_1$, or $\boldsymbol{G}_2$, we focus on characterizing the relative importance of student and family, neighborhood, school, and commuting zone levels in determining outcomes. To this end, the columns of Panels A, B, and C of Table 4 present estimates from (23) of the effects of a one standard deviation change in $\boldsymbol{X}_i\boldsymbol{B}$, $\boldsymbol{X}_n\boldsymbol{G}_1^N$, $\boldsymbol{X}_s\boldsymbol{G}_1^S$, $\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S$, and $\boldsymbol{Z}_{2c}^C\boldsymbol{G}_2^C$ on the latent probit indices that determine the education outcomes, as well as the corresponding standard deviations of the random effects. Panel D presents a corresponding set of estimates for $\ln(wage)$ at about age 25. Zero values for point estimates of these standard deviations indicate that the bias correction led to a negative value. The estimates of the interaction coefficients are somewhat noisy, so we also present the average of the point estimates for the NELS sample and the ELS ZIP code sample in the columns labeled "N + E", along with standard errors for this average.

The results reconfirm that the observed individual-level factors are powerful determinants of educational attainment. The neighborhood, school and commuting zone level components of the model have quite small standard deviations relative to $sd(v_i - v_n)$, which is normalized to 1, and $sd(\boldsymbol{X}_i\boldsymbol{B})$, which averages about 0.60 for high school, 0.90 for ENROLL and about 0.86 for COLLBA.[36] As we will demonstrate in Figure 1a below, the NELS results indicate that one can predict educational attainment quite well using observed student and family characteristics even in 8$^{\text{th}}$ grade. A one standard deviation increase in $\boldsymbol{X}_i\boldsymbol{B}$ raises the permanent wage by 0.163 (0.006) in the case of ELS and 0.134 (0.007) in the case of NELS, while $\widehat{sd}(v_i - v_n)$ is around twice as large, ranging from 0.274 (.008) to 0.318 (.007). These values are substantial relative to the estimated standard deviations of the permanent wage, which are 0.338 for ELS and 0.361 for NELS. The mid-20's wage used here does not fully capture the divergence that will occur later in life.

The estimates for the observed neighborhood index, $\widehat{sd}(\boldsymbol{X}_n\boldsymbol{G}_1^N)$, indicate that the ZIP code and block group characteristics taken from Census Bureau and LODES data have relatively little pre-

---

[35]The differences in across-neighborhood means of $\boldsymbol{X}_i$ and x_i^U of students drawn to a given school from different quality neighborhoods may understate differences in the mean of $\boldsymbol{X}_i$ and x_i^U of the full set students from the neighborhoods they come from. We do not have a good way to quantify this sample design effect.

[36]We obtain similar estimates of these parameters when we exclude interaction terms (see Online Appendix Table B5).

dictive power across outcomes and datasets. Indeed, once the bias correction has been applied, in 5 of the 12 education cases the estimated variance of the neighborhood index is less than one would expect due to sampling variance (so $\widehat{sd}(\boldsymbol{X}_n\boldsymbol{G}_1^N)$ is set to 0). $\widehat{sd}(v_n - v_s)$ is typically modest in size. For all three outcomes, it lies between 0.153 and 0.181 at the block group level in ELS and between 0.096 and 0.134 at the ZIP code level for both ELS and NELS. For the wage, $\widehat{sd}(v_n - v_s)$ is .049 (.006) log points in the ELS block group case. It is around .03 (.004) at the ZIP code level for both ELS and NELS.

In contrast to the neighborhood component of the control function, the index $\boldsymbol{X}_s\boldsymbol{G}_1^S$ exhibits sizable and statistically significant standard deviations between .1 and .2 for all education outcomes and datasets (except for COLLBA in ELS, where it is around .07), showing that the school-level demographic averages are absorbing a meaningful share of the outcome variation. They are similar in relative magnitude for the wage.

Next, we turn to the indices of observed group-level inputs that are not averages of individual inputs, $\boldsymbol{Z}_{2s}^S$ and $\boldsymbol{Z}_{2c}^C$. The ability of these indices to predict outcomes, along with 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 in the next section.

The estimates of $sd(\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S)$ are concentrated around 0.10 for HSGRAD and ENROLL. They are about 0.08 for COLLBA. The ELS estimates tend to be slightly smaller when using block group rather than ZIP code to define neighborhoods. The estimates of $sd(v_s - v_c)$ are around 0.1 across data sets and specifications for HSGRAD, around 0.14 for ENROLL, and around 0.065 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.

The estimates of $sd(\boldsymbol{Z}_{2c}^C\boldsymbol{G}_2^C)$ are relatively robust across datasets, alternative neighborhood definitions, and inclusion/exclusion of interaction terms for both high school graduation and four-year college enrollment. They are concentrated between .11 and .16, with slightly larger NELS88 values for college enrollment. Standard deviations for college graduation are somewhat smaller, between .062 and .091. Standard errors are typically around .03. The estimates are typically about one quarter larger than the estimates of $sd(\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S)$. The standard deviation of v_c is typically half to two thirds as large as $\widehat{sd}(v_s - v_c)$. Overall, observed and unobserved school inputs are slightly more important for education outcomes than CZ inputs. But as we will see in Section 8, $\widehat{sd}(\boldsymbol{Z}_{2c}^C\boldsymbol{G}_2^C)$ and $\widehat{sd}(v_c)$ are large enough to imply that shifting from low to high quality commuting zone environments substantially increases educational attainment.

Neither $\boldsymbol{Z}_{2s}^S$ nor $\boldsymbol{Z}_{2c}^C$ predicts log wages particularly well, with $\widehat{sd}(\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S)$ generally between .01 and .02 and $\widehat{sd}(\boldsymbol{Z}_{2c}^C\boldsymbol{G}_2^C)$ between .02 and .04. In comparison the values of $\widehat{sd}(\boldsymbol{X}_i\boldsymbol{B})$ are between 0.134 and 0.163. The values of $\widehat{sd}(\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S)$ are only reach statistical significance when NELS and ELS are pooled. The estimates of $sd(v_s - v_c)$ and $sd(v_c)$ are also relatively small—about .02 (.002) and 0.013 (.002) respectively. This may reflect the early age at which log wages are measured in both NELS and ELS.

7.2.2 Estimates of the Interaction Coefficients

Table 4 also reports estimates of the coefficients r_1^N, r_1^S, $\boldsymbol{r}_2^S$, and $\boldsymbol{r}_2^C$ on interactions between student-level variables and the standardized regression indices $\boldsymbol{X}_n \hat{\boldsymbol{G}}_1^N / sd(\boldsymbol{X}_n \hat{\boldsymbol{G}}_1^N)$, $\boldsymbol{X}_s \hat{\boldsymbol{G}}_1^S / sd(\boldsymbol{X}_s \hat{\boldsymbol{G}}_1^S)$, $\boldsymbol{Z}_{2s}^S \hat{\boldsymbol{G}}_2^S / sd(\boldsymbol{Z}_{2s}^S \hat{\boldsymbol{G}}_2^S)$, and $\boldsymbol{Z}_{2c}^C \hat{\boldsymbol{G}}_2^C / sd(\boldsymbol{Z}_{2c}^C \hat{\boldsymbol{G}}_2^C)$. In the case of the interactions between $\boldsymbol{X}_i \boldsymbol{B}$ (also in standard deviation units) and the neighborhood composition index $\boldsymbol{X}_n^N \boldsymbol{G}_1^N$, the estimates of r_1^N are almost always negative, suggesting that disadvantaged students disproportionately benefit from living in a stronger neighborhood. For *ENROLL* and *COLLBA*, the pooled estimates are negative and significant at the 0.15 and 0.01 levels, respectively. However, we do not make too much of this result because the bias corrected estimates of the main effect of $\boldsymbol{X}_n^N \boldsymbol{G}_1^N$ are weak, making identification of interaction effects suspect. Furthermore, the negative interaction is not present for high school graduation or for ELS log wages.

The coefficient r_1^S capturing interactions between $\boldsymbol{X}_i \boldsymbol{B}$ and $\boldsymbol{X}_s \boldsymbol{G}_1^S$ varies in sign. The pooled estimates and standard error are each around .02 for the three education outcomes. The estimate is essentially 0 for log wages.

The estimates of the interactions between $\boldsymbol{X}_i \boldsymbol{B}$ and the non-peer school inputs and policies captured in $\boldsymbol{Z}_{2s}^S \hat{\boldsymbol{G}}_2^S / sd(\boldsymbol{Z}_{2s}^S \hat{\boldsymbol{G}}_2^S)$ show consistent negative signs across all educational outcomes. The pooled estimates are -0.048 (0.023) and -0.059 (0.024) for *ENROLL* and *COLLBA*, respectively. In the case of *ENROLL*, the pooled estimate implies that the effect on the probit index of a one standard deviation increase in $\boldsymbol{Z}_{2s}^S \hat{\boldsymbol{G}}_2^S$ is -0.048 (0.023) lower for a student 1 standard deviation above the mean of $\boldsymbol{X}_i \boldsymbol{B}$ than for a student at the mean. This difference is small compared to the pooled estimate of 0.901 (0.020) of the main effect of $\boldsymbol{X}_i \boldsymbol{B}$, although one could not rule out a difference of -0.063 at the 90% level. This value is about half of the main effect of $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$. The estimates do suggest increased sensitivity to external inputs for disadvantaged students. The relatively low predictive power of $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$ in the case of wages inhibits identification of interactions with $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$.

The pooled estimates of the interactions between $\boldsymbol{X}_i \boldsymbol{B}$ and the commuting zone index $\boldsymbol{Z}_{2c}^C \hat{\boldsymbol{G}}_2^C / sd(\boldsymbol{Z}_{2c}^C \hat{\boldsymbol{G}}_2^C)$ are positive for the education outcomes and negative for log wages, and always small and statistically insignificant.

The interactions between the indicators for female, minority status, and low income with $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$ or $\boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C$ tend to be statistically insignificant as well as inconsistent in sign across datasets and outcomes. In the case of minority status and low income status, none of the pooled estimates are significant at the 10% level, and the point estimates are typically small relative to the main effects of $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$ and $\boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C$.

Overall, we do not find strong evidence of interaction effects. The standard errors are sufficiently large that we cannot rule out moderately important interactions. But we can rule out interactions that would substantially alter the relationship between $\boldsymbol{X}_i \boldsymbol{B}$ and the latent variables for education and for log wages. In the education case, the non-linearity of the probit function may be enough to capture the differential sensitivity of students with higher versus lower observed inputs to group-level inputs. We investigate this further in Sections 8.1.1 and 8.2.1.

8 The Impact of Shifts in School and Commuting Zone Quality

This section reports estimates of the effects of 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 X_i variables but briefly discuss estimates using the basic set, which are typically larger.

8.1 High School Graduation

The row labeled "School only" in Panel A of Table 5 reports the average treatment effect for the full student population of attending a 90th quantile school instead of a 10th quantile school. The rows labeled "CZ only" and "School+CZ" report the impact of a 10th to 90th quantile shift in the commuting zone environment and a 10th to 90th quantile shift in combined school and commuting zone quality, respectively. Standard errors are in parentheses. As in prior tables, NELS, ELS block group, and ELS ZIP code results are denoted by the column headings NELS, ELSbg, and ELSz respectively.

Focusing first on the NELS results, we see that shifting from a 10th to a 90th quantile school/commuting zone increases the high school graduation probability from 0.816 to 0.904, an increase of 0.088 (0.014). This is very large relative to the dropout rate. A corresponding shift in the school environment only (while receiving a random draw from the distribution of commuting zone inputs) changes the graduation probability from 0.831 to 0.891, an increase of 0.06 (0.014). The effect of a 10th-to-90th quantile shift in commuting zone inputs is 0.067 (0.017).

The estimates for ELS are somewhat smaller. In the ELS block group case, the "School and CZ" treatment raises the graduation probability by 0.062 (0.011), from 0.896 to 0.958. The school treatment raises it by 0.045 (0.010), and the CZ treatment raises it by 0.041 (0.009). The values are slightly smaller when we use ZIP code as the neighborhood measure. 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. Online Appendix Table B7 displays estimates of the impacts of 10th-to-50th quantile shifts for the same treatments. Despite the nonlinearity in the probit function, the estimates are only slightly more than half as large as the corresponding 10th-to-90th shifts, so the shift from the 50th-to-90th quantile matters almost as much as the shift from the 10th-to-50th quantile.

The estimates for the combined "School and CZ" treatment for ELS are essentially identical to the school-level estimates from the full specification in AM that implicitly incorporate CZ inputs as well. The NELS estimates are about .04 smaller than the corresponding estimates in AM, indicating some sensitivity to the changes in specification introduced in this paper: the exclusion of weights, the addition of interactions, and the addition of neighborhood-level control function variables X_n and CZ-level observables Z_{2c}^C.

While the estimates are somewhat noisy, 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 $(\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S, \boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C, (v_s - v_c), v_c)$ that we ascribe to school and commuting zone inputs comprise relatively small fractions of the outcome variance. Row 8 of Appendix Table A1 reports that the variance of $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S + \boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C + (v_s - v_c) + v_c$, which drives the "School and CZ" treatment, comprises just 2.2 percent of the full latent index variance in NELS, and only 3.0 percent in the ELS block group specification.[37] Rows 9 and 10 report that the school inputs that matter for the "School Only" treatment and the commuting zone inputs that matter for the "CZ Only" treatment each account for only about 1 to 1.5 percent of the latent index variance in NELS.

Why 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-level inputs. The second reason is that for binary outcomes even a small shift in the 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 $\boldsymbol{X}_i$ variables in Online Appendix Table B9 are typically larger by less than 1 percentage point for each treatment. The NELS estimates are each around 1 percentage point larger. As we have discussed in Section 5.3, the estimates based on the full set are more conservative. Both estimates exclude peer effects captured by $\boldsymbol{X}_s \hat{\boldsymbol{G}}_1^S$ and $\boldsymbol{X}_n \hat{\boldsymbol{G}}_1^N$.

8.1.1   Subpopulation Results for High School Graduation

We now examine heterogeneity in treatment effects. Treatment effect heterogeneity across groups arises through three pathways. First, the model includes interactions involving female, underrepresented minority, and low income indicator variables. As we discussed in Section 7.2.2, the coefficients on the interactions between female, minority status, or low income and either $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$ or $\boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C$ tend to be statistically insignificant as well as inconsistent in sign across datasets and outcomes.

The second and third pathways both arise from differences in the distributions of $\boldsymbol{X}_i \boldsymbol{B}$ across subpopulations. The second pathway arises because the model contains interactions between $\boldsymbol{X}_i \boldsymbol{B}$ and $\boldsymbol{Z}_{2s}^S \boldsymbol{G}_2^S$ and $\boldsymbol{Z}_{2c}^C \boldsymbol{G}_2^C$. As we have already discussed, the interactions between $\boldsymbol{X}_i \boldsymbol{B}$ and the non-peer school inputs and policies captured in $\boldsymbol{Z}_{2s}^S \hat{\boldsymbol{G}}_2^S / sd(\boldsymbol{Z}_{2s}^S \hat{\boldsymbol{G}}_2^S)$ show consistent negative signs across all educational outcomes but are small for high school graduation. The interactions between $\boldsymbol{X}_i \boldsymbol{B}$ and $\boldsymbol{Z}_{2c}^C \hat{\boldsymbol{G}}_2^C / sd(\boldsymbol{Z}_{2c}^C \hat{\boldsymbol{G}}_2^C)$ are typically too small to matter much, regardless of the outcome. Conse-

---

[37]Note that these variance components are computed based on the specification in Columns 1-3 of Table 4. 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.

quently, the interaction terms are not a primary source of differences across groups in treatment effects, at least for HSGRAD.

The third pathway arises because differences in the locations of the $X_i B$ distribution across subgroups produces treatment effect heterogeneity for binary outcomes even in the absence of explicit interaction terms. Due to the nonlinearity of the probit function that links Y_i to the binary outcome indicators for education, the sensitivity to school and commuting zone quality is higher for subpopulations with values of $X_i \hat{B}$ 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 $X_i B$ value, they would all attain the positive outcome regardless of their school/commuting zone environment. We think this is main source of heterogeneity in the treatment effects, to which we now turn.

Table 6 displays the impacts on high school graduation of the 10th-to-90th quantile shifts in quality for each of our three counterfactuals for six subpopulations, which are indicated by the column headings. We consider Hispanic students, non-Hispanic black students, non-Hispanic 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.

We start with NELS. White students dominate the sample, and so estimates are similar to but a bit smaller than those for the full sample. Black students are moderately more sensitive to each of the three treatments than the population at large. Hispanic students and non-Hispanic white students with a single mother with a high school education or less have the lowest mean $X_i B$ values and are the most sensitive. For non-Hispanic blacks, the "School and CZ" 10th-to-90th quality shift increases the predicted high school graduation probability by 0.095 (0.017), from 0.806 to 0.901. For children of less-educated white single mothers, the increase is 0.140 (0.021), from 0.669 to 0.810. In contrast, non-Hispanic white students with two college-graduate parents can only expect a 0.040 (0.007) increase in graduation probability from the same shift in school/commuting zone inputs. We obtain similar estimates of NELS 10th-90th treatment effects using the models without interactions[38] This suggests that the chief source of treatment effect heterogeneity is the nonlinearity of the probit function rather than the estimated interactions, though the interaction effects do cause the estimates to diverge from the mean by less than the probit function alone would predict.

The ELS results with either block group or ZIP code as the neighborhood are smaller. The "School and CZ" effects are 0.087 (0.015) and 0.076 (0.015), respectively, for Hispanic and non-Hispanic black students using the block group estimates, and only about 0.021 (0.004) for white students with two college-graduate parents. The smaller ELS results are primarily due to fact that the ELS sample average for HSGRAD is 0.92.

[38]The values are 0.088 for blacks, 0.130 for white students with less-educated single mothers, and 0.038 for white students with two college-graduate parents. The table is available upon request.

Differences across subpopulations in the effects of the "School only" and "CZ only" treatments follow a similar pattern. For these counterfactuals, the 10th-90th treatment effects for the subpopulations vary by a few percentage points around the "School only" and "CZ only" 10th-90th treatment effects for the full sample.

Figures 1a and 1b and Figure 2 illustrate more generally how the predicted outcome changes as the quantile of the $X_i B$ index increases. The solid dark line in Figure 1a (1b) shows how in the NELS (ELS block group) case the probability of high school graduation (vertical axis) varies with the quantile of $X_i B$ 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 2 graphs 10th-to-90th treatment effects against the $X_i B$ quantile. We exclude confidence intervals to reduce clutter, but Online Appendix Table B8 reports both estimates and standard errors of treatment effects for students at the 10th, 50th and 90th quantiles of $X_i B$.

The impact of each 10th-to-90th treatment decreases monotonically with the $X_i B$ quantile. Specifically, students in NELS (light grey line) at the 10th quantile of the $X_i B$ distribution move from a 0.592 chance of graduating to a 0.762 chance when the quality of the combined school/commuting zone environment shifts from the 10th to the 90th quantile, whereas students at the 90th quantile only move from a 0.970 graduation rate to a 0.991. So 8th grade school and commuting zone environment seems to be very 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 over the range of $X_i B$.

8.2 Enrollment in a Four-Year College

Panel B of Table 5 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 quantile shift in combined school/commuting zone quality raises the enrollment probability by 0.171 (0.018) (0.236 to 0.407) for 8th graders in NELS. The corresponding shifts in school inputs only and commuting zone inputs only raise enrollment by 0.125 (0.017) and 0.114 (0.017). The estimates based on ELS ZIP code and ELS block group specifications are quite close to the NELS estimates. Both the NELS and ELS "School and CZ" estimates are within .01 of the corresponding full specification estimates from AM. Estimates of the impact of 10th-to-50th quantile shifts in Online Appendix Table B7 are only slightly less than half of the corresponding 10th-to-90th shift estimates.

Note that the impacts of school and commuting zone on this outcome are considerably larger than those for high school graduation. For NELS a partial explanation lies in the fact that the fraction of the variance in Y_i (the latent index) that is attributed to school and commuting zone inputs is 0.033 for ENROLL and 0.022 for HSGRAD (Appendix Table A1). However, the variance

40

components underlying the ELS block group and ELS ZIP code counterfactuals are very similar in magnitude to those used for the high school graduation counterfactuals (0.030 of $Var(Y_i)$ in each case). Instead, the larger estimated impacts for college enrollment primarily reflect the fact that the college enrollment probability (0.327 in NELS and 0.422 in ELS, from Table 1-1) is much closer to 0.5 than for high school graduation (0.853 in NELS and 0.919 in ELS). The probit model assumes a normal distribution of unobserved inputs, so that an outcome mean of .5 suggests 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 X_i are about 0.04 larger in ELS and about 0.02 larger in NELS (Online Appendix Table B9, columns 4-6).

8.2.1 Subpopulation Results for College Enrollment

Panel B of Table 6 reports the impacts of our counterfactual shifts for particular subgroups of NELS 8th graders. Because the population mean for college enrollment is below 0.5, 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 on the enrollment probability range from 0.116 to 0.213. The "School only" treatment displays slightly higher effects than the "CZ only" treatments with similar degrees of heterogeneity: impacts range from 0.086 to 0.154 for "School only" and 0.075 to 0.142 for the "CZ only" treatment. The estimates of the treatment effects for the ELS ZIP code and block group specifications follow the same pattern as in NELS, although the subgroup estimates fall in a narrower range.

Figure 3 shows the relationship between the treatment effects on college attendance probability and the quantiles of $X_i B$. The impact of the counterfactual treatments is non-monotonic in $X_i B$, increasing and then decreasing as one moves through the distribution of $X_i B$. For example, for the "School and CZ" counterfactual in NELS, the estimated impact for the 10th, 75th (near the maximum impact), and 90th quantiles of $X_i B$ are 0.062, 0.253, and 0.218, respectively.

The pattern occurs despite the fact that in the NELS case the coefficient on the interaction between $X_i B$ and $Z_{2,3}^S G_2^S$ is substantial: -0.069 (0.034). Part of the decline at the top of the $X_i B$ distribution may be due to this. However, the pattern is primarily due to a striking fact about the variation in predicted outcome probabilities across $X_i B$ 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 90th quantiles of $X_i B$ are 0.042, 0.136, 0.429, and 0.703, 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. Consequently, very few students near the bottom of the $X_i B$ distribution are in a position

41

where higher quality school inputs would be enough to push them over the decision margin, even if they are slightly more sensitive to school quality than the average student.

In fact, the ELS ZIP code specification provides a good opportunity to assess the importance of the interaction terms in the model. First, $\widehat{sd}(\mathbf{Z}_{2c}^C \mathbf{G}_2^C)$ is substantial, so that there is more signal contained in this regression index with which to identify true interactions. Second, the interaction coefficients on $1(LowInc) \times \mathbf{Z}_{2c}^C \mathbf{G}_2^C$ and $1(URM) \times \mathbf{Z}_{2c}^C \mathbf{G}_2^C$ are seemingly non-negligible in size (albeit still noisily estimated): 0.050 and 0.054 respectively (Table 4 Panel B). This suggests greater sensitivity of students from modest backgrounds and minority students to CZ inputs. Third, the variance components for the "CZ only" treatment, $Var(\mathbf{Z}_{2c}^C \mathbf{G}_2^C + v_c)$, contribute 1.26% of $Var(Y_i)$ when interactions are included and 1.24% of $Var(Y_i)$ when they are excluded. This means that we can directly compare treatment effects based on the models with and without interactions essentially holding fixed the importance of the main effects. Indeed, the mean impact of the 10th-90th "CZ only" treatment on ENROLL is 0.118 for both specifications.

As it turns out, none of the subgroups we consider displays more than a 0.002 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

Panel C of Table 5 reports the estimated impacts of the same counterfactual treatments on a randomly selected student's probability of graduating college by age 25. The NELS results indicate that replacing a combined school/commuting zone at the 10th quantile of the quality distribution with one at the 90th quantile increases the college graduation probability by 0.085 (0.016) (from 0.309 to 0.394). Shifting only school inputs or only CZ inputs from their respective 10th to 90th quantiles increases the graduation probability by 0.077 (0.18) and 0.046 (0.019), respectively. The small increase in impact from shifting school and CZ quality together relative to the impact of the separate shifts is partly due to an estimated small negative covariance between $\mathbf{Z}_{2s}^S \mathbf{G}_2^S$ and $\mathbf{Z}_{2c}^C \mathbf{G}_2^C$. It also reflects the fact that doubling a variance only increases the standard deviation by a factor of $\sqrt{2}$.

The ELS block group estimates are only 0.090 (0.012), 0.071 (0.015), and 0.073 (0.017), respectively, with the ELS ZIP code estimates displaying slightly smaller magnitudes. Switching from the full specification to the basic specification (displayed in Online Appendix Table B9) increases the estimates for each dataset by about 0.015.

The three treatments have a smaller effect on the college graduation probability than the enrollment rate. We do not have a complete explanation for this, but three possibilities are worth mentioning. First, School and CZ inputs may play a bigger role in inducing students to attend a four-year college than to successfully complete college. The second and related point is that marginal students

who are induced to enroll in a four-year college as a result of superior school and CZ inputs may graduate at lower rates. Finally, there is considerable attrition in the sample between the survey in which ENROLL is measured (2nd and 3rd Follow-ups for NELS and ELS, respectively) and the survey in which COLLBA is measured (3rd and 4th follow-ups). Differential attrition could lead the sample of students remaining when COLLBA is measured to be less sensitive to school and CZ inputs than in earlier samples.[39]

8.3.1 Subpopulation Results for College Graduation

Table 6 reveals moderate heterogeneity in estimated impacts from the counterfactual treatments across subgroups for NELS 8th graders. A 10th-to-90th shift in combined school/CZ input quality increases the probability of college graduation by between 0.056 (0.011) and 0.101 (0.018) with the smallest impacts occurring for white students with a single mother with a high school degree or less and the largest for white students with two parents with bachelor's degrees.

The variance components attributed to school inputs ($Var(\mathbf{Z}_{2s}^{S}\mathbf{G}_{2}^{S})$ and $Var(v_s - v_c)$) are substantially larger in NELS than those attributed to CZ inputs ($Var(\mathbf{Z}_{2c}^{C}\mathbf{G}_{2}^{C})$ and $Var(v_c)$). As a result, the counterfactual impacts are substantially larger for the "School only" specification than the "CZ only" specification, particularly for subgroups featuring high values of $\mathbf{X}_i\mathbf{B}$. For example, a 10th-to-90th quantile shift in the quality of school inputs increases enrollment by 0.081 for the white subpopulation, while the corresponding shift in the quality of commuting zone inputs only produces a 0.049 increase. The "School Only" and "CZ Only" treatment effects are much more similar to another in ELS.

Once again, examining the variation in treatment impacts across $\mathbf{X}_i\mathbf{B}$ quantiles reveals a greater degree of heterogeneity (Figure 4). For NELS, the "School and CZ" impacts for the 10th, 75th, and 90th quantiles of $\mathbf{X}_i\mathbf{B}$ are 0.031, 0.124, and 0.098, respectively, demonstrating the same non-monotonicity of treatment effects across the $\mathbf{X}_i\mathbf{B}$ distribution as was observed for the college enrollment outcome. Figure 1a shows that in NELS the probability of college graduation is 0.046 for a student at the 10th quantile of $\mathbf{X}_i\mathbf{B}$ and 0.754 at the 90th (evaluated at the median value of the "School and CZ" treatment).

8.4 Log Wage Rates

Columns 10-12 of Table 5 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 quantile combined school/commuting zone treatment is 0.111 (0.025) log points (which corresponds to an 11.7% wage increase). A one standard deviation improvement in the treatment would raise log wages by 0.045 (a 4.6% wage increase). In comparison, the standard deviation of the permanent component of the wage is 0.361. The values for ELS are slightly smaller.

---

[39] As discussed in Section 5, the mean of ENROLL among sample members of the 3rd follow-up survey in NELS is .353 vs. .327 among those in the 2nd follow-up (when the outcome is determined), suggesting some differential attrition is taking place.

The estimate of the combined school/community 10th-90th treatment effect is only .104 (0.024) log points in the block group specification (an 11% wage increase). Estimates using the basic specification in Online Appendix Table B9 are very 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 $\boldsymbol{M}_i$ and $\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S$ and $\boldsymbol{Z}_{2c}^C\boldsymbol{G}_2^C$. 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.

Finally, both school and commuting zone inputs seem to be important, and with comparable magnitudes. However, this conclusion is more tentative, because we cannot fully distinguish between unobservable school inputs and unobservable commuting zone inputs.

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

44

educational attainment and early career wages, and the degree to which this relative importance differs across students from different backgrounds. We extend the identification results of Altonji and Mansfield (2018) to models with multiple group levels and with interactions between observed individual factors and both observed and unobserved group level factors. The identification results are based on the idea that group averages of individual-level observables can fully control for sorting bias from group averages of individual-level unobservables. A natural but challenging extension to our analysis would be to incorporate interactions with unobserved individual characteristics.

Our theoretical results demonstrate the existence of 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/commuting zone quality from the 10th to the 90th quantile of their respective distributions. We also produce estimates for particular student subpopulations that exploit the treatment effect heterogeneity accommodated 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 school/commuting zone combination at the 10th quality quantile to a 90th quantile combination increases the probability of high school graduation by at least 0.06-0.09 across datasets and specifications, the probability of enrollment at a four-year college by at least 0.17-0.18, and the probability of college graduation by at least 0.08-0.09. 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. A 10th-90th quantile shift in school/commuting zone quality increases wages by 10-11 percent, with commuting zone factors playing a somewhat larger role relative to school inputs and policies.

High school graduation rates of disadvantaged populations are considerably more sensitive to school and commuting zone inputs, while college graduation rates are more sensitive to group-level inputs for advantaged populations. However, the heterogeneity in treatment effects is driven by the 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 usually small (though somewhat 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 commut-

ing zone-level sorting and a very small role for neighborhood sorting. It is possible, of course, that the neighborhood component may be understated by the use of block group or ZIP code 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. Obtaining observed neighborhood characteristics that are plausible sources of neighborhood-level causal effects would be particularly valuable.

References

Aaronson, Daniel. 1998. Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources* 33(4), 915-946.

Alexander, Karl and Stephen L. Morgan. 2016. The Coleman Report at fifty: Its legacy and implications for future research on equality of opportunity. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(5), 1-16.

Altonji, Joseph G. and Thomas A. Dunn. 1996. Using siblings to estimate the effect of school quality on wages. *Review of Economics and Statistics* 78(4), 665-671.

Altonji, Joseph G. and Richard K. Mansfield. 2018. Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects. *American Economic Review* 108(10), 2902-46.

Altonji, Joseph G. and Richard K. Mansfield. 2011. The role of family, school, and community characteristics in inequality in education and labor market outcomes. In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, ed. Greg J. Duncan and Richard J. Murnane. Russell Sage Foundation.

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

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 L. S. 2016. The effects of school integration: Evidence from a randomized desegregation program. Working Paper No. 6119, CESifo Group Munich.

Berry, Steven T. 1994. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics* 25(2), 242-262.

Betts, Julian R. 1995. Does school quality matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics* 77(2), 231-250.

Browning, Martin, Pierre-Andre 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. 2018a. The effects of neighborhoods on intergenerational

mobility I: Childhood exposure effects. *Quarterly Journal of Economics* 133(3), 1107-1162.

——, 2018b. The effects of neighborhoods on intergenerational mobility II: County level estimates. *Quarterly Journal of Economics* 133(3), 1163-1228.

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. *American Economic Review* 106(4), 855-902.

Chyn, Eric. 2018. Moved to opportunity: The long-run effect of public housing demolition on labor market outcomes of children. *American Economic Review* 108(10), 3028-56.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederick D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity*. US Department of Health, Education & Welfare.

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 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, Thomas J. Kane, and Douglas O. Staiger. 2014. School choice, school quality, and postsecondary attainment. *American Economic Review* 104(3), 991-1013.

Dobbie, Will and Roland G. Fryer. 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.

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. *Annual Review of Economics* 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*

48

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* 64(4), 251-262.

Goldstein, Harvey. 2011. *Multilevel Statistical Models*. John Wiley & Sons.

Graham, Bryan S. 2018. Identifying and estimating neighborhood effects. *Journal of Economic Literature* 56(2), 450-500.

Harding, David J., Lisa Gennetian, Christopher Winship, Lisa Sanbonmatsu, and Jeffrey R. Kling. 2011. Unpacking neighborhood influences on education outcomes: Setting the stage for future research. In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, ed. Greg J. Duncan and Richard J. Murnane. Russell Sage Foundation.

Jacob, Brian A. 2004. Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in Chicago. *American Economic Review* 94(1), 233-258.

Jencks, Christopher S. and Susan E. Mayer. 1990. The social consequences of growing up in a poor neighborhood. In *Inner-City Poverty in the United States*, ed. National Research Council. National Academies Press.

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* (1), 629-662.

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

Laliberte, Jean-William P. 2018. Long-term contextual effects in education: Schools and neighborhoods. Working Paper.

Lindenlaub, Ilse. 2017. Sorting multidimensional types: Theory and application. *Review of Economic Studies* 84(2), 718-789.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. 2016. Matching, sorting, and wages. *Review of Economic Dynamics* 19, 63-87.

Lucas, Samuel R. 2016. First- and second-order methodological developments from the Coleman Report. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(5), 117-140.

McFadden, Daniel L. 1984. Econometric analysis of qualitative response models. *Handbook of Econometrics* 2, 1395-1457.

Meghir, Costas and Steven Rivkin. 2011. Econometric methods for research in education. *Handbook of the Economics of Education* 3, 1-87.

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

Oreopoulos, Philip. 2003. The long-run consequences of living in a poor neighborhood. *Quarterly Journal of Economics* 118(4), 1533-1575.

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage.

Rosen, Sherwin. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82(1), 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* 28, 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.

Todd, Petra E. and Kenneth I. 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.

TABLE 1-1

SUMMARY OF OUTCOME VARIABLES

Outcome Variable	NELS		ELS	
	Mean	Std. Dev.	Mean	Std. Dev.
Education outcome:				
1(HSGRAD)	0.853	0.354	0.919	0.272
1(ENROLL)	0.327	0.469	0.422	0.494
1(COLLBA)	0.353	0.478	0.376	0.484
Log hourly wage (2009 $)	2.544	0.534	2.666	0.500

NOTE.—Statistics calculated on respective regression samples based on the "full" specification. The dependent variables are indicators for high school graduation (HSGRAD), enrollment in a 4 year college within 2 years after graduation 1(ENROLL), attainment of a BA degree (COLLBA), and the log hourly wage rate, which is measured at about age 25 and adjusted to 2009 dollars.

Table 1-2

The Number of Individuals, ZIP Codes or Block Groups, Schools, and Commuting Zones in the Samples (by Outcome and Data Set)

	Sample Size	Block Groups	ZIP Codes	Schools	CZs
NELS specification:					
HSGRAD	13300	· · ·	3080	930	300
ENROLL	12900	· · ·	3050	930	300
COLLBA	10800	· · ·	2800	910	300
LOG WAGE	8800	· · ·	2470	880	300
ELS specification:					
HSGRAD	14600	9130	4140	700	260
ENROLL	13400	8630	3970	700	260
COLLBA	12400	8110	3810	700	260
LOG WAGE	11600	7720	3670	700	260

Note.—Statistics calculated on respective regression samples based on the "full" specification. The dependent variables are high school graduation (HSGRAD), enrollment in a 4 year college within 2 years after graduation (ENROLL), and attainment of a BA degree (COLLBA). The log hourly wage rate is for about age 25 and adjusted to 2009 dollars. All values are rounded to the nearest 10.

	A. High School Graduation				B. College Enrollment			
	NELS	ELSbg	ELSz	N+E	NELS	ELSbg	ELSz	N+E
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables:								
sd($X_i B$)	0.619***	0.588***	0.580***	0.600***	0.856***	0.946***	0.945***	0.901***
	(.072)	(.028)	(.025)	(.038)	(.026)	(.025)	(.03)	(.02)
sd($X_{1n} G_1^N$)	0.000	0.000	0.100**	0.050*	0.000	0.060*	0.000	0.000
	(.044)	(.044)	(.034)	(.028)	(.032)	(.031)	(.034)	(.023)
sd($X_{1s} G_1^S$)	0.162***	0.148***	0.117***	0.140***	0.155***	0.173***	0.186***	0.171***
	(.041)	(.038)	(.032)	(.026)	(.034)	(.028)	(.033)	(.024)
sd($Z_{2s}^S G_2^S$)	0.096***	0.101**	0.113***	0.105***	0.123***	0.088***	0.117***	0.120***
	(.036)	(.04)	(.031)	(.024)	(.029)	(.029)	(.031)	(.021)
sd($Z_{2c}^C G_2^C$)	0.134***	0.132***	0.111***	0.123***	0.165***	0.121***	0.136***	0.151***
	(.042)	(.035)	(.032)	(.026)	(.027)	(.031)	(.032)	(.021)
$X_i B$ x $X_{1n} G_1^N$ (r_1^N)	-0.025	0.024	0.040*	0.008	-0.028	-0.030	-0.049	-0.038
	(.035)	(.033)	(.023)	(.021)	(.035)	(.03)	(.031)	(.024)
$X_i B$ x $X_{1s} G_1^S$ (r_1^S)	0.031	0.012	0.018	0.025	0.054**	-0.036	-0.028	0.013
	(.033)	(.028)	(.028)	(.022)	(.024)	(.035)	(.035)	(.021)
$X_i B$ x $Z_{2s}^S G_2^S$ (r_{21}^S)	-0.007	-0.008	-0.015	-0.011	-0.069**	-0.017	-0.028	-0.048**
	(.036)	(.034)	(.033)	(.024)	(.034)	(.033)	(.032)	(.023)
$X_i B$ x $Z_{2c}^C G_2^C$ (r_{21}^C)	0.014	0.008	-0.014	0.000	-0.010	0.023	0.017	0.004
	(.025)	(.029)	(.029)	(.019)	(.027)	(.026)	(.025)	(.018)
Female x $Z_{2s}^S G_2^S$ (r_{22}^S)	0.013	-0.063	-0.070	-0.028	-0.012	-0.020	-0.010	-0.011
	(.043)	(.054)	(.054)	(.035)	(.031)	(.029)	(.03)	(.022)
Minority x $Z_{2s}^S G_2^S$ (r_{23}^S)	-0.075	0.122**	0.125**	0.025	0.041	-0.058	-0.052	-0.005
	(.047)	(.055)	(.053)	(.035)	(.05)	(.041)	(.038)	(.031)
LowInc x $Z_{2s}^S G_2^S$ (r_{24}^S)	0.055	0.028	0.007	0.031	-0.013	0.048	0.054	0.020
	(.054)	(.06)	(.064)	(.042)	(.051)	(.04)	(.039)	(.032)
Female x $Z_{2c}^C G_2^C$ (r_{22}^C)	-0.012	-0.010	0.038	0.013	-0.041	-0.035	-0.039	-0.040*
	(.039)	(.042)	(.042)	(.029)	(.029)	(.03)	(.032)	(.021)
Minority x $Z_{2c}^C G_2^C$ (r_{23}^C)	0.060	0.008	0.009	0.034	-0.024	0.070	0.054	0.015
	(.062)	(.06)	(.062)	(.044)	(.042)	(.043)	(.04)	(.029)
LowInc x $Z_{2c}^C G_2^C$ (r_{24}^C)	-0.015	-0.046	-0.042	-0.028	-0.079***	0.019	0.050	-0.015
	(.046)	(.044)	(.043)	(.031)	(.03)	(.049)	(.048)	(.028)
(Intercept)	1.337***	1.794***	1.778***	1.557***	-0.640***	-0.279***	-0.279***	-0.459***
	(.037)	(.046)	(.046)	(.029)	(.043)	(.038)	(.039)	(.029)
Random effects:								
sd($v_n - v_s$)	0.123***	0.181***	0.134***	0.128***	0.122***	0.158***	0.111***	0.117***
	(.017)	(.03)	(.017)	(.012)	(.018)	(.03)	(.018)	(.013)
sd($v_s - v_c$)	0.083***	0.118***	0.097***	0.090***	0.142***	0.146***	0.147***	0.145***
	(.013)	(.017)	(.016)	(.01)	(.021)	(.027)	(.022)	(.015)
sd(v_c)	0.060***	0.057***	0.059***	0.059***	0.046***	0.104***	0.101***	0.073***
	(.011)	(.012)	(.011)	(.008)	(.01)	(.03)	(.028)	(.015)

ESTIMATES OF THE EDUCATION AND WAGE PROBIT
MODEL PARAMETERS (FULL SET OF STUDENT VARIABLES)

	C. College Graduation				D. Log Wage			
	NELS	ELShg	ELSz	N+E	NELS	ELShg	ELSz	N+E
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Variables:								
$sd(X_i B)$	0.890***	0.835***	0.839***	0.864***	0.134***	0.163***	0.163***	0.148***
	(.027)	(.026)	(.027)	(.019)	(.007)	(.006)	(.006)	(.005)
$sd(X_{1n} G_1^N)$	0.040	0.095***	0.048	0.044*	0.042***	0.015	0.032**	0.037***
	(.042)	(.028)	(.032)	(.026)	(.014)	(.011)	(.015)	(.01)
$sd(X_{1s} G_1^S)$	0.139***	0.072***	0.071**	0.105***	0.055***	0.037***	0.029***	0.042***
	(.035)	(.027)	(.028)	(.022)	(.012)	(.009)	(.01)	(.008)
$sd(Z_{2s}^S G_2^S)$	0.088***	0.073***	0.074***	0.081***	0.017	0.012	0.018	0.018*
	(.03)	(.024)	(.026)	(.02)	(.013)	(.011)	(.012)	(.009)
$sd(Z_{2c}^C G_2^C)$	0.062**	0.091***	0.073***	0.067***	0.037***	0.028***	0.027**	0.032***
	(.026)	(.026)	(.026)	(.018)	(.01)	(.01)	(.011)	(.007)
$X_i B$ x $X_{1n} G_1^N$ (r_1^N)	-0.070*	-0.079***	-0.068*	-0.069**	-0.016	0.005	0.003	-0.006
	(.04)	(.026)	(.04)	(.028)	(.01)	(.007)	(.007)	(.006)
$X_i B$ x $X_{1s} G_1^S$ (r_1^S)	0.060**	-0.023	-0.021	0.019	-0.004	0.000	-0.002	-0.003
	(.027)	(.028)	(.029)	(.02)	(.009)	(.008)	(.008)	(.006)
$X_i B$ x $Z_{2s}^S G_2^S$ (r_{21}^S)	-0.042	-0.073**	-0.075**	-0.059**	0.006	0.002	0.002	0.004
	(.034)	(.033)	(.033)	(.024)	(.008)	(.006)	(.007)	(.005)
$X_i B$ x $Z_{2c}^C G_2^C$ (r_{21}^C)	0.058*	-0.013	-0.025	0.016	-0.017*	0.002	0.003	-0.007
	(.03)	(.03)	(.038)	(.024)	(.01)	(.007)	(.007)	(.006)
Female x $Z_{2s}^S G_2^S$ (r_{22}^S)	-0.043	0.032	0.034	-0.004	0.009	0.003	-0.004	0.002
	(.039)	(.03)	(.03)	(.024)	(.014)	(.015)	(.014)	(.01)
Minority x $Z_{2s}^S G_2^S$ (r_{23}^S)	-0.020	0.028	0.021	0.001	-0.024	0.005	0.001	-0.011
	(.05)	(.046)	(.044)	(.033)	(.016)	(.012)	(.012)	(.01)
LowInc x $Z_{2s}^S G_2^S$ (r_{24}^S)	-0.001	0.056	0.031	0.015	0.007	-0.003	-0.005	0.001
	(.046)	(.061)	(.058)	(.037)	(.015)	(.013)	(.014)	(.01)
Female x $Z_{2c}^C G_2^C$ (r_{22}^C)	-0.033	0.006	0.017	-0.008	0.017	0.026	0.020	0.018
	(.032)	(.03)	(.031)	(.023)	(.016)	(.016)	(.017)	(.011)
Minority x $Z_{2c}^C G_2^C$ (r_{23}^C)	-0.034	0.002	-0.005	-0.020	-0.026	-0.002	0.004	-0.011
	(.046)	(.049)	(.049)	(.034)	(.02)	(.013)	(.013)	(.012)
LowInc x $Z_{2c}^C G_2^C$ (r_{24}^C)	-0.040	0.118**	0.107*	0.033	0.020	0.016	0.011	0.016
	(.051)	(.049)	(.056)	(.038)	(.015)	(.012)	(.013)	(.01)
(Intercept)	-0.525***	-0.426***	-0.432***	-0.478***	2.545***	2.664***	2.666***	2.606***
	(.042)	(.035)	(.036)	(.028)	(.008)	(.009)	(.009)	(.006)
Random effects:								
$sd(v_i)$					0.318***	0.274***	0.278***	0.298***
					(.007)	(.008)	(.009)	(.006)
$sd(v_n - v_s)$	0.096***	0.153***	0.103***	0.099***	0.029***	0.049***	0.031***	0.030***
	(.012)	(.023)	(.014)	(.009)	(.004)	(.006)	(.004)	(.003)
$sd(v_s - v_c)$	0.070***	0.065***	0.059***	0.064***	0.021***	0.020***	0.020***	0.020***
	(.008)	(.01)	(.009)	(.006)	(.003)	(.002)	(.002)	(.002)
$sd(v_c)$	0.025***	0.042***	0.039***	0.032***	0.014***	0.011***	0.013***	0.013***
	(.005)	(.006)	(.006)	(.004)	(.002)	(.002)	(.002)	(.001)

NOTE.—Bootstrap standard errors are in parentheses. The dependent variables are high school graduation (HSGRAD) in Panel A, enrollment in a 4 year college within 2 years after expected high school graduation (ENROLL) in Panel B, attainment of a BA degree (COLLBA) in Panel C and the log hourly wage rate at about age 25 in Panel D. Panels A, B, and C refer to the latent index of an MME probit specification. Panel D is based on an MME regression specification. The model is

equation (23). The column heading NELS refers to the NELS data. ELSbg represents the ELS data set with the neighborhood specification of block group. ELSz represents the ELS data set with the neighborhood specification of ZIP code. Column 4 (8) report the average and standard error of the average of the NELS and ELSz estimates in columns 1 and 3 (5 and 7). The parameter vectors $\boldsymbol{B}$, $\boldsymbol{G}_1^N$, $\boldsymbol{G}_1^S$, $\boldsymbol{G}_2^S$, and $\boldsymbol{G}_2^C$ that define the explanatory index variables $\boldsymbol{X}_i\boldsymbol{B}$, $\boldsymbol{X}_{1n}\boldsymbol{G}_1^N$, $\boldsymbol{Z}_{2s}^S\boldsymbol{G}_2^S$ and $\boldsymbol{Z}_{2c}^C\boldsymbol{G}_2^C$ are estimated in a first step. The first 4 rows report the estimated standard deviations of these indices. The model includes the "full" set of $\boldsymbol{X}_i$ variables (student level) and the corresponding "full" set of $\boldsymbol{X}_s$ variables (school means). See Appendix A1 for a list of $\boldsymbol{X}_i$, $\boldsymbol{X}_{1n}$, $\boldsymbol{X}_s$, $\boldsymbol{Z}_{2s}^S$, and $\boldsymbol{Z}_{2c}^C$ variables and Appendix table B2 for summary statistics. The index variables in the interaction terms are standardized to be mean 0 and sd 1. Standard deviation of v_i is 1 in probit specifications. Names of interaction coefficients are next to the variables. See Section 6.2 for details about the estimation and bootstrap standard error procedures.

* $p < 0.1$
** $p < 0.05$
*** $p < 0.01$

TABLE 5

SAMPLE AVERAGES OF THE TREATMENT EFFECTS ON EDUCATION AND
WAGES OF A 10TH TO 90TH PERCENTILE SHIFT IN SCHOOL QUALITY AND
COMMUTING ZONE QUALITY (FULL SAMPLE)

	A. HS Grad			B. College Enrollment		
	NELS	ELSbg	ELSz	NELS	ELSbg	ELSz
School+CZ	0.088	0.062	0.060	0.171	0.173	0.183
	(.014)	(.011)	(.011)	(.018)	(.019)	(.019)
School only	0.060	0.045	0.043	0.125	0.120	0.133
	(.014)	(.01)	(.011)	(.017)	(.019)	(.019)
CZ only	0.067	0.041	0.037	0.114	0.112	0.118
	(.017)	(.009)	(.01)	(.017)	(.022)	(.021)
	C. College Grad			D. Log Wage		
	NELS	ELSbg	ELSz	NELS	ELSbg	ELSz
School+CZ	0.085	0.090	0.083	0.111	0.104	0.103
	(.016)	(.012)	(.015)	(.025)	(.024)	(.027)
School only	0.077	0.071	0.068	0.069	0.060	0.069
	(.018)	(.015)	(.016)	(.027)	(.023)	(.024)
CZ only	0.046	0.073	0.060	0.102	0.078	0.075
	(.019)	(.017)	(.016)	(.024)	(.024)	(.026)

NOTE.—The row School+Cz reports the average effect of moving students from a school and commuting zone at the 10th percentile of the distribution of the sum of school and commuting zone quality to the 90th percentile value. The row School only (CZ only) reports the average effect of a shift from the 10th percentile of school (commuting zone) quality to the 90th. The estimates are based on the model estimates in Table 4, which are for the full set of X_i variables. See Section 6.4 and Online Appendix B3.3 for the details of the treatment effect calculations. Column heading indicate the outcome, the data set, and whether the neighborhood definition is block group or ZIP code. ELSbg represents the ELS data set with the neighborhood specification of block group. ELSz represents the ELS data set with the neighborhood specification of ZIP code. Bootstrap standard errors are in parentheses.

TABLE 6

THE TREATMENT EFFECTS ON EDUCATION AND WAGES OF A 10TH TO 90TH PERCENTILE SHIFT IN SCHOOL QUALITY AND COMMUTING ZONE QUALITY, BY POPULATION SUBGROUP (FULL SAMPLE)

	A. HS Grad					B. College Enrollment				
	white	black	Hispanic	sg mother, hs deg	both par wh, col deg	white	black	Hispanic	sg mother, hs deg	both par wh, col deg
i. NELS / ZIP Code:										
Sch+CZ	0.086	0.095	0.113	0.140	0.040	0.177	0.167	0.138	0.116	0.213
	(.013)	(.017)	(.019)	(.021)	(.007)	(.018)	(.018)	(.017)	(.014)	(.02)
Sch only	0.058	0.066	0.077	0.093	0.027	0.129	0.121	0.101	0.086	0.154
	(.014)	(.016)	(.02)	(.022)	(.007)	(.017)	(.018)	(.015)	(.013)	(.018)
CZ only	0.066	0.071	0.085	0.106	0.031	0.118	0.110	0.091	0.075	0.142
	(.017)	(.019)	(.021)	(.027)	(.007)	(.017)	(.017)	(.016)	(.012)	(.02)
ii. ELS / Block Group:										
Sch+CZ	0.053	0.076	0.087	0.094	0.021	0.183	0.166	0.145	0.158	0.184
	(.01)	(.015)	(.015)	(.015)	(.004)	(.02)	(.019)	(.017)	(.021)	(.019)
Sch only	0.039	0.056	0.062	0.067	0.016	0.127	0.115	0.101	0.112	0.127
	(.009)	(.013)	(.014)	(.014)	(.004)	(.02)	(.019)	(.017)	(.021)	(.018)
CZ only	0.035	0.050	0.058	0.062	0.014	0.118	0.106	0.093	0.103	0.120
	(.008)	(.013)	(.013)	(.011)	(.003)	(.023)	(.021)	(.019)	(.022)	(.021)
iii. ELS / ZIP Code:										
Sch+CZ	0.051	0.074	0.083	0.089	0.020	0.193	0.177	0.153	0.167	0.194
	(.01)	(.015)	(.015)	(.015)	(.004)	(.02)	(.019)	(.017)	(.02)	(.019)
Sch only	0.043	0.037	0.054	0.060	0.065	0.140	0.129	0.111	0.120	0.139
	(.01)	(.014)	(.015)	(.015)	(.004)	(.02)	(.019)	(.017)	(.02)	(.017)
CZ only	0.031	0.045	0.052	0.055	0.013	0.125	0.112	0.097	0.106	0.126
	(.009)	(.014)	(.015)	(.013)	(.004)	(.022)	(.02)	(.018)	(.022)	(.021)

	C. College Grad					D. Log Wage				
	white	black	Hispanic	sg mother, hs deg	both par wh, col deg	white	black	Hispanic	sg mother, hs deg	both par wh, col deg
i. NELS / ZIP Code:										
Sch+CZ	0.090	0.074	0.067	0.056	0.101	0.111	0.112	0.111	0.111	0.110
	(.017)	(.015)	(.014)	(.011)	(.018)	(.025)	(.026)	(.025)	(.026)	(.025)
Sch only	0.081	0.067	0.061	0.051	0.090	0.069	0.068	0.069	0.067	0.069
	(.018)	(.016)	(.015)	(.012)	(.019)	(.027)	(.027)	(.027)	(.027)	(.027)
CZ only	0.049	0.039	0.036	0.029	0.056	0.102	0.104	0.102	0.104	0.101
	(.02)	(.017)	(.015)	(.012)	(.023)	(.024)	(.025)	(.024)	(.025)	(.023)
ii. ELS / Block Group:										
Sch+CZ	0.097	0.077	0.072	0.082	0.102	0.104	0.105	0.105	0.101	0.104
	(.013)	(.011)	(.011)	(.013)	(.013)	(.024)	(.024)	(.024)	(.024)	(.024)
Sch only	0.076	0.061	0.058	0.063	0.078	0.060	0.060	0.060	0.061	0.061
	(.016)	(.014)	(.013)	(.016)	(.015)	(.023)	(.023)	(.023)	(.023)	(.023)
CZ only	0.079	0.062	0.057	0.065	0.083	0.078	0.078	0.078	0.079	0.078
	(.018)	(.014)	(.013)	(.017)	(.018)	(.024)	(.024)	(.024)	(.024)	(.024)
iii. ELS / ZIP Code:										
Sch+CZ	0.090	0.072	0.067	0.074	0.094	0.103	0.102	0.103	0.103	0.103
	(.016)	(.013)	(.013)	(.016)	(.015)	(.027)	(.027)	(.027)	(.027)	(.027)
Sch only	0.073	0.059	0.055	0.061	0.076	0.069	0.069	0.068	0.067	0.069
	(.017)	(.015)	(.014)	(.017)	(.015)	(.024)	(.024)	(.024)	(.024)	(.024)
CZ only	0.065	0.051	0.048	0.055	0.069	0.075	0.076	0.076	0.075	0.076
	(.017)	(.014)	(.013)	(.017)	(.017)	(.026)	(.026)	(.026)	(.026)	(.026)

NOTE.—The row Sch+CZ reports the average effect of moving students from a school and commuting zone at the 10th percentile of the distribution of the sum of school and commuting zone

quality to the 90th percentile value. The row Sch Only (CZ Only) reports the average effect of a shift from the 10th percentile of school (commuting zone) quality to the 90th. The estimates are based on the model estimates in Table 4, which are for the full set of X_i variables. See Section 6.4 and Online Appendix B3.3 for the details of the treatment effect calculations. The panel headings indicate the outcome, the data set and whether the neighborhood definition is block group or ZIP code. The column heading identify the subgroup. "white" are white non-Hispanic students. "black" and "Hispanic" are non-Hispanic black and Hispanic students. "sg mother, hs deg" are students with a single mother who has a high school degree or less. "both par wh, col deg" are white students with two resident parents with college degrees. Bootstrap standard errors are in parentheses.
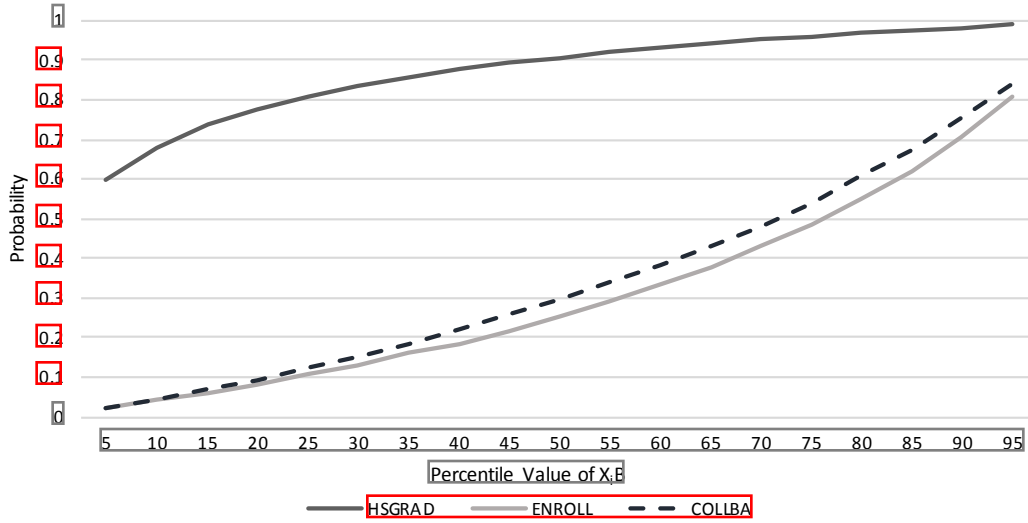
FIG. 1A.—The probability of high school graduation, college enrollment, and BA degree attainment by percentile of X_iB at the median value of school and commuting zone quality in NELS. The figure displays the average predicted value for the NELS grade 8 sample from the probit function evaluated at the indicated quantile of the X_iB index with school and commuting zone contributions set to the median value. The average is taken over neighborhood and unobserved individual contributions, as well as the control function, which represents a mix of school contributions and school-means of individual contributions. See Section 6 for details about the estimating equation.
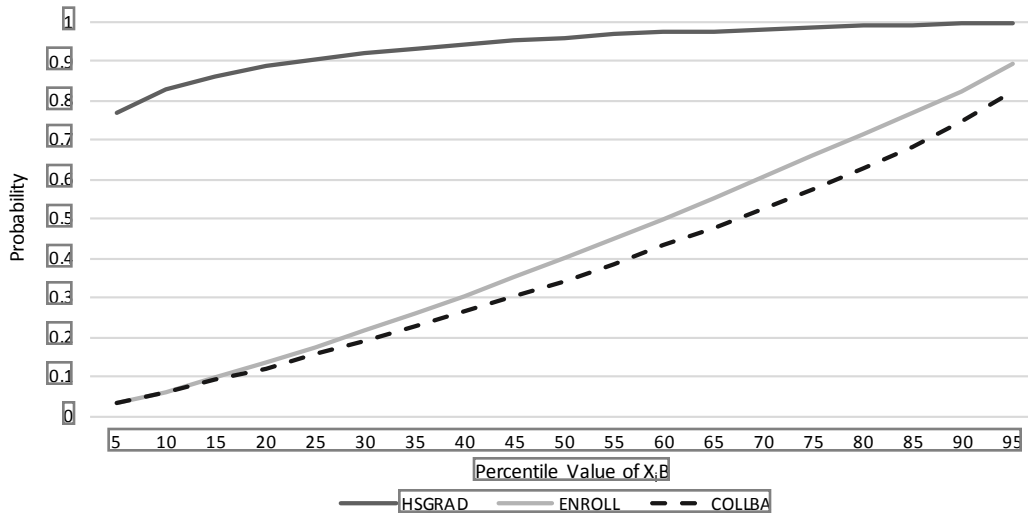

FIG. 1B.—The probability of high school graduation, college enrollment, and BA degree attainment by percentile of X_iB at the median value of school and commuting zone quality in ELS. The figure displays the average predicted value for the ELS block group sample from the probit function evaluated at the indicated quantile of the X_iB index with school and commuting zone contributions set to the median value. The average is taken over neighborhood and unobserved individual contributions, as well as the control function, which represents a mix of school contributions and school-means of individual contributions. See Section 6 for details about the estimating equation.
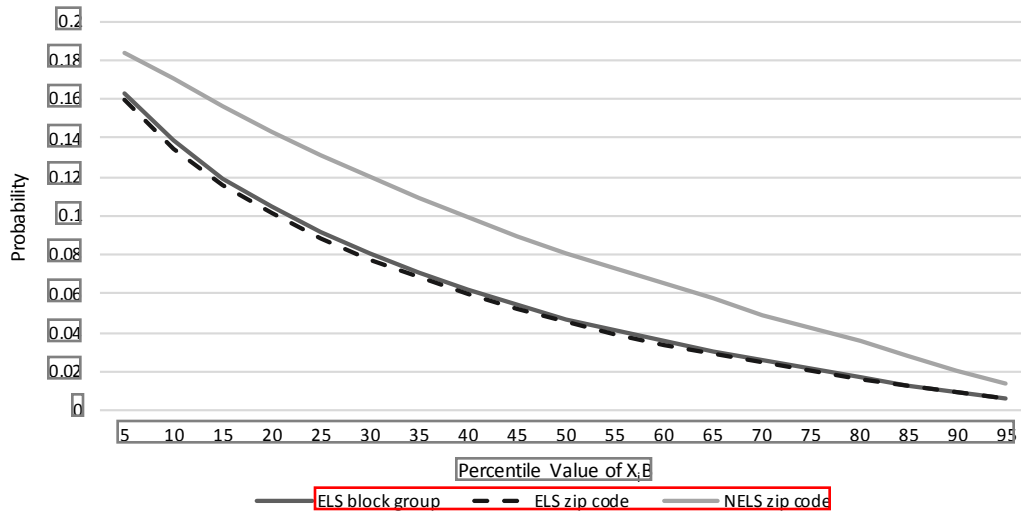
61

FIG. 2.—Treatment effect on high school graduation of a 10th to 90th percentile shift in school quality + commuting zone quality, by percentile of X_iB index. The figure displays the average predicted change in the probability of high school graduation for a student at the indicated quantile of the X_iB index from growing up in combined school/commuting zone environment at the 90th quantile of quality versus the 10th quantile of combined school/commuting zone quality. See Section 6.4 for details about how these values are calculated. The three lines display average predicted changes in probability for students in the NELS grade 8 sample, the ELS sample with ZIP codes designating neighborhoods, and the ELS sample with block groups designating neighborhoods.
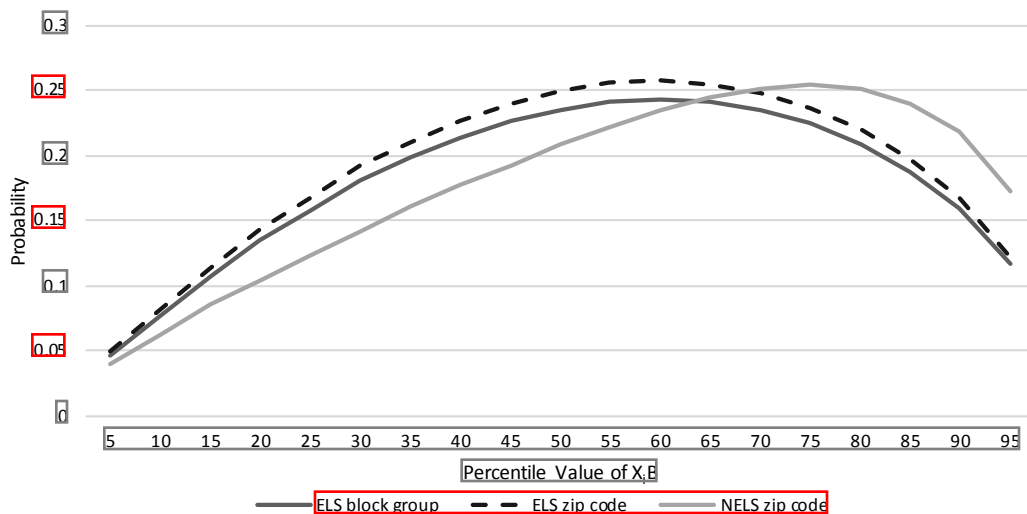
FIG. 3.—Treatment effect on college enrollment of a 10th to 90th percentile shift in school quality + commuting zone quality, by percentile of X_iB index. The figure displays the average predicted change in the probability of enrollment at a four-year college for a student at the indicated quantile of the X_iB index from growing up in combined school/commuting zone environment at the 90th quantile of quality versus the 10th quantile of combined school/commuting zone quality. See Section 6.4 for details about how these values are calculated. The three lines display average predicted changes in probability for students in the NELS grade 8 sample, the ELS sample with ZIP codes designating neighborhoods, and the ELS sample with block groups designating neighborhoods.
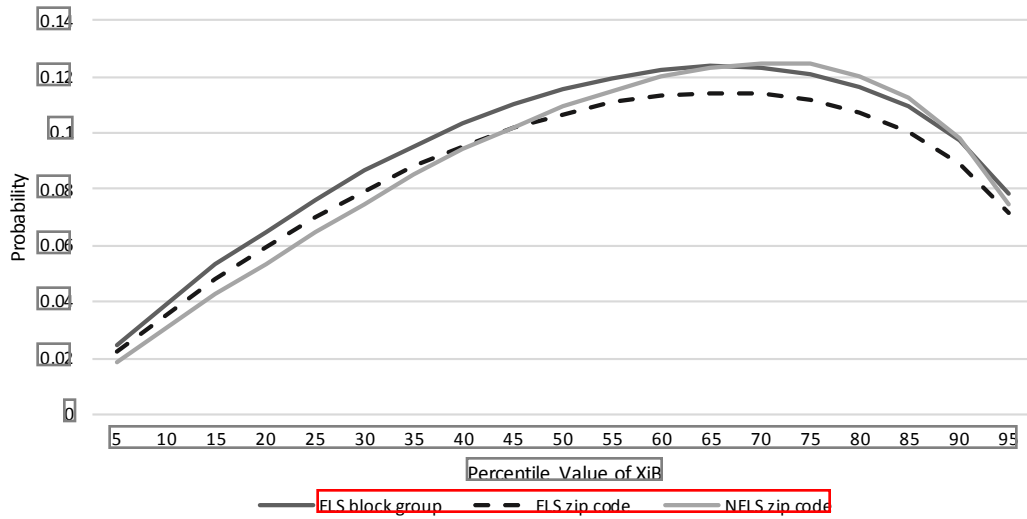
FIG. 4.—Treatment effect on college BA of a 10th to 90th percentile shift in school quality + commuting zone quality, by percentile of X_iB index. The figure display the average predicted change in the probability of graduation from a four-year college for a student at the indicated quantile of the X_iB index from growing up in combined school/commuting zone environment at the 90th quantile of quality versus the 10th quantile of combined school/commuting zone quality. See Section 6.4 for details about how these values are calculated. The three lines display average predicted changes in probability for students in the NELS grade 8 sample, the ELS sample with ZIP codes designating neighborhoods, and the ELS sample with block groups designating neighborhoods.

Appendices

A1   List of Variables

Student Characteristics (included in X_i): 1(Female), 1(Non-Hispanic black), 1(Hispanic), 1(Asian), and 1(White) (excluded) and 1(Other Race) (ELS only), 1(Immigrant), 1(Native English Speaker) and 1(Athletic). In addition, the full model includes Math score, Reading score, Weekly homework hours, 1(Parent checks homework), Hours/per week spent reading, 1(Attends class without a pencil) and 1(Fought in school).

Parent and Family Characteristics (included in X_i): SES Index, Number of siblings, 1(Does not live with both mother and father), Father's years education, Mother's years education, 1(Mother's education missing), Ln(family income), 1(Mother or father is an immigrant), religion dummies (1(Catholic), 1(Other Christian), 1(Other Religion), 1(Religion Missing), where 1(Protestant) is excluded); mother's occupation dummies (1(Clerical), 1(Manager/accountant/nurse/business owner/teacher), 1(Sales or service), 1(Other or homemaker), 1(Missing)), father's occupation dummies (1(Accountant/ nurse/teacher/manager/dentist/lawyer/business owner), 1(Military/security/craftsman/technician), 1(Service/ clerical/sales/missing/other/homemaker), 1(Farmer, laborer, operative)), Home environment index (first principal component), Parental school engagement index (first principal component), Parent's years of education desired for child.

Neighborhood Characteristics (included in X_n variables): NELS and ELS ZIP code specification: % Black, % Hispanic, % Non-married households, % Foreign born, % Some college or associate degree, % Four-year college degree or higher, Log(median income), Gini coefficient (ELS only), % SSI or welfare recipients, Log(median house value) (ELS only), and % Housing properties occupied (ELS only). ELS block group specification: Proportion of jobs in agriculture, mining, oil, utility, construction, or manufacturing, Proportion of jobs in information, finance, insurance, real estate, professional, or science, Proportion of jobs in management, administration, waste management, Proportion of jobs in education, other services and public administration, and Proportion of jobs in transportation and warehousing. They also include % Black, % Hispanic, % Other, % Non-married households, % Foreign born, % Some college or associate degree, % Four-year college degree or higher, Log(median income), Gini coefficient, % SSI or welfare recipients, Log(median house value) and % Housing properties occupied.

School Characteristics (included in X_s): % Minority, % Limited English proficient, % Free/reduced lunch. The full specification also includes % Special ed, % Remedial reading and % Remedial math.

School Characteristics (included in Z_{2s}^G): Variables common to both ELS and NELS: 1(Catholic school), 1(Private school), School enrollment, Student-teacher ratio, % Minority teachers, and school security policies indices (first and second principal components).

In addition, the ELS specification also includes Teacher turnover rate, % Teachers with certification, Teacher evaluation policy index (first principal component), Teacher incentive pay index (first principal component), Teacher technology access index (first principal component), School physical environment index (first principal component), administrator's reported school facilities index (first principal component), 1(Rural, not in MSA), 1(Rural, in MSA), 1(town), 1(Suburb of medium city), 1(Medium city), (Large city), and School administrator's assessment of crime in the school neighborhood.

In addition, the NELS specification includes % Teachers with a masters degree, 1(Collective bargaining), Log(minimum teacher salary), 1(Urban), 1(Suburban), 1(Gifted program), Junior High (JH) school minutes per year, JH course assignment index (first principal component), JH student retainment policy index (first principal component), JH student activities index (first principal component), JH school environment index (first principal component), JH movement in school policy index (first principal component), JH counseling policy index, JH school uniform policy index, JH student problems index (first principal component), and JH student punishment policy index (first principal component).

Commuting Zone Characteristics (included in Z_{2c}^C): Household income per capita, Theil racial segregation index, Log population density, % Black, Income segregation index, Social capital index, Poverty rate, Unemployment rate, Fraction of Children with Single Mothers, Gini coefficient, HS dropout rate (income adjusted), College graduate rate (income adjusted), Number of Colleges per Capita, Chetty-Hendren estimate of commuting zone causal effect on college attendance from age 18-23, Chetty-Hendren commuting zone causal effect on rank in national income distribution at age 26, 1(Northeast region), 1(Midwest region) and 1(West region).