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

By JOSEPH G. ALTONJI AND RICHARD K. MANSFIELD

We consider the classic problem of estimating group treatment effects when individuals sort based on observed and unobserved characteristics. Using a standard choice model, we show that controlling for group averages of observed individual characteristics potentially absorbs all *the across-group variation in* unobservable *individual characteristics. We use this insight to bound the treatment effect variance of school systems and associated neighborhoods for various outcomes. Across multiple datasets, we find that a 90th versus 10th percentile school/neighborhood increases the high school graduation probability and college enrollment probability by at least 0.04 and 0.11 and permanent wages by 13.7%.*

Supplementary Material: For Online Publication Only

Table of Contents:

Appendices:

Tables:

A2. Spanning Condition Examples

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Since $\tilde{\mathbf{\Theta}}$ is now rank-deficient, there is no matrix $\mathbf{R}$ such that $\mathbf{R}\tilde{\mathbf{\Theta}} = \mathbf{\Theta^U}$. In particular, for any matrix $\mathbf{R}$, each entry in column 2 be zero, but the second column of $\mathbf{\Theta^U}$ contains non-zero entries. Similarly, if both X_1 and X_2 affect WTP for A_1 and A_2 in the same proportion (and are each uncorrelated with $\mathbf{X^U}$, so that $\mathbf{\Pi_{X^UX}} = 0$, a rank-deficiency will also occur:

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

Here, an incremental unit of X_1 or X_2 will affect WTP for A_2 by twice as much as it will affect WTP for A_1. As in the previous example, there is no matrix $\mathbf{R}$ such that $\mathbf{R}\tilde{\mathbf{\Theta}} = \mathbf{\Theta^U}$. For any choice of $\mathbf{R}$, in each row of $\mathbf{R}\tilde{\mathbf{\Theta}}$ the second column will always be twice as large as the first column, but the second row of $\mathbf{\Theta^U}$ has a first column entry that is only half as large as its second column entry. Both these examples violate the spanning condition. If the row rank of $\tilde{\mathbf{\Theta}}$ is less than the row rank of $\mathbf{\Theta^U}$, then the row space of $\mathbf{\Theta^U}$ cannot possibly be a subspace of the row space of $\tilde{\mathbf{\Theta}}$.

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

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

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

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

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

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

There is no scalar R such that $R\tilde{\mathbf{\Theta}} = \mathbf{\Theta^U}$, since any value of R will preserve the one-to-one ratio between the first and second entries in $\mathbf{\Theta}$, while $\mathbf{\Theta^U}$ has a one-to-two ratio between its first and second entries. The spanning condition also fails in this case because the row space of $\mathbf{\Theta^U}$ is not a subspace of the row space of $\tilde{\mathbf{\Theta}}$. This example demonstrates that if the set of factors that individuals

consider when choosing groups is large, one will generally need an equally large set of observable characteristics in order to satisfy the spanning condition in Proposition 1.

Finally, suppose that both X_1 and X_1^U only affect willingness to pay for A_1 (Q may affect taste for A_2, so that A_2 is still relevant for school choice):

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

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

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

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

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

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

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

More generally, suppose Assumption 5.1 is nearly satisfied, so that a small number of amenity factors drive the vast majority of the variation in $\mathbf{X_s}$, but elements of $\mathbf{X_i}$ slightly influence tastes for several other amenities. Our simulations in section A6 suggest that such minor departures from the Assumptions 5.1 and 5.2 have little impact on the ability of $\mathbf{X_s}$ to effectively control for the unobservable between-school variation $\mathbf{X_s^U}$. But in such contexts, a small number of amenity factors should account for a very large fraction of the variation in $\mathbf{X_s}$, with only a very small amount of unexplained residual variation.

We test these predictions by performing principal components analysis (PCA) on $\mathbf{X_s}$. Because the sample school averages of observable characteristics $\hat{\mathbf{X}}_\mathbf{s}$ are noisy measures of the expected values $\mathbf{X_s} \equiv \mathbf{E}[\mathbf{X_i}|s(i) = s]$, we do not fit the PCA model to $\hat{\mathbf{X}}_\mathbf{s}$ directly. Instead, we estimate the underlying true covariance matrix $\mathbf{Var}(\mathbf{X_s})$,[1] and then directly perform the principal components analysis on the estimated covariance matrix.[2]

The results are in Online Appendix Table A5. Panel A reports, for each dataset we use, the number of principal components necessary to explain 75%, 90%, 95%, 99%, and 100% of the sum $\sum_{\ell=1}^{L} Var(X_{s\ell})$ of the variances of the standardized values of the L characteristics in $\mathbf{X_s}$, respectively. This is the standard output from a factor analysis. In Panel B, we also provide the number of principal components necessary to explain 75%, 90%, 95%, 99%, and 100% of the variance in $\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1}$, the regression index formed by using the estimated coefficients on school-level averages from our empirical analysis.

Both Panel A and Panel B provide strong evidence that $rank(\tilde{\mathbf{\Theta}}) \geq \dim(\mathbf{A_s^X})$, implying that Assumption 5.1 for the spanning condition $\mathbf{\Theta^U} = \mathbf{R}\tilde{\mathbf{\Theta}}$ is satisfied in the datasets we use. Specifically, in each dataset, $\mathbf{Var}(\mathbf{X_s})$ is found to be rank deficient. For example, in the full specification using ELS2002, only 33 latent factors are needed to explain all of the variance in $\mathbf{X_s}$ (Panel A, Row 6, Column 6), compared to $L = 51$ elements of $\mathbf{X_s}$. Similarly, in the NELS88 full specification, only 32 factors fully explain the variance in the 49 factors of $\mathbf{X_s}$.

Furthermore, the PCA analysis also suggests that a much smaller number of factors can account for the vast majority of the variation in either $\sum_{\ell=1}^{L} Var(X_{s\ell})$ or $Var(\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1})$. In the ELS2002 full specification, only 19 and 15 factors are needed to explain 95% of the variation in $\sum_{\ell=1}^{L} Var(X_{s\ell})$ and $Var(\mathbf{X_s}\hat{\mathbf{G}}_\mathbf{1})$, respectively (Panels A and B, Row 4, Column 6). For NELS88, only 20 and 13 factors are needed to explain 95% of the variation in the corresponding two measures (Panels A and B, Row 4, Column 4). The number of latent factors required to explain a given percentage of the sum of the variances of the elements of $\mathbf{X_s}$ is larger in the full specification, which contains more variables. This would be expected in the presence of sampling error in $\widehat{\mathbf{Var}}(\mathbf{X_s})$. However, it might also indicate that there are in fact additional amenity factors that play a small role in driving sorting (and thus have small eigenvalues) that can be picked up by the additional elements of $\mathbf{X_s}$ in the full specification.

Note, though, that because we only observe small samples of students in each school in our panel surveys and only have a sample of schools, the covariance matrix $\widehat{\mathbf{Var}}(\mathbf{X_s})$ that is decomposed by PCA is merely a consistent estimate of the population covariance matrix $\mathbf{Var}(\mathbf{X_s})$, and thus contains sampling error. The assumption underlying the spanning condition pertains to the rank of the population matrix $\mathbf{Var}(\mathbf{X_s})$. We address this issue in two ways. First, Panel A and B of Online

---

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

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

Appendix Table A5 report 90% bootstrap confidence interval estimates of the number needed to explain the specified percentages of $\sum_l^L Var(X_{sl})$ and $Var(\mathbf{X_s}\hat{\mathbf{G}_1})$. They are fairly tight.

Second, we also implement the formal test of rank proposed by Kleibergen and Paap (2006). Building on Cragg and Donald (1997) and Robin and Smith (2000), this test exploits the fact that a rank deficient matrix will have a subset of its singular values equal to 0, and tests whether the smallest singular values are farther from zero than one would expect based on sampling error.[3] The test compares the null hypothesis that $rank(\mathbf{Var}(\mathbf{X_s})) = q$, for some $q < L$, against the alternative that $rank(\mathbf{Var}(\mathbf{X_s})) > q$. Thus, Table A6 report the p-value from this test for each possible rank $1, \ldots, L-1$ for each of our panel survey datasets for our baseline specification. Table A7 displays the corresponding p-values across datasets for our full specification.

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

Across tests and datasets, the results are broadly quite consistent with the PCA results reported above. In particular, not only do the tests consistently fail to reject rank values well below the number of observables, but in fact the p-values generally converge to values indistinguishable from 1 as the numbers of factors being tested nears the number of principal components identified in Table A5. In sum, the Kleibergen/Paap tests provide no evidence against the null hypothesis that the number of factors that drive sorting on the observables $\mathbf{X_i}$ is substantially small than dimension of $\mathbf{X_i}$.

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

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

$$(39) \qquad\qquad \mathbf{E}[\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}|\mathbf{W_i}] = \mathbf{E}^*[\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}|\mathbf{W_i}] + e_i^{\tilde{X}^U}$$

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

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

Proposition 1A: *Assume that Assumptions A1, A2, A3, and A5 hold.*

---

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

Then the expectation $\mathbf{X_s^U}$ *is*

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

(41)
$$-\mathbf{E}[e_i^X | s(i) = s][\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})] + \mathbf{E}[e_i^{\tilde{X}^U} | s_i = s]$$

A.   Proof of Proposition 1A:

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

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

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

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

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

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

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

(44)
$$= \mathbf{E}^*[\mathbf{X_i} | \mathbf{W_i}]\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}).$$

Using

(45)
$$\mathbf{E}[\mathbf{X_i^U} | \mathbf{W_i}] = \mathbf{E}[\mathbf{X_i} | \mathbf{W_i}]\boldsymbol{\Pi_{X^U X}} + \mathbf{E}[\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U} | \mathbf{W_i}].$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- (A5'): There exists an $L^{U1} \times L$ matrix $\mathbf{R_1}$ such that $\boldsymbol{\Theta}_{11}^{\mathbf{U}} = \mathbf{R_1}\tilde{\boldsymbol{\Theta}}_1$.

- (A5"): There exists an $L^{U2} \times L$ matrix $\mathbf{R_2}$ such that $\boldsymbol{\Theta}_{21}^{\mathbf{U}} = \mathbf{R_2}\tilde{\boldsymbol{\Theta}}_1$.

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

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

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

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

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

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

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

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

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

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

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

$$= \left\{ \begin{array}{cc} \tilde{\boldsymbol{\Theta}}_1'\mathbf{R_1}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{1i}}^{\mathbf{U}}) + \tilde{\boldsymbol{\Theta}}_1'\mathbf{R_2}'\mathbf{Cov}(\tilde{\mathbf{X}}_{\mathbf{2i}}^{\mathbf{U}'},\tilde{\mathbf{X}}_{\mathbf{1i}}^{\mathbf{U}}) & \tilde{\boldsymbol{\Theta}}_1'\mathbf{R_1}'\mathbf{Cov}(\tilde{\mathbf{X}}_{\mathbf{1i}}^{\mathbf{U}'},\tilde{\mathbf{X}}_{\mathbf{2i}}^{\mathbf{U}}) + \tilde{\boldsymbol{\Theta}}_1'\mathbf{R_2}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{2i}}^{\mathbf{U}}) \\ \boldsymbol{\Theta}_{22}^{\mathbf{U}'}\mathbf{Cov}(\tilde{\mathbf{X}}_{\mathbf{2i}}^{\mathbf{U}'},\tilde{\mathbf{X}}_{\mathbf{1i}}^{\mathbf{U}}) & \boldsymbol{\Theta}_{22}^{\mathbf{U}'}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{2i}}^{\mathbf{U}}) \end{array} \right\}$$

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

Similarly, we have:

(49) $$\mathbf{Cov}(\mathbf{W_i}',\mathbf{X_i}) = \mathbf{Cov}(\tilde{\boldsymbol{\Theta}}'\mathbf{X}_{\mathbf{i}}',\mathbf{X_i}) = \tilde{\boldsymbol{\Theta}}'\mathbf{Var}(\mathbf{X_i}) =$$

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

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

(51)

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

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

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

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

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

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

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

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

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

However, we can gain additional insight by re-considering the continuous version of the model analyzed in Altonji and Mansfield (2014). In that context we assumed a continuum of schools and therefore a continuous joint distribution of amenity vectors. In Appendix A3 of Altonji and Mansfield (2014), we solve for an explicit unique equilibrium mapping between $\mathbf{A_s}$ and $\mathbf{W_s}$ under the assumptions that a) $[\mathbf{X_i}, \mathbf{X}_\mathbf{i}^\mathbf{U}, \mathbf{Q_i}]$ and $\mathbf{A}_{s(\mathbf{i})}$ are each jointly normally distributed (with variance matrices for W_i and $A_{s(i)}$ of $\mathbf{\Sigma_W}$ and $\mathbf{\Sigma_A}$ respectively), b) the ε_{is} are 0, and c) the equilibrium allocation takes a linear form: $\mathbf{A}_{s(\mathbf{i})} = \mathbf{\Psi}\mathbf{W_i}'$. The unique linear equilibrium mapping is

(56) $$\mathbf{\Psi} = \mathbf{\Sigma}_{\mathbf{W}'}^{-1/2}(\mathbf{\Sigma}_{\mathbf{W}'}^{1/2}\mathbf{\Sigma_A}\mathbf{\Sigma}_{\mathbf{W}'}^{1/2})\mathbf{\Sigma}_{\mathbf{W}'}^{-1/2} .$$

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

Since every positive definite matrix is invertible, we can also express the vector $\mathbf{W_i}$ for any indi-

vidual as a linear function of the amenity vector of their chosen school:

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

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

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

Substituting (58) into (61) leads to

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

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

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

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

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

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

(61)
$$\tilde{\mathbf{X}}_{\mathbf{s}}^{\mathbf{U}} = \mathbf{X_s}\{\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}}) + [\mathbf{X_s}\mathbf{\Theta} + \tilde{\mathbf{X}}_{\mathbf{1s}}^{\mathbf{U}}\mathbf{\Theta}_{\mathbf{1}}^{U} + \tilde{\mathbf{X}}_{\mathbf{2}}^{\mathbf{U}}\mathbf{\Theta}_{\mathbf{2}}^{U} + \mathbf{Q_s}\mathbf{\Theta}^{Q}]\mathbf{Var}(\mathbf{W_i})^{-1}\left\{\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_{\mathbf{22}}^{\mathbf{U\prime}} \end{array}\right\}\mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})$$

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

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

If we take iterated expectations of equations (47), (48), and (62) conditional on school $s(i)$ and

replace $\mathbf{W_s}$ with $(\mathbf{\Psi}^{-1}\mathbf{A_{s(i)}})'$, we obtain:

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

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

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

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

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

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

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

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

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

$$(69) \qquad \mathbf{A_s}^{\mathbf{X}'} = \mathbf{X_s}\mathbf{\Pi_{A^XX_s}} + \tilde{\mathbf{A}}_\mathbf{s}^{\mathbf{X}'} = \mathbf{X_s}\mathbf{\Pi_{A^XX_s}}$$

$$(70) \qquad \mathbf{A_s}^{\mathbf{U2}'} = \mathbf{X_s}\mathbf{\Pi_{A^{U2}X_s}} + \tilde{\mathbf{A}}_\mathbf{s}^{\mathbf{U2}'}$$

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

This implies:

$$\tilde{\mathbf{X}}_\mathbf{s}^\mathbf{U} = \mathbf{X_s}[\mathbf{Var}(\mathbf{X_i})^{-1}\mathbf{R}'\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}) + \tilde{\mathbf{\Theta}}'\mathbf{Var}(\mathbf{W_i})^{-1}\begin{Bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_{\mathbf{22}}^{\mathbf{U}'} \end{Bmatrix}\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U}) + \mathbf{M}]$$

$$(71)$$

$$+ \{\mathbf{0}, \tilde{\mathbf{A}}_\mathbf{s}^{\mathbf{U2}'}\}\mathbf{\Psi}^{-1}\mathbf{Var}(\mathbf{W_i})^{-1}[\mathbf{\Theta}^{\mathbf{U}'}\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})\mathbf{\Theta}^{\mathbf{U}} + \mathbf{\Theta}^{\mathbf{Q}'}\mathbf{Var}(\mathbf{Q_i})\mathbf{\Theta}^{\mathbf{Q}}]\mathbf{Var}(\mathbf{W_i})^{-1}\begin{Bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_{\mathbf{22}}^{\mathbf{U}'} \end{Bmatrix}\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})$$

where the matrix $\mathbf{M}$ is
$$(72)$$

$$\mathbf{M} = \{\mathbf{\Pi_{A^XX_s}}, \mathbf{\Pi_{A^{U2}X_s}}\}\mathbf{\Psi}^{-1}\mathbf{Var}(\mathbf{W_i})^{-1}[\mathbf{\Theta}^{\mathbf{U}'}\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})\mathbf{\Theta}^{\mathbf{U}} + \mathbf{\Theta}^{\mathbf{Q}'}\mathbf{Var}(\mathbf{Q_i})\mathbf{\Theta}^{\mathbf{Q}}]\mathbf{Var}(\mathbf{W_i})^{-1}\begin{Bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_{\mathbf{22}}^{\mathbf{U}'} \end{Bmatrix}\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})$$

While cumbersome, the second term in (71) provides an expression for the component of $\tilde{\mathbf{X}}_\mathbf{s}^\mathbf{U}$ that cannot be predicted by $\mathbf{X_s}$ (and thus may be a source of bias in our lower bound estimates of the variance in school/neighborhood treatment effects). The variance in this component depends on the following five factors: a) the full joint distribution of amenities (through $\boldsymbol{\Psi}$); b) the joint distribution of the WTP vector $\mathbf{W_i}$ (entering via the covariance matrix $\mathbf{Var}(\mathbf{W_i})$); c) the WTP co-efficient matrix $\boldsymbol{\Theta}^\mathbf{U}$ capturing the effect of $\mathbf{X}^U$ on willingness to pay for particular amenities; d) the joint distribution of the residual component of unobserved outcome-relevant student character-istics (entering via the covariance matrix $\mathbf{Var}(\tilde{\mathbf{X}}_\mathbf{i}^\mathbf{U})$; and e) the joint distribution of the unobserved outcome-irrelevant (but school choice-relevant) student characteristics weighted by effects on WTP (entering via $\boldsymbol{\Theta}^{\mathbf{Q}\prime}\mathbf{Var}(\mathbf{Q_i})\boldsymbol{\Theta}^\mathbf{Q}$).

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

A6. Monte Carlo Evidence on the Properties of the Control Function Estimator

This section describes a set of monte carlo simulations designed to explore the performance of our control function estimator across a number of key dimensions. We do not attempt to fully characterize the performance of our estimator.[4] Instead, our simulations center around a stylized test case that is calibrated to represent a plausible description of the school/neighborhood choice context. We focus on sensitivity to deviations among a set of key parameters designed to reveal the strengths and weaknesses of our approach. In the first set of simulations, we restrict attention to cases in which the conditions of Proposition 1 are satisfied in an infinite population, and consider the sensitivity of the performance of the control function approach in removing bias from sorting on unobservables to various parameters capturing the structure of tastes, amenities, school sizes, and survey sampling design. Then, in a second set of simulations, we fix the parameters considered in the first set of simulations at a set of baseline values, and examine the sensitivity of our approach to violations of the key spanning condition in Proposition 1 that vary in nature and degree. Section A6.A lays out the simulation methodology, while section A6.B presents and interprets the results.

A. Methodology

The stylized test case we consider is one in which:

1) The elements of $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$ are jointly normally distributed; the elements of $\mathbf{Q_i}$ are inde-pendent of each other and $[\mathbf{X_i}, \mathbf{X_i^U}]$, and each pair of characteristics in $[\mathbf{X_i}, \mathbf{X_i^U}]$ features a .25 correlation.[5]

---

[4]A full characterization is a daunting task given the large number of parameters that determine the full spatial equilibrium sorting of students to schools. The parameters include those characterizing the joint distribution of the individual characteristics affecting choice $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$, the joint distribution of the amenities $\mathbf{A_s}$, and the distribution of the idiosyncratic tastes ε_{si}. The parameters also include the $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}^\mathbf{U}$, and $\boldsymbol{\Theta}^\mathbf{Q}$ matrices that capture how observed and unobserved characteristics affect WTP.

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

2) The latent amenity vectors $\mathbf{A_s}$ are normally distributed with a .25 correlation between any pair of amenities across schools.

3) The matrices of taste parameters $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ represent draws from a multivariate normal distribution in which (a) $corr(\Theta_{k\ell}, \Theta_{jm}) \equiv \rho_\Theta$ if $j = k$ or $\ell = m$, and 0 otherwise, (b) $corr(\Theta_{k\ell}^U, \Theta_{jm}^U) = \rho_\Theta$ if $j = k$ or $\ell = m$, and 0 otherwise, and (c) $corr(\Theta_{k\ell}, \Theta_{jm}^U) = \rho_\Theta$ if $\ell = m$, and 0 otherwise.

4) The number of elements of $\mathbf{Q}_i$ is equal to the number of elements of $\mathbf{A}_s$. $\mathbf{\Theta}^Q$ is the identity matrix.

5) The variances of the elements of $\mathbf{A_s}$, $[\mathbf{X_i}, \mathbf{X_i^U}, \mathbf{Q_i}]$, and $\varepsilon_{i,s}$ (i.i.d. draws from a normal distribution) are chosen to create interclass correlations for $\mathbf{X_i}$ and $\mathbf{X_i^U}$ of between .1 and .25 across specifications. These values are in line with the range observed across the datasets used in the empirical analysis.

6) There are no school/neighborhood effects, so that $Y = \mathbf{X_i}\boldsymbol{\beta} + x_i^U$, where $x_i^U \equiv \mathbf{X_i^U}\boldsymbol{\beta}^U$. Consequently, our estimating equation also omits the school level controls $\mathbf{Z_{2s}}$ that are not averages of student characteristics. These simplifications allow us to focus attention exclusively on the extent to which a vector of group averages of observable individual characteristics can absorb between-school variation in the outcome contributions of unobservable individual characteristics.

7) All the observable and unobservable characteristics in $\mathbf{X_i}$ and $\mathbf{X_i^U}$ are equally important in determining the outcome, so that each characteristic features the same (unit) variance, $\beta_\ell = 1 \,\forall\, \ell$, and $\beta_\ell^U = 1 \,\forall\, \ell$.

Our test case implies considerable sorting into schools along many dimensions of school amenities and along many observable and unobservable dimensions of student quality. It represents a conservative case because one might expect that in reality a few key observable (and unobservable) individual level factors (e.g. parental income, education, and wealth) and a few key school/neighborhood amenities (e.g. ethnic composition, crime, principal quality) drive most of the systematic sorting of students to schools. Given restrictions 1-7, we complete the model by choosing particular sets of seven remaining parameters. The first parameter is the number of students per school. For simplicity, we impose that each school has capacity equal to a common student/school ratio.[6] The student/school ratio is denoted "# Stu" in Online Appendix Table A8. The second parameter is the total number of school/neighborhood combinations available (denoted "# Sch").

The parameter #Con is the number of schools in the consideration set for each household. This captures the possibility that most parents only realistically consider a limited number of possible locations. We implement this by distributing schools uniformly throughout the unit square, and drawing a random latitude/longitude combination for each household. The households then con-

---

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

sider the preset number of schools that are closest to their location. Thus, consideration sets of different households are overlapping.

The fourth and fifth parameters (denoted "# Ob." and "# Un.") specify the number of observed and unobserved student characteristics that affect outcomes. The sixth parameter is the dimension of the amenity vector over which households have preferences (#Am). In most of the specifications we assume that it is less then or equal to the number of observed characteristics and that the rows of $\mathbf{\Theta^U}$ form a linear subspace of the rows of $\tilde{\mathbf{\Theta}}$, as required by Proposition 1.

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

In addition, in a second set of simulations we hold fixed these seven parameters at their baseline values, and consider additional specifications that illustrate the degree to which our control function approach is robust to various failures of the spanning condition from Proposition 1 (i.e. cases in which $\mathbf{\Theta^U} \neq \mathbf{R}\tilde{\mathbf{\Theta}}$ for any $\mathbf{R}$). These simulations consider robustness of the control function approach to changes in the structure of the three matrices that determine whether a one-to-one mapping from a vector of group-average unobservables to a vector of group-average observables exists at the population level: (1) the projection matrix $\mathbf{\Pi_{X^UX}}$, which captures the degree to which individual-level unobservables project onto the space of individual-level observables, (2) the taste matrix $\mathbf{\Theta}$, which captures the degree to which each of the student-level observables affects tastes for each of the school/neighborhood amenities, and (3) the corresponding taste matrix for unobservable student characteristics, $\mathbf{\Theta^U}$.

We have two related metrics for evaluating the effectiveness of our control function approach. The first is the fraction of the between-group variance in the outcome contribution of unobservable individual-level characteristics $(Var(x_s^U) \equiv Var(\mathbf{X_s^U}\boldsymbol{\beta}^U))$ that can be predicted using group-averages of observable characteristics (after adjusting for the degrees of freedom absorbed by the vector of observables). This is the adjusted R^2 from a regression of the potential bias from unobservable sorting, x_s^U, on the vector $\mathbf{X_s}$. In cases where the conditions of Proposition 1 are satisfied, R_{adj}^2 should converge to 1 as the number of students per school gets large. However, the rate at which it does so is important for the efficacy of the control function approach.

The second metric is $[(1 - R_{adj}^2)Var(x_s^U)]/Var(Y_{si})$, which is the fraction of the total variance in the outcome Y_{si} attributable to the variance of the residual component of x_s^U not accounted for by $\mathbf{X_s}$. In Appendix tables A8 and A9, we denote our measures "Adj-R-sq" and "Resid" (short for "residual sorting variance fraction").

We present values of adjusted R-squared and the residual sorting variance fraction from specifications where the full population of students is used to calculate the school averages of observables

$\hat{\mathbf{X}}_\mathbf{s}$ that compose the control function (denoted "Adj-R-sq (All)" and "Resid (All)", respectively), as well as values from specifications in which random samples of 10, 20, or 40 students from each school are used to calculate $\hat{\mathbf{X}}_\mathbf{s}$ (these values are denoted "Adj-R-sq (10/20/40)" and "Resid (10/20/40)", respectively, in our tables).

We draw $\mathbf{X_i}$, $\mathbf{X_i^U}$, $\mathbf{Q_i}$, and $\{\varepsilon_{is}\}$ from the distributions described above to calculate the WTP of each household for each school.[7] Since our method does not require observation of the equilibrium price function $P(\mathbf{A})$, rather than iterating on an excess demand function to find the equilibrium matching, we instead exploit the fact that a perfectly competitive market will always lead to a pareto efficient allocation. The problem of allocating students to schools to maximize total consumer surplus can be written as a linear programming problem and solved quickly at relatively large scale using the simplex method combined with sparse matrix techniques.[8]

B.   Simulation Results

The simulation results are presented in online Appendix Table A8. Row (1) presents the base parameter set to which other parameter sets will be compared. It features 1000 students per school and 50 schools in the area, all of which are considered by each family when the school choice is made. It also features 5 amenities, 10 observable student characteristics, and 10 unobservable student characteristics. The variances of these characteristics are all identical, so that sorting on unobservables is as strong as sorting on observables. This is probably a conservative choice. Finally, the within-row and within-column correlation ρ_Θ among the elements of the random matrices from which the taste weight matrices $\mathbf{\Theta}$ and $\mathbf{\Theta^U}$ are drawn is assumed to be .25.

The first takeaway from Row (1) is that the control function approach is extremely effective even with reasonably-sized schools of 1000 students each (most of the schools in the North Carolina sample enroll between 250 and 2000 students) and a moderate number of available schools: 99.8 percent of the variance in the school-level contribution of unobserved student characteristics can be predicted by a linear combination of school-average observable characteristics (Column 9). Furthermore, the control function only leaves three hundredths of a percent of the variance in the outcome Y_{si} that can be attributed to residual between-school sorting (Column 10).

The second insight from Row (1) is that the performance of the control function may suffer somewhat when estimation is based on small subsamples of students at each school. We see that the adjusted R-squared falls from .998 to .869 when school averages are merely approximated based on samples of 10 students (top entry in Column (11)). Increasing the sample size to 20 students per school (middle entry in Column (11)) raises the adjusted R-squared to .926, while increasing it further to 40 students per school (bottom entry in Column (11)) raises the adjusted R-squared to .959. Column (12) shows that the fraction of the outcome variance consisting of residual between-school

---

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

[8]The problem can actually be classified as a binary assignment problem (a subset of linear programming problems), but we were unable to implement the standard binary assignment algorithms at scale.

sorting unabsorbed by the control function is .016/.009/.005 when 10/20/40 student samples, respectively, are used to construct the vector of school averages, $\mathbf{X_s}$.

Rows (2) and (3) illustrate the impact of adapting the specification in Row (1) by decreasing or increasing the number of individuals per group. Decreasing school sizes from 1000 to 500 decreases the R-squared from .998 to .996, while increasing from 1000 to 2000 increases the R-squared to .999 (Column (9)). Perhaps not surprisingly, more individuals per school has almost no impact on the effectiveness of the control function if the larger number of individuals are not used to construct the group averages of individual characteristics, $\mathbf{X_s}$. In Columns (11) and (12), the adjusted R-squared values and residual sorting variance fraction when samples of 10, 20 and 40 students are used to construct $\mathbf{X_s}$ are nearly identical across Rows (1) - (3).

Comparing Row (4) to Row (1), we see that increasing the number of schools from 50 to 100 has almost no impact on the performance of the control function when the full population of students is used to construct school averages. Interestingly, reducing the number of schools does not exacerbate problems posed by using small samples of students from each school to construct $\mathbf{X_s}$ (Column (11)). Similarly, Row (5) shows that restricting the number of schools in each household's consideration set from 50 to 10 reduces the control function's ability to absorb unobservable sorting, but only negligibly. The adjusted R-squared is effectively unchanged when the full population of students is used to construct $\mathbf{X_s}$, but drops slightly from Row (1) to Row (5) when samples of 10, 20, or 40 students are used instead. Nonetheless, the high adjusted R-squared and low variance of the residual sorting component in Row (5) reveals that our approach works well even if households only consider a relatively small number of schools.

Row (6) illustrates the impact of doubling both the number of observable and unobservable outcome relevant characteristics. By increasing the numbers of both observable and unobservable characteristics symmetrically, we can show the impact of utilizing a richer control set while holding fixed the strength of sorting on observables relative to unobservables.[9] Doubling the number of elements of $\mathbf{X_i}$ and $\mathbf{X_i^U}$ increases the adjusted R-squared from .9979 in Row (1) to .9993, and decreases the fraction of outcome variance attributable to the residual sorting component to two hundredths of a percentage point. This very small increase understates the importance of the richness of the control set, since the control function was already nearly perfectly effective for the baseline parameter set. Column 11 shows that when only 10 students are used to construct sample school averages, doubling the control set from 10 to 20 characteristics increases the adjusted R-squared from .869 to .897. This highlights the importance of collecting data on a wide variety of student/parent inputs that capture different dimensions of taste (as the panel surveys we use do).

Row (7) shows that doubling the number of amenity factors from 5 to 10 very slightly reduces the effectiveness of the control function, dropping the adjusted R-squared from .9979 in Row (1)

---

[9]In all of these simulations, we assumed that the strength of sorting on unobservables mirrored the strength of sorting on unobservables. In results not shown, we also experimented with weakening the degree of sorting on unobservables by making $\mathbf{\Theta^U}$ smaller in magnitude and increasing the variance of $\mathbf{Q_i}$ to compensate. While the control function absorbs a slightly smaller *fraction* of the between-school variance of the regression index of unobservable outcome-relevant characteristics when sorting on these characteristics is weak, this is precisely the case when the *magnitude* of the between-school variance in outcome-relevant unobservables is small. Thus, there is very little potential bias to be absorbed.

to .9933. The impact of doubling the number of amenities is also small when small samples of students are used to construct school averages. Comparing Row (8) to Row (6) reveals that the performance of the control function really depends on the dimension of the amenity space *relative* to the dimension of $\mathbf{X_s}$, rather than the absolute number of amenities: when $\mathbf{X_s}$ has 20 elements, the fraction of absorbed sorting bias barely changes as the number of amenities rises from 5 to 10.

Finally, Row (9) displays the results of a specification in which all of the $\Theta_{k\ell}$ and $\Theta^U_{k\ell}$ elements are drawn independently ($\rho_\Theta = 0$). Compared to Row (1), the adjusted R-squared for the full population falls slightly (.9979 to .9941), but the adjusted R-squared when samples of (10/20/40) are used to construct $\mathbf{X_s}$ falls more substantially, from (.87/.93/.97) to (.65/.78/.87). However, removing correlation among the elements of $\boldsymbol{\Theta}$ also reduces the amount of sorting on unobservables to be explained. When the school averages of the various unobservables become more weakly correlated with one another, their contributions to student outcomes are more likely to cancel each other out. Consequently, the fraction of between-school outcome variation that can be attributed to residual school-level differences in unobservable student characteristics that is unpredictable based on the vector of school-average observables $\mathbf{X_s}$ remains quite small (Row 9, Col. 11).

Overall, the results in Online Appendix Table A8 indicate that the control function approach could potentially work extremely well even in settings where 1) individuals have idiosyncratic tastes for particular groups, 2) there are only a moderate number of total groups to join, and 3) only a subset of these are considered by any given individual.[10] The simulations suggest that the control function works well even when only a small sample of individuals is observed in each group. In Online Appendix A11, we use the North Carolina administrative data to directly assess the effect of using smaller samples of students to construct $\mathbf{X_s}$ for some of the outcomes and characteristics we actually consider. We find that our main results are relatively insensitive to restricting school sample sizes to match the distribution of sample sizes observed in the NLS72, NELS88, and ELS2002 datasets.

PERFORMANCE OF THE CONTROL FUNCTION WHEN THE SPANNING CONDITION FAILS

All the specifications in Online Appendix Table A8 consider cases in which the assumptions of Proposition 1 are satisfied, so we should expect the control function to perfectly absorb sorting on observables as the number of students per school gets sufficiently large. However, there also may be many contexts in which the set of observables is not sufficiently rich to make the spanning condition plausible. Thus, we are also interested in the extent to which the addition of group-averages of individual characteristics can substantially reduce bias from sorting on unobservables, even if it cannot completely eliminate the bias. Online Appendix Table A9 considers a number of such scenarios.

Recall from the discussion in Section II.A that $\tilde{\boldsymbol{\Theta}}$ can be represented as the sum of $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}_{X^U X} \boldsymbol{\Theta^U}$.

---

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

The dependence on Θ indicates that the mapping from $\mathbf{X_s^U}$ to $\mathbf{X_s}$ is generated partly because ob-served characteristics $\mathbf{X_i}$ and unobserved characteristics $\mathbf{X_i^U}$ directly affect WTP for overlapping sets of amenities (which creates a degree of overlap in the row spaces of Θ and $\Theta^{\mathbf{U}}$). The term $\mathbf{\Pi_{X^U X}\Theta^U}$ captures the part of the mapping that arises because $\mathbf{X_i}$ indirectly predicts WTP for the amenities for which $\mathbf{X_i^U}$ predicts WTP through the correlation between $\mathbf{X_i}$ and $\mathbf{X_i^U}$ (thereby creat-ing further overlap in the row spaces of Θ and $\Theta^{\mathbf{U}}$). The spanning condition ($\Theta^{\mathbf{U}} = \mathbf{R}\tilde{\Theta}$ for some $L^U \times L$ matrix $\mathbf{R}$) is satisfied whenever these two pathways, working in combination, produce a preference matrix $\tilde{\Theta}$ whose row space is a linear superspace of the row space of $\Theta^{\mathbf{U}}$.

Thus, before investigating the impact of violations of the spanning condition, we illustrate the im-portance of both pathways by considering specifications in which one or the other pathway is shut down. Row (1) is identical to Row (1) of Online Appendix Table A8, and represents the baseline case against which the other specifications are compared. Row (2) considers the case in which the entire vector of unobservable characteristics $\mathbf{X_i^U}$ is independent of the vector of observables $\mathbf{X_i}$, so that $\mathbf{\Pi_{X^U X}}$ converges to the zero matrix as school sizes become large. However, $\mathbf{X_i}$ and $\mathbf{X_i^U}$ predict tastes for a common set of amenities ($A_1 - A_5$), so that Θ has (full) rank K and the row space of $\Theta^{\mathbf{U}}$ is a linear subspace of the row space of Θ. The results in Row (2) suggest that the control function approach still works quite well when large populations of students at each school are available (adjusted R-squared of .965), but suffers somewhat when school averages are constructed using subsamples of 10, 20 or 40 students: adjusted R-squared values fall to .49/.61/.72 (Column 10), with substantial residual bias from sorting on unobservables left uncaptured by the control function $\hat{\mathbf{X}}_\mathbf{s}$ (Column 11).

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

The remaining rows of Online Appendix Table A9 consider cases in which the spanning condition fails (the row space of $\Theta^{\mathbf{U}}$ is not a linear subspace of the row space of $\tilde{\Theta} = \Theta + \mathbf{\Pi_{X^U X}\Theta^U}$). Row (4) presents results from the worst-case scenario: (a) the entire vector of unobservable characteristics is independent of the entire vector of observable characteristics ($\mathbf{\Pi_{X^U X}}$ converges to $\mathbf{0}$ as school sizes become large), and (b) the unobservable characteristics only predict WTP for an amenity (A_5) that the observable characteristics do not affect taste for (they exclusively weight $A_1 - A_4$). Thus, Θ and $\Theta^{\mathbf{U}}$ have orthogonal row spaces as well. Since the group averages of the observables and

unobservables are functions of disjoint sets of amenities, it comes as no surprise that only 15% of the variance in $\mathbf{X_s^U}$ is predictable given $\mathbf{X_s}$, even when the universe of students at each school is observed (Column 8).[11]

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

These two scenarios are quite pessimistic, however. If WTP for an amenity is unaffected by the entire vector $\mathbf{X_i}$, then it seems likely that a subset of the unobservables may not predict WTP for this amenity either. Thus, we consider two additional scenarios in which WTP for the last amenity (A_5) is only affected by one of the ten components of the unobserved vector $\mathbf{X_i^U}$. In Row (6), $X_{i,10}^U$ affects WTP for A_5 only. In Row (7), $X_{i,10}^U$ predicts willingness to pay for all amenities A_1 to A_5. Rows (6) and (7) reveal that our control function performs quite well in these scenarios: it absorbs around 95% of the variation in $\mathbf{X_s^U}$ in each case.

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

We conclude that our control function approach may be quite robust to the violations of the spanning condition that are arguably the most plausible: namely, cases in which just a few components of the subvector of $\mathbf{X_i^U}$ that is orthogonal to $\mathbf{X_i}$ affect WTP for just a few additional amenities for which $\mathbf{X_i}$ does not affect WTP.

A7. Proof of Proposition 2

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

$$(73) \qquad \Delta\mathbf{X_i^U} = \Delta\mathbf{X_i}\mathbf{\Pi_{\Delta X^U \Delta X}} + \widetilde{\Delta\mathbf{X_i^U}}$$

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

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

---

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

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

(74) $$\mathbf{B} = \boldsymbol{\beta} + \boldsymbol{\Pi}_{\Delta \mathbf{X}^U \Delta \mathbf{X}} \boldsymbol{\beta}^U + \boldsymbol{\Pi}_{\eta_{si}^U \mathbf{X_i}}$$

(75) $$\mathbf{G_1} = [-\boldsymbol{\Pi}_{\Delta \mathbf{X}^U \Delta \mathbf{X}} + \boldsymbol{\Pi}_{\mathbf{X}^U \mathbf{X}} + \mathbf{Var}(\mathbf{X_i})^{-1} \mathbf{R}' \mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})] \boldsymbol{\beta}^U + \boldsymbol{\Gamma}_1 + \boldsymbol{\Pi}_{z_s^U \mathbf{X_s}}$$

(76) $$\mathbf{G_2} = \boldsymbol{\Gamma}_2 + \boldsymbol{\Pi}_{z_s^U \mathbf{Z_{2s}}}$$

(77) $$v_s = \tilde{z}_s^U + \xi_s$$

(78) $$v_{si} - v_s = \tilde{x}_{si}^U + \tilde{\eta}_{si} + \xi_i$$

Proof: Recall the projection equation (9):

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

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

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

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

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

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

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

Now, consider the between-school regression (80). By Proposition 1, $\mathbf{X_s^U} = \mathbf{X_s}[\boldsymbol{\Pi}_{\mathbf{X}^U \mathbf{X}} + \mathbf{Var}(\mathbf{X_i})^{-1} \mathbf{R}' \mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})]$. Post-multiplying both sides by $\boldsymbol{\beta}^{\mathbf{U}}$, we obtain:

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

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

(82) $$\boldsymbol{\Pi}_4 = \mathbf{0}$$

(83) $$\boldsymbol{\Pi}_5 = [\boldsymbol{\Pi}_{\mathbf{X}^U \mathbf{X}} + \mathbf{Var}(\mathbf{X_i})^{-1} \mathbf{R}' \mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})] \boldsymbol{\beta}^{\mathbf{U}}$$

(84) $$\boldsymbol{\Pi}_6 = \mathbf{0}$$

(85) $$\tilde{x}_s^U = 0$$

Adding together $\mathbf{\Pi_1}$ and $\mathbf{\Pi_4}$, $\mathbf{\Pi_2}$ and $\mathbf{\Pi_5}$, and $\mathbf{\Pi_3}$ and $\mathbf{\Pi_6}$ yields:

$$(86) \qquad \mathbf{\Pi}_{x_i^U \mathbf{X_i}} = \mathbf{\Pi}_{\Delta \mathbf{X}^U \Delta \mathbf{X}} \boldsymbol{\beta}^U$$

$$(87) \qquad \mathbf{\Pi}_{x_i^U \mathbf{X_s}} = [-\mathbf{\Pi}_{\Delta \mathbf{X}^U \Delta \mathbf{X}} + \mathbf{\Pi}_{\mathbf{X}^U \mathbf{X}} + \mathbf{Var}(\mathbf{X_i})^{-1} \mathbf{R}' \mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})] \boldsymbol{\beta}^U$$

$$(88) \qquad \mathbf{\Pi}_{x_i^U \mathbf{Z_{2s}}} = \mathbf{0}$$

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

$$(89) \qquad \mathbf{B} = \boldsymbol{\beta} + \mathbf{\Pi}_{\Delta \mathbf{X}^U \Delta \mathbf{X}} \boldsymbol{\beta}^U + \mathbf{\Pi}_{\eta_{si}^U \mathbf{X_i}}$$

$$(90) \qquad \mathbf{G_1} = [-\mathbf{\Pi}_{\Delta \mathbf{X}^U \Delta \mathbf{X}} + \mathbf{\Pi}_{\mathbf{X}^U \mathbf{X}} + \mathbf{Var}(\mathbf{X_i})^{-1} \mathbf{R}' \mathbf{Var}(\tilde{\mathbf{X}}_{\mathbf{i}}^{\mathbf{U}})] \boldsymbol{\beta}^U + \mathbf{\Gamma_1} + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}$$

$$(91) \qquad \mathbf{G_2} = \mathbf{\Gamma_2} + \mathbf{\Pi}_{z_s^U \mathbf{Z_{2s}}}$$

$$(92) \qquad v_s = \tilde{z}_s^U + \xi_s$$

$$(93) \qquad v_{si} - v_s = \tilde{x}_{si}^U + \tilde{\eta}_{si} + \xi_i$$

This concludes the proof.

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

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

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

   A.   *Proof that under Assumptions A1-A5, $\mathbf{G_1} = \mathbf{\Gamma_1} + \mathbf{\Pi}_{z_s^U \mathbf{X_s}}$ in the Absence of Sorting on $\mathbf{X_i^U}$*

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

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

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

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

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

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

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

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

A.   Proof of Proposition 3:

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

A6.1:

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

A6.2:

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

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

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

$$Var(\mathbf{Z_s \Gamma} + z_s^U) \geq Var(\mathbf{Z_{2s} G_2}) + Var(v_s).$$

Recall the definition $\mathbf{Z_s \Gamma} \equiv \mathbf{X_s \Gamma}_1 + \mathbf{Z_{2s} \Gamma}_2$. Also, under the assumptions in Proposition 1, Proposition 2 establishes that $\mathbf{G_2} = \mathbf{\Gamma}_2 + \mathbf{\Pi}_{z_s^U Z_{2s}}$ and $v_s = \tilde{z}_s^U + \xi_s$. Using these three equations, we can rewrite the previous inequality as

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

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

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

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

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

$$(97) \quad + Var(\mathbf{Z}_{2s}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{\tilde{z}_s^U Z_{2s}}]) + Var(\tilde{z}_s^U) \geq Var(\mathbf{Z}_{2s}[\mathbf{\Gamma}_2 + \mathbf{\Pi}_{\tilde{z}_s^U Z_{2s}}]) + Var(\tilde{z}_s^U) + Var(\xi_s)$$

Cancelling common terms from both sides yields A6.2:

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

In the case of our more conservative estimator, $Var(\mathbf{Z}_{2s}\mathbf{G}_2)$, the proof follows the exact same logic, except that the terms $Var(\tilde{z}_s^U)$ and $Var(\xi_s)$ do not appear on the right side in (95), and thus $Var(\tilde{z}_s^U)$ does not cancel and $Var(\xi_s)$ need not be subtracted in (98). This leaves the inequality A6.1:

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

This concludes the proof.

B.   Analysis of Assumption 6

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

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

This condition is stronger than A6.1 (since it omits $Var(\tilde{z}_s^U)$), and is equivalent to A6.2 in contexts where common shocks either do not exist (such as high school graduation) or are considered part of the group treatment effect component $\tilde{z}_s^U$ (since individuals who choose different schools will receive different common shocks). Because 1) we discuss common shocks elsewhere, 2) we introduced the conservative estimator $\widehat{Var}(\mathbf{Z}_{2s}\mathbf{G}_2)$ to eliminate their influence, and 3) our sorting model provides no guidance about their size, we focus attention here on the case where common shocks do not exist ($\xi_s = 0 \ \forall \ s$).

From a theoretical standpoint, note that standard sorting patterns would suggest a positive rather than a negative covariance between the observable peer effect term $\mathbf{X}_s\mathbf{\Gamma}_1$ and the observed school input term $\mathbf{Z}_{2s}\mathbf{\Gamma}_2$. Without loss of generality, suppose that each element $\mathbf{X}_i$ has been defined so that higher values increase Y_i (i.e. each element of $\boldsymbol{\beta}$ is positive). Most evidence suggests that concentrations of better prepared students and parents (high values of $\mathbf{X}_s$) are likely to provide stronger peer effects relative to concentrations of less prepared students and parents, suggesting

that $\mathbf{X_s}\boldsymbol{\Gamma}_1$ would be positive when $\mathbf{X_s}$ values are high. And wealthier, more educated parents with more able children (presumed to be positive $\mathbf{X_i}$ inputs) tend to be willing to pay more for neighborhoods featuring schools with better inputs. Thus, standard assumptions about sorting would predict that $\mathbf{X_s}\boldsymbol{\Gamma}_1$ would display a positive covariance with the direct school inputs captured by $\mathbf{Z_{2s}}(\boldsymbol{\Gamma}_2 + \boldsymbol{\Pi}_{z_s^U \mathbf{Z_{2s}}})$. And because z_s^U is likely to partially represent peer effects associated with unobserved school characteristics $\mathbf{X_s^U}\boldsymbol{\Gamma}_1^U$, and $\mathbf{X_s^U}$ projects fully onto $\mathbf{X_s}$ under Proposition 1, $\mathbf{X_s}\boldsymbol{\Pi}_{z_s^U \mathbf{X_s}}$ is likely to also capture peer effects associated with concentrations of parents/students with high values of unobserved characteristics. So we would also expect $Cov(\mathbf{X_s}\boldsymbol{\Pi}_{z_s^U \mathbf{X_s}}, \mathbf{Z_{2s}}(\boldsymbol{\Gamma}_2 + \boldsymbol{\Pi}_{z_s^U \mathbf{Z_{2s}}}))$ to be positive as well.

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

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

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

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

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

Define the following scalar parameters:

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

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

potential overstatement of the true variance in percentage terms) in terms of only ρ and μ:

$$\frac{Var(\mathbf{Z_s}\mathbf{\Gamma} + z_s^U) - Var(\mathbf{Z_{2s}}\mathbf{G_2} + v_s)}{Var(\mathbf{Z_{2s}}\mathbf{G_2} + v_s)}$$

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

$$= \frac{\mu Var(\mathbf{Z_{2s}}\mathbf{G_2} + v_s) + 2\rho\sqrt{\mu}Var(\mathbf{Z_{2s}}\mathbf{G_2} + v_s)}{Var(\mathbf{Z_{2s}}\mathbf{G_2} + v_s)}$$

$$= \mu + 2\rho\sqrt{\mu}$$

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

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

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

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

least the half the size of (and often exceeding) that of $Var(\mathbf{Z_{2s}G_2})$, suggesting that the sorting on unobservables component $\mathbf{X_s}\mathbf{\Pi}_{\tilde{x}_i^U \mathbf{X_s}}$ would need to be quite substantial for μ to be low enough to be consistent with a violation of Assumption 6.

Indeed, as we pointed out in Section 8 of the paper, under the stronger but common selection-on-observables-only assumption, $\mathbf{X_s}\mathbf{B}$ would fully capture student sorting, $\mathbf{\Pi}_{\tilde{x}_i^U \mathbf{X_s}} = \mathbf{0}$ and $\mathbf{G_1} = \mathbf{\Gamma_1} + \mathbf{\Pi}_{\zeta_s^U \mathbf{X_s}}$. Thus, under selection-on-observables, $2Cov(\mathbf{X_sG_1}, \mathbf{Z_{2s}G_2}) + Var(\mathbf{X_sG_1}) = Var(\mathbf{Z_s}\mathbf{\Gamma} + \xi_s) - Var(\mathbf{Z_{2s}G_2} + v_s)$, capturing directly the degree to which our estimator $Var(\mathbf{Z_{2s}G_2} + v_s)$ understates or overstates the true neighborhood/school effect variance. And this sum is positive in every specification and outcome we use, and usually substantially so. Thus, the limited evidence that our empirical estimates provide about Assumption 6 strongly suggest that it holds in our data.

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

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

A9. Estimation of Model Parameters

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

A. Years of Postsecondary Academic Education

Parameter estimation is most straightforward in the case of years of postsecondary academic education. Recall that $\mathbf{Z_s}$ is comprised of two components: $\mathbf{Z_s} = [\mathbf{X_s}; \mathbf{Z_{2s}}]$. $\mathbf{Z_{2s}}$ consists of school

and neighborhood characteristics for which direct measures are available, such as student/teacher ratio, city size, and school type. $\mathbf{X_s}$ consists of school wide averages for each variable in $\mathbf{X_i}$, such as parental education or income, which we do not observe directly but must estimate from sample members at each school. Consequently, the makeup of $\mathbf{X_s}$ differs across specifications that use different $\mathbf{X}$ vectors. $\mathbf{G_1}$ and $\mathbf{G_2}$ are the corresponding subsets of the coefficients in $\mathbf{G}$. We replace $\mathbf{X_s}$ with $\mathbf{\hat{X}_s}$, where $\mathbf{\hat{X}_s}$ is the average of $\mathbf{X_i}$ computed over all available students from the school, leading to the equation below.[12] The regression model is

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

using the appropriate panel weights from the surveys. We estimate the model parameters using restricted maximum likelihood (REML). We treat v_s as a random effect and assume the error components are normally distributed, ignoring the error component $[\mathbf{X_s} - \mathbf{\hat{X}_s}]\mathbf{G_1}$. REML accounts for degrees of freedom in estimating $Var(v_s)$ and $Var(v'_{si})$, while maximum likelihood does not.[13] Computations were performed using the STATA Version 14 procedure *mixed* with the REML option. Because we experienced computational difficulties when using panel weights, the estimates are unweighted.

B. Permanent Wage Rates

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

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

In the above equation Y_{si} is i's "permanent" log wage (given that he/she attended high school s) as of the time by which most students have completed education and spent at least a couple of years in the labor market, which we take to be 1979 in the case of NLS72. ς_{sit} is a stationary component that evolves as a result of luck in the job search process or within a company, changes in motivation or productivity due to health and other short term factors that may persist for up to 7 years. It also includes measurement error.[14] The determination of Y_{si} is given by (8) which leads to the regression equation (12). After substituting for Y_{si} and replacing $\mathbf{X_s}$ with $\mathbf{\hat{X}_s}$, the wage equation

---

[12]A substantial number of students who appear in the base year of the surveys can be used to construct $\mathbf{\hat{X}_s}$ but cannot be used to estimate (102) because some variables, such as test scores, are missing, or because the students are not included in the follow-up surveys that provide the measure of Y_{si}. As we discuss in Section V, we impute missing values for most of our explanatory variables prior to estimating $\mathbf{B}$ and $\mathbf{G}$, but we do not use the imputed values when constructing the school averages.

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

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

is

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

We estimate the model by REML under the assumption that the error components are normally distributed and that $Cov(\varsigma_{si1979}, \varsigma_{si1986})$ is 0.[15]

C. High School Graduation and College Enrollment

For binary outcomes such as high school graduation we reinterpret Y_{si} to be the latent variable that determines the indicator for whether a student graduates, $HSGRAD_{si}$. That is,

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

or, after substituting for Y_{si},

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

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

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

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

FORMULAE FOR ESTIMATING $Var(v_i)$ AND $Var(v_s)$ IN THE CASE OF BINARY OUTCOME VARIABLES

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

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

---

[15]In reality, we also include a vector $\mathbf{T_{it}}$ consisting of a dummy indicator for the year 1979 (relative to 1986), years of work experience of i at time t, and experience squared. Let $\mathbf{\Upsilon}$ be the corresponding vector of wage coefficients. The term $\mathbf{T_{it}\hat{\Upsilon}}$ does not contribute to Y_{si} and does not play a role in the variance decompositions. The estimate of $\mathbf{\hat{\Upsilon}}$ depends on whether tests, postsecondary education, or both are in $\mathbf{X_i}$. We report results with and without these variables. In our main specification, we exclude postsecondary education from $\mathbf{X_i}$.

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

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

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

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

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

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

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

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

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

$$SSR_b = \sum_{s=1}^{S} (\overline{Y}_s - \mathbf{X_s}\hat{\mathbf{B}} - \mathbf{X_s}\hat{\mathbf{G}_1} - \mathbf{Z_{2s}}\hat{\mathbf{G}_2})^2.$$

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

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

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

estimate), in the binary case we approximate $\widehat{Var}(v_s)_{ML}$ as:

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

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

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

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

(113)

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

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

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

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

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

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

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

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

D.   Estimating the Variances and Covariances of the Components of the Regression Index

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

$$(116) \qquad Var(\mathbf{Z_{2s}}\hat{\mathbf{G}}_2) = \frac{1}{N}\sum_i(\mathbf{Z_{2s(i)}}\hat{\mathbf{G}}_2\hat{\mathbf{G}}_2'\mathbf{Z}_{2s(i)}')$$

$$(117) \qquad = \frac{1}{N}\sum_i(\mathbf{Z_{2s(i)}}\mathbf{G}_2\mathbf{G}_2'\mathbf{Z}_{2s(i)}') + \frac{1}{N}\sum_i\mathbf{Z_{2s(i)}}[\hat{\mathbf{G}}_2 - \mathbf{G}_2][\hat{\mathbf{G}}_2 - \mathbf{G}_2]'\mathbf{Z}_{2s(i)}'.$$

In the above equation we have made the dependence of s on i explicit. The expectation of the first term on the right is $Var(\mathbf{Z_{2s}}\mathbf{G}_2)$. The second term is the effect of sampling error in $\hat{\mathbf{G}}_2$, conditional on $\mathbf{Z_{2s(i)}}$. It has expectation

$$(118) \qquad \frac{1}{N}\sum_i\mathbf{Z_{2s(i)}}\mathbf{Var}(\hat{\mathbf{G}}_2)\mathbf{Z}_{2s(i)}'.$$

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

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

In the above formula $\widehat{\mathbf{Var}}(\hat{\mathbf{G}}_2)$ is the estimator based on the formula for the asymptotic variance associated with the estimator $\hat{\mathbf{G}}_1$, which depends on the outcome. We do not account for the use of imputed data. In practice, we report population weighted variances, so sample weights appear in the two sums. The estimators of $Var(\mathbf{X_i}\mathbf{B})$, $Var(\mathbf{X_s}\mathbf{B})$ and $Var(\mathbf{Z_{2s}}\mathbf{G_2})$ are almost exactly analogous.

To estimate $Cov(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2})$, we first estimate $Var(\mathbf{Z_s}\mathbf{G}) \equiv Var(\mathbf{X_s}\mathbf{G_1} + \mathbf{Z_{2s}}\mathbf{G_2})$ via:

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

Then we generate $\widehat{Cov}(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2})$ via:

$$(120) \qquad \widehat{Cov}(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2}) = \frac{\widehat{Var}(\mathbf{Z_s}\mathbf{G}) - \widehat{Var}(\mathbf{X_s}\mathbf{G_1}) - \widehat{Var}(\mathbf{Z_{2s}}\mathbf{G_2})}{2}$$

We use an analogous procedure for $Cov(\mathbf{X_s}\mathbf{B}, \mathbf{X_s}\mathbf{G_1})$ and $Cov(\mathbf{X_s}\mathbf{B}, \mathbf{Z_{2s}}\mathbf{G_2})$.

A10. Decomposing the Variance in Educational Attainment and Permanent Wages

In this section we discuss an analysis of variance based on (23) that can be used to place a lower bound on the importance of factors that are common to students from the same school.[16] As with parameter estimation, the details of our procedure depend upon the outcome. We begin with years of postsecondary education and permanent wages.

A. Years of Postsecondary Education and Wages

We start with years of education. One may decompose $Var(Y_{si})$ into its within and between school components

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

where Y_s is the average outcome for students from s. From (21) we obtain

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

and

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

Thus, one may express the outcome variance as[17]

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

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

(123) $2Cov(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2}) + Var(\mathbf{Z_{2s}}\mathbf{G_2}) + Var(v_s)].$

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

(124) $\widehat{Var}(Y_{si}) \equiv [\widehat{Var}((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + \widehat{Var}(v_{si} - v_s)] +$

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

(126) $2\widehat{Cov}(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2}) + \widehat{Var}(\mathbf{Z_{2s}}\mathbf{G_2}) + \widehat{Var}(v_s)].$

as the estimator of $Var(Y_{si})$. We express the variance and covariance estimates as fractions of $Var(Y_{si})$ by dividing the variance and covariance terms by $\widehat{Var}(Y_{si})$. For example, we compute $\widehat{Var}(\mathbf{Z_{2s}}\mathbf{G_2})/\widehat{Var}(Y_s)$. The fractions are reported in Table 6, Columns 1 and 2 and Appendix Table A22, Columns 1 and 2.

The procedure for decomposing the variance of the permanent log wage component Y_{si} is essen-

---

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

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

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

B. High School Graduation and College Enrollment

For both of our binary outcomes, high school graduation and enrollment in a four-year college, we decompose the latent variable that determines the outcome. Given that there is no natural scale to the variance of the latent variable, we normalize $Var(v_{si} - v_s)$ to one, and define the total variance of the latent variable to be

$$
(127) \qquad Var(Y_{si}) \equiv Var((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + 1 +
$$
$$
[Var(\mathbf{X_s B}) + 2Cov(\mathbf{X_s B}, \mathbf{X_s G_1}) + 2Cov(\mathbf{X_s B}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{X_s G_1}) +
$$
$$
(128) \qquad 2Cov(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + Var(\mathbf{Z_{2s} G_2}) + Var(v_s)]
$$

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

$$
(129) \qquad \widehat{Var}(Y_{si}) \equiv \widehat{Var}((\mathbf{X_i} - \mathbf{X_s})\mathbf{B}) + 1
$$
$$
(130) \qquad [\widehat{Var}(\mathbf{X_s B}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{X_s G_1}) + 2\widehat{Cov}(\mathbf{X_s B}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{X_s G_1}) +
$$
$$
(131) \qquad 2\widehat{Cov}(\mathbf{X_s G_1}, \mathbf{Z_{2s} G_2}) + \widehat{Var}(\mathbf{Z_{2s} G_2}) + \widehat{Var}(v_s)].
$$

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

C. Calculation of Standard Errors

We calculate bootstrap standard errors for each of our point estimates and bound estimates based on re-sampling schools with replacement using 500 replications. We bootstrap the entire estimation procedure, including imputation of missing data, estimation of model parameters, variance decompositions, and treatment effects. To preserve the size distribution of the samples of students from particular schools, we divide the sample into five school sample size classes and re-sample schools within class.

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

The monte carlo simulations in Online Appendix A6 suggest that estimation based on subsamples of 20 students per school (similar to those used to construct $\mathbf{X_s}$ in the three panel survey datasets) could diminish the ability of school-average observables to capture sorting on unobservables. However, these simulations are based on particular assumptions about the dimensionality of the underlying desired amenities, the joint distribution of the observable and unobservable characteristics, and the degree to which these characteristics predict tastes for schools/neighborhoods.

In this appendix, we assess the potential for bias in our survey-based estimates more directly by drawing samples of students from North Carolina schools using either the NLS72, NELS88, or ELS2002 sampling schemes and re-estimating the model for high school graduation using these samples. By comparing the results derived from such samples to the true results based on the universe of students in North Carolina, we can determine which if any of the survey datasets is likely to produce reliable results. To remove the chatter produced by a single draw from these sampling schemes, we computed estimate averages over 200 samples drawn from each sampling scheme.

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

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

A12. Construction and Use of Weights

In the NLS72 analyses of four-year college enrollment and postsecondary years of education, we use a set of panel weights (w22) designed to make nationally representative a sample of respondents who completed the base-year and fourth-follow up (1979) questionnaires. For the NLS72 wage

analysis, we chose a set of panel weights (comvrwt) designed for all 1986 survey respondents for whom information exists on 5 of 6 key characteristics: high school grades, high school program, educational attainment as of 1986, gender, race, and socioeconomic status. Since there are very few 1986 respondents who did not also respond in 1979, this weight matches the wage sample fairly well. For the NELS88 sample, we use a set of weights (f3pnlwt) designed to make nationally representative the sample of respondents who completed the first four rounds of questionnaires (through 1994, when our outcomes are measured). For the ELS02 sample, we use a set of weights (f2bywt) designed to make nationally representative a sample of respondents who completed the second follow up questionnaire (2006) and for whom information was available on certain key baseline characteristics (gathered either in the base year questionnaire or the first follow-up). This seemed most appropriate given that our outcomes are measured in the 2006 questionnaire and we require non-missing observations on key characteristics for inclusion in the sample.

We use panel weights in the estimation when possible for a number of reasons. The first is to reduce the influence of choice-based sampling, which is an issue in NELS88. It is also a potential issue for the wage analysis based on NLS72, but we had difficulty computing weighted estimates. The second is to correct for non-random attrition from follow-up surveys. The third is a pragmatic adjustment to account for the possibility that the link between the observables and outcomes involves interaction terms or nonlinearities that we do not include. The weighted estimates may provide a better indication of average effects in such a setting. Finally, various populations and school types were oversampled in the three datasets, so that applying weights makes our sample more representative of the universe of American 8th graders, 10th graders, and 12th graders, respectively. For all outcomes, including wages, we employ sample weights when using the regression model parameters to construct estimates of $Var(\mathbf{Z}_{2s}\mathbf{G}_2)$, $Var(\mathbf{X}_s\mathbf{G}_1)$. Note, though, that we do not adjust weights for item non-response associated with the key variables required for inclusion in our sample. As discussed in section IV.A, due to computational difficulties, for our continuous outcomes (years of postsecondary education and log wages) we do not incorporate weights into the REML procedure used to estimate the coefficients $\mathbf{B}$, $\mathbf{G}_1$, and $\mathbf{G}_2$ and the error component variances $Var(v_{si} - v_s)$ and $Var(v_s)$. However, conditional on the estimated coefficient vectors and error variances, panel weights are still used to compute the variances of the various regression indices (such as $Var(\mathbf{Z_{2s}G_2})$), and they are still used to average over the population of students when estimating the impact of 10th-to-90th quantile shifts in school quality.

A13.   Other Applications: Estimating Teacher Value-Added

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

A. *Sorting of Students Across Class Rooms*

Assume for now that the administrator has already determined which teachers to allocate to which courses for which periods of the day, so that a classroom c can be effectively captured by a vector of amenity values $\mathbf{A_c}$. Some of the amenities are likely to reflect the demographic makeup of the class and thus are endogenous to classroom assignment. Others can be considered exogenous to the principal's student-to-classroom allocation problem. These would include the principal's percep-tions of various teacher attributes or skills, but could also include classroom amenities unrelated to teacher quality that might capture whether the heating works, the quality of classroom technology in the room, the time in the day that the class is held, or the difficulty level of the class.

We can then adapt the utility function featured in (2) to model the payoff that the principal obtains from assigning student i to class c (simply replace all s subscripts with c subscripts). As before, $\mathbf{X_i}$ is a vector of student characteristics that are observed by the econometrician and are relevant for the outcome Y_{si}, the student's end-of-year standardized test score. Similarly, $\mathbf{X_i^U}$ is a vector of student characteristics that are unobserved by the econometrician but are observed by the principal and are relevant for test score performance, and $\mathbf{Q_i}$ represents a vector of student characteristics that are unobserved by the econometrician and observed by the principal, but do not affect test score performance. The $\mathbf{\Theta}$, $\mathbf{\Theta^U}$ and $\mathbf{\Theta}^Q$ matrices might capture a principal's belief about which types of students are most likely to benefit from a better teacher or difficulty level. $\mathbf{\Theta}$, $\mathbf{\Theta^U}$, and $\mathbf{\Theta}^Q$ might also reflect the desire to placate parents or students, where students/parents with certain values of $\mathbf{X_i}$, $\mathbf{X_i^U}$, or $\mathbf{Q_i}$ are more likely to advocate for particular classroom assignments. Some parental or student characteristics may predict a stronger preference for a particular difficulty level or time of day, while others predict a stronger preference for teacher quality. Similarly, the idiosyncratic match value ε_{ic} might capture, for example, the desire to fulfill a particular family's request that their child be assigned to the same teacher that his brother had. Thus, we model parent and student preferences as affecting choice through their impact on principal preferences.[18]

Let $\mathscr{I}$ represent the set of students to be allocated, and let $\mathscr{C}$ represent the set of available class-rooms (each of which has an associated teacher). First we consider the special case in which none of the amenities reflect the demographic makeup of the class and thus $\mathbf{A_c}$ can be considered ex-ogenous to the principal's student-to-classroom allocation problem. The principal's problem is to choose the mapping $c : \mathscr{I} \to \mathscr{C}$ from students to classrooms that maximizes the sum of student utilities, subject to the constraints that each classroom cannot exceed its capacity and every stu-dent (or perhaps student-subject combination at the high school level) can only be assigned to one classroom:

---

[18]Rothstein (2009) provides a useful classroom assignment model in which principals assign students to classrooms based on student characteristics that are observable to both the principal and the econometrician $\mathbf{X_i}$ and student characteristics that are only available to the principal (part of $\mathbf{X_i^U}$). He discusses bias in VAM models that include $\mathbf{X_i}$ and possibly other controls. He does not consider the potential for $\mathbf{X_c}$ to control for $\mathbf{X_c^U}$.

$$\max_{c:\mathscr{I}\to\mathscr{C}} \sum_{i\in\mathscr{I}} U_{ic(i)}$$

$$s.t. \sum_{c'} \mathbb{1}(c(i) = c') = 1 \ \forall \ i$$

(132)
$$s.t. \sum_{i'} \mathbb{1}(c(i') = c) = \overline{C}_c \ \forall \ c \in \mathscr{C}$$

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

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

$$\max_{\mathbf{d}} \ \mathbf{a} * \mathbf{d}$$

$$s.t. \ \mathbf{M}_i * \mathbf{d} = 1 \ \forall \ i \in \mathscr{I}$$

$$s.t. \ \mathbf{N}_c * \mathbf{d} = \overline{C}_c \ \forall \ c \in \mathscr{C}$$

(133)
$$s.t. \ \mathbf{d} \in \{0,1\}$$

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

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

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

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

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

$\mathbf{M}_i$ consists of a $1 \times I*C$ row vector capturing the number of classrooms to which each student (or

student-subject combination) is assigned (restricted here to be $1 \; \forall \; i$):

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

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

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

Koopmans and Beckmann (1957) show that the solution to this binary integer program problem can be sustained by a one-sided set of prices for classrooms $\{P_c\}$.[19] This means that the optimal assignment for each individual is also the solution to his/her utility maximization problem:

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

Notice that the structure of this utility maximization problem is isomorphic to that of the decentralized school choice problem from Section I. Consequently, if the spanning condition $\mathbf{\Theta}^{\mathbf{U}} = \mathbf{R}\tilde{\mathbf{\Theta}}$ is satisfied for some matrix $\mathbf{R}$, $\mathbf{X_c}$ will be a linear function of $\mathbf{X_c^U}$.

Exogeneity of the amenity vector may be a reasonable assumption in some high school and college contexts in which students submit course preferences and a schedule-making algorithm assigns students to classrooms. However, in the elementary and middle school contexts, it is likely that some elements of $\mathbf{A_c}$ reflect the student makeup of the class. Anticipated peer effects complicates the principal's problem, since now the utility from assigning a given student to a classroom would depend on the other students assigned to the classroom. The classroom sorting problem differs from the school/neighborhood sorting problem in that the principal would internalize the effect that allocating a student to c has on $\mathbf{A_c}$, while parents would take $\mathbf{A_s}$ as given. We have not yet extended Proposition 1 to a classroom assignment problem with endogenous amenities.

B. Implications for Estimation of Teacher Value Added

Suppose that the true classroom contribution to a given student i's test scores can be captured by $\mathbf{Z_c}\mathbf{\Gamma} + z_c^U + \eta_{ci}$, mirroring (8). As before, partition the vector of observed classroom characteristics into two parts $\mathbf{Z_c} = [\mathbf{X_c}, \mathbf{Z_{2c}}]$, where $\mathbf{X_c}$ captures classroom averages of observed student character-

---

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

istics and $\mathbf{Z_{2c}}$ represents other observed classroom characteristics. Consider the classroom version of our estimating equation (21):

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

When past test scores are elements of $\mathbf{X_i}$ and a design matrix $\mathbf{D_{c(i)}}$ indicating which classrooms were taught by which teachers is included in $\mathbf{Z_{2c}}$, (135) represents a standard teacher value-added specification.[20]

Suppose that Proposition 1 can be extended to the classroom choice setting (as proven in the exogenous amenities case) and that the corresponding spanning condition is satisfied, so that $\mathbf{X_c}$ and $\mathbf{X_c}^U$ are linearly dependent. Suppose in addition that the principal's perception of teacher quality is noisy, so that $\mathbf{D_c}$ is not collinear with $\mathbf{A_c}$ (and therefore not collinear with $\mathbf{X_c}$). Then our analysis in Section III.C suggests that $\mathbf{G_2} = \boldsymbol{\Gamma_2} + \boldsymbol{\Pi}_{\mathbf{Z_c}^U \mathbf{Z_{2c}}}$. Since $\mathbf{Z_{2c}}$ includes the teacher design matrix $\mathbf{D_{c(i)}}$, we see that including classroom averages of student characteristics $\mathbf{X_c}$ in teacher value-added regressions will purge estimates of individual teachers' value-added from any bias from non-random student sorting on either observables or unobservables. Any remaining bias $\boldsymbol{\Pi}_{\mathbf{Z_c}^U \mathbf{Z_{2c}}}$ stems from the possible correlation between the assignment of the chosen teacher to the classroom and other aspects of the classroom environment. Note that $\mathbf{G_1}$ should be allowed to differ across schools or districts if the preference parameters $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}^U$ are believed to differ.

However, suppose that all unobserved classroom factors that are inequitably distributed across teachers are either being used as a basis for student allocation to classrooms (i.e. are elements of $\mathbf{A_c}$) or are directly included as other controls in $\mathbf{Z_{2c}}$. If in addition the mapping from $\mathbf{A_c}$ to $\mathbf{X_c}$ is linear, then the analysis in Section III.C reveals that including classroom averages of observed student characteristics will also purge teacher value-added estimates $\mathbf{G_2}$ of any omitted variables bias driven by inequitable access to advantageous classroom environments (the subvector of $\boldsymbol{\Pi}_{\mathbf{Z_c}^U \mathbf{Z_{2c}}}$ corresponding to the teacher design matrix $\mathbf{D_c}$ will equal $\mathbf{0}$).

Of course, our simulations suggest that the effectiveness of the control function approach depends on observing moderately large samples of students with each teacher. And in practice there may be classroom factors ignored by students and principals that do not even out across teachers. While these caveats should be kept in mind, our analysis may partially explain the otherwise surprising finding that non-experimental OLS estimators of teacher quality produce nearly unbiased estimates of true teacher quality as ascertained by quasi-experimental and experimental estimates (Chetty et al. (2014), Kane and Staiger (2008)).

---

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

References

Altonji, Joseph G and Richard K Mansfield, "Group-Average Observables as Controls for Sorting on Unobservables when Estimating Group Treatment Effects: the Case of School and Neighborhood Effects," Working Paper 20781, National Bureau of Economic Research 2014.

Bai, Jushan and Serena Ng, "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 2002, *70* (1), 191–221.

Baker, Michael and Gary Solon, "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records," *Journal of Labor Economics*, 2003, *21* (2), 267–288.

Chetty, Raj, John Friedman, and Jonah Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014, *104* (9), 2593–2632.

Cragg, John G and Stephen G Donald, "Inferring the Rank of a Matrix," *Journal of Econometrics*, 1997, *76* (1), 223–250.

Haider, Steven J, "Earnings Instability and Earnings Inequality of Males in the United States: 1967–1991," *Journal of Labor Economics*, 2001, *19* (4), 799–836.

Harvey, Walter Robert, *User's Guide for LSML76: Mixed Model Least-squares and Maximum Likelihood Computer Program*, Ohio State University, 1977.

Jencks, Christopher S and Marsha D Brown, "Effects of High Schools on Their Students," *Harvard Educational Review*, 1975, *45* (3), 273–324.

Kane, Thomas J and Douglas O Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," Working Paper No. 14607, National Bureau of Economic Research 2008.

Kleibergen, Frank and Richard Paap, "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 2006, *133* (1), 97–126.

Koopmans, Tjalling C and Martin Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica: journal of the Econometric Society*, 1957, *25* (1), 53–76.

Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 2004, *72* (1), 1–32.

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

Rothstein, Jesse, "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables," *Education*, 2009, *4* (4), 537–571.

Table A1—: Estimates of the Contribution of School Systems and Neighborhoods to High School Graduation Decisions Under the Assumption that Only Observables $\mathbf{X_i}$ Drive Sorting

Appendix Tables

Panel A: Fraction of Latent Index Variance Determining Graduation Attributable to School/Neighborhood Quality						
Lower Bound	NC		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full (1)
	(1)	(2)	(3)	(4)	(5)	(6)
No Unobs. Sort.	0.052	0.042	0.090	0.083	0.049	0.032
$Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(0.017)	(0.011)	(0.009)	(0.009)	(0.011)	(0.010)

Panel B: Effect on Graduation Probability of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile						
Lower Bound	NC		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
No Unobs. Sort.: 10th-90th	0.181	0.162	0.183	0.175	0.095	0.077
Based on $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(0.025)	(0.015)	(0.015)	(0.019)	(0.009)	(0.009)
No Unobs. Sort.: 10th-50th	0.100	0.089	0.111	0.105	0.056	0.044
Based on $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$	(0.017)	(0.010)	(0.024)	(0.013)	(0.006)	(0.006)
Sample Mean	0.769	0.769	0.827	0.827	0.897	0.897

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

Panel A reports lower bound estimates of the fraction of variance in the latent index that determines high school graduation that can be directly attributed to school/neighborhood choices for each dataset. The label "No Unobs. Sort." reports $Var(\mathbf{X_s G_1} + \mathbf{Z_{2s} G_2} + v_s)$, which captures the variance in true school/neighborhood contributions under the assumption that sorting is driven only by $\mathbf{X_i}$.

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

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

For each data set the variables in the baseline and full models are specified in Table 1.

The full variance decompositions underlying these estimates are presented in Online Appendix Table A20.

Online Appendices A9 and A10 discuss estimation of model parameters and the variance decompositions. Section IV.D discusses estimation of the 10-50 and 10-90 differentials.

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

Panel A: Fraction of Latent Index Variance Determining Enrollment Attributable to School/Neighborhood Quality						
Lower Bound	NLS72		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
No Unobs. Sort.	0.062	0.046	0.076	0.071	0.068	0.043
$Var(\mathbf{X_sG_1}+\mathbf{Z_{2s}G_2}+v_s)$	(0.012)	(0.006)	(0.009)	(0.007)	(0.007)	(0.005)

Panel B: Effect on Enrollment Probability of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile						
Lower Bound	NLS72		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
No Unobs. Sort.: 10th-90th	0.216	0.184	0.261	0.246	0.264	0.204
Based on $Var(\mathbf{X_sG_1}+\mathbf{Z_{2s}G_2}+v_s)$	(0.017)	(0.016)	(0.018)	(0.016)	(0.018)	(0.015)
No Unobs. Sort.: 10th-50th	0.097	0.084	0.117	0.112	0.124	0.097
Based on $Var(\mathbf{X_sG_1}+\mathbf{Z_{2s}G_2}+v_s)$	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)
Sample Mean	.267	.267	.310	.310	.365	.365

Bootstrap standard errors based on resampling at the school level are in parentheses.
The notes to Table A1 apply, except that Table A2 reports results for enrollment in a 4-year college two years after graduation.
The column headed NLS72 refers to a variance decomposition that uses the 12th grade school as the group variable.

Table A3—: Lower Bound Estimates of the Contribution of School Systems and Neighborhoods to Four Year College Enrollment Decisions (Naive OLS Specification: School-Averages $\mathbf{X_s}$ omitted from estimating equation)

Panel A: Fraction of Latent Index Variance Determining Enrollment Attributable to School/Neighborhood Quality						
Lower Bound	NLS72		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs	0.028	0.017	0.023	0.022	0.024	0.016
$Var(\mathbf{Z_{2s}G_2})$	(0.007)	(0.004)	(0.005)	(0.005)	(0.010)	(0.007)
LB w/ unobs	0.066	0.049	0.075	0.073	0.072	0.050
$Var(\mathbf{Z_{2s}G_2} + v_s)$	(0.015)	(0.008)	(0.009)	(0.009)	(0.023)	(0.016)

Panel B: Effect on Enrollment Probability of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile						
Lower Bound	NLS72		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.141	0.111	0.138	0.136	0.152	0.124
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.013)	(0.011)	(0.014)	(0.013)	(0.014)	(0.013)
LB w/ unobs: 10th-90th	0.219	0.187	0.253	0.246	0.266	0.218
Based on $Var(\mathbf{Z_{2s}G_2} + v_s)$	(0.016)	(0.014)	(0.014)	(0.013)	(0.017)	(0.015)
LB no unobs: 10th-50th	0.066	0.052	0.065	0.064	0.073	0.060
Based on $Var(\mathbf{Z_{2s}G_2})$	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)
LB w/ unobs: 10th-50th	0.098	0.085	0.114	0.111	0.124	0.104
Based on $Var(\mathbf{Z_{2s}G_2} + v_s)$	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.007)
Sample Mean	.267	.267	.310	.310	.365	.365

"Naive OLS Specification" refers to a specification in which school-averages of individual characteristics $\mathbf{\overline{X}_s}$ are omitted from the estimating equation (or equivalently, the coefficient vector $\mathbf{G_1}$ is constrained to be equal to $\mathbf{0}$).
The notes to Table 2 apply, except that Table A3 reports results for enrollment in a 4-year college two years after graduation, and the naive OLS specification and estimates are used, as described in Section VI.E

Table A4—: Lower Bound Estimates of the Contribution of School Systems and Neighborhoods to Completed Years of Postsecondary Education in NLS72 data (Naive OLS Specification: School-Averages $\mathbf{X_s}$ omitted from estimating equation)

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

Panel B: Effects on Years of Postsecondary Education of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile		
Lower Bound	Yrs. Postsec. Ed.	
	Baseline	Full
	(1)	(2)
LB no unobs: 10th-90th Based on $Var(\mathbf{Z_{2s}G_2})$	0.353 (0.052)	0.227 (0.045)
LB w/unobs: 10th-90th Based on $Var(\mathbf{Z_{2s}G_2} + v_s)$	0.679 (0.059)	0.526 (0.056)
LB no unobs: 10th-50th Based on $Var(\mathbf{Z_{2s}G_2})$	0.176 (0.026)	0.114 (0.022)
LB w/unobs: 10th-50th Based on $Var(\mathbf{Z_{2s}G_2} + v_s)$	0.339 (0.029)	0.263 (0.028)
Sample Mean	1.62	1.62

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

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

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

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

	Panel A: Fraction of Total Variance in $\mathbf{X_s}$ Explained by Various Numbers of Principal Components					
	NLS72		NELS88 gr8		ELS2002	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
(1) # of Variables in $\mathbf{X_s}$	32	34	39	49	40	51
# Factors Needed to Explain:						
(2) 75% of Total $\mathbf{X_s}$ Var.	7 [7,8]	7 [8,8]	7 [7,8]	9 [8,9]	6 [6,7]	8 [7,8]
(3) 90% of Total $\mathbf{X_s}$ Var.	12 [11,12]	12 [12,13]	13 [11,13]	16 [14,15]	11 [11,12]	14 [14,15]
(4) 95% of Total $\mathbf{X_s}$ Var.	15 [14,15]	15 [14,15]	17 [14,16]	20 [18,19]	14 [14,15]	19 [17,19]
(5) 99% of Total $\mathbf{X_s}$ Var.	20 [18,19]	21 [17,18]	22 [19,21]	26 [23,25]	20 [18,20]	25 [23,25]
(6) 100% of Total $\mathbf{X_s}$ Var.	24 [21,23]	25 [18,19]	27 [23,26]	32 [29,31]	26 [23,25]	33 [29,31]

	Panel B: Fraction of Variance in the Regression Index $\mathbf{X_s}\hat{\mathbf{G}}_1$ Explained by Various Numbers of Principal Components					
	NLS		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
(1) # of Variables in $\mathbf{X_s}$	32	34	39	49	40	51
# Factors Needed to Explain:						
(2) 75% of $Var(\mathbf{X_s}\mathbf{G_1})$	3 [3,5]	3 [3,6]	6 [3,7]	5 [5,8]	2 [2,3]	5 [4,7]
(3) 90% of $Var(\mathbf{X_s}\mathbf{G_1})$	8 [5,9]	7 [5,10]	10 [6,11]	10 [9,14]	5 [3,7]	11 [8,14]
(4) 95% of $Var(\mathbf{X_s}\mathbf{G_1})$	10 [8,13]	9 [7,11]	13 [9,14]	13 [12,17]	7 [5,11]	15 [11,17]
(5) 99% of $Var(\mathbf{X_s}\mathbf{G_1})$	14 [13,17]	15 [10,15]	19 [13,19]	20 [19,24]	14 [11,16]	22 [17,23]
(6) 100% of $Var(\mathbf{X_s}\mathbf{G_1})$	24 [21,23]	25 [18,19]	27 [23,26]	32 [29,31]	26 [23,25]	33 [29,31]

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

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

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

Under the conditions laid out in Proposition 1 of the paper, the rank of the co-variance of $\mathbf{X_s}$ reveals the number of amenity factors driving sorting. See Online Appendix A3 for details. Each element in the table reports a p-value from a test based on Kleibergen and Paap (2006) of the null that the rank of the covariance matrix of school-averages of observable student characteristics $\mathbf{X_s}$ is equal to value associated with the row label, against the alternative hypothesis that the rank exceeds this value. "Het. Only" refers to the heteroskedasticity-robust (but unclustered) version of the test. "Cluster" refers to the more general test that is robust to arbitrary correlation in sampling error within clusters. We cluster at the school level. Each test is implemented via the STATA ranktest.ado code provided by Kleibergen and Paap (2006).

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

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

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

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

Under the conditions laid out in Proposition 1 of the paper, the rank of the covariance of $\mathbf{X_s}$ reveals the number of amenity factors driving sorting. See Online Appendix A3 for details. Each element in the table reports a p-value from a test based on Kleibergen and Paap (2006) of the null that the rank of the covariance matrix of school-averages of observable student characteristics $\mathbf{X_s}$ is equal to value associated with the row label, against the alternative hypothesis that the rank exceeds this value. "Het. Only" refers to the heteroskedasticity-robust (but unclustered) version of the test. "Cluster" refers to the more general test that is robust to arbitrary correlation in sampling error within clusters. We cluster at the school level. Each test is implemented via the STATA ranktest.ado code provided by Kleibergen and Paap (2006).

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

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

Table A8—: Monte Carlo Simulation Results: Cases in which the Spanning Condition in Proposition 1 is Satisfied ($\Theta^U = \mathbf{R}\Theta$ For Some $\mathbf{R}$)

Row	# Stu.	# Sch.	# Con.	# Ob.	# Un.	# Am.	ρ_Θ	$\frac{Var(\mathbf{X_s^U B^U})}{Var(Y)}$	Adj-R-Sq (All)	Resid (All)	Adj-R-Sq (10/20/40)	Resid (10/20/40)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	1000	50	50	10	10	5	0.25	.122	0.9979	.0003	.869 .926 .959	.016 .009 .005
(2)	500	50	50	10	10	5	0.25	.123	0.9961	.0005	.869 .926 .960	.016 .009 .005
(3)	2000	50	50	10	10	5	0.25	.122	0.9989	.0001	.869 .925 .959	.016 .009 .005
(4)	1000	100	50	10	10	5	0.25	.122	0.9979	.0002	.868 .924 .959	.017 .009 .006
(5)	1000	50	10	10	10	5	0.25	.100	0.9976	.0001	.835 .908 .950	.016 .009 .005
(6)	1000	50	50	20	20	5	0.25	.122	0.9993	.0002	.897 .944 .971	.014 .007 .003
(7)	1000	50	50	10	10	10	0.25	.136	0.9933	.0009	.872 .923 .952	.018 .010 .006
(8)	1000	50	50	20	20	10	0.25	.135	0.9988	.0002	.909 .951 .973	.014 .007 .003
(9)	1000	50	50	10	10	5	0	.048	0.9941	.0003	.649 .779 .867	.016 .010 .006

Stu.: Number of students per school

Sch.: Total number of schools

Con.: Number of schools in each family's consideration set

Ob: Number of observable student characteristics

Un: Number of unobservable student characteristics

Am.: Number of latent amenity factors valued by families

ρ_Θ: Correlation in Θ_{lk} taste parameters across student characteristics for a given amenity and across amenities for a given student characteristic

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

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

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

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

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

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

Row	X/X^U Corr. Structure	WTP for A1-A4 Depends On	WTP for A5 Depends On	Assu. (A5) Satisfied	$\Theta^U = R^A\Theta$?	$\Theta^U = R^B\Pi^{X^UX}\Theta^U$?	$\frac{Var(X_b^U B^U)}{Var(Y)}$	Adj-R-Sq (All)	Resid (All)	Adj-R-Sq (10/20/40)	Resid (10/20/40)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Corr = .25 for each pair of (obs. or unobs.) char.	All elements of X_i and X_i^U	All elements of X_i and X_i^U	Yes	Yes	Yes	.122	0.998	.0003	.869 .926 .959	.016 .009 .005
(2)	Elements of X^U independent of elements of X	All elements of X_i and X_i^U	All elements of X_i and X_i^U	Yes	No	Yes	.101	0.965	.0035	.492 .607 .717	.052 .040 .029
(3)	Corr = .25 for each pair of (obs. or unobs.) char.	All elements of X_i	All elements of X_i^U	Yes	Yes	No	.049	0.967	.0015	.622 .720 .795	.019 .014 .010
(4)	Elements of X^U independent of elements of X	All elements of X_i	All elements of X_i^U	No	No	No	.069	0.148	.0591	.114 .128 .139	.062 .060 .059
(5)	Elements of X^U independent of elements of X	All elements of X_i and X_i^U	All elements of X_i^U	No	No	No	.098	0.524	.0465	.325 .374 .421	.065 .062 .057
(6)	Elements of X^U independent of elements of X	All elements of X_i and X_i^U	$X_{i,10}^U$ only	No	No	No	.109	0.948	.0051	.589 .682 .760	.045 .035 .026
(7)	Elements of X^U independent of elements of X	All obs. and unobs. char. except $X_{i,10}^U$	$X_{i,10}^U$ only	No	No	No	.095	0.952	.0050	.580 .677 .756	.040 .030 .023
(8)	Corr = .25 for each pair of obs. or unobs. char. except $X_{i,10}^U$ (independent)	All elements of X_i and X_i^U	$X_{i,10}^U$ only	No	No	No	.117	0.997	.0003	.906 .947 .971	.013 .006 .004
(9)	Corr = .25 for each pair of obs. or unobs. char. except $X_{i,10}^U$ (independent)	All obs. and unobs. char. except $X_{i,10}^U$	$X_{i,10}^U$ only	No	No	No	.131	0.997	.0003	.893 .942 .969	.013 .006 .004

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

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

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

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

Table A10—: Bias from Observing Subsamples of Students from Each School: Comparing Results from the Full North Carolina Sample to Results from Subsamples Mirroring the Sampling Schemes of NLS72, NELS88, and ELS2002 - Baseline Specification

Panel A: Fractions of Total Outcome Variance

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

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

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

The column "Full NC Sample" reports variance decompositions based on the full North Carolina sample. They are the same as the estimates reported for NC sample in Online Appendix Table A20.

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

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

Table A11—: Bias from Observing Subsamples of Students from Each School: Comparing Results from the Full North Carolina Sample to Results from Subsamples Mirroring the Sampling Schemes of NLS72, NELS88, and ELS2002 - Full Specification

Panel A: Fractions of Total Outcome Variance

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

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

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

The column "Full NC Sample" reports variance decompositions based on the full North Carolina sample. They are the same as the estimates reported for NC sample in Online Appendix Table A20.

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

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

Table A12—: Summary Statistics for Student Characteristics in NLS72

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

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

Table A13—: Summary Statistics for School Characteristics in NLS72

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

*School characteristics treated as elements of $\mathbf{X_s}$ are included to reduce measurement error in school sample averages of student characteristics. They do not contribute to the estimated lower bound on contributions of schools/neighborhoods.

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

Table A14—: Summary Statistics for Student Characteristics in NELS88

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

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

Table A15—: Summary Statistics for School Characteristics in NELS88

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

*School characteristics treated as elements of $\mathbf{X_s}$ are included to reduce measurement error in school sample averages of student characteristics. They do not contribute to the estimated lower bound on contributions of schools/neighborhoods.

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

Table A16—: Summary Statistics for Student Characteristics in ELS2002

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

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

Table A17—: Summary Statistics for School Characteristics in ELS2002

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

*School characteristics treated as elements of $\mathbf{X_s}$ are included to reduce measurement error in school sample averages of student characteristics. They do not contribute to the estimated lower bound on contributions of schools/neighborhoods.

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

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

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

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

Variable	% Imputed	Mean	Std. Dev.
School Characteristics (Treated as elements of $\mathbf{Z_{2s}}$)			
# of Books Per Student	0.41	10.85	6.74
1(Magnet School)	0.00	.064	.244
1(Charter School)	0.00	.007	.083
% of Teachers with Advanced Degrees	0.79	.249	.079
% of Classrooms Taught by "High Quality" Teachers	0.03	.956	.060
Teacher Turnover Rate	0.87	.214	.081
Total School Enrollment	0.03	1323	581
Student-to-Teacher Ratio	0.03	15.5	2.02
Neighborhood Characteristics			
1(Remote Rural)	0.00	.028	.166
1(Distant Rural)	0.00	.160	.366
1(Fringe Rural)	0.00	.284	.451
1(Remote Town)	0.00	.006	.078
1(Distant Town)	0.00	.075	.263
1(Fringe Town)	0.00	.050	.218
1(Small Suburb)	0.00	.006	.076
1(Mid-Sized Suburb)	0.00	.049	.216
1(Large Suburb)	0.00	.096	.295
1(Small City)	0.00	.072	.259
1(Midsize City)	0.00	.086	.281

Table A20—: Decomposition of Variance in Latent Index Determining High School Graduation from the NC, NELS88, and ELS2002 Datasets (Baseline and Full Specifications)

Fraction of Variance	NC Baseline	NC Full	NELS88 gr8 Baseline	NELS88 gr8 Full	ELS2002 Baseline	ELS2002 Full
Within School:						
Total	0.913	0.913	0.812	0.812	0.891	0.904
$Var(Y_i - Y_s)$	(0.021)	(0.014)	(0.019)	(0.019)	(0.014)	(0.015)
Observable Student-Level (Within):	0.123	0.229	0.145	0.201	0.125	0.213
$Var((\mathbf{X_i} - \mathbf{X_s})\mathbf{B})$	(0.004)	(0.005)	(0.013)	(0.014)	(0.012)	(0.015)
Unobservable Student-Level (Within)	0.790	0.684	0.667	0.611	0.767	0.690
$Var(v_{si} - v_s)$	(0.018)	(0.010)	(0.020)	(0.019)	(0.016)	(0.017)
Between School:						
Total	0.087	0.087	0.188	0.188	0.109	0.097
$Var(Y_s)$	(0.021)	(0.014)	(0.019)	(0.019)	(0.014)	(0.015)
Observable Student-Level:	0.018	0.033	0.057	0.069	0.031	0.054
$Var(\mathbf{X_s}\mathbf{B})$	(0.001)	(0.003)	(0.012)	(0.012)	(0.005)	(0.007)
Student-Level/						
School-Level Covariance	0.017	0.012	0.042	0.035	0.029	0.010
$2*Cov(\mathbf{X_s}\mathbf{B}, \mathbf{X_{2s}}\mathbf{G_1} + \mathbf{Z_{2s}}\mathbf{G_2})$	(0.004)	(0.005)	(0.018)	(0.019)	(0.010)	(0.013)
School-Avg. Student-Level/						
School Char. Covariance	-0.016	-0.007	0.016	0.015	0.007	0.006
$2*Cov(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2})$	(0.011)	(0.005)	(0.007)	(0.007)	(0.011)	(0.011)
School-Avg. Student-Level	0.017	0.011	0.025	0.024	0.010	0.006
$Var(\mathbf{X_s}\mathbf{G_1})$	(0.010)	(0.005)	(0.010)	(0.007)	(0.012)	(0.011)
School Char.	0.018	0.010	0.011	0.006	0.012	0.009
$Var(\mathbf{Z_{2s}}\mathbf{G_2})$	(0.008)	(0.004)	(0.008)	(0.007)	(0.010)	(0.009)
Unobservable School-Level	0.033	0.027	0.038	0.038	0.023	0.012
$Var(v_s)$	(0.013)	(0.008)	(0.008)	(0.008)	(0.002)	(0.000)

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

The rows labels indicate the variance component.

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

Online Appendices A9 and A10 discuss estimation of model parameters and the variance decompositions.

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

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

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

For each data set the variables in the baseline model and the full model are specified in Table 1

Table A21—: Decomposition of Variance in Latent Index Determining Enrollment in a Four-Year College from the NLS72, NELS88, and ELS2002 Datasets (Baseline and Full Specifications)

| | NLS72 | | NELS88 gr8 | | ELS2002 | |
Fraction of Variance	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.861	0.861	0.775	0.777	0.786	0.793
$Var(Y_{is} - Y_s)$	(0.012)	(0.012)	(0.017)	(0.017)	(0.069)	(0.066)
Observable Student-Level (Within):	0.170	0.354	0.170	0.261	0.180	0.327
$Var((\mathbf{X_{si}} - \mathbf{X_s})\mathbf{B})$	(0.010)	(0.012)	(0.012)	(0.013)	(0.257)	(0.201)
Unobservable Student-Level (Within)	0.691	0.507	0.606	0.517	0.606	0.466
$Var(v_{si} - v_s)$	(0.014)	(0.013)	(0.018)	(0.016)	(0.189)	(0.137)
Between School:						
Total	0.139	0.139	0.225	0.223	0.214	0.207
$Var(Y_s)$	(0.012)	(0.012)	(0.017)	(0.017)	(0.069)	(0.066)
Observable Student-Level:	0.041	0.062	0.071	0.104	0.073	0.122
$Var(\mathbf{X_s}\mathbf{B})$	(0.005)	(0.006)	(0.015)	(0.015)	(0.025)	(0.040)
Student-Level/						
School-Level Covariance	0.036	0.032	0.078	0.047	0.072	0.042
$2*Cov(\mathbf{X_s}\mathbf{B}, \mathbf{X_s}\mathbf{G_1} + \mathbf{Z_{2s}}\mathbf{G_2})$	(0.007)	(0.009)	(0.014)	(0.016)	(0.024)	(0.016)
School-Avg. Student-Level/						
School Char. Covariance	-0.003	-0.004	0.006	0.008	-0.001	-0.001
$2*Cov(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2})$	(0.006)	(0.005)	(0.006)	(0.005)	(0.008)	(0.006)
School-Avg. Student-Level	0.018	0.012	0.021	0.025	0.017	0.007
$Var(\mathbf{X_s}\mathbf{G_1})$	(0.006)	(0.005)	(0.006)	(0.006)	(0.011)	(0.007)
School Char.	0.027	0.018	0.017	0.015	0.019	0.014
$Var(\mathbf{Z_{2s}}\mathbf{G_2})$	(0.006)	(0.005)	(0.006)	(0.005)	(0.010)	(0.008)
Unobservable School-Level	0.021	0.020	0.032	0.024	0.033	0.023
$Var(v_s)$	(0.006)	(0.006)	(0.006)	(0.005)	(0.008)	(0.005)

The table reports fractions of the total variance of the latent index that determines enroll-
ment in a 4-year college two years after high school graduation.
The rows labels indicate the variance component.
Bootstrap standard errors based on resampling at the school level are in parentheses.
NLS72 refers to a variance decomposition that employs NLS72 data and uses the 12th
grade school as the group variable.
See the note to Online Appendix Table A20 for additional details.

Table A22—: Decomposition of Variance in Years of Post-Secondary Education and Adult Log Wages using NLS72 (Baseline and Full Specifications)

Fraction of Variance	Yrs. Postsec. Ed.		Perm. Wages No Post-sec Ed.		Perm. Wages w/ Post-sec Ed.	
	Baseline	Full	Baseline	Full	Baseline	Full
Within School:						
Total	0.899	0.901	0.831	0.826	0.844	0.839
$Var(Y_{is} - Y_s)$	(0.008)	(0.008)	(0.021)	(0.021)	(0.024)	(0.012)
Observable Student-Level (Within):	0.146	0.271	0.138	0.184	0.113	0.137
$Var((\mathbf{X_{si}} - \mathbf{X_s})\mathbf{B})$	(0.006)	(0.007)	(0.012)	(0.014)	(0.012)	(0.012)
Unobservable Student-Level (Within)	0.753	0.630	0.694	0.642	0.731	0.703
$Var(v_{si} - v_s)$	(0.009)	(0.008)	(0.023)	(0.024)	(0.026)	(0.013)
Between School:						
Total	0.101	0.099	0.169	0.174	0.156	0.161
$Var(Y_s)$	(0.008)	(0.008)	(0.021)	(0.021)	(0.024)	(0.012)
Observable Student-Level:	0.038	0.056	0.044	0.056	0.029	0.036
$Var(\mathbf{X_s}\mathbf{B})$	(0.003)	(0.004)	(0.006)	(0.007)	(0.005)	(0.006)
Student-Level/						
School-Level Covariance	0.036	0.027	0.031	0.029	0.020	0.021
$2*Cov(\mathbf{X_s}\mathbf{B}, \mathbf{X_{s}}\mathbf{G_1} + \mathbf{Z_{2s}}\mathbf{G_2})$	(0.005)	(0.006)	(0.009)	(0.010)	(0.009)	(0.009)
School-Avg. Student-Level/						
School Char. Covariance	0.001	0.001	0.005	0.007	0.003	0.007
$2*Cov(\mathbf{X_s}\mathbf{G_1}, \mathbf{Z_{2s}}\mathbf{G_2})$	(0.002)	(0.002)	(0.010)	(0.009)	(0.013)	(0.011)
School-Avg. Student-Level	0.013	0.007	0.019	0.013	0.018	0.015
$Var(\mathbf{X_s}\mathbf{G_1})$	(0.004)	(0.002)	(0.011)	(0.009)	(0.013)	(0.005)
School Char.	0.002	0.000	0.025	0.028	0.032	0.033
$Var(\mathbf{Z_{2s}}\mathbf{G_2})$	(0.002)	(0.001)	(0.010)	(0.011)	(0.012)	(0.005)
Unobservable School-Level	0.011	0.009	0.045	0.042	0.053	0.048
$Var(v_s)$	(0.003)	(0.002)	(0.016)	(0.015)	(0.019)	(0.006)

The table reports fractions of the total variance of years of postsecondary education, permanent wages controlling for year of post secondary education, and permanent wages not controlling for years of post secondary education.
Bootstrap standard errors based on re-sampling at the school level are in parentheses.
See the note to Online Appendix Table A20 for additional details.

Table A23—: Potential Bias from Violations of Assumption 6.2

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