

Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition Among Jurisdictions

Jesse M. Rothstein*

First version: November 2002

This version: January 2006

ABSTRACT:

School choice may improve productivity if parents choose well-run schools, but not if parents primarily choose schools for their peer groups. Theoretically, high income families cluster near preferred schools in housing market equilibrium; these need only be effective schools if effectiveness is highly valued. If it is, “effectiveness sorting” will be more complete in markets offering more residential choice. Although effectiveness is unobserved to the econometrician, I test for an observable implication of effectiveness sorting. I find no evidence of a choice effect on sorting, indicating a small role for effectiveness in preferences and suggesting caution about choice’s productivity implications.

(*JEL* H7, I2, L1)

School choice policies may, by aligning administrators’ incentives with parental demand, yield improved efficiency in educational production (Friedman, 1962; Chubb and Moe, 1990). But Hanushek (1981) cautions: “If the efficiency of our school systems is due to poor incentives for teachers and administrators *coupled with poor decision-making by consumers*, it would be unwise to expect much from programs that seek to strengthen ‘market forces’ in the selection of schools” (emphasis added). Poor decision making is not required; parents may rationally choose schools with “pleasant surroundings, athletic facilities, cultural advantages,” (*ibid.*, p. 34) over those that most efficiently pursue academic performance; they may prefer poorly run schools with good peer groups over those that are more effective but enroll worse students (Willms and Echols, 1992; 1993); or they may simply be unable to identify effective schools (Kane and Staiger, 2002). Any factor that leads parents to choose any but the most effective available schools will tend to dilute the incentives for efficient management that choice might otherwise create.

* Department of Economics, Princeton University, Princeton NJ 08544 (e-mail: jrothst@princeton.edu). I thank Alan Auerbach, Tom Davidoff, Caroline Hoxby, Justin McCrary, Rob McMillan, John Quigley, Cecilia Rouse, Emmanuel Saez, Till von Wachter, four anonymous referees, seminar participants at several institutions, and especially David Card for helpful suggestions. I am grateful to the College Board for essential data and to a National Science Foundation Graduate Research Fellowship and the Fisher Center for Real Estate and Urban Economics at UC Berkeley for funding.

This study examines the distribution of student outcomes across schools within metropolitan housing markets for evidence on parental demand. Economists have long noted that parents' choices among residential locations are potentially informative about how more complete choice systems may operate (Tiebout, 1956; Borland and Howsen, 1992; Hoxby, 2000; Rothstein, forthcoming).¹ I ask whether school effectiveness plays a sufficiently important role in these decisions to create meaningful incentives for more productive school management.

I adopt a specific understanding of "effectiveness" appropriate to the question at hand. A substantial portion of between-school differences in student performance can be attributed to differences in student body composition. This portion includes the effects of individual students' characteristics on their own test scores, any direct peer group effects, and any indirect effects of a school's composition on the quality of its instruction. If wealthy schools attract better teachers (Antos and Rosen, 1975) or more parental involvement, this is for my purposes a peer effect; it depends on the quality of the school's administration only via school composition. A school administrator cannot attract demand by offering a school with wealthy students, as these can only be offered if wealthy families demand the school in the first place. Only to the remaining portion of a school's contribution to test scores is "effectiveness."² Choice will not yield improved school performance unless parents demand schools that are effective by this definition.

If parents do demand effective schools, a school's peer group will be correlated with effectiveness in housing market equilibrium, as willingness-to-pay for a demanded school is correlated with the characteristics that produce positive peer effects. This sorting is an obvious source of bias in cross-sectional peer effects estimates, and it similarly confounds direct estimates of

¹ See Rouse (1998); Cullen et al. (2005); Mayer et al. (2002); and Krueger and Zhu (2004) for analyses of several existing non-residential choice programs, though none focuses on the particular issue studied here.

² This definition avoids the need to find observable determinants of effectiveness, which have proved elusive (Hanushek, 1986). My definition, however, ignores a school's contribution to non-test outcomes (e.g. sports or music). If such outcomes motivate parental choices, I will conclude that parents do not demand effective schools. Of course, in this case we would expect schools to compete by improving their non-academic programs, not their test scores.

the relative demand for school effectiveness and peer groups.

The sorting process also provides information, however: The most desired schools, regardless of what makes them desirable, should have the highest housing prices and—under the conventional “single crossing” property—should attract the families with the highest willingness-to-pay. The most desired schools are the most effective ones only if parents attach great importance to effectiveness. As a result, the equilibrium effectiveness-price and -income correlations are increasing functions of the importance of school effectiveness to parental decisions. Moreover, as the number of communities expands, coordination failures that keep high-income families in communities with ineffective schools become less common, and if effectiveness is demanded the income-effectiveness correlation also rises with the number of choices.

This correlation produces a positive bias in naïve, cross-sectional estimates of peer group effects on student performance, so apparent peer effects should be larger in high-choice markets if parents prefer effective to ineffective schools. I test this admittedly indirect implication using data from the National Educational Longitudinal Survey (NELS), a random sample of 8th grade students from roughly 750 metropolitan schools, and from the SAT college entrance exam. The SAT sample is by far the larger—with observations from nearly every high school—though the potential endogeneity of SAT participation may introduce bias.

I find no evidence in either data set that the school-level association between student characteristics and outcomes is stronger in high-choice markets. This result is robust to nonlinearity in the causal peer effect; to several measures of choice and of peer-group quality; to a variety of alternative specifications; to instrumental variables methods that address the potential endogeneity of market structure; and to multiple strategies for dealing with sample selection in the SAT data.

The indirect tests proposed here cannot conclusively determine parental demand. The results nevertheless suggest that effectiveness is not a primary determinant of parental choices, perhaps because variation in effectiveness, as defined here, is not an important determinant of

student performance;³ because parents prefer other neighborhood or school attributes to effectiveness; or because parents cannot distinguish effective from ineffective schools. Any of these would imply that the Tiebout marketplace does not reliably sanction unproductive schools and that Tiebout choice does not create meaningful incentives toward more effective school administration.

I. Allocation of effective schools in Tiebout equilibrium

The basic prediction to be tested derives from a multicommodity model in the spirit of those examined in greater detail by, e.g., Epple et al. (2001), Epple and Sieg (1999), and Fernandez and Rogerson (1996, 1997). I attempt to develop a “best case” for Tiebout choice, and I ignore complications such as private schools; childless families; and non-school locational amenities, such as views, home size, crime, and air quality. I focus on the static allocation of a collection of schools in a metropolitan area with exogenously determined school effectiveness, but I also discuss potential dynamic effects on effectiveness production.

A. A multicommodity model with exogenous effectiveness

A region with population of measure N contains J jurisdictions. Each jurisdiction j contains n identical houses (with $n(J-1) < N \leq nJ$) and a unique, exogenous school effectiveness parameter, μ_j . All houses are owned by absentee landlords, (perhaps a previous generation of parents) who have no current use for them, and they rent for the lowest non-negative market-clearing price.

Family i 's exogenous income is $x_i \in X \subset \mathbf{R}^+$. The metropolitan income distribution function is F . Family i gets utility $U(x_i - b_j, q_j)$ if it rents in jurisdiction j , where the first argument is numeraire non-housing consumption (and b_j is the rental price of a house in community j) and the second is the perceived quality of the jurisdiction's schools. Perceived quality is $q_j = \bar{x}_j \delta + \mu_j$, with \bar{x}_j the jurisdiction's average income and δ the relative importance of peer groups. The externality that derives from the effect of a family's choice on the average income of its chosen community is unpriced. I assume that U is twice differentiable everywhere, with $U_1 > 0$ and $U_2 > 0$. I also assume

³ This would be consistent with the results of Kane and Staiger's (2002) study of school accountability measures.

that the relative marginal utility of quality is increasing in consumption (“single crossing”):

$$\frac{\partial}{\partial c} \left(\frac{U_2(c, q)}{U_1(c, q)} \right) = \frac{U_{12}U_1 - U_{11}U_2}{(U_1)^2} > 0. \quad (1)$$

Given $\delta, J, \{\mu_1, \dots, \mu_J\}$, and F , market equilibrium consists of a set of housing prices $\{b_1, \dots, b_J\}$ and an allocation rule $G: X \mapsto \mathbf{Z}_J$ assigning families to communities such that each housing market clears and each family is satisfied with its assignment, taking other families’ assignments as fixed. Letting $\bar{x}_j \equiv E[x | G(x) = j]$, the conditions are:

EQ1 *Market clearing*. No district is has more residents than houses, and less-than-full districts

have housing prices of zero: $\int \mathbf{1}(G(x) = j) dF(x) \leq \nu'_N$ for each j and

$$\int \mathbf{1}(G(x) = j) dF(x) < \nu'_N \Rightarrow b_j = 0.$$

EQ2 *Nash equilibrium*. No family would prefer another community over the one to which it is

assigned: $U(x_i - b_{G(x_i)}, q_{G(x_i)}) \geq U(x_i - b_k, q_k)$ for all i and all k .

EQ3 *No ties in realized quality*. For any $j \neq k$, $q_j \neq q_k$.⁴

An allocation rule is *admissible* if there exist prices with which it is an equilibrium. In an appendix, I show that there is always at least one admissible rule (and therefore at least one equilibrium), and that a rule is admissible if and only if it produces perfect quality sorting: $q_{G(y)} > q_{G(w)}$ for all y and w where $y > w$ and $G(y) \neq G(w)$. I also show that:

Proposition 1. In any equilibrium, rankings of communities by quality, rent, or incomes are all identical: The n highest-income families live in the highest-quality, highest-rent community; the next n in the second-highest-quality, second-highest-rent community; and so on.

Proposition 2. If $\delta = 0$ there is a unique admissible rule, which sorts families by effectiveness.⁵

⁴ This corresponds to Fernandez and Rogerson’s (1996; 1997) “local stability” notion, and ensures that the equilibrium is stable in the face of small perturbations to communities’ effectiveness or peer quality.

The admissible set expands with δ : Any rule admissible with δ_0 is also admissible with $\delta > \delta_0$.

B. Graphical description of equilibrium allocation

To illustrate the relationships between J , δ , and the equilibrium allocations of peers and effectiveness, Figure 1 presents several sample markets. In each, $x \sim \mathbf{N}(1, 1)$, $\mu_j = j/J$, and $n = N/J$; $J=3$ in the two upper panels and $J=10$ in the lower panels. By Proposition 1, we need only consider allocation rules that permute J quantiles of the income distribution among the J communities. The four panels present two such allocation rules for each J . In each panel, the thin solid line illustrates the allocation of school effectiveness to families of different incomes; the dashed line the allocation of community mean incomes (an increasing function in any admissible rule); and the thick solid line the allocation of q_j when $\delta=1.5$. (Note that when $\delta=0$, $q_j \equiv \mu_j$.) Given δ , admissibility requires that q_j be non-decreasing in x_j .

Panels A and C illustrate the effectiveness-sorted allocations that are the only admissible ones when $\delta=0$. These assign the highest-income quantile to community J , the next to community $J-1$, and so on. These allocations remain admissible when $\delta=1.5$, though now the rent premia associated with higher-numbered communities must be larger to reflect the larger quality disparities.

With positive δ , other allocations become admissible as well. Panel B depicts the “reverse sorted” allocation, in which higher-income students attend schools that are uniformly less effective than those enrolling poorer students, for $J=3$. This is admissible for any $\delta \geq 0.31$, as with this weighting the higher average incomes in districts 2 and 3 dominate their effectiveness deficiencies in parental preferences.⁶ Indeed, for $\delta \geq 0.61$ any permutation of the income terciles is admissible.

Between-decile differences in average income are much smaller than between-tercile

⁵ With a discrete income distribution, there are infinitely many price vectors that support G as an equilibrium but all generate the same ordinal ranking of communities by housing prices. My empirical analysis neglects prices entirely, and focuses solely on the allocation of schools and peers in equilibrium. Bayer et al. (2003) use price data along with a parameterization of the utility function to estimate a model much like this one within a single housing market.

⁶ Income differences between adjacent terciles are 1.1; admissibility of the reverse-sorted rule requires $1.1*\delta > 1/3$.

differences. Thus, with $J=10$ there are some inadmissible permutations whenever $\delta < 3.6$.⁷ Panel D depicts one allocation that is inadmissible with $\delta=1.5$. The third decile of the income distribution is assigned to a community that, because its schools are so ineffective and its students only slightly better, is seen as inferior to that where the second decile resides, violating Proposition 1.

The contrast between the three-district and the ten-district cases indicates a general tendency: Imperfectly effectiveness-sorted allocations—low or negative rank-order correlations between \bar{x} and μ —are admissible when jurisdictions are few and large but not when J grows. Imperfect sorting occurs when families who care about both peers and effectiveness are unwilling to leave an underperforming jurisdiction in favor of a better performer with worse peers. Increased parental choice means closer competitors in income space, limiting the amount of underperformance that high-income families will accept before moving.

C. Comparative statics in J and δ

I use simulations of toy economies like those illustrated above to further illustrate the relationship between preferences, choice, and the central tendency of equilibrium allocations. For each of several (J, δ) combinations, I simulated 5,000 markets. In each simulation, effectiveness parameters for the J communities were drawn independently from a standard normal distribution. I then randomly chose one from among the admissible rules, treating each as equally likely. Figure 2 shows the average effectiveness allocated to families at each income quantile for $J = 3$ and $J = 10$ under each of four values of the parental valuations parameter ($\delta = 0, 1.5, 3,$ and 6 in Panels A, B, C, and D, respectively).

When $\delta = 0$, parents care only for school effectiveness and only perfect effectiveness-sorting is admissible. Panel A thus graphs order statistics for 3 or 10 draws from the μ distribution. As δ grows in the remaining panels, allocations with progressively less complete effectiveness sorting become admissible, and the mean effectiveness experienced at any particular point in the income distribution approaches the unconditional mean of zero. Importantly—see Panels B and C—this

⁷ Average income in the 5th and 6th deciles differs by only 0.25, while their effectiveness might differ by as much as -0.9.

happens faster with 3 than with 10 districts. As δ grows further in Panel D, the difference disappears along with any semblance of sorting on effectiveness in either type of market.⁸

Figure 2 indicates that effectiveness sorting decreases with δ , and that for moderate δ there is more sorting the higher is J . To further illustrate this tendency, I performed the simulations for several additional (J, δ) combinations, for each combination pooling the simulated markets and estimating a market-fixed-effect regression of effectiveness on average income. The coefficients from these regressions are plotted in Figure 3. When δ is small, effectiveness sorting is substantial regardless of J ; when δ is large the coefficients are uniformly small. For moderate δ , the coefficients are larger the more “choice” the market offers.

There is one important caveat to this result: In these simulations, the across-school variance of effectiveness is invariant to choice. Choice might lead to either increases or reductions in the heterogeneity of school effectiveness, depending on whether effective or ineffective schools are most responsive to competition. Changes in heterogeneity affect effectiveness-sorting, so competitive impacts of this sort could confound the choice effect on sorting depicted in Figure 3. I discuss below observable implications of a choice effect on effectiveness production.

II. Estimation

The above simulations suggest that we might assess the magnitude of δ by examining the relationship between the number of school districts serving a market and the income-effectiveness correlation. Without a measure of effectiveness, this correlation cannot be examined directly. But it does have observable implications: The extent of omitted variables bias in a regression of test scores on income depends directly on the correlation of income with unobserved effectiveness.

A. Educational production

I assume an additive reduced-form educational production function. If t_{ijm} is the test score

⁸ Non-monotonicities in Panels C and D derive from the larger differences in average income between adjacent deciles at the extremes of the normal income distribution than near the middle.

(or other outcome measure) of student i when he or she attends school j in market m , I assume that

$$t_{ijm} = a_m + x_{ijm}\beta + \bar{x}_{jm}\gamma + \mu_{jm} + \varepsilon_{ijm}, \quad (2)$$

where a_m is a market-specific intercept capturing unobserved differences between regions' populations or educational systems; x_{ijm} is an index of the student's background characteristics; and \bar{x}_{jm} and μ_{jm} are the average background index of students and effectiveness at school j , respectively. ε_{ijm} is uncorrelated with x_{ijm} , \bar{x}_{jm} , and μ_{jm} , but need not be independent within schools.⁹

Test-score maximizing parents with perfect information will rank schools according to $\bar{x}_{jm}\gamma + \mu_{jm}$ (i.e. will set $\delta = \gamma$). This requires partialling out the portion of the school average,

$$\bar{t}_{jm} = a_m + \bar{x}_{jm}(\beta + \gamma) + \mu_{jm} + \bar{\varepsilon}_{jm}, \quad (3)$$

that is due to β . Parents may use $\delta \neq \gamma$ if they have preferences beyond their children's scores or if they lack sufficient information to perform this partialling out.¹⁰

B. Observable implications of effectiveness sorting

A single-market estimate of equation (3) that omits μ_{jm} yields an \bar{x}_{jm} coefficient that is biased upward by $\theta_m = \text{cov}_m(\mu_{jm}, \bar{x}_{jm}) / \text{var}_m(\bar{x}_{jm})$ relative to the causal effect $\beta + \gamma$. If we could observe θ_m from several markets, we might project it onto a measure of choice, c_m , and a vector of control variables, Z_m :

$$\theta_m = \varphi_0 + c_m\varphi_1 + Z_m\varphi_2 + \omega_m, \text{ with } c'\omega=0 \text{ and } Z'\omega=0. \quad (4)$$

⁹ My notation appears to permit only peer effects that depend upon \bar{x}_{jm} , not those that depend on \bar{t}_{jm} ("endogenous" peer effects, in Manski's (1993) terminology). The latter are nevertheless permitted, as in reduced form they appear as \bar{x}_{jm} effects plus a school-level component of ε_{ijm} .

¹⁰ Researchers vary in their assessments of the relative importance of $\bar{x}_{jm}(\beta + \gamma)$ and μ_{jm} in explaining cross-sectional variation in performance. Chubb and Moe (1990) and Hoxby (1999a) note parents' inability to enforce administrative effort, only an important factor if effectiveness is. On the other hand, Kane and Staiger's (2002) analysis suggests that sampling error and demographic composition swamp effectiveness in the across-school distribution of scores.

The simulations in Section II suggest that $\varphi_1 > 0$ if δ is neither very small nor very large.

Using the projection of μ_{jm} on \bar{x}_{jm} , $\mu_{jm} = \lambda_m + \bar{x}_{jm}\theta_m + v_{jm}$, equation (3) can be re-cast in terms of observables and orthogonal errors:

$$\bar{t}_{jm} = (a_m + \lambda_m) + \bar{x}_{jm}(\beta + \gamma + \theta_m) + (v_{jm} + \bar{\varepsilon}_{jm}) \quad (5A)$$

$$= (a_m + \lambda_m) + \bar{x}_{jm}(\beta + \gamma + \varphi_0) + \bar{x}_{jm}c_m\varphi_1 + \bar{x}_{jm}Z_m\varphi_2 + (\bar{x}_{jm}\omega_m + v_{jm} + \bar{\varepsilon}_{jm}) \quad (5B)$$

This is my basic specification. I estimate regressions of school average test scores on market fixed effects, a measure of the school's peer group quality, and interactions of peer quality with choice and with a vector of market-level controls. I report clustered standard errors that allow for the error structure implied by (5B) (Kézdi, 2002). Note that the resulting estimate of φ_1 reflects only the relationship between choice and the *within-market* allocation of effectiveness; *across-market* variation in average effectiveness or student background is absorbed by the fixed effects.

C. Likely biases

The above specification assumes that the causal peer effect is constant across markets, and in particular that it does not vary with choice. There is some evidence that the educational labor market is more liquid in markets that have many districts competing for teachers' talent than in those with more concentrated governance (Luizer and Thornton, 1986). Choice may thus facilitate teacher sorting by making it easier for a high- \bar{x}_{jm} school to attract good teachers. This is simply one channel by which a school's composition determines output, so for my purposes is a peer effect. It would mean that the causal peer effect γ increases with c_m ; φ_1 would capture this, so might be positive even in the absence of the effectiveness sorting discussed above.

Similarly, mismeasurement of \bar{x}_{jm} likely biases the estimate of φ_1 upward relative to the peer group main effect $\beta_0 + \gamma + \varphi_0$. In single-market estimates of the peer effect, measurement error would attenuate the estimated effect in proportion to the degree to which the reliability of \bar{x}_{jm} is reduced. Choice increases stratification—a clear implication of the model, and demonstrated

empirically in the online appendix—and stratification makes \bar{x}_{jm} more reliable.¹¹ The proportional attenuation will therefore tend to decrease with choice. In my pooled model, so long as the reliability of \bar{x}_{jm} increases with choice, measurement error leads to an upward bias in the estimated φ_1 relative to that in $(\beta_0 + \gamma + \varphi_0)$.¹² I present a specification below in which \bar{x}_{jm} is instrumented with an independent measure (as are its interactions), with little effect on $\hat{\varphi}_1$.

D. Supply-side effects

As noted earlier, competition may affect the variance of effectiveness. If this effect is negative, choice might not have positive effects on θ even when δ is small.¹³ Note that

$$\text{var}_m(\mu_{jm}) = \text{var}_m(\bar{x}_{jm})\theta_m^2 + \text{var}_m(v_{jm}). \quad (6)$$

Structural assumptions about the causal peer effect and about the variance components in (5A) allow calibration of the components of (8). A natural approximation is that the within-MSA residual variance of school mean scores is attributable to $\text{var}_m(v_{jm})$ and $\text{var}_m(\bar{\varepsilon}_{jm})$, with the latter inversely proportional to the within-school sample size.¹⁴ The coefficients from (5A) can then be combined with assumptions about γ to estimate θ_m and, via it, $\text{var}_m(\mu_m)$ and $\rho_m \equiv \text{corr}_m(\bar{x}_{jm}, \mu_{jm})$. I discuss an

¹¹ Stratification implies a higher true variance of the peer group, and therefore a larger signal component of the signal-to-noise ratio. Also, schools in more stratified markets are more internally homogeneous and school-level averages are thus more reliable. Finally, in markets that are more heavily stratified unobserved student characteristics are likely to be more strongly associated with observed characteristics at the school level, making the observed variables better indicators of the true peer-group quality.

¹² There is one channel by which measurement error might reduce the reliability of \bar{x}_{jm} and bias $\hat{\varphi}_1$ downward:

Average school size declines slightly with my choice measure. I present a specification below that controls directly for the peer group interaction with a polynomial in the school-level sample size, with no impact on $\hat{\varphi}_1$.

¹³ In Hoxby's (1999b) model, competition levels-up the lowest-quality schools, reducing the variance of quality and raising the average. Tests of the latter prediction have yielded at best mixed results (Rothstein, forthcoming).

¹⁴ This implicitly rules out Mankiw's (1993) endogenous and correlated peer effects, each of which would introduce a school-level error component beyond effectiveness.

analysis along these lines in Section V.

III. Data

The empirical strategy outlined above requires data on the joint distribution of peer groups and outcomes *across* schools *within* educational markets that differ in the amount of Tiebout choice. I approximate educational markets by 1990 Metropolitan Statistical Areas (MSAs). I use two data sets for information about student outcomes. First, the National Educational Longitudinal Survey (NELS) of 1988 surveyed and tested approximately 25 8th-grade students from each of about 750 metropolitan schools. I focus on two outcomes: The 8th-grade composite score and an indicator for whether the student was still in school or had graduated at the time of the 1992 follow-up survey.

Second, I use a data set consisting of observations on 330,000 metropolitan SAT-taker observations from the cohort that graduated from high school in 1994. The SAT is an entrance exam required by many colleges, so is taken by a large fraction of college-bound students. I use a sample of about one third of SAT-takers from the 1994 cohort.¹⁵ The data contain students' SAT scores along with high school indicators and several family background measures.

The SAT data offer the important advantage that the sample includes a substantial number of students from nearly every high school in the United States, whereas the NELS offers only two or three schools from each MSA and small samples within each school. Moreover, parents are likely to be particularly concerned with a school's effect on students' SAT scores, as these scores have personal consequences that the NELS tests do not. Finally, because SAT-takers' parents are presumably above average in both their financial resources and their involvement in their children's education, this population likely has a high willingness-to-pay for a house near a high-quality school.

On the other hand, endogenous selection into SAT-taking may introduce biases. I take several steps to try to reduce selection bias in the SAT analyses. First, I limit my sample to

¹⁵ SAT-takers who reported their ethnicity were sampled with probability one if they were black or Hispanic or if they were from California or Texas, and with probability one-quarter otherwise. Due to an error in the processing of the file, students who did not report an ethnicity are excluded. In recent years, these comprise around 12 percent of SAT-takers.

metropolitan areas in 23 “SAT states,” where most college-bound students take the SAT.¹⁶ Second, I include the metropolitan SAT-taking rate in Z_m in all SAT regressions. Finally, I present several alternative specifications designed to locate selection bias in the SAT analyses.

For data reasons, I use test score levels rather than so-called “gain scores” to measure student achievement. My models should thus be thought of as reduced forms for the cumulative effects of peer groups and school effectiveness through grade 8 (NELS) or 12 (SAT).

MSA-level control variables are drawn from the 1990 Census, Summary Tape File 3C (U.S. Census Bureau, 1993). One key variable, the degree of equalization built into the state school finance rule—which may affect the benefits of attending a school in a wealthy neighborhood—is not available from the Census. I use two measures of finance equalization from Card and Payne (2002): An indicator for whether the state had a Minimum Foundation Plan financing rule in 1991, and the coefficient from a state-specific regression of school district per-capita operating expenditures on median family income.¹⁷

Table 1 reports summary statistics at the metropolitan, school, and individual levels. Columns A and B report statistics for the full set of MSAs and for the NELS data, while columns D and E report the same statistics for MSAs in SAT states and for the SAT sample. Code for variable construction is available in an online appendix; additional information about the algorithm used to assign schools to MSAs is available from the author.

A. Measuring the peer group

It is helpful to have a one-dimensional index of student quality at each school. To create one, I estimate a flexible regression of individual NELS and SAT scores on student characteristics, controlling for school fixed effects:

¹⁶ In the remaining states, SAT-takers are primarily students hoping to attend out-of-state colleges (Clark, 2003). Among the SAT states, there is no state-level correlation between participation rates and average scores.

¹⁷ I am grateful to David Card for providing these variables, which I allocate to MSAs spanning several states on the basis of population shares. Because they are unavailable for Alaska and Hawaii, Anchorage and Honolulu are excluded from all analyses. Estimates that include these MSAs but exclude the finance variables produce similar results.

$$t_{jm} = \psi_{jm} + W_{jm} \beta_W + e_{jm}.^{18} \quad (7)$$

I next define an index of peer quality as the school-level average of the predicted values (excluding the estimated school effect) from (9), $\bar{x}_{jm} = \bar{\psi} + \overline{W}_{jm} \hat{\beta}_W$. This construction implicitly normalizes β from equation (2) to one, and the peer-group index has the interpretation of the school sample's predicted average performance at a nationally representative school.¹⁹ Table 1 reports summary statistics for the index. Specification checks reported below explore the sensitivity of the results to deviations from the single-index assumption.

B. Measuring Tiebout choice

With some exceptions, children may not attend public schools outside their home districts. Most districts operate multiple schools, and establish attendance zones for each school. Thus, families may in principle exercise Tiebout choice among both districts and schools.

Following previous authors (Borland and Howsen, 1992; Hoxby, 2000), I focus on district-level choice. One reason for this is that within-district attendance-zone boundaries are more flexible and less permanent than district boundaries, reducing the scope for school choice via residential location.²⁰ A second reason is less principled: School size varies relatively little across metropolitan

¹⁸ In the SAT analysis, W includes 12 ethnicity-gender indicators, the interactions of 10 maternal with 10 paternal education categories, and the interactions of six ethnicity with 12 family income categories. The NELS version includes 15 income categories and the interaction of race with gender and with two parental education dummies. Note that the β_W coefficients are identified only from within-school variation in W_{jm} , while equation (5B) is identified only from across-school variation.

¹⁹ An index calculated from 1995 SAT data (with independent β_W estimates) correlates 0.94 with the 1994 version at the school level, indicating high reliability. I also explored allowing β_W to vary with region. The resulting peer quality indices were very similar and produced similar results in the analyses below. Finally, note that

$\bar{t}_{jm} \equiv \hat{\psi}_{jm} + \overline{W}_{jm} \hat{\beta}_W \equiv (\hat{\psi}_{jm} - \bar{\psi}) + \bar{x}_{jm}$. Thus, were I to replace school average test scores in (5B) with the estimated fixed effects $\hat{\psi}_{jm}$, the \bar{x}_{jm} main effect would decline by precisely one but interaction coefficients would be unaffected.

²⁰ On desegregation remedies, which frequently modify within-district school attendance zones but almost always respect district boundaries, see Welch and Light (1987), Orfield (1983), and *Milliken v. Bradley* 418 U.S. 717, 1974. Kane et al.

areas, and the number of schools varies almost perfectly with metropolitan population.

I use Hoxby's (2000) Herfindahl-style index of the concentration of public enrollment in the largest districts to measure choice. If e_{dm} is the enrollment of district d in MSA m and e_m the total enrollment in the MSA, the choice index is $c_m = 1 - \sum_{d \in m} (e_{dm} / e_m)^2$.²¹ The online appendix presents evidence that the choice index captures meaningful variation in parents' ability to exercise Tiebout choice: Controlling for other metropolitan characteristics, c is negatively associated with private enrollment rates and positively associated with racial stratification across schools. Columns C and F of Table 1 report the correlations of all other variables with the choice index. High-choice markets are larger; have fewer blacks and Hispanics, higher incomes, and less inequality; and are in states with less equitable school finance.

IV. Empirical Results: Within-market sorting

I begin my analysis of the within-MSA relationship between peer quality and average scores with nonparametric estimates that allow the peer effect to be a nonlinear function of mean student quality. I group MSAs into quartiles by c_m and estimate separate school-level kernel regressions of test scores on the student background index for each quartile. Figure 4 displays the estimated functions for NELS 8th-grade composite scores (left panel) and SAT scores (right panel).²² Neither indicates important differences among quartiles in reduced-form educational production functions, nor any substantial nonlinearity in the peer group-test score relationship. I therefore impose a linear structure throughout the remainder of the analysis.

Table 2 reports basic regression results. Column 1 reports a simple specification for NELS

(forthcoming) discuss frequent desegregation-related changes in attendance zone boundaries within one large school district, while Reber (2005) discusses cross-district residential responses to within-district desegregation policies.

²¹ Following Urquiola (2005), I use only enrollment in grades 9-12 for this computation.

²² NELS scores are scaled throughout to the same mean and student-level standard deviation as SAT scores.

8th-grade scores in which the Z vector (equation 5B) contains only census division effects.²³ The peer-group main effect—which estimates $\beta + \gamma + \varphi_0$, with β standardized to one—is a surprisingly large 1.70. The combination of peer effects and effectiveness sorting in zero-choice MSAs is thus 70% as large as the effect of own characteristics on students’ test scores. There is no indication, however, that high-choice MSAs exhibit a stronger apparent peer effect: $\hat{\varphi}_1$ is -0.19. Columns 2 and 3 add additional MSA-level control variables, each interacted with the background index.²⁴ Standard errors grow with the additional controls, but $\hat{\varphi}_1$ is quite stable.

Columns D through F report similar analyses for the NELS retention rate, the fraction of a school’s 8th-grade sample that is still in school at the time of the follow-up survey four years later. (The sample average is 0.84.) $\hat{\varphi}_1$ is quite negative here. Columns G through I turn to the SAT data, with an interaction of the MSA SAT-taking rate with the background index included in each specification. Standard errors are much smaller in the SAT data, but point estimates are similar.

Nothing in Table 2 offers evidence of a meaningful positive choice interaction, and eight of the nine point estimates are negative. The positive coefficient is very small and insignificant, and moreover derives from a saturated specification with 18 MSA-level controls and only 177 MSAs. In the SAT sample, where estimates are sufficiently precise to distinguish interesting hypotheses, even at the upper limit of the confidence interval the choice effects (φ_1) are small relative to the implied

²³ Unless otherwise stated, schools are weighted by the sum of individual sampling weights, adjusted at the MSA level to weight MSAs by their 17-year-old populations. Z_m and c_m are demeaned before being interacted with \bar{x}_{jm} . Standard errors are clustered to allow for arbitrary correlation among schools in the same MSA (or CMSA, in the largest cities). I do not adjust for the sampling variation in the β_w coefficients: A bootstrap analysis of the background index construction indicates reliability ratios of about 0.96 (SAT) and 0.93 (NELS), suggesting that understatement in my reported standard errors is likely small (Murphy and Topel, 1985).

²⁴ The Z variables added in columns 2, 5, and 8 are the log of population, the fractions black and Hispanic, log mean household income, the gini coefficient for household income, the fraction of adults with college degrees, and the two school finance variables. Columns 3, 6, and 9 also add the log of population density, the fraction high school dropouts, and the square of the white population share.

average peer-group effect ($\gamma + \varphi_0$) of 0.68.

Table 3 presents a variety of specification tests. Estimates are presented for NELS 8th-grade scores in Columns 1 and 2 and for SATs in Columns 3 and 4. Row A repeats the baseline specifications from Columns 2 and 8 of Table 2. Row B adds interactions of the peer group with quadratics in the school sample size and its inverse. Because high-choice MSAs have somewhat smaller schools, mismeasurement of the peer group in small schools may bias the peer group-choice interaction in row A downward. The size controls would absorb any such attenuation, but the choice effect is essentially unchanged when they are included.

Rows C, D, and E explore alternative specifications of the peer effect, first allowing it to depend on the standard deviation of student background as well as the mean, then allowing the racial composition of the school to have a separate effect from that of the background index, and finally treating average family income, rather than the full background index, as the key sorting variable.²⁵ None of these has much impact on the choice-background interaction coefficient.

Rows F, G, and H explore alternative samples, first excluding private schools (enrollment in which does not depend on residential location), then excluding the 26 zero-choice (i.e. one-district) MSAs, and finally excluding as well schools from central city school districts, which may not be realistic choices for the high-income parents whose preferences determine the equilibrium allocation. None indicates a sizable positive choice effect, and indeed the latter two samples in the SAT data produce large negative coefficients, one significant. Rows I through L take a different approach, attempting to measure directly the number of choices available to parents who are willing to consider only schools offering a minimum peer group quality. I modify the “market shares” used for computation of the choice index, in row I using districts’ shares of SAT-takers—a proxy for college-bound students—and in row J using shares of families with incomes above the MSA median. Rows K and L then repeat these using only schools with above-average SAT participation rates.

²⁵ This specification includes main effects for both income and the background index, but interactions only between income and ϵ_m and Z_m .

None of these specifications indicates a positive choice effect.

The next rows present instrumental variables estimates. I first investigate the possibility that the choice index is endogenous to school effectiveness or to the sorting process. This might be true if, as Hoxby (2000) proposes, school districts have been less willing to consolidate with their neighbors in areas with low average effectiveness. Hoxby proposes generating exogenous variation from differences across MSAs in conditions influencing the pre-consolidation governance structure. In Row M, I instrument for the choice index with the number of streams flowing through the metropolitan area (Rothstein, forthcoming), which might capture geographic differences in the optimal district layout before modern transportation networks were developed. Row N uses a stronger instrument, a choice index computed using information on the number of school districts operating in the MSA in 1942 (Gray, 1944). By the logic underlying the streams approach, this historical measure should be a valid instrument if school quality is at all transitory. The IV results are noisy, with standard errors nearly double those from the OLS model, and one coefficient is positive.²⁶ None of the estimates come close to rejecting either zero effect or the OLS point estimate, however, and there seems little evidence of bias in the OLS estimates.

Row O presents another IV model, in which I instrument the background index itself and its interactions. I construct an independent estimate of the index for SAT schools using data from the 1995 cohort of SAT-takers. The IV estimate of the peer group main effect is slightly larger than OLS, consistent with measurement error in the index, but there is no effect on the coefficient of interest.

The final rows of Table 3 present specifications designed to look for signs of bias in the SAT analyses from endogenous SAT participation. Row P reports a model that includes a control for the school-level participation rate. Here, the choice coefficient is slightly positive but very small.²⁷ Row Q takes another approach to variation in school participation rates, dropping from the sample any

²⁶ *F* statistics for tests of the exclusion of the instruments from an MSA-level first stage are 11.7 and 129.4, respectively.

²⁷ The school SAT-taking-rate coefficient is 57 (s.e. 10), the opposite of the expected sign if selection is positive.

Analyses that replace the school- or metropolitan-level participation rates with inverse Mills ratios yield similar results.

school with a rate below the MSA average, again with little effect.

In Row R, I take advantage of a variable on the SAT data characterizing each student's high school class rank (reported as the first or second decile, or by quintile below that). I discard all observations reporting a rank in the bottom 40 percent of their class, then reweight the remaining observations so that the weighted rank distribution at each school is balanced, with 1/6 of the sample in each of the top two deciles and 1/3 in each of the next two quintiles. If selection into SAT-taking is uncorrelated with potential scores conditional on school and rank, this re-weighting should recover estimates that would be obtained were scores available for all highly-ranked students. Results from the reweighted sample are again similar to those obtained earlier.

The next two rows turn from selection in the dependent variable to the effect of sample selection on the background index, which is estimated only over sample individuals and may not accurately measure the school peer group.²⁸ There is no alternative source of detailed school-level background information. District-level mean incomes are available, however, from Census data.

Row S estimates the basic model on data collapsed to the district level, still using the sample background index. (Private schools and single-district MSAs are excluded.) Standard errors are larger, but results are otherwise similar to the school-level specification in the SAT data. In the NELS data, however, the choice coefficient is large and negative. Row T adds a control for the log of median household income in the district (standardized to the same standard deviation as the background index) and uses this variable in the interaction terms, retaining a single control for the background index main effect. This has little effect in the NELS data, but in the SAT data the income-choice coefficient is positive and large, though insignificant.

The change in results in the SAT sample between rows S and T might indicate that reliance on a background index estimated from SAT-takers alone biases $\hat{\phi}_1$ downward. This provides reason

²⁸ It is plausible that SAT-takers' friends are drawn disproportionately from other SAT-takers at the school, which might make the SAT-taker background index average a better peer-group measure for the current purpose than would be a schoolwide measure.

for caution in interpretation of the SAT-based results. However, the large negative coefficient from a similar specification in the NELS—imprecisely estimated, but with a small enough confidence region to reject the SAT point estimate—suggests that we should not be too quick to conclude that the choice-background index interaction would be positive if only we had a superior background measure, particularly given the stability of $\hat{\phi}_1$ across rows P-R, each of which should have ameliorated selection bias in \bar{x}_{jm} . Given the full set of results in Tables 2 and 3, it seems reasonable to conclude tentatively that the interaction coefficient of interest is approximately zero.

V. Effectiveness supply

If competition leads to reduced variation in effectiveness, θ_m might not be larger in high-choice markets regardless of parental demand. This would imply that the residual variance of effectiveness after regressing it on the peer group— $\text{var}_m(v_{jm})$ in (6)—should be unambiguously lower in high-choice markets. To evaluate this, I estimated an MSA-by-MSA regression of test scores on the peer group. I next regressed the within-MSA residual variance from this regression— $\text{var}_m(v_{jm} + \bar{\varepsilon}_{jm})$ —on choice and the usual vector of other MSA-level controls.²⁹ The choice coefficient in this regression was negative and significant, indicating that a one-unit increase in choice reduces the across-school standard deviation of residual scores by 20 percent of its zero-choice average. This is at least consistent with the claim that high-choice MSAs exhibit somewhat less variability in school effectiveness, which complicates the interpretation of the earlier results as informative about parental demand.

Both $\text{var}_m(\mu_{jm})$ and $\rho_m = \text{corr}_m(\mu_{jm}, \bar{x}_{jm})$ are functions of the observable data and the unknown peer-effect parameter γ . For several γ values, I computed each and tested for choice effects. For reasonable γ , choice’s effect on $\text{var}_m(\mu_{jm})$ was comparable to that on residual variance,

²⁹ The regression also included a control for the MSA average of N_{jm}^{-1} , where N_{jm} is the number of observations at school j , to absorb differences in $\text{var}_m(\bar{\varepsilon}_{jm})$.

while choice had a positive but small and statistically insignificant effect on ρ_m . These results cast doubt on my maintained assumption of constant effectiveness variance, though they suggest that deviations from this assumption are likely reasonably small.

VI. Conclusion

The effects of choice policies on the incentives faced by school administrators depend crucially on how parents choose. If parents have strong preferences for well-run, productive schools that focus on academic skills related to test performance, we might expect administrators to compete for students by implementing policies that lead to increased scores. If parents look for other characteristics in schools, however, incentives toward productive management can be diluted. In particular, if the peer group is important to parental preferences, coordination failures can prevent the market from rewarding school effectiveness.

Strong parental preferences for effective schools produce a correlation between effectiveness and the peer group in Tiebout equilibrium, a correlation that is stronger when parents have more ability to buy their way into a desired school. This correlation produces an upward bias in cross-sectional estimates of the peer effect.³⁰ Tests of the relationship between inter-district choice and a reduced form estimate of the peer effect offer no evidence that the peer group coefficient varies systematically with choice. Even at the upper extreme of the estimated confidence intervals (in the SAT data), the performance gap between more- and less-desirable schools is not meaningfully larger in markets with decentralized governance than in those with less Tiebout choice. Moreover, although the analysis relies on observational rather than experimental variation in choice, the coefficient of interest does not seem particularly sensitive to the choice of control variables or to reasonable modifications to the sample or specification. Although some analyses indicate the

³⁰ The choice effect on “effectiveness sorting” disappears if parents attach zero weight to either effectiveness or the peer group. The latter is implausible, both because parents seem to believe that peer effects are important and because parents are likely unable to correct for the mechanical correlation between average test scores and peer group quality that arises from the effect of students’ own characteristics on their performance.

possibility of bias in the SAT data arising from endogenous selection into SAT-taking, estimates from un-selected NELS data are similar, if less precise.

There is some evidence that choice is associated with reductions in the dispersion of effectiveness across schools, which weakens the theoretical predictions. This effect appears to be small, however, consistent with recent evidence that the choice effect on average effectiveness is negligible (Rothstein, forthcoming). In light of this, I tentatively conclude that choice does not have sufficiently strong effects on the production of school effectiveness to invalidate my primary analysis. This, however, bears further investigation.

The most plausible explanations for the current results are that parents place a low weight on school effectiveness in their preferences over neighborhoods;³¹ that parents value effectiveness greatly but lack the information needed to identify effective schools; or that variation in “effectiveness” as defined here, encompassing only school characteristics not causally dependent upon the enrolled population, is responsible for only a small share of the across-school variation in student outcomes. Under any of these, there is little theoretical support for the claim that Tiebout-choice markets create strong incentives for school administrators to exert greater effort to raise student performance (Chubb and Moe, 1990). Of course, under the second explanation the provision of more complete information—e.g., new accountability measures that better distinguish effectiveness from the peer group—might materially change the nature of Tiebout equilibrium, as other evidence indicates that parents do respond to accountability scores (Figlio and Lucas, 2004).

Great caution is required in generalizing from this paper’s results to choice markets that de-link school assignment from residential location, as choices may be sensitive to factors (non-school neighborhood amenities, for example) that have not been considered here. Moreover, voucher programs that encourage market entry may provide more choice than is achievable in even the most decentralized of governmental structures. It nevertheless seems unlikely that choice programs can

³¹ With heterogeneous preferences, choice can of course increase the match quality between parents and schools. Nothing in the analysis here can reject the claim that *some* parents use the opportunity to select effective schools, although it does suggest that most do not.

produce substantial market pressure toward greater school productivity without careful attention in their design to the role of peer groups in parental choices.

References

- Antos, Joseph R. and Rosen, Sherwin.** "Discrimination in the Market for Public School Teachers." *Journal of Econometrics*, 1975, 3(2), pp. 123-50.
- Bayer, Patrick; Ferreira, Fernando and McMillan, Robert.** "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." Discussion paper #872, Economic Growth Center, Yale University, 2003.
- Borland, Melvin V. and Howsen, Roy M.** "Student Academic Achievement and the Degree of Market Concentration in Education." *Economics of Education Review*, 1992, 11(1), pp. 31-39.
- Card, David and Payne, A. Abigail.** "School Finance Reform, the Distribution of School Spending, and the Distribution of SAT Scores." *Journal of Public Economics*, 2002, 83(1), pp. 49-82.
- Chubb, John and Moe, Terry M.** *Politics, Markets, and America's Schools*. Washington, D.C.: The Brookings Institution, 1990.
- Clark, Melissa.** "Selection Bias in College Admissions Tests." Mimeo, Princeton University, 2003.
- Cullen, Julie Berry; Jacob, Brian A. and Levitt, Steven D.** "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools." *Journal of Public Economics*, 2005, 89(5-6), pp. 729-760.
- Epple, Dennis ; Romer, Thomas and Sieg, Holger.** "Interjurisdictional Sorting and Majority Rule: An Empirical Analysis." *Econometrica*, 2001, 69(6), pp. 1437-65.
- Epple, Dennis and Sieg, Holger.** "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy*, 1999, 107(4), pp. 645-81.
- Fernandez, Raquel and Rogerson, Richard.** "Income Distribution, Communities, and the Quality of Public Education." *Quarterly Journal of Economics*, 1996, 111(1), pp. 135-64.
- _____. "Keeping People Out: Income Distribution, Zoning, and the Quality of Public Education." *International Economic Review*, 1997, 38(1), pp. 23-42.
- Figlio, David N. and Lucas, Maurice E.** "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review*, 2004, 94(3), pp. 591-604.
- Friedman, Milton.** *Capitalism and Freedom*. Chicago: University of Chicago Press, 1962.
- Gray, E. R.** *Governmental Units in the United States 1942*. Washington: Bureau of the Census, 1944.
- Hanushek, Eric A.** "Throwing Money at Schools." *Journal of Policy Analysis and Management*, 1981, 1(1), pp. 19-41.
- _____. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, 1986, 24(3), pp. 1141-77.
- Hoxby, Caroline M.** "The Effects of School Choice on Curriculum and Atmosphere," in S. E. Mayer and P. E. Peterson, eds., *Earning and Learning: How Schools Matter*. Washington D.C.: Brookings Institution Press, 1999a, 281-316.
- _____. "The Productivity of Schools and Other Local Public Goods Producers." *Journal of Public Economics*, 1999b, 74(1), pp. 1-30.
- _____. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 2000, 90(5), pp. 1209-38.
- Kane, Thomas J. and Staiger, Douglas O.** "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, 2002, 16(4), pp. 91-114.

- Kane, Thomas; Staiger, Douglas O. and Cellini, Stephanie R.** “School Quality, Neighborhoods, and Housing Prices: The Impacts of School Desegregation.” *American Law and Economics Review*, forthcoming.
- Kézdi, Gábor.** “Robust Standard Error Estimation in Fixed-Effects Panel Models.” Mimeo, University of Michigan, February 2002.
- Krueger, Alan B. and Zhu, Pei.** “Another Look at the New York City School Voucher Experiment.” *American Behavioral Scientist*, 2004, 47(5), pp. 658-98.
- Luizer, James and Thornton, Robert.** “Concentration in the Labor Market for Public School Teachers.” *Industrial and Labor Relations Review*, 1986, 39(4), pp. 573-84.
- Manski, Charles F.** “Identification of Endogenous Social Effects: The Reflection Problem.” *Review of Economic Studies*, 1993, 60(3), pp. 531-42.
- Mayer, Daniel P.; Peterson, Paul E.; Myers, David E.; Tuttle, Christina Clark and Howell, William G.** “School Choice in New York City after Three Years: An Evaluation of the School Choice Scholarships Program.” Mathematica Policy Research, Inc., 2002.
- Murphy, Kevin M. and Topel, Robert H.** “Estimation and Inference in ‘Two-Step’ Econometric Models.” *Journal of Business and Economic Statistics*, 1985, 3(4), pp. 370-379.
- Orfield, Gary.** *Public School Desegregation in the United States, 1968-1980*. Washington, D.C.: Joint Center for Political Studies, 1983.
- Reber, Sarah.** “Court-Ordered Desegregation: Successes and Failures in Integration since *Brown vs. Board of Education*,” *Journal of Human Resources*, 2005, 40(3), pp. 559-590.
- Rothstein, Jesse M.** “Does Competition among Public Schools Benefit Students and Taxpayers? Comment.” *American Economic Review*, forthcoming.
- Rouse, Cecilia Elena.** “Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program.” *The Quarterly Journal of Economics*, 1998, 113(2), pp. 553-602.
- Tiebout, Charles M.** “A Pure Theory of Local Public Expenditures.” *Journal of Political Economy*, 1956, 64(5), pp. 416-24.
- U.S. Census Bureau.** “Census of Population and Housing, 1990 [United States]: Summary Tape File 3c.” 1993.
- Urquiola, Miguel.** “Does School Choice Lead to Sorting? Evidence from Tiebout Variation.” *American Economic Review*, 2005, 95(4), pp. 1310-1326.
- Welch, Finis and Light, Audrey.** *New Evidence on School Desegregation*. Washington, D.C.: United States Commission on Civil Rights, 1987.
- Willms, J. Douglas and Echols, Frank H.** “Alert and Inert Clients: The Scottish Experience of Parental Choice of Schools.” *Economics of Education Review*, 1992, 11(4), pp. 339-50.
- _____. “The Scottish Experience of Parental School Choice,” in E. Rasell and R. Rothstein, eds., *School Choice: Examining the Evidence*. Washington, D.C.: Economic Policy Institute, 1993.

Figure 1: Illustrative allocations of school effectiveness and community desirability.

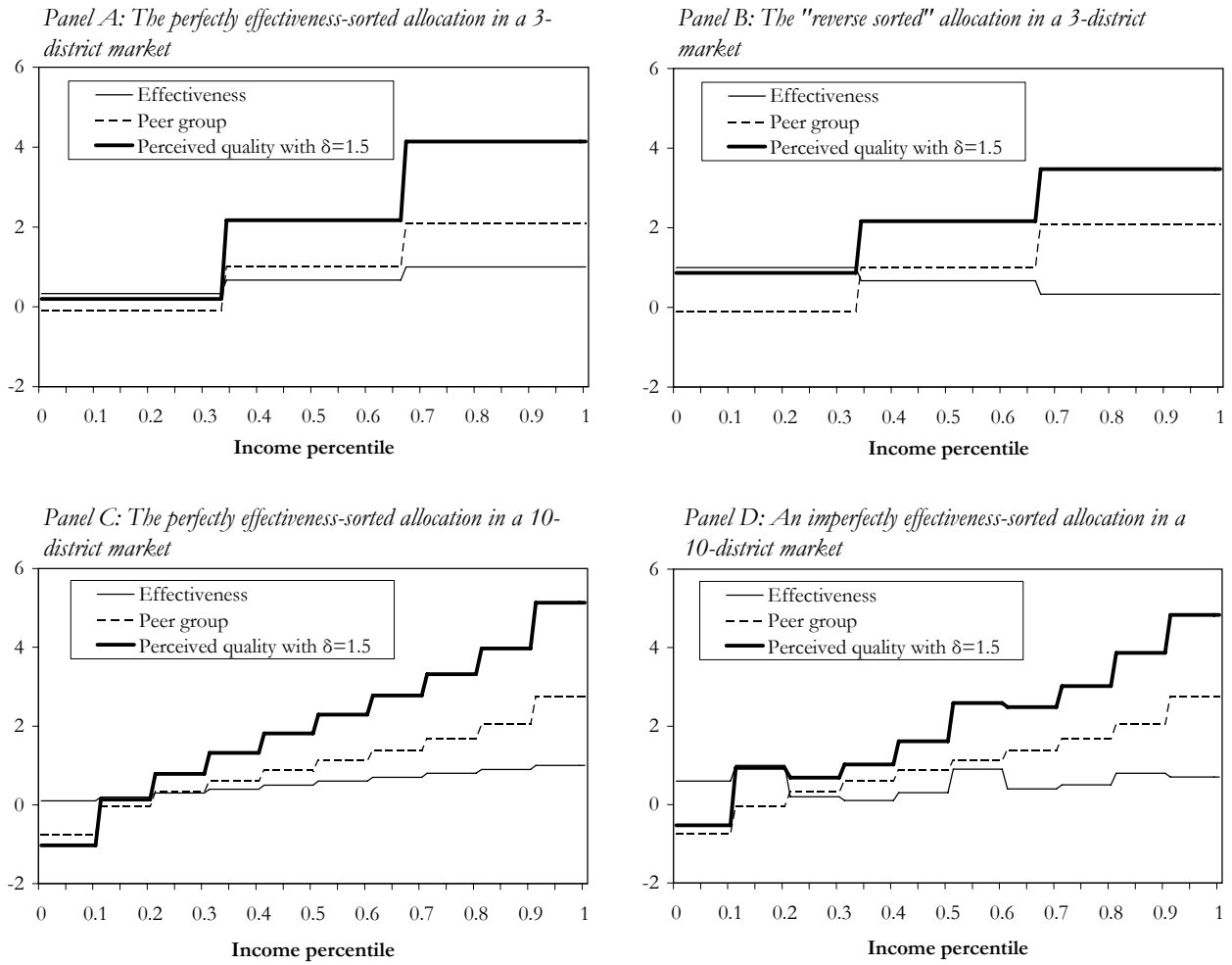
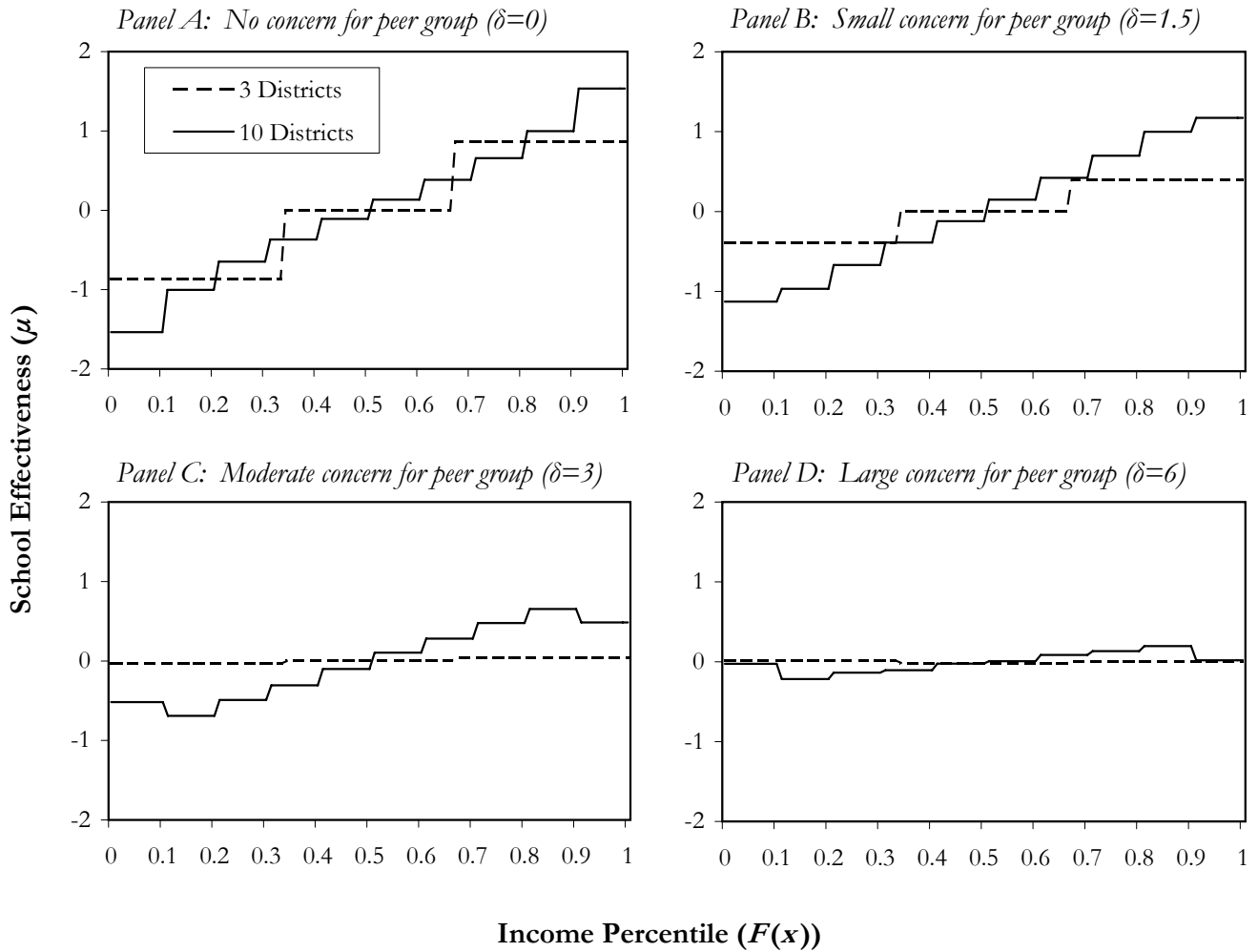
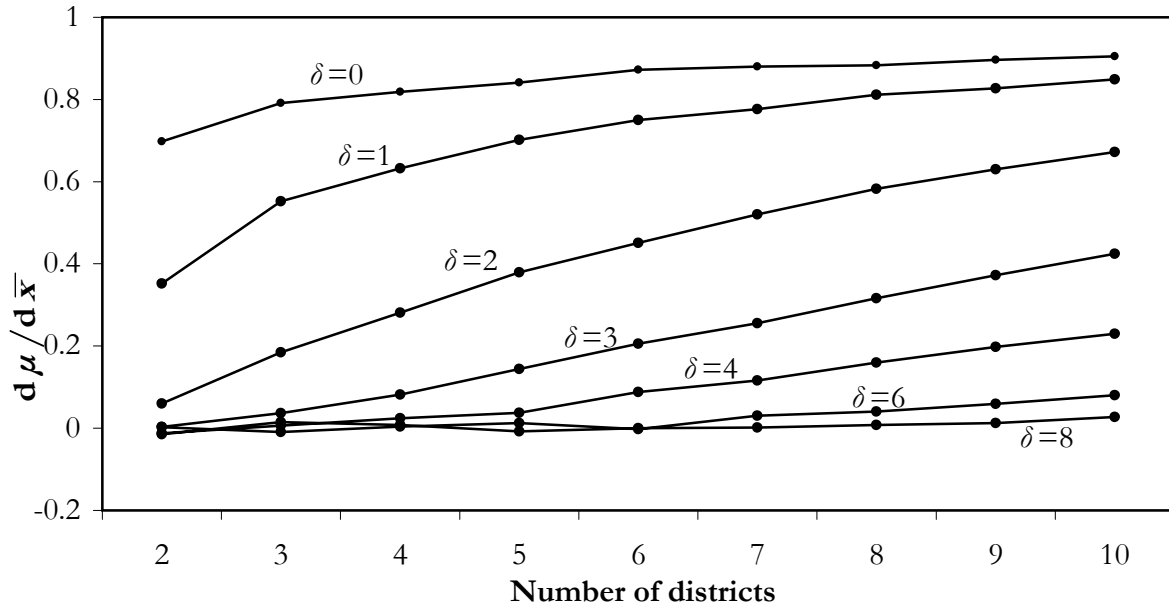


Figure 2: Effectiveness experienced in average simulation equilibrium, by income percentile, number of districts, and parental concern for peer group.



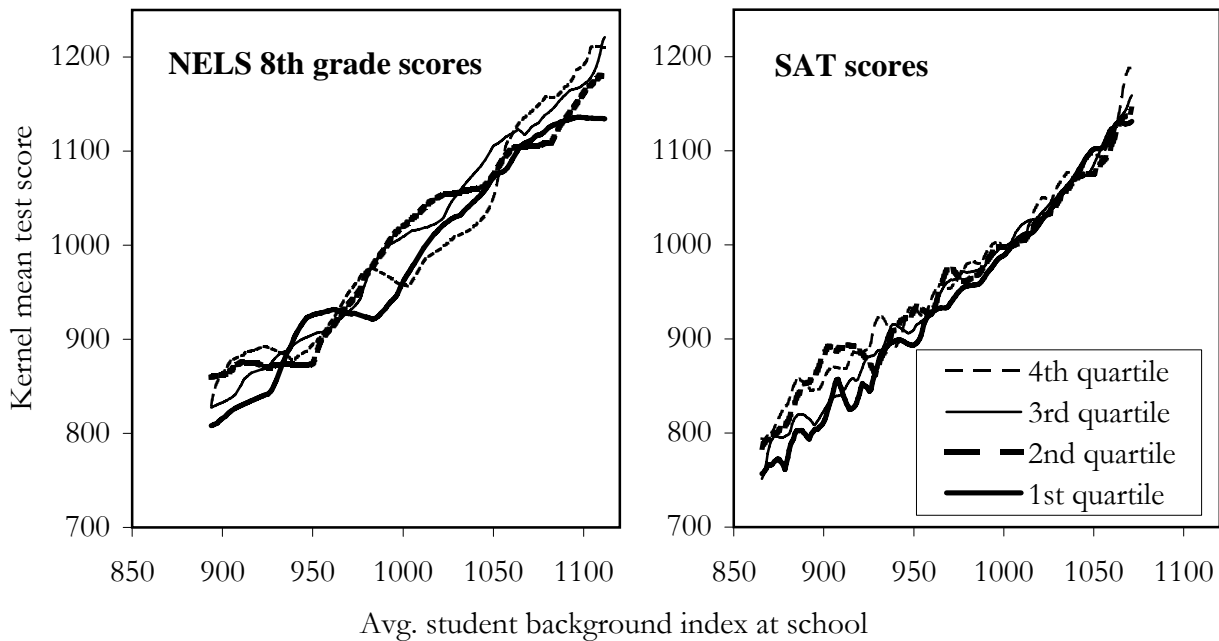
Notes: Each horizontal line segment represents the effectiveness of schools attended by families in the indicated income range, averaged over 5,000 simulated equilibria. See text for details.

Figure 3: Slope of effectiveness with respect to average income in simulated equilibria, by number of districts and δ .



Notes: Each point represents the coefficient from a district-level regression of effectiveness on equilibrium average income, using as data 5,000 simulated markets (with the indicated number of districts in each market and with parental preferences characterized by the indicated δ) and including a fixed effect for each market. See text for details.

Figure 4: Kernel estimates of the student background-test score relationship, by choice quartile.



Notes: "1st quartile" is MSAs with choice index above 0.92; 2nd is indices 0.85-0.92; 3rd is 0.74-0.85; and 4th contains MSAs with choice indices below 0.74. Graphs show school-level kernel means within each quartile. An Epanechnikov kernel (bandwidth 10 in NELS and 3 in SAT data) is used, and schools are weighted by the sum of NELS/SAT sampling weights, adjusted at the MSA level to weight MSAs by their 17-year-old populations. Series are truncated to exclude the top and bottom 2 percent of school-level background index values.

Table 1: Summary statistics for MSAs, individuals, and schools.

	All MSAs			MSAs in SAT states		
	Mean	S.D.	MA-level correlation	Mean	S.D.	MA-level correlation
			with choice			with choice
(A)	(B)	(C)	(D)	(E)	(F)	
<i>Panel A: MSA-level variables</i>						
	<i>N = 320</i>			<i>N = 179</i>		
Choice index (over districts' HS enroll.)	0.76	0.25	1.00	0.75	0.26	1.00
ln(Population)	14.00	1.21	0.20	14.17	1.18	0.10
Pop.: Fr. Black	0.13	0.09	-0.18	0.12	0.09	-0.25
Pop.: Fr. Hispanic	0.11	0.14	-0.19	0.14	0.15	-0.18
Mean log HH income	10.23	0.19	0.34	10.27	0.20	0.36
Gini, HH income	0.43	0.03	-0.39	0.43	0.03	-0.45
Pop: Fr. BA+	0.22	0.06	0.16	0.23	0.06	0.21
Finance: Foundation Plan rule	0.71	0.44	-0.07	0.65	0.46	-0.03
Finance: d(oper. exp.)/d(median inc.)	3.17	2.70	0.19	3.09	2.93	0.11
South	32%		-0.30	36%		-0.29
SAT-taking rate	0.28	0.14	0.08	0.36	0.08	0.30
Private enrollment share (HS)	0.11	0.05	-0.10	0.11	0.05	-0.19
Racial dissimilarity index, high schools	0.49	0.14	0.36	0.47	0.13	0.28
Racial isolation index, high schools	0.28	0.16	0.22	0.28	0.14	0.17
1942 choice index	0.92	0.20	0.35	0.91	0.23	0.41
Number of streams in MSA	271	222	0.71	283	214	0.73
<i>Panel B: Individual samples</i>						
	<i>NELS (N = 15,589)</i>			<i>SAT-takers (N = 330,688)</i>		
NELS 8th-grade test / SAT	1012	203	0.26	995	201	0.29
NELS continuation rate	85%		0.04			
Black	15%		-0.22	12%		-0.32
Hispanic	12%		-0.21	12%		-0.23
Asian	5%		-0.18	10%		-0.04
Female	50%		-0.06	55%		-0.28
Family income (\$000s)	\$43	\$39	0.09	\$46	\$26	0.43
Background index	1012	77	0.23	995	82	0.40
<i>Panel C: School samples</i>						
	<i>NELS (N = 748)</i>			<i>SAT-takers (N = 5,779)</i>		
Test score (NELS/SAT) mean	1012	104	0.26	995	95	0.29
Size (students per grade)	263	430	-0.07	387	211	-0.44
Mean background index	1012	52	0.23	995	48	0.40
Public	0.84		0.09	0.90		0.42
Number of SAT-takers				179	116	-0.19
SAT-taking rate				0.49	0.1947	0.31

Notes: MSA-level statistics are weighted by the MSA 17-year-old population. Individual NELS and SAT-taker data are weighted by inverse sampling probabilities, and school means by the sum of individual weights; both are adjusted at the MSA level to weight MSAs in proportion to their 17-year-old populations.

Table 2: Choice and the within-MSA average outcome-peer group gradient.

	NELS						SAT data		
	8th-grade score			HS continuation (per 1000)			SAT score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average student background index	1.70 (0.06)	1.78 (0.07)	1.78 (0.07)	1.67 (0.13)	1.71 (0.18)	1.71 (0.18)	1.74 (0.02)	1.68 (0.02)	1.68 (0.01)
<i>Interactions of average background index with MSA-level:</i>									
* Choice index	-0.19 (0.17)	-0.07 (0.51)	-0.15 (0.47)	-0.76 (0.44)	-1.30 (1.02)	-1.28 (1.30)	-0.24 (0.08)	-0.04 (0.13)	0.05 (0.13)
* MSA SAT-taking rate							1.49 (0.59)	0.86 (0.43)	0.95 (0.39)
* basic demographic controls (8)	n	y	y	n	y	y	n	y	y
* additional controls (3)	n	n	y	n	n	y	n	n	y
# schools / MSAs	741 / 209			723 / 207			5,711 / 177		
R ²	0.79	0.80	0.80	0.54	0.58	0.58	0.78	0.78	0.78

Notes: Dependent variable is the weighted mean score (or continuation rate in columns D-F) at the school. NELS scores are standardized to the same individual mean (1000) and standard deviation (200) as SATs. Schools are weighted by the sum of individual sampling weights, with an MSA-level adjustment to weight MSAs by their 17-year-old populations. MSA-level explanatory variables are demeaned before being interacted with the school-average background index. All models include MSA fixed effects and interactions of the background index with demeaned census division dummies, and standard errors are clustered at the (C)MSA level (Kezdi, 2002).

Table 3: Alternate specifications.

	NELS (8th grade scores)				SAT			
	Peer gp. main effect		Peer gp.- choice interaction		Peer gp. main effect		Peer gp.- choice interaction	
	(1)	(2)	(3)	(4)	(3)	(4)	(3)	(4)
A. Basic model	1.78	(0.07)	-0.07	(0.51)	1.68	(0.02)	-0.04	(0.13)
<i>Alternative background controls</i>								
B. Interact bkgd. with school sample size poly.	1.62	(0.10)	-0.17	(0.51)	1.65	(0.02)	-0.02	(0.13)
C. Control for S.D. of peer group at school	1.77	(0.07)	-0.06	(0.50)	1.68	(0.01)	-0.10	(0.12)
D. Control for school racial composition	2.02	(0.08)	-0.08	(0.46)	2.09	(0.08)	-0.08	(0.10)
E. Sorting based on avg. income			-0.03	(0.48)			-0.01	(0.11)
<i>Alternative samples</i>								
F. Sample excludes private schools	1.68	(0.08)	0.02	(0.61)	1.66	(0.02)	-0.04	(0.13)
G. Sample excludes 1-district MSAs	1.78	(0.07)	-0.11	(0.61)	1.71	(0.01)	-0.36	(0.15)
H. Sample also excludes central city districts	1.76	(0.10)	-0.13	(0.66)	1.79	(0.03)	-0.48	(0.27)
<i>Alternative choice measures</i>								
I. SAT-taker choice index	1.79	(0.07)	-0.43	(0.52)	1.68	(0.02)	-0.01	(0.14)
J. SES choice index	1.78	(0.07)	-0.08	(0.56)	1.68	(0.02)	-0.03	(0.13)
K. SAT-taker choice indx, high SAT partic. skls					1.69	(0.02)	-0.14	(0.21)
L. SES choice indx, high SAT partic. schls					1.69	(0.02)	-0.18	(0.19)
<i>Instrumental variables estimates</i>								
M. Streams as instrument for choice	1.78	(0.07)	-0.29	(0.88)	1.69	(0.02)	-0.41	(0.33)
N. 1942 choice as instrument for choice	1.82	(0.08)	-1.28	(1.00)	1.68	(0.02)	0.22	(0.22)
O. 1995 bkgd. avg. as instrument for 1994 avg.					1.74	(0.02)	-0.03	(0.14)
<i>Exploration of selection bias in SAT sample</i>								
P. Control for school participation rate					1.54	(0.03)	0.03	(0.13)
Q. Sample excludes low participation schools					1.69	(0.02)	-0.19	(0.17)
R. Sample reweighted using class rank					1.68	(0.02)	0.06	(0.14)
S. District level	1.65	(0.09)	-0.61	(0.46)	1.65	(0.03)	-0.09	(0.22)
T. District level; sorting on ln(median income)			-0.64	(0.47)			0.33	(0.20)

Notes: "Basic model" in row A is that from Table 2, Column 2 or 8. Remaining rows reestimate this model with small changes to the specification or sample. Rows M and N instrument the student background-choice interaction with the interaction of background and the listed instrument; row O instruments the background main effect and all interactions with corresponding measures from the next year's sample of SAT-takers. Rows S and T include interactions of the background index with the MSA private enrollment rate and its square, in addition to the usual controls. Standard errors, clustered on the (C)MSA, in parentheses.