WORKING PAPER # 3
PRINCETON UNIVERSITY
EDUCATION RESEARCH SECTION
OCTOBER 2003
<http://arks.princeton.edu/ark:/88435/dsp01f7623c618>

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

Jesse M. Rothstein*
Princeton University

October 2003

* I am grateful to Alan Auerbach, David Lee, John Quigley, Emmanuel Saez, and especially David Card for their advice. Tom Davidoff, Caroline Hoxby, Justin McCrary, Rob McMillan, Till von Wachter, and participants in seminars at several institutions provided helpful comments and suggestions. This research would not have been possible without access to confidential data from the College Board (via the Mellon Foundation, David Card, and Alan Krueger) and the National Center for Education Statistics. Financial support from a National Science Foundation Graduate Research Fellowship and from the Fisher Center for Real Estate and Urban Economics at U.C. Berkeley is gratefully acknowledged. Any errors, opinions, findings, conclusions, or recommendations expressed in this publication are mine alone and do not necessarily reflect the views of the NSF, the Fisher Center, or any of my advisors.

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

ABSTRACT

School choice policies aim to improve school productivity by rewarding administrators of schools that parents prefer. Parental choice may not create incentives for effective administration if parents prefer schools with desirable peer groups to those with inferior peers but better policies and instruction. I examine families' revealed preferences in "Tiebout choice" residential location markets for evidence on the importance of school effectiveness to parental choices. In a multicommodity-style model, wealthy parents cluster together in any Tiebout equilibrium, and cluster near effective schools if effectiveness is an important component of school desirability. Moreover, decentralization of educational governance—choice among several local school districts—facilitates this residential sorting. Thus, if parents strongly prefer effective schools, average income correlates with school effectiveness in high-choice-market equilibrium. I use a large sample of SAT-takers to examine the joint distribution of student background and outcomes across schools within metropolitan housing markets that differ in the structure of educational governance. I find strong evidence that choice facilitates residential sorting, but little evidence of the sorting that is predicted if parents choose neighborhoods for school characteristics other than peer groups. Moreover, average SAT scores are no higher in high-choice than in low-choice markets. These results suggest caution about the potential to induce improvements in educational productivity through expansions of parental choice.

Jesse M. Rothstein
Economics Department
Princeton University
Princeton, NJ 08544
(609) 258-4045
jrothst@princeton.edu

1. Introduction

Monopoly public education providers may operate inefficiently (Hanushek, 1981). One policy that may create incentives toward efficient production is to allow parents to choose among several schools, with compensation of school administrators and staff linked to parental demand (Friedman, 1962; Chubb and Moe, 1990; Hoxby, 1994).

Hanushek 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” (1981, p. 34-5; emphasis added). Poor decision making is not required; if fully-rational parents choose schools with “pleasant surroundings, athletic facilities, cultural advantages,” (ibid, p. 34) or easy access to resources over those that most efficiently pursue academic performance, choice will not create strong incentives for teachers and school administrators to increase academic output.

Formal choice programs are both too new and too rare to permit decisive evaluation.¹ Economists have long argued, however, that housing markets represent a long-established, potentially informative form of school choice, as parents influence their children’s school assignments via their residential location decisions (Tiebout, 1956; Brennan and Buchanan, 1980; Borland and Howsen, 1992; Hoxby, 2000). One can infer the effects of so-called *Tiebout choice* by comparing U.S. metropolitan housing markets that differ in the control over school assignment afforded by the location decision (Belfield and Levin, 2001, review studies of this sort). This promises to “demonstrate general properties of school choice that are helpful for thinking about reforms,” (Hoxby, 2000, p. 1209).

Existing research on Tiebout choice studies the supply side of the market, focusing on differences in average performance across markets. The interpretation of these estimates is not straightforward, particularly when outcomes are averaged only across students in public schools

¹ See Rouse (1998); Cullen, Jacob, and Levitt (2000); Mayer, Peterson, Myers, et al. (2002); and Krueger and Zhu (2003) for analyses of several existing choice programs.

(Hsieh and Urquiola, 2002), as choice likely has important effects on the distribution of students across schools and sectors. In this paper, I study instead the distribution of student outcomes *within* metropolitan housing markets for evidence on parental demand. Looking at the allocation of effective schools in static cross-section, I ask whether school effectiveness plays a sufficiently important role in the residential location decision to create meaningful incentives for more productive school management.

As an alternative, I consider the possibility that peer groups may be the primary determinant of parental preferences over schools. Test score maximizing parents may rationally prefer schools with good peer groups to those that are more effective if the peer group—a classic externality—is an important input to educational production. Moreover, parents may have direct preferences for a good peer group in the school or neighborhood, or may be unable to identify effective schools distinct from those with advantaged students.² If parents’ revealed preferences fail to place substantial weight on school effectiveness, even fully implemented choice systems will not reliably reward administrators for effective instruction.

The primary hurdle to be overcome in assessing the relative importance of school effectiveness and peer group in parental choices is that effectiveness is difficult to measure. Prominent research (Hanushek, 1986) casts doubt on the educational importance of many of the observable characteristics of schools—class size, teacher qualifications, etc.—that might be thought to indicate effective administration. A common fallback approach is to measure effectiveness as the residual in a regression of student outcomes on individual and peer group characteristics. The latter are badly endogenous in this regression, however: Schools are an important component of the residential location decision, and their characteristics may be capitalized into housing prices. As a result, any unobserved characteristic of a school that makes it desirable—including, potentially, its effectiveness—will raise the price of local housing, and thereby attract a better peer group.

² Real estate agents typically report the average outcomes of students at a local school. These are arguably as informative about the peer group as about the school’s effectiveness. In California’s school accountability program, for example, the primary rankings correlate 0.93 with the average education of students’ parents (Technical Design Group, 2000; see also Rothstein, forthcoming).

Rather than assume away any correlation of effectiveness with student composition, I look for evidence of such a correlation as an indication of parental preferences. I develop a multicomunity-style model of the metropolitan housing market, in the spirit of those reviewed by Ross and Yinger (1999). In this model, high-income families always cluster together in equilibrium, but the location of these clusters depends on both parental preferences and market structure. If parents primarily choose effective schools, high-income families will cluster in effective schools' attendance areas. If location decisions are primarily motivated by other factors—potentially including the peer group—high-income clusters may just as easily appear near ineffective schools.

Clustering near effective schools will tend to magnify naïve estimates of the peer effect that do not account for the Tiebout sorting process, as it induces a positive correlation between the peer group and unobserved school effectiveness. Importantly, the model predicts that for plausible parental preferences, this magnification effect will be larger in housing markets that offer more opportunity for Tiebout choice, those that more readily permit a family with the available means to purchase access to its desired school via the residential location decision.

I test the predictions of the model with a unique data set, with observations on more than 300,000 SAT-takers from a single cohort, matched to the high schools that they attended. This sample is large enough to permit accurate estimation of both peer quality and average performance for the great majority of public and private high schools in each of 179 metropolitan housing markets. Figure 1 suggests the estimation strategy. The first panel depicts average SAT scores and average student characteristics across the 179 Metropolitan Statistical Areas (MSAs). By construction, a one point increase in a student's background index (without any change in school assignment or peer group) corresponds, on average, to one additional SAT point.³

The across-MSA slope of SAT scores with respect to student characteristics is considerably greater than one. As there is not likely to be much endogeneity of the peer group to school quality at this level of aggregation—researchers typically assume, reasonably, that families search within a

³ The student background index is the fitted value from a within-school regression of SAT scores on student background characteristics. The data, sample, and background index construction are described below.

metropolitan area for good schools, but select among metropolitan areas for exogenous reasons—the estimated slope in this panel of 1.45 indicates a peer effect nearly half as large as the effect of individual characteristics on own test scores. MSAs may well differ, however, in other determinants of performance (like school funding) that correlate with average family background, however, and recent estimates of peer effects eschew this sort of across-market variation in favor of more experimental assignment.⁴

Panel B performs the same exercise across schools within metropolitan markets, graphing school deviations from MSA-average SAT scores against similar deviations of student characteristics. This is the “MSA fixed effect” estimator of the importance of peer characteristics. If parents exercise Tiebout choice, this estimator may be biased relative to the true peer effect: Wealthy families are likely to disproportionately purchase houses near effective schools, creating a positive correlation between the peer group and unobserved school effectiveness. The *Tiebout sorting* that creates this bias should be more complete, however, in areas with a strong residential location-school assignment link than in areas where this link is weaker. The basic strategy taken in this paper, then, is to ask whether the relationship depicted in Panel B is systematically stronger in high-choice areas, where it reflects the combination of peer effects and housing market sorting, than in low-choice areas, where sorting is more difficult to accomplish.

To preview the results, I find no evidence that the school-level association between student characteristics and outcomes is stronger in high-choice than in low-choice markets. This result is robust to nonlinearity in the causal peer effect; to several vectors of control variables and several measures of peer group quality; to several strategies for dealing with sample selection in the SAT data; and to potential endogeneity of the choice measure. Moreover, although there is no other suitable data set with nearly the coverage of the SAT sample, the basic results are supported by estimates from the much sparser National Educational Longitudinal Survey (NELS) sample. Supporting analyses indicate that the choice measure used here captures meaningful variation in

⁴ The across-market estimate is quite similar to the peer effects estimator proposed by Evans, Oates, and Schwab (1992), which uses MSA averages as instruments for school-level peer group characteristics. Section 2 addresses the role of other educational inputs (i.e. financing), which may correlate with family background both across and within MSAs.

parents' ability to sort themselves into desirable schools, and that differences between high- and low-choice areas in the variance of school effectiveness, while significant, are not large enough to mask a large association between choice and sorting on school effectiveness.

There are two plausible explanations for the pattern of results presented here. First, it may be that school policies are not responsible for a large share of the extant variation in student performance. We would not then expect to observe any sorting on school effectiveness in the distribution of SAT scores. A second explanation is that effectiveness does matter for student performance, but that it does not matter greatly to parental residential choices. This could be because peer groups swamp effectiveness in parental preferences or because the latter is difficult to observe directly. In either case, the Tiebout marketplace does not seem to reliably sanction administrators who pursue unproductive policies and, as a result, Tiebout choice appears unlikely to create meaningful incentives toward greater school effectiveness. Consistent with this, across-MSA analyses of average scores, similar to those in the literature, indicate that average performance is no higher in high-choice markets once student characteristics are controlled.

2. Tiebout sorting and the peer group: Intuition

In this section, I describe the Tiebout choice process and its observable implications in the context of a simple educational technology with peer effects. This is used to develop intuition for the formal model, a simple extension of existing multicomunity models developed in Section 3, and for the empirical approach, in Section 7. Let

$$t_{ij} = x_{ij}\beta + \bar{x}_j\gamma + \mu_j + \varepsilon_{ij} \quad (1)$$

be a reduced-form representation of the educational production function, where t_{ij} is the test score (or other outcome measure) of student i when he or she attends school j ; x_{ij} is an index of the student's background characteristics; $\bar{x}_j$ is the average background index of students at school j ; and μ_j —which need not be orthogonal to $\bar{x}_j$ —measures the effectiveness of school j , its policies and practices that contribute to student performance.

Many of the characteristics and policies often thought of as indicative of a high-quality school are functions of the enrolled population: Schools with advantaged students may necessarily have more active parents, better teachers (Antos and Rosen, 1975), more resources, and even—if parental “voice” is effectual (Hirschman, 1970)—harder-working administrators. To the extent that these depend on the enrolled population and not on fixed characteristics of the school or of its immobile staff, these are externalities, best thought of as indirect peer effects rather than as school effectiveness and included as part of γ in (1).⁵

For analyses of the incentives created by choice, effectiveness— μ_j in equation (1)—should encompass only differences across schools that do not depend on the characteristics of enrolled students. They might include differences in administrative quality or effort beyond those forced by parental voice, as well as any idiosyncratic differences in costs, resources, or productive efficiency. In principle, decentralized markets can create incentives for increased production of these effectiveness attributes.⁶ It is not clear, however, how important effectiveness so defined is in educational production. Chubb and Moe (1990) and Hoxby (1999a) argue that many of the problems in American education derive from parents’ inability to enforce administrative effort through either voice or exit, implying a large role for effectiveness. On the other hand, if the apparent effects of family background in Figure 1 are taken at face value, there may be little cross-sectional variation in school effectiveness (see also Kane and Staiger, 2002). Productivity arguments for choice policies, of course, rely crucially on the former view that school productivity has large effects on student outcomes and can vary substantially.

Now consider the character of Tiebout equilibrium when location decisions depend only on school characteristics— $\bar{x}_j$ and μ_j —and housing prices. Some communities will offer more

⁵ This paper considers only peer effects that load upon students’ observable characteristics. I need not determine whether these occur because t_{ij} depends on $\bar{x}_j$ or on $\bar{t}_j$, a distinction that Manski (1993) points out is not identified. Equation (1) appears to allow for only the former—Manski’s “exogenous” effects—but the latter—“endogenous” effects—reduce to the former plus a school-level term that appears as residual variance in my analysis.

⁶ By contrast, administrative effort cannot increase the overall stock of good peers. Even if peer effects are nonlinear, so that the distribution of peer groups has efficiency consequences, a complex centralized pricing scheme would likely be required to produce the efficient allocation as a market equilibrium.

attractive $(\bar{x}_j, \mu_j)$ packages than will others, and parents will bid up their housing prices. If parents' willingness to pay for desired schools is positively correlated with x , high- x families will tend to win the bidding for homes in these communities, bringing high- $\bar{x}$ peer groups to the local schools.

The allocation of peer groups does not depend on parental preferences: The most desirable communities will necessarily be those with the highest $\bar{x}$, even if $\bar{x}$ is irrelevant to parental evaluations of schools. Parental preferences for effectiveness, however, do have observable implications for the *location* of these high- $\bar{x}$ communities. If effectiveness is parents' primary concern, high- μ communities will be the ones whose housing prices are bid up, and will attract disproportionately high- x families (and thus high values of $\bar{x}$). In this situation, Tiebout choice induces *effectiveness sorting*, a positive correlation between μ_j and $\bar{x}_j$. On the other hand, if parents place too little relative value on effectiveness to choose a high- μ , low- $\bar{x}$ school over the reverse, equilibrium need not exhibit any correlation between μ and $\bar{x}$, as high income families may be stuck in wealthy communities with ineffective schools by their unwillingness to abandon the peer groups that they offer.

Although effectiveness is not itself observable by the researcher, effectiveness sorting is: It magnifies the apparent within-market impact of peer groups on student outcomes in analyses, such as those in Figure 1, that neglect administrative quality. Formally, (1) implies that

$$E[\bar{l}_j | \bar{x}_j] = \bar{x}_j(\beta + \gamma) + E[\mu_j | \bar{x}_j], \quad (2)$$

or, simplifying to a linear projection,

$$E^*[\bar{l}_j | \bar{x}_j] = \bar{x}_j(\beta + \gamma + \theta^*), \quad (3)$$

where $\theta^* \equiv \text{cov}(\bar{x}_j, \mu_j) / \text{var}(\bar{x}_j)$ represents the degree of effectiveness sorting in the market. (For notational simplicity, I neglect intercepts.) The stronger are parental preferences for effective schools relative to those for schools with other attributes, the more reliably will high- μ schools attract high- $\bar{x}_j$ students, and the larger will θ^* tend to be in Tiebout equilibrium.

Importantly, the extent of Tiebout choice in the local market will interact with parental preferences in determining θ^* whenever parents care both about school effectiveness and peer groups. The reasoning is simple: In most models of Tiebout choice—including the one developed below—equilibrium stratification is an increasing function of the degree of choice. If there are only a small number of local districts, then, differences in $\bar{x}$ among competing districts will be large, and the peer group sacrifice needed for a resident of a wealthy district to obtain a more effective school will frequently be prohibitive. This derives from families’ inability to coordinate, as the residents of that high- $\bar{x}$ district could collectively outbid poorer families for the most effective schools. Such coordination failures are less likely to result in mis-sortings in markets with more choices, as in these markets there are alternative districts that are relatively similar in the peer group offered, and parents are able to select among several districts without paying a steep price in reduced peer quality.⁷

The number of jurisdictions-effectiveness sorting relationship does not hold for extreme parental preferences: When parents are concerned only with school effectiveness, high- μ schools attract high- x families regardless of the market structure, and θ^* need not vary with the availability of choice. Similarly, when parental concern for peer group is large enough, coordination failures persist even in highly competitive markets, and again θ^* is largely independent of market structure.

It is the potential relationship between effectiveness sorting (θ^*) and choice that I use to distinguish this sorting from the causal effects of individual and peer characteristics (β and γ) in (1). Let $\theta = \theta(c, \delta) = E[\theta^* | c, \delta]$ be the average effectiveness sorting of markets characterized by the parameters c and δ , where c is the degree of Tiebout choice and δ is the importance that parents place on the peer group relative to effectiveness. The argument above, formalized in the next section, implies that $\partial\theta/\partial\delta < 0$; that $\partial\theta/\partial c > 0$ for moderate values of δ ; and that $\partial\theta/\partial c = 0$ when δ is zero or large (i.e. when parents care only about school effectiveness or not at all about it). To the extent that θ^* tends to increase with choice, then, we can infer that parents’ preferences over the peer

⁷ This is the mechanism at work in Hoxby’s (1999b) model, in which student types are discrete and few, and there are assumed to be several districts competing for each possible peer group type.

group are small enough to prevent a breakdown of effectiveness sorting in high-choice markets. On the other hand, if θ^* is no larger in high-choice than in low-choice cities, parental valuations may be characterized either by very small or very large δ .

2.1. Static allocation and dynamic responses

In the demand-side model developed here, housing-price-mediated choice increases inequality in two ways. First, as in all Tiebout models, residential choice promotes social stratification, with increased separation of high- and low-income students into different schools and therefore increased inequality of peer group allocations. Second, if parents prefer effective schools, Tiebout choice allocates the most effective schools to families with the highest willingness-to-pay, leaving low-income students to attend the least effective schools.

Competitive pressures—likely strongest when parents attach great weight to school effectiveness—may induce a supply response that raises effectiveness at all schools, potentially offsetting the negative allocative consequences of choice for poor students. As these allocative effects are informative about parental demand, I primarily consider within-market sorting with a fixed effectiveness distribution. My empirical implementation of the sorting model, in Section 7, includes fixed effects for each market to absorb any differences in average effectiveness. I then consider two forms of supply responses. First, I ask whether choice reduces the variation in effectiveness, which would appear as a smaller θ^* for any fixed peer group-effectiveness correlation and could mask effectiveness sorting differences between markets. Second, I present estimates similar to those in the literature of the effect of choice on average test scores, which might be positive if parental demand for effectiveness induces greater effectiveness production.

3. A multicommodity model with exogenous community attributes

My sorting model ignores several characteristics of real-world Tiebout choice. There are no private schools. These appear to be an inferior alternative to residential choice among public

schools (see Section 6), and they complicate the theoretical analysis.⁸ I also assume that all families agree about the relative desirability of different communities. This rules out, for example, families without children who are unconcerned about the quality of local schools. The attitudes of such families toward school quality are not well understood (Black, Gates, and Sanders, 2002; Barrow, 2002; Hilber and Mayer, 2002). Finally, I omit all non-school amenities like beaches, parks, views, home size, and air quality. I develop here a “best case” for Tiebout choice, where schools are the only determinants of community desirability.⁹

As in other models, equilibrium must be as stratified as the community structure allows, with wealthy families always residing in communities that are preferred to those where low-income families live. If parents are concerned only with school effectiveness, there is a unique equilibrium with perfect effectiveness sorting. When the peer group enters parental preferences, however, other imperfectly sorted equilibria are possible. In general, the weaker are parental preferences for effectiveness relative to peer group, the lower the correlation between effectiveness and average community income that can be supported in equilibrium. I follow the formal model with simulations that vary the number of communities. These demonstrate that, for fixed valuations, effectiveness sorting increases with Tiebout choice.

Turning to the formal exposition, assume that a metropolitan housing market contains J jurisdictions and N families, $N \gg J$. Each jurisdiction, indexed by j , contains n identical houses (with $n(J-1) < N \leq nJ$) and is endowed with an exogenous effectiveness parameter, μ_j . I assume that there are no ties in effectiveness: $\mu_j \neq \mu_k$ for $j \neq k$. All homes are owned by absentee landlords, perhaps a previous generation of parents, who have no current use for them. There is no possibility for collusion among landlords, and houses rent for the lowest non-negative market-clearing price. Each jurisdiction’s housing supply is thus perfectly inelastic.

⁸ My empirical implementation includes both public and private schools, although I also present specifications demonstrating that the results are insensitive to the exclusion of the latter.

⁹ The importance of schools to location decisions is supported by, among many others, Reback (2001), Bogart and Cromwell (2000), Figlio and Lucas (2000), and Black (1999).

Family i 's exogenous income is $x_i > 0$. No two families have the same income, and the metropolitan distribution function is F . Families spend their incomes on housing and on numeraire consumption. Let $\bar{x}_j$ denote the (endogenous) mean income of families in community j , and let b_j be the rental price of local housing. The utility that family i obtains in jurisdiction j is, then, $U_{ij} = U(x_i - b_j, \bar{x}_j\delta + \mu_j)$, with U twice differentiable everywhere and U_1 and U_2 both positive.

I make the usual single-crossing assumption, $U_{12}U_1 - U_{11}U_2 > 0$. Under single-crossing, if a high income and a low income family live in the same district, the wealthier family will always bid more for the opportunity to move to an alternative, higher-quality (i.e. higher $\bar{x}_j\delta + \mu_j$) district than will the lower-income family. This assumption drives the stratification results.

Families choose the district that offers them the highest utility, taking community composition and housing prices as given. Market equilibrium is defined as a set of housing prices and an allocation of families (or, equivalently, family incomes) to communities such that all community housing markets clear and each family is satisfied with its community assignment, taking other families' assignments as fixed:

Definition: An equilibrium for a market defined by $\delta; j; \{\mu_1, \dots, \mu_J\}$; and F consists of non-negative housing prices $\{b_1, \dots, b_J\}$ and an allocation rule $G: [F^{-1}(0), F^{-1}(1)] \mapsto \mathbf{Z}_J$ that satisfy the following conditions (where $\bar{x}_j \equiv E[x | G(x) = j] \equiv \frac{\int \mathbf{1}(G(x) = j)x dF(x)}{\int \mathbf{1}(G(x) = j)dF(x)}$):

- EQ1 *Nash equilibrium:* Given housing prices and the existing allocation rule, no family would prefer a district other than the one to which it is assigned:
 $U(x_i - b_{G(x_i)}, \bar{x}_{G(x_i)}\delta + \mu_{G(x_i)}) \geq U(x_i - b_k, \bar{x}_k\delta + \mu_k)$ for all i and all k .
- EQ2 *Approximate market clearing:* No district is over-full (for each j , $\int \mathbf{1}(G(x) = j)dF(x) \leq \eta_N$), and less-than-full districts have housing prices of zero ($\int \mathbf{1}(G(x) = j)dF(x) < \eta_N \Rightarrow b_j = 0$).
- EQ3 *No ties in realized quality:* For any j, k , $\bar{x}_j\delta + \mu_j \neq \bar{x}_k\delta + \mu_k$.¹⁰

The following results are proved in an appendix:

¹⁰ Condition 3 corresponds to Fernandez and Rogerson's (1996, 1997) "local stability" notion. Allocations that satisfy EQ1 and EQ2 but not EQ3 are unstable, and perturbations in one of the tied communities' effectiveness or peer group lead to non-negligible differences between the communities as families adjust.

Theorem 1. Any equilibrium is perfectly stratified by income: The n highest-income families live in the highest-quality (i.e. highest $\bar{x}\delta + \mu$), highest-rent community; the next n families live in the second-highest-quality community, which has the second-highest-rents; and so on. Thus, rankings of communities by rent, quality, or the incomes of their residents are all identical.

Theorem 2. Equilibrium always exists.

Theorem 3. If $\delta = 0$ —if families are unconcerned with peer group—there is a unique equilibrium allocation rule G .¹¹ This rule incorporates perfect effectiveness sorting: The ranking of communities by equilibrium rent or income is identical to that by effectiveness.

Theorem 4. When $\delta > 0$, there may be multiple equilibrium allocation rules. For any perfectly stratified (in the sense of Theorem 1) rule G such that $\bar{x}_{G(y)}\delta + \mu_{G(y)} \geq \bar{x}_{G(w)}\delta + \mu_{G(w)}$ whenever $y > w$, with strict inequality whenever $G(y) \neq G(w)$, there exist housing prices with which G is an equilibrium.

Theorems 1 and 4 together provide necessary and sufficient conditions for an allocation of families to communities to be consistent with equilibrium. Hereafter, I neglect housing prices as an epiphenomenon, necessary for the market to clear but not to the identification of an equilibrium allocation rule.¹²

4. Comparative statics in parental preferences and market structure

By Theorem 4, there can be equilibria in which the wealthiest, highest-rent communities are not those with the most effective schools, so long as parental preference for the peer group offered is large enough to offset these communities' effectiveness deficiencies. In this section, I explore the relationship of these imperfectly sorted equilibria with the strength of parental preferences for effective schools and with the number of jurisdictions in the local market

4.1. Graphical depiction of illustrative equilibria

By Theorem 1, any equilibrium allocation rule can be summarized by a one-to-one mapping from J quantiles of the income distribution to the effectiveness parameters of the communities in

¹¹ 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. With a continuous, connected income distribution, the housing price vector would be unique (up to specification of the minimum housing price if $nJ = N$).

¹² With data on housing prices and a parameterization of the utility function, it is possible to estimate the model without reliance on the across-market strategy used here. See Bayer, Ferreira, and McMillan (2003) for a thorough implementation of this approach.

which families in each quantile reside. Consider a simple case in which there are three equally-sized jurisdictions ($J = 3$) just large enough to accommodate the population ($n = N/3$). Any equilibrium for this market must assign the wealthiest third of families to the most desirable district, the middle third in the second-most desired district, and the poorest third in the remaining district. The dashed line in Panel A of Figure 2 depicts the resulting peer group allocation when income is normally distributed with unit mean and standard deviation.¹³

Suppose that the three jurisdictions' effectiveness parameters are equally spaced on the unit interval ($\mu_1 = 1/3$, $\mu_2 = 2/3$, $\mu_3 = 1$). If parents care only about effectiveness (i.e. if $\delta = 0$), district 3 is necessarily the most desirable and district 1 the least. Thus (Theorem 2), there is only one equilibrium allocation. The thin solid line in Figure 2, Panel A graphs the effectiveness (and desirability) allocation in this case.

Now consider the impact of changing parental valuations to incorporate moderate concern for the peer group ($\delta = 1.5$). The bold line in Panel A depicts the new mapping from family income to desirability in the perfectly sorted allocation. This δ parameter, however, admits all permutations of the three income terciles among the three communities as additional equilibria. Note that the average incomes in these terciles are approximately -0.1, 1, and 2.1; differences among any pair of communities in $\bar{x}_j\delta$, never smaller than 1.65, are therefore large enough to overwhelm any differences in μ_j , which are never larger than $2/3$. Panel B displays the three mappings for the reverse-sorted allocation, in which wealthy districts are uniformly less effective than those where lower-income families reside. Even here, the necessarily increasing income-peer group function is important enough to create a non-decreasing income-desirability mapping, by Theorem 4 the condition for an equilibrium allocation.

¹³ The scale parameters are chosen arbitrarily to make the graphs readable. Log-normality of income would be more realistic, but the asymmetry of this distribution can be distracting. For the purposes of the model, a log-normal distribution could effectively be transformed to the normal distribution by allowing utility to depend on the school's average log income. In the empirical application, I use an index of students' family background in place of actual income; this index is approximately normally distributed.

Next, consider altering the toy economy to allow for greater Tiebout choice, moving from three districts to ten ($\mu_j = j/10$, $j = 1, \dots, 10$). Return to the case of $\delta = 0$. The perfectly sorted effectiveness allocation, graphed as the thin solid line in Panel C of Figure 2, is the unique equilibrium. The dashed line again displays the only possible allocation of peers.

With ten districts, the differences among adjacent deciles in average income are much smaller than among adjacent terciles in the three district case. As a result, when we allow for concern for peer group with $\delta = 1.5$, not all allocations of deciles to communities are equilibria. Panel D displays an imperfectly sorted allocation that is an equilibrium, while Panel E offers one that is not. In the latter case, the fourth decile of the income distribution lives in a community that, because its schools are so ineffective while its students only slightly better, is considered inferior to that where the fifth decile resides, violating Theorem 1.

The contrast between the three-district and ten-district cases illustrates the principle described above: Imperfectly sorted equilibria are relatively easy to maintain when jurisdictions are few and large. As jurisdictions shrink, the increase in parental choice reduces the need for coordination, permitting some families to escape an underperforming school and reducing the possibility for imperfect sorting in equilibrium.

4.2. Simulation

To indicate the generalizability of the illustrative equilibria in Figure 2, I present here results of simulations of similar toy economies. For purposes of simulation, I abandon the fixed distribution of effectiveness parameters and instead draw them from the same normal distribution used for income. For specified (δ, J) , then, a simulation consists of a draw of J independent effectiveness parameters, one for each community. The J quantiles of the income distribution are randomly permuted among the J communities, and each community's desirability ($\bar{x}\delta + \mu$) is computed. If each quantile is assigned to a more desirable community than are lower-income quantiles, the permutation is an equilibrium allocation. If desirability is not increasing in income, the

allocation is not an equilibrium, and quantiles are re-permuted among the communities until an equilibrium allocation is found.¹⁴

Figure 3 reports the average effectiveness experienced by families at each income quantile for two market structures ($J = 3$ and $J = 10$) and four values of the parental valuations parameter ($\delta = 0, 1.5, 3$, and 6), with averages computed over 5,000 simulated economies. Panel A depicts the case where parents are unconcerned about the peer group and care only for school effectiveness ($\delta = 0$). As equilibrium must be perfectly sorted in this case, this panel merely graphs estimated order statistics for 3 or 10 draws from the μ distribution.

As δ grows in the remaining panels, progressively less complete sortings become equilibria. As a result, the distribution of effectiveness parameters that might be experienced by any particular family approaches the unconditional distribution, and averages collapse toward the overall mean of 0. Importantly—see Panel C—this collapse is faster in the 3-district case than in the 10-district case. As δ grows further in Panel D, however, the differences between the two sorts of market disappear along with any semblance of sorting on school effectiveness.¹⁵

Figure 3 indicates that effectiveness sorting tends to increase with δ and, for moderate values like that in Panel C, with J . To provide another view of this, I estimated on the simulation data models similar to those implemented on real data in Section 7. For each (δ, J) combination I regressed μ on $\bar{x}$, using as data the $5,000 \times J$ communities from the 5,000 simulation draws and including a fixed effect for each draw. The resulting estimates of $\theta(\delta, J) = \text{cov}(\bar{x}_j, \mu_j) / \text{var}(\bar{x}_j)$ are displayed in Figure 4. The trends identified in Figures 2 and 3 are again clear: First, θ is well above zero when δ is small, indicating that the Tiebout choice process operates well when parents assess schools primarily by their effectiveness. When δ is large, θ is close to zero for all J , as no district structure reliably produces well-sorted equilibria when parents are largely unconcerned with school

¹⁴ This strategy treats all possible equilibria as equally likely. This might be refined with a theory of equilibrium selection, perhaps attaching higher probability to equilibria that are attracting points for larger ranges of initial assignments, or with a dynamic approach in which effectiveness parameters evolve slowly and a fraction of families move in each period. These are left for future work.

¹⁵ Non-monotonicities in Panels C and D derive from the tails of the normal income distribution, which produce larger differences in average income between adjacent quantiles at the extremes than near the middle.

effectiveness. For moderate δ , we observe more perfect sorting on effectiveness—and therefore larger slope parameters, θ —when there are many districts than when there are few. That is,

$$\frac{\partial \theta}{\partial J} > 0 \text{ for moderate } \delta.^{16}$$

5. Data

My test of parental valuations requires data describing the distribution of peer groups and outcomes *across* schools *within* housing markets that differ in the amount of Tiebout choice. The choice measure is defined at the level of the metropolitan area. I use student outcome data characterizing individual test scores, aggregated to the school level and matched to the MSA in which the school is located.

I use the June, 1990 MSA definitions (U.S. Census Bureau, 2001), which use observed commuting patterns to approximate local housing markets. The largest urbanizations form Consolidated MSAs (CMSAs), and are subdivided into several component parts, Primary MSAs (PMSAs). I treat several PMSAs within a larger area as distinct markets, reasoning that a move from, for example, Riverside to Ventura—both cities within the Los Angeles CMSA, but separated by about 125 miles—is more akin to a migration between metropolitan areas than to a within-market move. Most MSAs and PMSAs are defined along county boundaries; in New England, where town boundaries define MSAs, I use the alternative—and larger—county-based definitions.

Metropolitan-level data come primarily from the 1990 Decennial Census. I derive from the Summary Tape File 3C (U.S. Census Bureau, 1993) measures of each county’s demographic composition and educational and income distributions, then aggregate counties to form MSA totals. As the degree of equalization built into the state school finance rule may have direct effects on the benefits of attending a school in a wealthy neighborhood, it is important to control for this directly. I use for this a variable measuring the fraction of each MSA’s population residing in one of the 37

¹⁶ Figure 4 reveals a small effect of Tiebout choice on the effectiveness gradient θ even when $\delta = 0$ and equilibrium is unique, but this is sensitive to the simulation assumptions about the distribution of effectiveness as J grows. Simulations for positive δ are less sensitive.

states with a Minimum Foundation Plan rule—the most redistributive of Card and Payne’s (2002) classification—in 1991.

For the joint distribution of student characteristics and outcomes within each MSA, neither of the most commonly used datasets with observations on student outcomes, the National Educational Longitudinal Survey (NELS) and the National Longitudinal Survey of Youth (NLSY), is suitable for my analysis. The NELS uses a multi-stage sampling procedure and draws data from only three schools in the average MSA.¹⁷ The NLSY uses a neighborhood-based sampling design, so may include more schools, but students cannot be matched to the schools that they attended and in any case are not representative of those schools.

As an alternative, I use a restricted-access data set consisting of observations on 462,424 metropolitan SAT-taker observations from the cohort that graduated from high school in 1994. The sample includes about one third of SAT-takers from that cohort, and represents nearly 20 percent of 1994 high school graduates.¹⁸ Parents are likely to be particularly concerned with schools’ capacity to raise students’ SAT scores, as these scores have personal consequences that most other tests do not. SAT-takers are matched to schools using a crosswalk provided by the College Board that links to the National Center for Education Statistics school codes used in the 1993-94 Common Core of Data (CCD), a census of public schools, and the 1997-98 Private School Survey (PSS), a similar census of private schools. Schools are then linked to MSAs using county codes on the CCD and PSS.¹⁹ These surveys also provide grade-level enrollment figures, which I use to construct school SAT participation rates, and the racial composition of each school, used in analyses of racial stratification in Section 6.

The size of the SAT database permits precise estimation of school-level measures: I have at least ten observations per school from schools with 87 percent of enrollment in the MSAs studied.

¹⁷ I nevertheless present estimates for my basic model using the NELS data as a specification test in Section 5.

¹⁸ SAT-takers who report 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 College Board’s processing of the file, students who did not report an ethnicity are excluded from the sample. In data for 1999, in which I have a complete version of the file, these students comprise about 12% of SAT-takers.

¹⁹ 78 percent of schools (enrolling 90 percent of students) in the SAT sample are public.

SAT data have a serious limitation, however: Students self-select into taking the SAT, and there is evidence that at large geographic scales the SAT-taking rate is negatively correlated with average performance (Dynarski, 1987). A key source of variation in SAT-taking rates is the state public university system's preference for the SAT versus its competitor, the ACT. In "ACT states," only students who are applying to out-of-state colleges need take the SAT, inducing significant positive selection into the sample of observed SAT scores (Clark, 2003). To partially remedy this, I use only observations from MSAs wholly contained within the 23 states with SAT-taking rates above one third.²⁰ The remaining sample consists of 330,688 SAT-takers from 179 MSAs in "SAT states." This sample is likely representative of the college-bound population within the areas under consideration, although all analyses of the SAT data control for the MSA SAT-taking rate.²¹ Exploratory analyses with more involved selection corrections—reported in Section 7 and in an earlier version of this paper, available from the author—suggest that the resulting estimates are not seriously biased by within-school selection into SAT-taking.

It is helpful to have a one-dimensional index of student quality at each school. To create one, I estimated a flexible regression of individual SAT scores on student characteristics, controlling for school fixed effects. The model included effects for 100 parental education categories (ten for mother's education interacted with ten for father's education, each including a non-response category) and the interactions of six ethnicity indicators with two gender categories (eleven parameters) and with twelve family income bins (66 additional parameters).²² The sample is large enough to permit relatively precise estimation of even this flexible model, and effect standard errors are generally below ten SAT points. I defined an index of peer quality by averaging the fitted values

²⁰ SAT-taking rates use 12th-grade enrollment at schools which successfully match to the SAT data as the denominator, although other definitions produce the same sample. The selection rule is insensitive to the particular cutoff used, and among states above the cutoff, there is no correlation between SAT-taking rates and average scores.

²¹ Roughly 45% of the relevant national cohort enrolled in college after graduation, although only about two-thirds of enrollment is at four-year institutions (National Center for Education Statistics, 1999, Tables 101, 173 and 184).

²² The model explained 78 percent of the cross-sectional variance in individual SAT scores; school effects alone explain 67 percent.

(excluding the estimated school effect) of this regression over all students at each school.²³ This index has the interpretation of the school SAT-taking population’s predicted average performance at a nationally representative school. Specification checks reported in Section 7 explore the sensitivity of the results to the particular index used.

By using SAT data to describe each school’s peer group, I necessarily exclude the characteristics of students who do not take the SAT. The average characteristics of SAT-takers are arguably a more accurate measure of the peer group for college-bound students than would be averages over the entire student population, as students at many schools are tracked into college-preparatory and non-college-preparatory courses with little interaction between students in the two groups, and it seems plausible that parents distinguish between the groups in their evaluations of schools. Absent microdata for non-SAT-taking students, however, I am unable to test this restriction.

The leftmost columns of Table 1 report summary statistics for the primary metropolitan, school, and individual-level variables used in the current analysis.

6. Measuring Tiebout choice

The theoretical model’s idealized division of housing markets into discrete, equally sized communities does not have a direct real-world analogue. As a result, measuring the degree of Tiebout choice available to residents of a particular housing market is not a trivial task. The appropriate measure should capture two concepts: The strength of the link between residential location and school assignment, and the amount of variation across neighborhoods in important school characteristics.

Public education in the United States is provided by local school districts, typically independent governmental units wholly contained within a county. With some exceptions, residents of a particular district must attend that district’s schools or opt out of the public education system entirely. Most districts operate multiple schools, and establish attendance zones for each school that

²³ Endogenous selection into schools may bias individual characteristic coefficients, but does not create a problem for my estimation strategy as long as the bias affects all background variables equally: The only role for these coefficients is to assign relative weights to the individual variables, and the scale of the background index is irrelevant. An identical index calculated from 1995 data correlates 0.94 with the 1994 version at the school level, indicating high reliability.

regulate the within-district school assignment of district residents based on residential location.

Thus, families may in principle exercise Tiebout choice among districts, and within districts among school attendance zones.

Direct peer effects almost certainly operate at the school level, as students are rarely exposed to students attending other schools within the same district. The indirect peer effects referred to in Section 2—parental involvement, resources, teacher quality—may operate at either the school or district level, however. Similarly, effectiveness may vary primarily at the school level, if its most important determinants are the quality and effort of the school administrator, or at the district level, if hiring policies or budgetary decisions—typically made by districts—are major components.

I follow Hoxby (2000), and use as measures of choice Herfindahl-style indices of the concentration of public enrollment in the largest schools and districts.²⁴ If e_{jm} is the enrollment of district or school j in MSA m (drawn from the CCD) as a share of total metropolitan enrollment, the choice index for market m is $c_m = 1 - \sum_{j \in m} e_{jm}^2$.

Columns C and D of Table 1 report the bivariate correlations of all other variables with the district and school choice indices. Column C indicates that high district-level choice markets tend to be larger, to have more Blacks and Hispanics, and to be outside the South. They are also more likely to be located in states with Foundation Plan financing schemes.

Previous authors (Borland and Howsen, 1992; Hoxby, 2000), incline toward the district choice measure as the more appropriate. One good reason for this is that within-district attendance zone boundaries are subject to relatively frequent revision and in any case programs like magnet schools, busing, and formal or informal choice programs often weaken the link between residential location and within-district school assignment.²⁵ These make it difficult to exercise Tiebout choice

²⁴ In many states, elementary (grades 1-8) and secondary (grades 9-12) schools are managed by different, overlapping school districts. Families cannot meaningfully be said to choose between districts or schools operating at different levels (Urquiola, 1999). My choice indices are calculated over enrollment in grades 9-12 only.

²⁵ On desegregation remedies, which cut across and frequently dictate 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.

over schools within a large district, as families may not move frequently enough to keep up with changing boundaries.

The private enrollment rate is a clear indicator of whether the local governance structure provides sufficient Tiebout choice to satisfy parents: If parents can satisfy their desires through location decisions, they will be less likely to avail themselves of the private school option as a substitute. Table 2 presents models for metropolitan private enrollment rates, running a “horse race” between district and school choice indices. In column A, the metropolitan private enrollment rate is regressed on the district choice index and a variety of MSA-level control variables. In Column B, the school choice index is added as well. Column C adds controls for measures of micro-level residential segregation in the MSA. These, measured at the level of the census tract (much smaller than either a district or a school attendance zone), may be endogenous to choice, but their inclusion is useful to ensure that the choice indices are not capturing the effect of unrelated, unobserved determinants of social stratification in the MSA. The regressions indicate that the district choice index is a consistently strong predictor of the private enrollment rate, while the school index is not.

The remaining columns of Table 2 present further horse races between the two choice indices. In these columns the dependent variables are measures of the racial stratification of students across schools. The most basic prediction of Tiebout-style models like that in Section 3 is that choice will facilitate residential stratification. Thus, a variable that captures variation in available Tiebout choice should be positively associated with measures of stratification across schools (Eberts and Gronberg, 1981, and Epple and Sieg, 1999). Using both isolation and dissimilarity indices (Cutler, Glaeser, and Vigdor, 1999) for the distribution of white students relative to non-whites—data on racial composition of schools are much better than those on income distributions—and defining these indices over all schools or just public schools, Columns D through G indicate, again,

that the district choice index has a large effect in the expected direction. The school choice index has much larger positive coefficients, though recall that there is much less variation in this index.²⁶

In light of these inconclusive results, I follow previous authors and use the district index as my primary measure of Tiebout choice. I also present below a specification test that uses the school index, although the resulting estimate is too imprecise to be informative.

7. Within-market sorting results

The above sorting model suggests that if parents choose neighborhoods largely for the effectiveness of the local schools, equilibrium effectiveness sorting will depend on the educational market structure. Specifically, the gap in effectiveness between the schools attended by advantaged and disadvantaged students will tend to be larger when Tiebout choice makes it easier for wealthy parents to select effective schools without accepting unwanted peers. As a result, naïve estimates of the peer effect that do not account for effectiveness sorting should be larger in high-choice than in low-choice markets.

My first test of this prediction uses nonparametric techniques to allow for a nonlinear peer effect. Absent evidence of nonlinearity, I next turn to regression estimates of linear specifications. I also present estimates from alternative data sets; these are imprecise but completely consistent with those derived from the SAT data. None of the analyses indicate that the apparent peer effect is meaningfully larger in high-choice than in low-choice areas.

7.1. Nonparametric estimates

If neither the effect of individual characteristics on own scores (β in equations 1, 2, and 3) nor the reduced-form peer effect (γ) varies systematically with the structure of local school governance, and if sorting on effectiveness is more complete in high-choice than in low-choice markets, a version of Figure 1 which included data only from markets with high choice indices would exhibit a steeper slope than that shown, while a version estimated only from low-choice

²⁶ There is a mechanical relationship between the school index and these school-level segregation measures, as across-school segregation will tend to increase with the number of schools for fixed residential patterns.

markets would be less steep. This might be confounded if the causal peer effect is nonlinear, however, as Section 6 indicates that choice influences the dispersion of peer groups.

The median MSA contains only 16 high schools, not enough to permit unrestricted nonparametric estimation. As an alternative, I group MSAs into quartiles by the district choice index and estimate separate school-level kernel regressions of test scores on student characteristics for each quartile. I use an Epanechnikov kernel and a bandwidth of five (about one-tenth of a school-level standard deviation). Figure 5 displays the estimated functions. The figure offers little evidence of any important differences in reduced-form educational production functions among quartiles. One can discern a slight upward slope in the most-choice quartile relative to the least-choice, but the pattern is reversed for the second and third quartiles.

7.2. Regression estimates of linear models

In light of the near linearity of the nonparametric estimates in Figure 5, and to incorporate controls for MSA variables that might have independent effects on the housing market or on the causal importance of the peer group, I revert to a linear model. Letting m index housing markets,

$$\bar{t}_{jm} = \alpha_m + \bar{x}_{jm}(\beta + \gamma) + \mu_{jm} + \bar{\varepsilon}_{jm}, \text{ with} \quad (4)$$

$$E[\mu_{jm} | \bar{x}_{jm}; m] = \psi_m + \bar{x}_{jm} \theta_m^*. \quad (5)$$

A well-sorted market assigns high- $\bar{x}_{jm}$ students to high- μ_{jm} schools, and corresponds to a high value of θ_m^* . In general, for fixed parental valuations, δ , the sort may vary both with choice and with other metropolitan characteristics, Z_m :

$$\theta(c_m, Z_m; \delta) \equiv E[\theta_m^* | c_m, Z_m; \delta] = \varphi_0 + c_m \varphi_1 + Z_m \varphi_2. \quad (6)$$

The model in Section 3 suggests that if the peer group is not too important to parents, effectiveness sorting will be more complete when there are more jurisdictions, so our hypothesis is that $\varphi_1 > 0$.

Combining (4), (5), and (6), we obtain an estimable equation:

$$\begin{aligned} E[\bar{t}_{jm} | \bar{x}_{jm}; c_m, Z_m; \delta] = & (a_m + \psi_m) + \bar{x}_{jm}(\beta + \gamma + \varphi_0) + \bar{x}_{jm} c_m \varphi_1 + \bar{x}_{jm} Z_m \varphi_2 + \\ & + (\bar{x}_{jm} \omega_m + (\mu_{jm} - E[\mu_{jm} | \bar{x}_{jm}; m]) + \bar{\varepsilon}_{jm}), \end{aligned} \quad (7)$$

where $\omega_m \equiv \theta_m^* - \theta_m(c_m, Z_m; \delta)$ is the residual from (6), which I assume is independent of the stratification of peer groups (i.e. of the distribution of $\bar{x}_{jm} - \bar{x}_m$). Deviating from MSA means, this becomes

$$E[\bar{t}_{jm} - \bar{t}_m | \bar{x}_{jm} - \bar{x}_m; c_m, Z_m; \delta] = (\bar{x}_{jm} - \bar{x}_m)(\beta + \gamma + \varphi_0) + (\bar{x}_{jm} - \bar{x}_m)c_m\varphi_1 + (\bar{x}_{jm} - \bar{x}_m)Z_m\varphi_2 + \xi_{jm}. \quad (8)$$

The effect of choice on effectiveness sorting can thus be estimated as the coefficient on the interaction of peer group ($\bar{x}_{jm}$) with the choice index (c_m) in a within regression for school average test scores that also controls for interactions of the peer group with MSA-level control variables. All standard errors reported here are clustered at the MSA level to allow for the likely heteroskedasticity and within-MSA autocorrelation of the error term, ξ_{jm} (Kezdi, 2002).

Possible biases in estimates of (8)

Several identifiable factors may bias the coefficient on the peer group-Tiebout choice interaction in specifications like (8). I discuss two here; I argue that each has the potential to produce an upward bias in $\hat{\varphi}_1$.

The first potential bias relates to measurement error in the peer group variable: There may not be enough observations at any particular high school to accurately estimate the school-level average; the data may omit important background variables; or the included variables may be imperfectly measured—likely a particular problem for family income in the SAT data, which high school students are not likely to report reliably. Any of these would attenuate the gradient of school average student outcomes with respect to peer group characteristics

$\bar{x}_{jm}$ should be more reliable in markets where schools are more stratified, for several reasons. First, stratification implies a higher true variance of the peer group, and therefore a larger signal component of the signal-to-noise ratio. In the denominator, schools in more stratified markets are more internally homogenous; as the sampling variance of the school average depends linearly on the within-school variance of individual characteristics, more homogenous schools imply less noise in school averages. Finally, 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, in markets that are more heavily stratified.

The above arguments imply that the peer group coefficient in a regression using data from a single MSA should be attenuated to a lesser degree when the MSA has more stratified schools. As choice is positively correlated with stratification, this would produce a tendency toward larger estimated coefficients (i.e. less bias toward zero) in high-choice MSAs. In fact, I do not estimate separate regressions for each MSA, but the general effect is the same: Unreliability of the peer group measure is likely to bias upward the effect of choice on the measured peer group gradient.

A second possible bias in φ_1 is economic. There is some evidence that the educational labor market is more liquid in MSAs that have many districts competing for teachers' talent than in those with more concentrated governance (Luizer and Thornton, 1986). This may make it easier for a high- $\bar{x}_{jm}$ school to attract good teachers in a high-choice market than in one with less choice, where teachers are likely to be assigned to schools by bureaucratic rules rather than by the market. Any such effect would imply a positive effect of choice on the reduced-form peer effect— γ in equations (1), (4), and (7)—which will appear as a positive contribution to φ_1 .²⁷

Either of these effects would imply upward bias in estimates of φ_1 relative to the effect of interest. The results presented here should be seen as upper bounds on the effect of Tiebout choice on parental sorting into effective schools.

Basic results

Table 3 contains the main empirical results of the paper. It presents OLS estimates of (8), using the district choice index as a measure of Tiebout choice. Schools are weighted by the sum of individual SAT-taker observations' inverse sampling probabilities, with an adjustment at the MSA level to weight MSAs in proportion to their 17-year-old populations.

²⁷ Note that this effect has nothing to do with parents' use of their power to choose: It arises from teachers moving to schools with students who are easy to teach, rather than from parents moving to schools with good teachers.

Column A displays a very restricted version of model (8) that excludes all interactions between the peer quality index and metropolitan area characteristics. (That is, it sets $\varphi_1 = \varphi_2 = 0$; this is the model depicted in Figure 1.) It indicates that when all MSAs in the sample are pooled, the gradient of school average SAT scores with respect to the characteristics of SAT-takers is 1.74. A one standard deviation (48 points) change in the mean background of students at a school corresponds, on average, with an 84 point (0.88 school-level standard deviations) change in average SAT scores. This, of course, reflects the combined influence of individual characteristics (β), peer effects (γ) and the within-MSA gradients of school effectiveness with respect to peer group (θ^* s).

Column B adds a single interaction of the peer group with a choice index. The estimate of φ_1 is nearly zero. The remaining columns add additional interactions of $\bar{x}_{jm}$ with several metropolitan-level controls (deviated from their sample means) that might capture other determinants of the sorting process, the distribution of school quality, the reduced-form peer effect, or the sample selection process. Moving from left to right, these include controls for the MSA-level SAT-taking rate; six census divisions; the log of the MSA population; and three combinations of additional demographic, income distribution, and institutional controls. In each specification but the last, the φ_1 point estimate is negative, although it is only significantly different from zero in columns D and E. By the rightmost columns, the model is somewhat overfit: there are 17 interactions of MSA variables with the $\bar{x}_{jm}$ index and only 179 MSAs in the sample. Hence, my preferred specification is that in Column F, which indicates a small, insignificant, negative effect of the choice index on the test score gradient.

Specification tests

Table 4 reports the results of several alternative specifications, each using the control variables from Column F of Table 3. Row 1 repeats the relevant coefficients from that specification. In Row 2, the peer effect is allowed to depend on the standard deviation of student characteristics as well as on their average level. The standard deviation term enters significantly, indicating that

heterogeneous schools produce higher scores than homogenous schools with the same average student background. The choice-peer group average interaction is slightly more negative than in Row 1.

Row 3 allows the racial and ethnic composition of SAT-takers to have an independent effect on average SAT scores. If there are cultural biases in SAT scores, for example, individual ethnicity may have a different effect than does the composition of the peer group. The coefficients on racial composition variables are large and significant, but again their inclusion has essentially no effect on the parameter of interest, the interaction of average peer quality with Tiebout choice.

Row 4 tests a different aspect of the specification, the assumption that the background characteristics predicting SAT scores are identical to those indexing willingness-to-pay for desirable schools. To test this, I allow willingness-to-pay to depend on students' self-reported family income, estimating the interaction between income and Tiebout choice while including the peer quality index to absorb peer effects. The interaction coefficient here is positive, but small and insignificant.

Row 5 tests the effect of replacing the district choice index with its counterpart defined over schools, as discussed in Section 6. Here, the interaction coefficient of interest is positive and nearly significant. Recall, however, that the school choice index exhibits far less variation than does the district index, so even the point estimate here indicates that the gradient of test scores with respect to student characteristics rises by only about 0.06 with a one standard deviation increase in the school choice index.

The remainder of the table returns to the specification in Row 1, and explores variations in the sample, weighting, and estimation procedure. Row 6 presents estimates from a re-weighted sample designed to locate biases deriving from selection into SAT-taking. I take advantage of a variable on the SAT data characterizing each student's high school class rank (reported as first or second decile, or as a quintile below that). I discard all observations reporting a rank in the bottom 40% of their class, then re-weight the remaining observations so that the weighted class rank distribution at each school is balanced, with one sixth of the overall weight in each of the top two deciles and one third in each of the next two quintiles. Under the assumption that selection into

SAT-taking is random conditional on school attended and rank, this re-weighting should recover estimates that would be obtained were SAT scores available for all students ranked in the top three quintiles of their schools. The estimates from the re-weighted data in row 6 are nearly identical to those from the base specification.

Additional estimates, not reported here, have used Heckman-style models in an attempt to avoid selection biases (Heckman, 1990). The difficulty is that there is no plausible instrument for SAT-taking that is exogenous to latent scores. Thus, I used the school SAT-taking rate as a summary of the selectivity of school's SAT-taker sample, including in the basic model a control function of the SAT-taking rate. The implied selectivity bias parameters are quite small, indicating essentially no relationship between SAT-taking and latent scores. A likely explanation is that the implicit functional form assumptions are insufficient to identify the effect of interest without an instrument for test-taking. However, one can test the sensitivity of the parameters of interest by adjust school SAT means for different assumed selectivity parameters. SAT averages are quite insensitive to assumptions about selectivity: The cross-sectional correlation between school SAT-taking rates and SAT averages is large and positive, and actual SAT averages correlate well above 0.95 with corrected averages even under large—correlations of 0.25 and higher—assumed selection parameters. Although this is not conclusive, it, together with the re-weighted sample estimate in Row 6, suggests that the basic results in Table 3 are not seriously biased by selection into SAT-taking.

Rows 7 and 8 of Table 4 explore the impact of varying the sample definition. In Row 7, the basic model is estimated on public schools only, while in row 8 the 20 MSAs that have only a single district are excluded.²⁸ Both (district) choice-peer group interaction coefficients are negative, in the latter case significantly so and with a larger point estimate than in the basic specifications.

The final rows of the table investigate the possibility that the district choice index is endogenous to school quality or to the sorting process. This might be true if, as Hoxby (2000) proposes, suburban school districts have been less willing to consolidate with their central-city

²⁸ The choice distribution is bimodal, and these 20 MSAs form a spike nearly three standard deviations from the mean.

neighbors in areas where the central-city schools are badly run.²⁹ Hoxby proposes generating exogenous variation from differences across MAs in the initial, pre-consolidation governance structure. In Row 9, I estimate an instrumental variables specification with an instrument similar to Hoxby’s “smaller streams” variable, derived from the U.S. Geologic Survey’s Geographic Names Information System, which might capture geographic differences in the optimal district layout before modern transportation networks were developed. Row 10 uses a stronger instrument, a choice index computed using information on the number of school districts operating in the MA in 1942 (Gray, 1944), just at the beginning of the great post-war wave of district consolidation. By the logic underlying the streams approach, this historical measure should be a valid instrument. The latter model produces a positive coefficient, but does not reject either zero effect or the OLS point estimate.³⁰

Although results are not presented here, I have estimated several additional specifications of the basic empirical test. The absence of a positive choice effect does not seem to derive from the particular weighting of the data used here—one might prefer to weight MSAs equally, or by the number of SAT-takers, rather than by their high-school-age populations—nor from the inclusion in the sample of schools with too few SAT-takers to permit accurate estimation of the school mean.

7.3. Evidence from the NELS

The SAT data are uniquely valuable for my empirical strategy, both because they span a large fraction of metropolitan high schools and because they describe an outcome that is important to parents. Nevertheless, it seems desirable to test the conclusions from the above analysis on other data sets that are not subject to the selection issues deriving from the SAT-taking decision.

I use test scores and continuation rates from the National Educational Longitudinal Survey (NELS) for this purpose. The NELS is not nearly as broad a sample as is provided by the SAT data,

²⁹ There was a great deal of school district consolidation in the United States in the post-war period, with the total number of districts falling by nearly 80% between 1950 and 1970. The latter year marked the approximate end of the consolidation wave (Kenny and Schmidt, 1994, and National Center for Education Statistics, 1973).

³⁰ In an MSA-level regression for the choice index, F statistics for tests of the exclusion of the instruments are 21 and 134, respectively. Formal Hausman tests are inappropriate, as observations are not assumed independent within MSAs and OLS is not efficient even if the choice index is exogenous.

so the estimates presented here are imprecise, but the point estimates are reassuringly similar. The NELS sampled about 23 eighth grade students from each of 815 public and 237 private schools in 1988, following up with portions of this original sample at two-year intervals thereafter. Using a confidential version of the NELS data, I am able to match 748 schools (564 public and 184 private) in the original NELS sample to the 211 MSAs in which they are located.

The first panel of Table 6 presents estimates using the composite test scores that students earned during the original wave of the NELS, when they were in 8th grade. (I continue to use the district choice index defined over high school enrollment in this analysis.) Column A presents the coefficient from a regression of school average scores on an index of student quality, pooling all metropolitan schools in the NELS sample and including a fixed effect for each MSA.³¹ As in the SAT data, peer effects and effectiveness sorting are together substantial, inflating the school-level background index coefficient by two thirds relative to the coefficient of a within-school regression of individual scores on own characteristics. When the peer group measure is interacted with the choice index—in Column B, and again with additional controls in the remaining columns—the coefficient is indistinguishable from zero, with a negative point estimate in every specification.

The second panel models school continuation—the fraction of students from the NELS 8th grade sample who remained in school at the time of the follow-up survey four years later—as this is a non-test outcome that may particularly concern parents.³² The background index used is the same as that used in Panel A; it is a strong predictor of continuation rates but there is no evidence that it is a stronger predictor in markets with more Tiebout choice.

Given the lack of precision in the NELS estimates, it is somewhat surprising how well they line up with those in Tables 3 and 4. As before, the choice effect is indistinguishable from zero, but point estimates suggest that the apparent peer effect is *smaller* in high-choice markets. There is nothing to indicate that the SAT-based results are an aberration.

³¹ The background measure is a weighted average of variables characterizing students' race and their parents' education, again using weights chosen to best predict student test scores within schools.

³² There is substantial sample attrition, not necessarily random, in the follow-up survey. Panel B uses only individuals who actually appear in both the original and follow-up surveys, with the survey longitudinal weights.

7.4. Does choice reduce variation in effectiveness?

The above results indicate that the school-level association between test scores and student characteristics is no larger in high-choice than in low-choice areas. The implied interpretation is that increases in Tiebout choice do not facilitate the allocation of the most effective schools to the highest-income families, and that these parents must not be using their choice to pursue effective schools. There is another possibility, however: A positive association of choice with effectiveness sorting may be masked by a negative association with the overall within-MSA variance of school effectiveness. That is, it may be that the correlation between effectiveness and peer group is increasing with choice but that the coefficient of a regression of the former on the latter— θ_m^* in (3) and (5)—is not, if $\text{var}(\mu_{jm})$ is decreasing. This might be the case if, as in Hoxby's (1999b) model with discrete student characteristics, competition serves to reduce variation in administrative effort.

If this claim is true, choice should have an unambiguously negative effect on the residual variance in (8), both because there is less variation to be explained and because the peer group absorbs more of it.³³ To examine this, I estimated the residual variance in each MSA for the model in Column F of Table 3 (after accounting for the portion due to small-sample measurement error in schools' average SAT scores) and regressed this on the MSA characteristics included in interaction terms in the original model. The choice coefficient is -2,533 with a standard error of 778. (The overall average residual variance is 3,184.) This lends some support to the hypothesis that schools in high-choice areas compete toward a common level of effectiveness.

Any resulting bias in estimates of effectiveness sorting must be quite small, however. If there are no influences on school average test scores except for peer group characteristics and the school's effectiveness, the effect of choice on the peer group-effectiveness correlation can be calculated from its effects on the test score gradient, the residual variance, and the across-school variance of peer groups. These calculations, available from the author, indicate that if the average peer group-effectiveness correlation is 0.3 in single-district markets (with choice indices of zero), it

³³ Note that the clustered standard errors reported in Tables 3 through 5 allow for arbitrary heteroskedasticity.

averages approximately 0.33 in markets where the index is one. (Ignoring the effect of choice on residual variance produces an estimate of 0.24 for the high-choice markets.) The residual variance adjustment has the greatest effect if zero-choice markets have a correlation around 0.6, which implies a competitive market correlation of 0.71. For baseline correlations below 0.25, even the adjusted choice effect on the effectiveness-peer group correlation is negative. The point estimates do not indicate that choice, as measured by the district concentration index, facilitates substantially more complete effectiveness sorting.

8. Choice and average scores

The result that Tiebout choice does not substantially increase the probability that effective schools attract students from advantaged backgrounds casts doubt on the claim that it can provide incentives that will lead administrators to exert greater effort. The argument (Brennan and Buchanan, 1980; Hoxby, 2000) that average school performance should be higher in markets with decentralized governance may thus not be expected to hold. I present here tests of that prediction from the SAT data.

Table 7 presents regression models for the average level of SAT scores across MSAs.³⁴ Column A includes only the district choice index as a regressor. It enters with a positive coefficient, implying that fully decentralized MSAs produce average SAT scores about forty points higher than do those with only a single district. Recall, however, that the choice index is strongly correlated with both SAT-taking rates and demographic characteristics. Column B adds controls for the SAT-taking rate and for the average background index of SAT-takers. The choice coefficient is negative and significant here, indicating that the positive effect in Column A derives from the omission of students' background characteristics. The MSA SAT-taking rate, which should have a negative effect on average scores if SAT-takers are positively selected, comes in with an effect indistinguishable from zero. Columns C through F add additional MSA-level regressors. Their

³⁴ Hoxby (2000) presents estimates of the same effect from individual-level data with linear controls for the peer group, arguing that these are more efficient than are estimates from aggregated data. I present the more conservative MSA-level results, although individual-level estimates are not substantially different.

inclusion does not substantially alter the estimated effect of choice: It remains negative and significant. Note that the SAT-taking rate takes on its expected sign in these columns, but that the estimated choice effect appears robust to the form in which the participation rate is controlled.

Finally, Columns H and I report instrumental variables estimates of the model for MSA-average SAT scores, using the same instruments as in Table 4. The streams instrument indicates a larger negative effect of choice, while the 1942 choice index produces a smaller effect that is indistinguishable from zero. A Hausman test using the former specification rejects the OLS model, but one based on the latter does not.

The negative effect of Tiebout choice on average SAT scores indicated by Table 7 is not very large: A one standard deviation (0.26) increase in the choice index in Column E corresponds with a reduction in mean scores of only about 3.8 points, about one-ninth of an MSA-level standard deviation. Moreover, when MSAs are weighted equally rather than by the 17-year-old population or by the number of SAT takers (not shown, but similar to the former weighting in Table 6), the choice effect is approximately zero. Nevertheless, there is no indication in any specification that Tiebout choice is associated with higher SAT scores once student background is controlled.³⁵ Moreover, the coefficient on the background index across MSAs—1.77 in Column C, and slightly larger in later columns—is quite similar to that found within MSAs (Table 3). This suggests that both coefficients may measure primarily the sum of individual and peer effects ($\beta + \gamma$), which might be the same across MSAs as within, rather than effectiveness sorting (θ), which we would expect to see within but not across MSAs.

The results across MSAs thus support those within MSAs: The evidence does not indicate that productive administrators are more likely to be rewarded with increased demand for local housing in high-choice than in low-choice areas, nor that Tiebout choice as measured here provides meaningful incentives for school administrators to improve productivity.

³⁵ Hoxby (2000), using much less rich data from the NELS and NLSY, finds a positive effect of choice on average scores across MSAs, one that is larger for “not low income” than for low income students. The SAT sample might be thought analogous to her “not low income” group. See Rothstein (2003) for further exploration of this discrepancy.

9. Conclusion

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

This paper has developed observable implications of parental valuations of school characteristics in the context of Tiebout choice, the selection of schools via housing location decisions. I have argued that strong parental preferences for school effectiveness produce “effectiveness sorting,” a correlation between the effectiveness of a school and the wealth of its students, in Tiebout equilibrium, and that this will magnify observational estimates of the peer effect. Moreover, this sorting (and magnification) can be expected to be more complete in housing markets that offer a great deal of Tiebout choice than in those where the link between residential location and school assignments is weaker.

Tests of this prediction offer no evidence that the gradient of school average test scores with respect to student characteristics varies systematically with measures of Tiebout choice. Even at the upper extreme of the estimated confidence intervals, the SAT gap between more- and less-desirable schools is not meaningfully larger in markets with decentralized governance than in those with less Tiebout choice.

Several candidate explanations for this non-result appear unsatisfactory. The choice index used seems to capture meaningful variation in the availability of Tiebout choice, as it is strongly predictive of both private enrollment and across-school racial stratification. Moreover, despite my use of observational rather than experimental variation in choice, the coefficients of interest do not seem particularly sensitive to the inclusion of additional control variables. Finally, although I find

some evidence that choice is associated with reductions in the dispersion of effectiveness across schools, this effect is too small to mask a large effect of choice on the allocation of effective schools.

One additional hypothesis merits mention. The tests developed here do not well distinguish the case where parents value school effectiveness to the exclusion of all else and that where they attach no value to effectiveness, as in neither case does the theoretical model predict differences in effectiveness sorting between high- and low-choice markets. The former is implausible, though, for two reasons. First, parents seem to believe that peer effects are important, so would be unlikely to ignore the peer group in their choices. Second, the performance measures available to parents—generally average outcomes at the school—load heavily on the peer group even in the absence of peer effects, and parents are likely unable to fully correct for this.

The three most plausible explanations for the current results are that parents place a low weight on school effectiveness in their preferences over neighborhoods; that parents value effectiveness greatly but lack the information needed to identify effective schools; or that the sorts of policies that I call “effectiveness,” those not causally dependent on the enrolled population, are relatively unimportant components 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 incentives for school administrators to exert greater effort to raise student performance (Chubb and Moe, 1990).³⁶ Consistent with this, though not with the earlier results of Hoxby (2000), I find no evidence that choice produces higher average SAT scores.

Great caution is required in generalizing from this paper’s results to choice markets that do not link school assignment to residential location. Under Tiebout choice, parents may have to give up desired neighborhood amenities—views, parks, air quality, or characteristics of neighbors—to obtain a more effective school. They may be unwilling to do this even though they would choose the better school were that choice separable from the residential location decision. Moreover, voucher programs that encourage the entry of new competitors may produce more options for

³⁶ Under the second explanation, the provision of new information might materially change the nature of Tiebout equilibrium. It will be interesting to study the effects of recent accountability programs, which purport to provide parents with annual measures of school performance, on metropolitan housing markets.

parents than even the most decentralized of district governance structures, reducing the potential for coordination failures and increasing the probability that even parents who value the peer group highly will choose effective schools.

It nevertheless seems likely that parental choices and equilibrium allocations will depend crucially on the particular design of any choice program. Such programs may need mechanisms—like compensatory funding, better and more widely available performance measures, or targeted vouchers—that reduce the importance of peer groups in parental choices in order to produce market incentives toward greater school productivity.

References

- Antos, Joseph R., and Sherwin Rosen (1975). “Discrimination in the Market for Public School Teachers.” *Journal of Econometrics* 3 (2), 123-150.
- Barrow, Lisa (2002). “School Choice Through Relocation: Evidence from the Washington, D.C. Area.” *Journal of Public Economics* 86, 155-189.
- Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2003). “A Unified Framework for Measuring Preferences for Schools and Neighborhoods,” mimeo, July.
- Belfield, Clive R., and Henry M. Levin (2001). “The Effects of Competition on Educational Outcomes: A Review of U.S. Evidence.” Occasional Paper No. 35, National Center for the Study of Privatization in Education, Teachers College, Columbia University. September.
- Black, Dan, Gary Gates, Seth Sanders, and Lowell Taylor (2002). “Why Do Gay Men Live in San Francisco?” *Journal of Urban Economics* 51, 54-76.
- Black, Sandra E. (1999). “Do Better Schools Matter? Parental Valuation of Elementary Education.” *Quarterly Journal of Economics* 114 (2), May, pp. 577-599.
- Bogart, William, and Brian Cromwell (2000). “How Much is a Neighborhood School Worth?” *Journal of Urban Economics* 47 (2), March, 280-305.
- Borland, Melvin V., and Roy M. Howsen (1992). “Student Academic Achievement and the Degree of Market Concentration in Education.” *Economics of Education Review* 11 (1), 31-39.
- Brennan, Geoffrey, and James Buchanan (1980). *The Power to Tax: Analytical Foundation of a Fiscal Constitution*. Cambridge: Cambridge University Press.
- Card, David, and A. Abigail Payne (2002). “School Finance Reform, the Distribution of School Spending, and the Distribution of SAT Scores.” *Journal of Public Economics* 83, 49-82.
- Chubb, John, and Terry M. Moe (1990). *Politics, Markets, and America's Schools*. Washington, D.C.: The Brookings Institution.
- Clark, Melissa (2003). “Selection Bias in College Admissions Test Data,” mimeo.

- Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt (2000). "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools." Working Paper #7888, National Bureau of Economic Research, September.
- Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor, (1999). "The Rise and Decline of the American Ghetto." *Journal of Political Economy* 107(3), 455-506.
- Dynarski, Mark (1987). "The Scholastic Aptitude Test: Participation and Performance." *Economics of Education Review* 6, 263-73.
- Eberts, R. W., and T. J. Gronberg (1981). "Jurisdictional Homogeneity and the Tiebout Hypothesis." *Journal of Urban Economics* 10, 227-239.
- Epple, Dennis N., and Holger Sieg (1999). "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107 (4), August, 645-681.
- Evans, William N., Wallace E. Oates, and Robert M. Schwab (1992). "Measuring Peer Group Effects: A Study of Teenage Behavior." *Journal of Political Economy* 100 (5), October, 966-991.
- Fernandez, Raquel, and Richard Rogerson (1996). "Income Distribution, Communities, and the Quality of Public Education." *Quarterly Journal of Economics* 111 (1), February, 135-164.
- _____, and _____ (1997). "Keeping People Out: Income Distribution, Zoning, and the Quality of Public Education." *International Economic Review* 38 (1), February, 23-42.
- Figlio, David N., and Maurice E. Lucas (2000). "What's in a Grade? School Report Cards and House Prices." Working Paper #8019, National Bureau of Economic Research.
- Friedman, Milton (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.
- Gray, E. R. (1944). *Governmental Units in the United States 1942*. Washington D.C.: Bureau of the Census.
- Hanushek, Eric A. (1981). "Throwing Money at Schools." *Journal of Public Policy Analysis and Management* 1 (1), 19-41.
- _____. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24 (3), September, 282-294.
- Heckman, James (1990). "Varieties of Selection Bias," *American Economic Review* 80 (2), 313-318.
- Hilber, Christian A. L., and Christopher J. Mayer (2002). "Why Do Households Without Children Support Local Public Schools? Linking House Price Capitalization to School Spending." Mimeo, December
- Hirschman, Albert O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Cambridge, Mass.: Harvard University Press.
- Hoxby, Caroline M. (1994). "Do Private Schools Provide Competition for Public Schools?" Working Paper #4978, National Bureau of Economic Research.
- _____. (1999a). "The Effects of School Choice on Curriculum and Atmosphere." In Mayer, Susan E., and Peterson, Paul E., eds., *Earning and Learning: How Schools Matter*. Washington, DC: Brookings Institution Press. Chapter 11, 281-316.
- _____. (1999b). "The Productivity of Schools and Other Local Public Goods Producers," *Journal of Public Economics* 74 (1), October, 1-30.

- _____. (2000). "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90 (5), December, 1209-38.
- Hsieh, Chang-Tai, and Miguel Urquiola (2002). "When Schools Compete, How Do They Compete? An Assessment of Chile's Nationwide School Voucher Program." Unpublished manuscript, January.
- Kane, Thomas J., and Douglas O. Staiger (2002). "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives* 16 (4), 91-114.
- Kenny, Lawrence W., and Amy B. Schmidt (1994). "The Decline in the Number of School Districts in the U.S.: 1950-1980." *Public Choice* 79 (1-2), April, 1-18.
- Kézdi, Gábor (2002). "Robust Standard Error Estimation in Fixed-Effects Panel Models," mimeo, University of Michigan, February.
- Krueger, Alan B., and Pei Zhu (2003). "Another Look at the New York City School Voucher Experiment." Princeton University: Education Research Section Working Paper #1, April.
- Luizer, James, and Robert Thornton (1986). "Concentration in the Labor Market for Public School Teachers." *Industrial and Labor Relations Review* 39 (4), 573-584.
- Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60, 531-542.
- Mayer, Daniel P., Paul E. Peterson, David E. Myers, Christina Clark Tuttle, and William G. Howell (2002). "School Choice in New York City After Three Years: An Evaluation of the School Choice Scholarships Program," Mathematica Policy Research, Inc.: Washington, DC.
- National Center for Education Statistics (1973). "Elementary and Secondary General Information System (ELSEGIS): Public School Universe Data, 1969-1970 Through 1972-1973." Inter-university Consortium for Political and Social Research, Study no. 2238.
- _____. (1999). *Digest of Education Statistics 1998*. Washington D.C.: U.S. Department of Education.
- Orfield, Gary (1983). *Public School Desegregation in the United States, 1968-1980*. Washington: Joint Center for Political Studies.
- Reback, Randall (2001). "Capitalization Under School Choice Programs: Are the Winners Really the Losers?" University of Michigan Office of Tax Policy Research Working Paper 2001-21, November.
- Ross, Steven, and John Yinger (1999). "Sorting and Voting." In Cheshire, Paul, and Edwin S. Mills, eds, *Handbook of Regional and Urban Economics* vol. 3. New York: North-Holland.
- Rothstein, Jesse M. (2003). "Does Competition Among Public Schools Really Benefit Students? A Reappraisal of Hoxby (2000)," mimeo, May.
- _____. (forthcoming). "College Performance Predictions and the SAT," *Journal of Econometrics*.
- Rouse, Cecilia Elena (1998). "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *The Quarterly Journal of Economics* 113 (2), May.
- Technical Design Group (2000). "Construction of California's 1999 School Characteristics Index and Similar Schools Ranks." PSAA Technical Report 00-1. Sacramento: California Department of Education.

- Tiebout, Charles M. (1956). “A Pure Theory of Local Public Expenditures.” *Journal of Political Economy*, 64 (5), October, 416–24.
- U.S. Census Bureau (1993). “Census of Population and Housing, 1990 [United States]: Summary Tape File 3C,” Inter-university Consortium for Political and Social Research, Study No. 6054.
- (2001). “Metropolitan Areas and Components, 1990 with FIPS Codes,” released November, 1998; revised March, 2001 (<http://www.census.gov/population/estimates/metro-city/90mfips.txt> and [90nfips.txt](http://www.census.gov/population/estimates/metro-city/90nfips.txt)).
- Urquiola, Miguel (1999). “Demand Matters: School District Concentration, Composition, and Educational Expenditure.” Working Paper #14, University of California, Berkeley, Center for Labor Economics, April.
- Welch, Finis, and Audrey Light (1987). *New Evidence on School Desegregation*. United States Commission on Civil Rights Clearinghouse Publication 92, June.

Appendix: Proofs of Propositions

It is useful to begin with the following Lemma:

Lemma 1. Assume two communities, $(\mu_j, \bar{x}_j, b_j)$ and $(\mu_k, \bar{x}_k, b_k)$, with the first community having both higher quality schools and higher housing prices: $\bar{x}_j\delta + \mu_j > \bar{x}_k\delta + \mu_k$ and $b_j > b_k$. If the single crossing property for preferences holds,

- When a family with income x_0 (weakly) prefers j to k , all families with incomes $x > x_0$ strictly prefer j to k .
- When a family with income x_0 (weakly) prefers k to j , all families with incomes $x < x_0$ strictly prefer k to j .

Proof of Lemma 1. I prove part I; the remainder follows directly by a similar argument. Define “quality” as $q_j \equiv \bar{x}_j\delta + \mu_j$. Begin by assuming that the communities have similar q ’s and b ’s, so that first-order Taylor expansions of the utility function are accurate, and consider an expansion around the utility that family x obtains in district k , evaluated at $(x - b_j, q_j)$:

$$U(x - b_j, q_j) - U(x - b_k, q_k) \approx -(b_j - b_k)U_1(x - b_k, q_k) + (q_j - q_k)U_2(x - b_k, q_k), \text{ or} \quad (A1)$$

$$\frac{U_2(x - b_k, q_k)}{U_1(x - b_k, q_k)} \approx \frac{1}{(q_j - q_k)} [U(x - b_j, q_j) - U(x - b_k, q_k) + (b_j - b_k)] \quad (A2)$$

Note that the derivative of $U_2(x - b_k, q_k)/U_1(x - b_k, q_k)$ with respect to x is

$(U_{21}U_1 - U_{11}U_2)/U_1^2 > 0$. Thus, for $x > x_0$,

$$\frac{U_2(x - b_k, q_k)}{U_1(x - b_k, q_k)} > \frac{U_2(x_0 - b_k, q_k)}{U_1(x_0 - b_k, q_k)}, \text{ and by (A2),} \quad (A3)$$

$$U(x - b_j, q_j) - U(x - b_k, q_k) > U(x_0 - b_j, q_j) - U(x_0 - b_k, q_k). \quad (A4)$$

By assumption, the right side of (A4) is non-negative, so the left side must be positive.

Now suppose that districts j and k are discretely different. The single-crossing property holds everywhere and is so named because it implies that two families’ indifference curves in q - b

space either exactly coincide (if the families have the same x) or cross no more than once.³⁷ Consider family x 's indifference curve ($x > x_0$) through (q_j, b_j) . I have shown that this curve is superior to that through $(q_j - \varepsilon, b_j - \nu)$ for small ε and ν such that $U(x_0 - b_j + \nu, q_j - \varepsilon) = U(x_0 - b_j, q_j)$. Because x 's curve crosses family x_0 's indifference curve at (q_j, b_j) , it cannot cross anywhere else, so in particular must remain strictly below family x_0 's at all points to the left of (q_j, b_j) . As by assumption (q_k, b_k) is one such point, and as family x_0 does not prefer (q_k, b_k) to (q_j, b_j) , family x must strictly prefer (q_j, b_j) to (q_k, b_k) . $\square$

Proof of Theorem 1. Let $r_G(j)$ be the mapping from community index numbers to quality rank in some allocation rule G , so that j has higher quality than k under G iff $r_G(j) < r_G(k)$. Also, define $\tilde{x}_j \equiv F^{-1}(1 - \frac{jn}{N})$ as the income of the j n wealthiest family. The Theorem states that the following are necessary conditions for G to be an equilibrium allocation with housing price vector $(b_1, \dots, b_J)$:

- i. $b_j > b_k$ whenever $r_G(j) < r_G(k)$,
- ii. If $x \geq \tilde{x}_1$, then $r_G(G(x)) = 1$, and
- iii. If $\tilde{x}_{j-1} > x \geq \tilde{x}_j$ for $j = 2, \dots, J$, then $r_G(G(x)) = j$.

I prove this by contradiction. First, suppose that (i) fails to hold, so that there exist j and k such that $r_G(j) < r_G(k)$ but $b_j \leq b_k$. In words, this means that j has higher quality and no lower housing prices than does k . It thus strictly dominates k in every family's preferences. Because every community must have at least one resident (by EQ2 and $N > n(J-1)$), there must be some family assigned to community k , violating EQ1.

Now consider the remaining conditions. If (ii) or (iii) fails to hold in equilibrium, there must be some empty house in a community that is not lowest in quality, or there must be at least some pair of incomes (x, y) such that $x < y$ but $r_G(G(x)) < r_G(G(y))$. The former cannot be true, by the above argument, because empty houses imply zero housing prices (EQ2). So suppose the latter. By revealed preference, family x prefers district $G(x)$ to $G(y)$, and the rank condition means that $\bar{x}_{G(x)}\delta + \mu_{G(x)} > \bar{x}_{G(y)}\delta + \mu_{G(y)}$. By Lemma 1, then, family y also prefers district $G(x)$ but lives in district $G(y)$, violating EQ2. $\square$

Proof of Theorem 2. I prove the Theorem by construction. First, without loss of generality, let the μ_j s be sorted in descending order: $\mu_j > \mu_{j+1}$ for all $j < J$. Define an allocation rule:

$$\tilde{G}(y) = \begin{cases} j & \text{whenever } F(x) \in [1 - \frac{jn}{N}, 1 - \frac{(j-1)n}{N}), j = 1, \dots, J; \\ 1 & \text{when } F(x) = 1. \end{cases} \quad (\text{A5})$$

This rule assigns the n highest-income families to district 1—the district with the highest μ —the next n families to district 2; and so on. To construct housing prices that make this allocation an equilibrium, let $\tilde{b}_J = 0$. For $j < J$, let

³⁷ Note that $U_2(x-b, q)/U_1(x-b, q)$ is the slope of x 's indifference curve through (b, q) , so single-crossing says that the slope of the curve through this point is increasing in family income. Since this is true for all (b, q) , curves that cross at one point cannot cross elsewhere.

$$\tilde{b}_j = \tilde{b}_{j+1} + (q_j - q_{j+1}) \frac{U_2(\tilde{x}_j - b_{j+1}, q_{j+1})}{U_1(\tilde{x}_j - b_{j+1}, q_{j+1})}, \quad (\text{A6})$$

where $\tilde{x}_j \equiv F^{-1}\left(1 - \frac{jn}{N}\right)$. These prices make family $\tilde{x}_j$ indifferent between districts j and $j+1$.

I demonstrate that $\tilde{G}(\cdot)$ and $\{\tilde{b}_1, \dots, \tilde{b}_J\}$ are an equilibrium. By Lemma 1, all families in districts $k > j$ —who under $\tilde{G}(\cdot)$ have incomes $x < \tilde{x}_j$ —will strictly prefer district $j+1$ to j under $\tilde{b}$, while all families in districts $k < j+1$ —other than the boundary family—will strictly prefer district j to $j+1$. Since this holds for all j , there cannot be any family who prefers another district to the one to which it is assigned by $\tilde{G}(\cdot)$, so EQ1 is satisfied.

For EQ2, note that $\int \mathbf{1}(G(x) = j) f(x) dx = F\left(F^{-1}\left(1 - \frac{(j-1)n}{N}\right)\right) - F\left(F^{-1}\left(1 - \frac{jn}{N}\right)\right) = \frac{n}{N}$ for each $j < J$ and that $\int \mathbf{1}(G(x) = J) f(x) dx \leq \frac{n}{N}$, the latter a direct result of $N \leq Jn$.

All that remains is to demonstrate EQ3. By definition of $\tilde{G}(\cdot)$, $\bar{x}_j > \bar{x}_k$ whenever $j < k$, which also implies that $\mu_j > \mu_k$. For any $\delta \geq 0$, then, $\bar{x}_j \delta + \mu_j > \bar{x}_k \delta + \mu_k$, so in particular $\bar{x}_j \delta + \mu_j \neq \bar{x}_k \delta + \mu_k$. $\square$

Proof of Theorem 3. By Theorem 1, equilibrium allocations amount to permutations of J bins of the income distribution among the J communities, with the restriction that wealthier families live in higher-quality districts. When $\delta = 0$, $q_j \equiv \bar{x}_j \delta + \mu_j \equiv \mu_j$, so the only possible quality ranking is the ranking by effectiveness and only one permutation is admissible. (When $\delta > 0$, a high-income population can allow an ineffective school to outrank an effective one.)

In order to maintain this permutation as an equilibrium, housing prices must separate the income bins, so that the lowest-income family in bin j prefers district j to the next-lower-ranked district, while the next-lower-income family has the opposite preferences. The price vector described in the proof of Theorem 1 accomplishes this. If the income distribution is continuous, any change to that vector would change some family's choices so would not be an equilibrium. With a discrete income distribution, the price vector could change by small amounts without altering any family's choices, and therefore many prices could exist in equilibrium. $\square$

Proof of Theorem 4. Let G be an allocation rule satisfying Theorem 1 such that $\bar{x}_{G(y)} \delta + \mu_{G(y)} \geq \bar{x}_{G(w)} \delta + \mu_{G(w)}$ whenever $y > w$ (with strict inequality when $G(y) \neq G(w)$), and let $r_G(j)$ be the quality rank of district j under G . Define housing prices $\tilde{b}_{r(J)} = 0$ and

$$b_{r(j)} = b_{r(j+1)} + \frac{U_2(\tilde{x}_j - b_{r(j+1)}, q_{r(j+1)})}{U_1(\tilde{x}_j - b_{r(j+1)}, q_{r(j+1)})} (q_{r(j)} - q_{r(j+1)}) \quad (\text{A7})$$

for $j < J$. These housing prices, together with G , form an equilibrium.

EQ2 and EQ3 follow directly from the assumptions. The housing prices are constructed to make the family with income $\tilde{x}_j$ indifferent between the j th and $(j+1)$ th ranked communities. By Lemma 1, families with $x > \tilde{x}_j$ prefer the j th community, while those with $x < \tilde{x}_j$ prefer the $(j+1)$ th. As this must hold for each j , EQ1 must hold globally. $\square$

Figure 1: Student characteristics-SAT score relationships, across and within MSAs

Panel A: Across MSAs

Panel B: Within MSAs, Across Schools

Notes: Each circle represents one MSA (in Panel A; N=179) or one school (in Panel B; N=5,779). Circle areas are proportional to the number of SAT-takers in the MSA/school, with different factors of proportionality in the two panels. Text labels indicate the slopes of the weighted least squares regression estimates.

Figure 2: Illustrative allocations of school effectiveness and community desirability in a toy market

Figure 3. 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 4: 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 5. Kernel estimates of the student background-test score relationship, by choice quartile

Notes: Sample MSAs are divided into quartiles by their district choice indices. The quartile thresholds are indices of 0.74, 0.85, and 0.91. Within each quartile, the indicated functions are estimated school-level kernel means, using an Epanechnikov kernel with a bandwidth of 5. Schools are weighted by their number of SAT-takers, adjusted at the MSA level to weight MSAs by their 17-year-old populations.

Table 1. Summary statistics for MSAs, SAT-takers, and schools

	Mean	S.D.	MA-level correlation with choice over:	
	(A)	(B)	Districts (C)	Schools (D)
<i>MSA-level measures of Tiebout choice (N=179 MSAs in SAT states)</i>				
District choice index	0.75	0.26	1.00	0.37
School choice index	0.96	0.05	0.37	1.00
<i>MSA covariates (N=179)</i>				
ln(Population)	14.17	1.18	0.10	0.79
Pop.: Fr. Black	0.12	0.09	-0.25	0.19
Pop.: Fr. Hispanic	0.14	0.15	-0.18	0.06
Mean log HH income	10.27	0.20	0.36	0.42
Gini, HH income	0.43	0.03	-0.45	0.11
Pop: Fr. BA+	0.23	0.06	0.21	0.36
Foundation Plan finance rule	65%		-0.03	-0.14
South	36%		-0.29	-0.21
SAT-taking rate	0.36	0.08	0.30	0.23
Private enrollment share (HS)	0.11	0.05	-0.19	0.33
Racial dissimilarity index, high schools	0.47	0.13	0.28	0.53
Racial isolation index, high schools	0.28	0.14	0.17	0.55
District choice index in 1942	0.91	0.23	0.73	0.35
Number of streams in MSA	283	214	0.41	0.38
<i>Individual SAT-taker sample (N=330,688)</i>				
SAT	995	201	0.29	-0.04
Black	12%		-0.32	0.20
Hispanic	12%		-0.23	0.04
Asian	10%		-0.04	0.28
Female	55%		-0.28	-0.07
Family income	\$46,389	\$25,987	0.43	0.09
Background index	995	82	0.40	-0.11
<i>SAT sample schools (N=5,779)</i>				
SAT mean	995	95	0.29	-0.04
Size (students per grade)	387	211	-0.44	0.15
Number of SAT-takers	179	116	-0.19	0.46
SAT-taking rate	0.49	0.19	0.31	0.39
Mean background index, SAT-takers	995	48	0.40	-0.11
Public	0.90		0.42	-0.24

Notes: SAT-taker statistics are calculated over students in 179 MSAs in SAT states (see text). MSA-level statistics are weighted by the MSA 17-year-old population. Individual SAT-taker data are weighted by inverse sampling probabilities, and school means by the sum of individual SAT-taker weights; both are adjusted at the MSA level to weight MSAs in proportion to their 17-year-old populations.

Table 2. Choice indices as predictors of private enrollment rates and of the racial segregation of schools

	Dependent Variable						
	MA Private Enrollment Rate			Measures of Racial Segregation			
				Dissimilarity		Isolation	
				All HS	Public HS	All HS	Public HS
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
District choice index	-0.029 (0.012)	-0.027 (0.013)	-0.029 (0.013)	0.105 (0.024)	0.109 (0.027)	0.075 (0.026)	0.083 (0.027)
School choice index		-0.012 (0.057)	-0.018 (0.057)	0.446 (0.105)	0.440 (0.115)	0.296 (0.111)	0.275 (0.115)
ln(Population) / 100	0.98 (0.35)	1.05 (0.49)	0.97 (0.52)	-1.07 (0.94)	-0.97 (1.03)	-0.66 (1.00)	-0.77 (1.03)
Pop.: Fr. Black	0.11 (0.04)	0.11 (0.04)	0.12 (0.05)	0.10 (0.09)	0.04 (0.10)	0.27 (0.09)	0.16 (0.10)
Pop.: Fr. Hispanic	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	0.06 (0.04)	0.07 (0.05)	0.08 (0.05)	0.06 (0.05)
Mean log HH income	0.10 (0.02)	0.10 (0.02)	0.10 (0.03)	0.01 (0.05)	-0.01 (0.05)	0.03 (0.05)	0.02 (0.05)
Gini, HH income	0.41 (0.15)	0.40 (0.15)	0.34 (0.16)	0.25 (0.29)	0.04 (0.31)	0.12 (0.30)	-0.08 (0.32)
Pop: Fr. BA+	-0.05 (0.05)	-0.06 (0.05)	-0.03 (0.06)	0.13 (0.12)	0.21 (0.13)	0.20 (0.12)	0.26 (0.13)
Foundation Plan finance rule	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.02 (0.01)	-0.02 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Tract-level segregation measures							
Dissimilarity Index			0.09 (0.06)	1.02 (0.11)	1.09 (0.12)	0.12 (0.11)	0.12 (0.12)
Isolation Index			-0.05 (0.06)	-0.24 (0.10)	-0.21 (0.11)	0.61 (0.11)	0.69 (0.11)
X-tract share of variance, log(HH income)			0.01 (0.08)	0.16 (0.16)	0.19 (0.17)	-0.08 (0.17)	-0.04 (0.17)
R ²	0.36	0.36	0.37	0.79	0.78	0.76	0.76

Notes: Observations are MSAs; N=320 (289 in Columns D-G, which exclude MSAs missing racial composition for schools with more than 25% of enrollment). All models include fixed effects for 8 census divisions.

Table 3. District choice index and other determinants of the within-MSA SAT score-peer group gradient

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Avg. student background index	1.74 (0.04)	1.74 (0.04)	1.74 (0.04)	1.74 (0.03)	1.70 (0.02)	1.69 (0.02)	1.69 (0.02)	1.69 (0.01)
<i>Interactions of avg. background index with MSA-level:</i>								
* District choice index		-0.01 (0.19)	-0.07 (0.20)	-0.27 (0.11)	-0.22 (0.11)	-0.14 (0.13)	-0.16 (0.14)	0.05 (0.13)
* MSA SAT-taking rate			0.70 (0.60)	1.63 (0.47)	1.56 (0.47)	1.18 (0.42)	1.26 (0.43)	0.92 (0.34)
* ln(Population)					0.09 (0.02)	0.03 (0.02)	0.02 (0.03)	0.03 (0.03)
* Pop.: Frac. Black						-0.19 (0.36)	-0.45 (0.36)	-2.91 (0.54)
* Pop.: Frac. Hispanic						0.06 (0.18)	-0.29 (0.24)	-1.88 (0.32)
* Mean log HH income						0.14 (0.21)	0.07 (0.19)	-0.05 (0.16)
* Gini, HH income						2.15 (1.30)	1.12 (1.29)	1.94 (1.26)
* Pop: Frac. BA+						1.51 (0.53)	2.25 (0.67)	1.71 (0.60)
* Foundation Plan finance rule						0.03 (0.06)	0.04 (0.06)	0.02 (0.05)
* ln(Pop. density)							0.03 (0.03)	0.00 (0.03)
* Pop: Frac. LTHS							1.41 (0.83)	0.72 (0.74)
* Pop: Frac. White ²								-1.61 (0.30)
* Census division effects	n	n	n	y	y	y	y	y
R ²	0.74	0.74	0.74	0.74	0.74	0.75	0.75	0.75

Notes: N=5,779 schools in 179 MSAs. Schools are weighted by the sum of individual sampling weights, with an MSA-level adjustment to weight MSAs by their 17-year-old populations. Dependent variable is the weighted mean SAT score at the school. All models include MSA fixed effects, and are estimated by "within" regressions with standard errors clustered at the MSA level (Kezdi, 2002). MSA-level explanatory variables are demeaned, then interacted with the school average background index.

**Table 4. Tiebout choice and the within-MA SAT score-peer group gradient:
Alternative specifications**

	Peer group main effect		Peer gp.-choice interaction		Number of schools
	Coeff.	S.E.	Coeff.	S.E.	
	(A)	(B)	(C)	(D)	(E)
1. Basic model	1.69	(0.02)	-0.14	(0.13)	5,779
2. Control for S.D. of peer group at school	1.69	(0.02)	-0.18	(0.12)	5,634
3. Control for school racial composition	2.09	(0.07)	-0.14	(0.10)	5,779
4. Student bkgd. is avg. income (\$1,000s)	5.44	(0.08)	0.06	(0.52)	5,742
5. Choice is school choice index	1.69	(0.02)	1.10	(0.65)	5,779
6. Sample re-weighted using class rank	1.68	(0.02)	-0.05	(0.15)	4,656
7. Sample excludes private schools	1.66	(0.02)	-0.05	(0.13)	4,498
8. Sample excludes 1-district MSAs	1.71	(0.02)	-0.31	(0.15)	5,460
9. IV: Streams as instrument	1.69	(0.02)	-0.30	(0.34)	5,779
10. IV: 1942 choice as instrument	1.68	(0.02)	0.19	(0.21)	5,711

Notes: "Basic model" in row 1 is that from Table 3, Column F. Other rows re-estimate this model with small changes to the specification, weighting, or sample; rows 9-10 instrument the student background-choice interaction with the interaction of student background with the listed instrument.

Table 5. Within-MSA sorting estimates from the National Education Longitudinal Survey (NELS:88)

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
<i>Panel A: Dependent variable is the school average composite test score in the 8th grade sample</i>								
Avg. student bkgd. index	1.66 (0.07)	1.67 (0.07)	1.67 (0.07)	1.71 (0.05)	1.78 (0.06)	1.80 (0.06)	1.80 (0.06)	1.81 (0.06)
* District choice index		-0.13 (0.20)	-0.13 (0.20)	-0.27 (0.19)	-0.39 (0.18)	-0.30 (0.34)	-0.19 (0.41)	-0.27 (0.43)
R ²	0.71	0.71	0.71	0.72	0.72	0.72	0.72	0.72
<i>Panel B: Dependent variable is the school continuation rate 4 yrs. after the 8th grade survey (multiplied by 100)</i>								
Avg. student bkgd. index	3.30 (0.38)	3.32 (0.35)	3.32 (0.35)	3.35 (0.23)	3.44 (0.30)	3.44 (0.30)	3.44 (0.30)	3.41 (0.30)
* District choice index		-1.04 (1.13)	-1.04 (1.13)	-1.75 (0.92)	-1.89 (0.97)	-3.02 (1.63)	-2.76 (1.85)	-2.29 (2.22)
R ²	0.28	0.28	0.28	0.34	0.34	0.34	0.34	0.34

Notes: N=748 schools in 211 MSAs (Panel A); 730 schools in 209 MSAs (Panel B). Specifications are identical to those in corresponding columns of Table 3. Schools are weighted by the sum of individual sampling weights (first wave weights in Panel A, and panel weights in Panel B), with an MSA-level adjustment to weight MSAs by their 17-year-old populations. The background index is the same in both panels, and weights NELS family background variables to best predict differences in 8th grade composite scores within schools.

Table 6. Across-MSA models for average SAT scores

	OLS						IV	
	(A)	(B)	(C)	(D)	(E)	(F)	Streams	1942 Choice
							(H)	(I)
District choice index	35.9 (9.0)	-10.9 (4.7)	-16.8 (4.5)	-14.3 (4.5)	-14.2 (4.5)	-20.3 (5.5)	-46.8 (16.8)	-9.8 (7.4)
MSA SAT-taking rate		2 (16)	-34 (17)	-77 (16)	17 (78)	-50 (82)	27 (90)	-56 (81)
MSA SAT-taking rate ²					-128 (104)	-20 (110)	-119 (119)	-12 (112)
Avg. student background		1.50 (0.06)	1.77 (0.06)	1.83 (0.13)	1.85 (0.13)	1.83 (0.13)	2.05 (0.18)	1.76 (0.13)
ln(Population)			5.86 (0.95)	1.61 (1.20)	1.36 (1.21)	1.12 (1.36)	4.87 (2.21)	0.45 (1.30)
Pop.: Fr. Black				40 (24)	42 (24)	122 (36)	77 (33)	30 (24)
Pop.: Fr. Hispanic				53 (15)	57 (15)	107 (23)	73 (19)	47 (15)
Mean log HH income				-5.9 (8.7)	-6.3 (8.7)	-2.1 (9.3)	-18.7 (11.7)	1.5 (8.6)
Gini, HH income				-97 (55)	-99 (55)	-127 (59)	-291 (113)	-44 (60)
Pop: Fr. BA+				150 (29)	147 (29)	157 (35)	128 (35)	142 (29)
Foundation Plan finance rule				-3.4 (2.8)	-3.9 (2.8)	-2.6 (2.8)	1.3 (4.1)	-6.0 (2.8)
ln(Pop. density)						2.1 (1.6)		
Pop: Fr. LTHS						19 (35)		
Pop: Fr. White ²						53 (18)		
Census divis. effects	n	n	y	y	y	y	y	y
R ²	0.08	0.80	0.86	0.91	0.91	0.92		

Notes: N=179 (177 in col. I) MSAs in SAT states. Dependent variable is the average SAT score across all sample students attending schools in the MSA.