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

Jesse M. Rothstein[*]
jrothst@econ.berkeley.edu

Department of Economics
University of California, Berkeley

JOB MARKET PAPER
November 20, 2002

Abstract

School choice policies aim to improve school productivity by rewarding administrators of schools that parents choose. These policies may not create incentives for effective administration, however, if parents prefer schools with desirable peer groups to those with inferior peers but better policies and instruction. I develop a model of the "Tiebout choice" residential housing market in which schools differ in both peer group and effectiveness. If parental preferences depend primarily on school effectiveness, we should expect both that wealthy parents disproportionately purchase houses near effective schools and that decentralization of educational governance—competition among local school districts—facilitates this residential sorting. On the other hand, if peer group dominates effectiveness in parental preferences, wealthy families will still cluster together in equilibrium but these clusters need not be near effective schools. I use a large sample of SAT-takers to examine the distribution of student outcomes across schools within metropolitan housing markets that differ in the structure of educational governance. I find little evidence that parents choose schools for characteristics other than peer groups. Moreover, average SAT scores are no higher in markets with many districts than in those with few. These results suggest caution about the potential to induce improvements in educational productivity through expansions of parental choice.

[*] I am indebted to my thesis advisor, David Card. I also thank Alan Auerbach, David Autor, Ken Chay, Tom Davidoff, Jonah Gelbach, David Lee, Justin McCrary, Rob McMillan, John Quigley, Joan Reller, Emmanuel Saez, Till von Wachter, Diane Whitmore, Eric Verhoogen, and participants in the UC Berkeley Public Finance seminar for helpful comments. Research 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 National Science Foundation, the Fisher Center, or any of my advisors.

1. Introduction

Many analysts have identified principal-agent problems as a major source of underperformance in public education. Public school administrators need not compete for customers and are therefore free of the market discipline that aligns producer incentives with consumer demand in private markets. Chubb and Moe, for example, argue that the interests of parents and students "tend to be far outweighed by teachers' unions, professional organizations, and other entrenched interests that, in practice, have traditionally dominated the politics of education," (1990, p. 31).[1] One proposed solution—advocated by Friedman (1962) and others—is to allow dissatisfied parents to choose another school, and to link school administrators' compensation to parents' revealed demand. This would strengthen parents relative to other actors, and might "encourage competition among schools, forcing them into higher productivity," (Hoxby, 1994, p. 1).

The mechanisms proposed to increase parental exit options—vouchers, charter schools, etc.—are not at present sufficiently widespread to permit decisive empirical tests of their effects on school productivity.[2] Nevertheless, economists have long argued that housing markets represent a long established, potentially informative form of school choice (Tiebout, 1956; Brennan and Buchanan, 1980; Oates, 1985; Hoxby, 2000a). Parents exert some control over their children's school assignment via their residential location decisions, and can exit undesirable schools by moving to a neighborhood served by a different school district. As U.S. metropolitan areas vary dramatically in the amount of control over children's school assignment that the residential decision affords to parents, one can hope to infer the effect of so-called *Tiebout choice* by comparing student outcomes across metropolitan housing markets (Borland and Howsen, 1992; Hoxby, 2000a).[3]

---

[1] Chubb and Moe also identify the school characteristics that parents would presumably choose, given more influence: "strong leadership, clear and ambitious goals, strong academic programs, teacher professionalism, shared influence, and staff harmony," (p. 187). See also Hanushek (1986) and Hanushek and Raymond (2001).

[2] Hsieh and Urquiola (2002) study a large-scale voucher program in Chile, but argue that effects on school productivity cannot be distinguished from the allocative efficiency effects of student stratification.

[3] Hoxby argues that this sort of analysis can "demonstrate general properties of school choice that are helpful for thinking about reforms," (2000a, p. 1209). Belfield and Levin (2001) review other, similar studies.

The potential effects of both Tiebout choice and less restricted choice systems like vouchers depend critically on what characteristics parents value in schools. Hanushek, for example, notes that parents might not choose effective schools over others that are less effective but offer "pleasant surroundings, athletic facilities, [and] cultural advantages," (1981, p. 34). This seems especially likely under Tiebout choice, where school and neighborhood characteristics are chosen jointly.

To the extent that parents choose productive schools, market discipline can induce greater productivity from school administrators and teachers. If parents primarily value other features, however, market discipline may be less successful. 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. 34-35; emphasis added).

In this paper, I use data on school assignments and outcomes of students across schools within different metropolitan housing markets to assess parents' revealed preferences. I focus in particular on the role of the peer group in parental choices: If students' outcomes depend importantly on the characteristics of their classmates (i.e. if so-called "peer effects" are important components of educational production), even rational, fully informed, test-score-maximizing parents may prefer schools with poor management but desirable peer groups to better managed competitors that enroll less desirable students, and administrators may be more reliably rewarded for enrolling the right peer group than for offering effective instruction. To preview the results, I find little evidence that parents use Tiebout choice to select effective schools over those with desirable peers, or that schools are on average more effective in markets that offer more choice.

Existing research does not conclusively establish the causal contribution of peer group characteristics to student outcomes (see, e.g., Coleman et al., 1966; Hanushek, Kain, and Rivkin, 2001; Katz, Kling, and Liebman, 2001). Anecdotal evidence suggests, however, that parents may substantially overvalue a good peer group relative to its importance in educational production. Realtor.com, a web site for house hunters, offers reports on several neighborhood characteristics that parents apparently value. These include a few variables that may be interpreted as measures of school resources or effectiveness (e.g. class size and the number of computers); detailed socioeconomic data (e.g. educational attainment and income); and the average SAT score at the local high school. Given similar average scores, test-score maximizers should prefer demographically unfavorable schools, as these must add more value to attain the same outcomes as their competitors with more advantaged students.[4] While it is possible that parents use the demographic data in this way, it seems more likely that home buyers prefer wealthier neighborhoods, even conditional on average student performance (Downes and Zabel, 1997).[5]

In modeling the effects of parental preferences on equilibrium outcomes under Tiebout choice, it is important to account for two key issues that do not arise under choice programs like vouchers. The first is that residential choice rations access to highly-demanded schools by willingness-to-pay for local housing.[6] As a result, both schools and districts in high-choice markets (those with many competing school districts) are more stratified than in low-choice markets. Increased stratification can have allocative efficiency consequences that confound estimates of the effect of choice on productive efficiency.

A second issue is that there is little or no threat of market entry when competition is among geographically-based school districts. In the absence of entry, administrators of undesirable districts are not likely to face substantial declines in enrollment. Indeed, a reasonable first approximation is that total (public) school and district enrollments are invariant to schools' relative desirability.[7] Instead, Tiebout choice works by rewarding the administrator of a preferred school with a better

[4] This does not rely on assumptions about the peer effect: The effect of individual characteristics on own test scores, distinct from any spillover effects, is not attributable to the school, and test-score-maximizing parents should penalize the average test scores of schools with advantaged students to remove this effect (Kain, Staiger, and Samms, 2002).

[5] Postsecondary education offers additional evidence of strong preferences over the peer group: Colleges frequently trumpet the SAT scores of their incoming students—the peer group—while data on graduates' achievements relative to others with similar initial qualifications, which would arguably be more informative about the college's contribution, are essentially non-existent. Along these lines, Tracy and Waldfogel (1997) find that popular press rankings of business schools reflect the quality of incoming students more than the schools' contributions to students' eventual salaries (but see also Dale and Krueger, 1999, who obtain somewhat conflicting results at the undergraduate level).

[6] Small-scale voucher programs may not have to ration desired schools. One imagines that broader programs will use some form of price system, perhaps by allowing parents to "top up" their vouchers (Epple and Romano, 1998).

[7] Poor school management can, of course, lead parents to choose private schools, lowering public enrollment. Similarly, areas with bad schools may disproportionately attract childless families. These are likely second-order effects. The private option, in any case, is not the mechanism by which residential choice works but an alternative to it: Inter-jurisdictional competition has been found to lower private enrollment rates (Urquiola, 1999; Hoxby, 2000a).

student body and with wealthier and more motivated parents. There are obvious benefits for educational personnel in attracting an advantaged population, and I assume throughout this paper that the promise of such rewards can create meaningful incentives for school administrators.

With several school characteristics over which parents choose, understanding which schools are chosen and which administrators are rewarded requires a model of residential choice. Building on the framework of so-called multicommunity models in the local public finance literature (Ross and Yinger, 1999), I introduce a component of school desirability that is exogenous to parental decisions, "effectiveness," which is thought of as the portion of schools' effects on student performance that does not depend on the characteristics of enrolled students. Parental preferences among districts depend on both peer group and effectiveness, and I consider the implications of varying the relative weights of these characteristics for the rewards that accrue in equilibrium to administrators of effective schools.

Hoxby (1999b) also models Tiebout choice of schools, but she assumes a discrete distribution of student types and allows parents to choose only among schools offering identical peer groups. I allow a continuous distribution of student characteristics, which forces parents to trade off peer group against effectiveness in their school choices. This seems a more accurate characterization of Tiebout markets, as the median U.S. metropolitan area has fewer than a dozen school districts from which to choose.

As in other multicommunity models, equilibrium in my model exhibits complete stratification: High-income families live in districts that are preferred to (and have higher housing prices than) those where low-income families live. That this must hold regardless of what parents value points to a fundamental identification problem in housing price-based estimates of parental valuations: Peer group and, by extension, average student performance are endogenous to unobserved determinants of housing prices.[8] One estimation strategy that accommodates this endogeneity is that taken by Bayer, McMillan, and Reuben (2002), who estimate simultaneous

[8] Shepard (1999) reviews hedonic studies of housing markets

equations for housing prices and community composition in San Francisco. I adopt a different strategy: I compare housing markets that differ in the strength of the residential location-school assignment link. This across-market approach has the advantage that it does not rely on strong exclusion restrictions; my primary assumption is that the causal effect of individual and peer characteristics on student outcomes does not vary systematically with the structure of educational governance.

Like Baker, McMillan, and Reuben (2002), I identify parental valuations by the *location* of clusters of high income families: If parental preferences over communities depend exclusively on the effectiveness of the local schools, the most desirable communities are necessarily those with the most effective schools. If peer group matters at all to parents, however, there can be "unsorted" equilibria in which communities with ineffective schools have the wealthiest residents and are the most preferred. These equilibria result from coordination failures: The wealthy families in ineffective districts would collectively have the highest bids for houses assigned to more effective schools, but no individual family is willing to move alone to a district with undesirable peers.

The more importance that parents attach to school effectiveness, the more likely we are to observe equilibria in which wealthy students attend more effective schools than do lower-income students. Moreover, if parental concern for peer group is not too large, the model predicts that this equilibrium *effectiveness sorting* will tend to be more complete in "high-choice" markets, those with many small school districts, than in markets with more centralized governance. This is because higher choice markets divide the income distribution into smaller bins, which reduces the "cost" (in peer quality) that families pay for moving to the next lower peer group district and thus reduces the probability that wealthy families will be trapped in districts with ineffective schools.

Effectiveness sorting should be observable as a magnification of the causal peer effect, as it creates a positive correlation between the peer group and an omitted variable—school effectiveness—in regression models for student outcomes.[9] This provides my identification: I look for evidence that the apparent peer effect, the reduced-form gradient of school average test scores with respect to student characteristics, is larger in high-choice than in low-choice markets. If parents select schools for effectiveness, wealthy parents should be better able to obtain effective schools in markets where decentralized governance facilitates the choice of schools through residential location, and student performance should be more tightly associated with peer characteristics in these markets. If parents instead select schools primarily for the peer group, there is no expectation that wealthy students will attend effective schools in equilibrium, regardless of market structure, and the peer group-student performance relationship should not vary systematically with Tiebout choice.

I use a unique data set consisting of observations on more than 300,000 metropolitan SAT takers from the 1994 cohort, matched to the high schools that students attended. The size of this sample permits accurate estimation of both peer quality and average performance for the great majority of high schools in each of 177 metropolitan housing markets. I find no evidence that the association between peer group and student performance is stronger in high-choice than in low-choice markets. This result is robust to nonlinearity in the causal effects of the peer group as well as to several specifications of the educational production function. Moreover, although there is no other suitable data set with nearly the coverage of the SAT sample, the basic conclusions are supported by models estimated both on administrative data measuring high school completion rates and on the National Educational Longitudinal Survey (NELS) sample.

This result calls the incentive effects of Tiebout choice into question, as it indicates that administrators of effective schools are no more likely to be rewarded with wealthy students in high-choice than in low-choice markets. To explore this further, I estimate models for the effect of Tiebout choice on mean scores across metropolitan areas. Consistent with the earlier results, I find

[9] Willms and Echols (1992, 1993) are the first authors of whom I am aware to note the distinction between preferences for peer group and for effective schools. They use hierarchical linear modeling techniques (Raudenbush and Willms, 1995; Raudenbush and Bryk, 2002), and estimate school effectiveness as the residual from a regression of total school effects on peer group. This is appropriate if there is no effectiveness sorting; otherwise, it may understate the importance of effectiveness in output and in parental choices.

no evidence that high-choice markets produce higher average SAT scores. Together with the within-market estimates, this calls into question Hoxby's (1999a, 2000a) conclusion that Tiebout choice induces higher productivity from school administrators.[10]

There are two plausible explanations for the pattern of findings 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 effectiveness sorting, regardless of its extent, in the distribution of student SAT scores. A second explanation is that effectiveness does matter for student performance, but that it does not matter greatly to parental residential choices.[11] This could be because effectiveness is swamped by the peer group in parental preferences or because it is difficult to observe directly. In either case, administrators who pursue unproductive policies are unlikely to be disciplined by parental exit and Tiebout choice can create only weak incentives for productive school management.

2. Tiebout Sorting and the Role of Peer Groups: Intuition

To fix ideas, in this section I describe the Tiebout choice process and its observable implications in the context of a very simple educational technology with peer effects. Let

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

be a reduced-form representation of the 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 among students at school j; and

---

[10] Hoxby argues that market structure is endogenous to school quality. Instrumenting for it and using relatively sparse data from the NELS and the National Longitudinal Survey of Youth, she finds a positive effect of choice on mean scores across markets. I discuss the endogeneity issue in Appendix B, and consider several instrumentation strategies. As none indicate substantial bias in OLS results, the main discussion here treats market structure as exogenous.

[11] In fact, the main empirical approach cannot well distinguish between the case where parents value effectiveness to the exclusion of all else and that where they ignore effectiveness in their residential choices, as in either case effectiveness sorting may not depend on the market structure. The former hypothesis seems implausible on prior grounds, and is further discredited by the evidence on mean scores across markets.

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.[12]

In view of the vast literature documenting the important role of family background characteristics—e.g. parental income and education—in student achievement (Coleman et al., 1966; Phillips et al., 1998; Bowen and Bok, 1998), I assume for now that x_{ij} is positively correlated with willingness-to-pay for educational quality, though this is not essential. In the empirical analysis below, I also estimate specifications that allow willingness-to-pay to depend on family income while other characteristics have direct effects on student achievement.

Since model (1) excludes school resources, the term $\bar{x}_j\gamma$ potentially captures both conventional peer group effects and other indirect effects associated with the family background characteristics of students at school j. For example, wealthy parents may be more likely to volunteer in their children's schools, or to vote for increased tax rates to support education. They may also be more effective at exerting "voice" to manage agent behavior, even without the exit option that school choice policies provide (Hirschman, 1970). Finally, student composition may operate as an employment amenity for teachers and administrators, reducing the salaries that the school must pay and increasing the quality of teachers that can be hired for any fixed salary (Antos and Rosen, 1975).[13]

The effectiveness parameter in (1), μ_j, encompasses the effects of any differences across schools that do not depend on the characteristics of students that they enroll. It may include, for example, the ability and effort levels of local administrators, or their effectiveness in resisting the

[12] In the empirical application in Section 5, I allow for more general technologies in which the effects of individual or peer characteristics are arbitrarily nonlinear or higher moments of the peer group distribution enter the production function. The key assumption is that all families agree on the relative importance of peer group and school effectiveness. This assumption of similar preference structures is common in studies of consumer demand, and in particular underlies both the multicommunity and hedonic literatures.

[13] The distinction between direct and indirect effects of school composition is not always clear in discussions of peer effects. Peer effects estimates that use transitory within-school variation in the composition of the peer group (Hoxby, 2000b; Hanushek, Kain, and Rivkin, 2001) likely estimate only the direct peer effect, while those that use the assignment of students to schools (Evans, Oates, and Schwab, 1992; Katz, Kling, and Liebman, 2001) likely estimate something closer to the full reduced-form effect of school composition.

demands of bureaucrats and teacher's unions.[14] It is worth noting that the relative magnitude of μ_j may be quite modest. Family background variables typically explain the vast majority of the differences in average student test scores across schools, potentially leaving relatively little room for efficiency (or school "value added") effects.[15] Nevertheless, most observers believe that public school efficiency is important, that it exerts a non-trivial role on the educational outcomes of students, and that it varies substantially across schools.

The potential efficiency-enhancing effects of increased Tiebout choice operate through the assumption that parents prefer schools with μ_j-promoting policies. To the extent that this is true, Tiebout choice induces a positive correlation between μ_j and $\bar{x}_j$, since high-x_i families will outbid lower-x_i families for the most preferred districts. Thus, active Tiebout choice can *magnify* the apparent impact of peer groups on student outcomes in analyses that neglect administrative quality. Formally,

$$E\left[\bar{t}_j \mid \bar{x}_j\right] = \bar{x}_j(\beta + \gamma) + E\left[\mu_j \mid \bar{x}_j\right], \tag{2}$$

or, simplifying to a linear projection,

$$E^*\left[\bar{t}_j \mid \bar{x}_j\right] = \bar{x}_j\left(\beta + \gamma + \theta^*\right), \tag{3}$$

where $\theta^* \equiv \text{cov}\left(\bar{x}_j, \mu_j\right) / \text{var}\left(\bar{x}_j\right)$ represents the degree of effectiveness sorting in the local market. (For notational simplicity, I neglect the intercept in both test scores and school effectiveness.) The stronger are parental preferences for effective schools (relative to schools with other desired attributes), the more actively will high-x_i families seek out neighborhoods in effective districts, and

[14] More precisely, ability and effort of school personnel is included in μ only to the extent that a good peer group does not enable a school to bid the best employees away from low-$\bar{x}$ schools. A wealthy, involved population may not ensure high-quality, high-effort staff if agency problems produce district hiring policies that do not reflect parents' preferences (Chubb and Moe, 1990), or if it is difficult to enforce contracts over unobservable components of administrator actions (Hoxby, 1999b).

[15] In the SAT data used here, a regression of school mean test scores on average student characteristics has an R^2 of 0.74. The correlation is substantially stronger in California data used to evaluate school performance (Technical Design Group, 2000). The argument below suggests, however, that these raw correlations may seriously overstate the causal importance of peer group if effectiveness sorting is substantial.

the larger will θ^* tend to be in Tiebout equilibrium. The weaker are parental preferences for μ_j relative to other factors, the smaller will θ^* tend to be.

Importantly, one would expect the degree of local competition in public schooling (i.e. the number of school districts in the local area among which parents can choose) to affect the magnitude of θ^* whenever parents care both about peer groups and school effectiveness. The reasoning is simple: If there are only a small number of local districts and parents value the peer group, they may be "stuck" with a high-$\bar{x}$ /low-μ school, even in housing market equilibrium, by their unwillingness to sacrifice peer group in a move to a more effective school district. These coordination failures are less likely in markets with more interjurisdictional competition, as in these markets there are always alternative districts that are relatively similar in the peer group offered, and parents are able to select effective schools without paying a steep price in reduced peer quality.

When parental concern for peer group is moderate, then, a high degree of public school choice is needed to ensure that high-μ schools attract high-x families, and θ^* tends to be larger in high-choice than in low-choice markets. On the other hand, 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 local competition. Similarly, when parental concern for peer group is large enough, even in highly competitive markets high-x families are not drawn to high-μ schools, and again θ^* is largely independent of market structure.

This idea forms the basis of my empirical strategy. In essence, I compare the sorting parameter θ^* in equation (3) across metropolitan housing markets with greater and lesser degrees of residential school choice. Let $\theta = \theta(c, \delta) = E[\theta^* \mid c, \delta]$ be the average effectiveness sorting of markets characterized by the parameters c and δ, where c is the degree of jurisdictional competition (i.e. the number of competing districts from which parents can choose, adjusted for their relative sizes) and δ is the importance that parents place on peer group relative to effectiveness. The argument above, supported by the theoretical model developed in the next section, predicts that

$\partial\theta/\partial c > 0$ for moderate values of δ but that $\partial\theta/\partial c = 0$ when $\delta = 0$ or when δ is large (i.e. when parents care only about effectiveness or only about peer group). To the extent that θ^* tends to increase with choice, then, we can infer that parents' peer group preferences are small enough to prevent a breakdown in high-choice markets of the sorting mechanism that rewards high-μ administrators with high-x students. On the other hand, if θ^* is no larger in high-choice than in low-choice cities it is more difficult to draw inferences about parental valuations, which may be characterized either by very small or very large δ. In either of case, however, we can expect little effect of expansions of Tiebout choice on school efficiency, as in the former even markets with only a few districts can provide market discipline and in the latter no amount of governmental fragmentation will create efficiency-enhancing incentives for school administrators.

3. A Model of Tiebout Sorting on Exogenous Community Attributes

In this section, I build a formal model of the Tiebout sorting process described above. As my interest is in the demand side of the market under full information, I treat the distribution of school effectiveness as exogenous and known to all market participants.[16] I demonstrate that Tiebout equilibrium must be stratified as much as the market structure allows: Wealthy families always attend schools that are preferred to those attended by low-income families. There can be multiple equilibria, however, and the allocation of effective schools is not uniquely determined by the model's parameters. Conventional comparative statics analysis is not meaningful when equilibrium is non-unique, as the parental valuation parameter affects the set of possible equilibria rather than altering a particular equilibrium. To better understand the relationships between parental valuations, market concentration, and the equilibrium allocation, the formal exposition of the model is followed by simulations of markets under illustrative parameter values.

[16] In other words, parents are "effectiveness-takers." This does not rule out administrative responses to the incentives created by parental choices, as these are a higher order phenomenon, deriving from competition among schools to attract students rather than from reactions of school administrators to the realized desirability of their schools.

My model is a much simplified version of so-called "multicommunity" models. I maintain the usual assumptions that the number of communities is fixed and finite, and that access to desirable communities is rationed through the real estate market.[17] Although some authors (i.e. Epple and Zelenitz, 1981) include a supply side of the housing market, I assume that communities are endowed with perfectly inelastic stocks of identical houses.[18] Communities differ in three dimensions: The average income of their residents and the rental price of housing, both endogenous, and the effectiveness of the local schools.[19]

Several characteristics of real-world Tiebout choice are neglected. First, I ignore private schools. A private sector complicates the sorting process substantially, as one must allow parental preferences to distinguish between the characteristics of neighbors and those of children's classmates. Relatedly, I 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. It is not obvious how to model such families' preferences: One might expect them to segregate into child-free communities without schools. While this sort of segregation is observed occasionally (Black, Gates, and Sanders, 2002; Barrow, 2002), it is rarely as stark as simple models predict.

A final omission is of all non-school exogenous amenities like beaches, parks, views, and air quality. I develop here a "best case" for Tiebout choice, where schools are the only factors in neighborhood desirability. Amenities could either increase or reduce the extent of effectiveness

[17] Where most models incorporate within-community voting processes for public good provision (Fernandez and Rogerson, 1996; Epple and Romano 1996; Epple, Filimon and Romer, 1993), income redistribution (Epple and Romer, 1991; Epple and Platt, 1998), or zoning rules (Fernandez and Rogerson, 1997; Hamilton, 1975), I simply allow for preferences over the mean income of one's neighbors. These preferences might derive either from the effects of community composition on voting outcomes or from reduced-form peer effects in education.

[18] Tiebout equilibria must evolve quickly to provide discipline to school administrators, whose careers are much shorter than the lifespan of houses. Inelastic supply is probably realistic in the short term, except possibly at the urban fringe. Nechyba (1997) points out that it is much easier to establish existence of equilibrium with fixed supply.

[19] The inclusion of even one exogenous component of community desirability is not standard in the multicommunity literature, which, beginning with Tiebout's (1956) seminal paper, has typically treated communities as *ex ante* interchangeable. This assumption leaves no room for managerial effort or quality, so is inappropriate for analyses of the incentives that mobility creates for public-sector administrators.

sorting relative to this pure case, though the latter seems more likely.[20] If, as the hedonics literature implies, schools are one of the more important determinants of neighborhood desirability (see, e.g., Reback, 2001; Bogart and Cromwell, 2000; Figlio and Lucas, 2000; Black, 1999), the existence of relatively unimportant amenities should not much alter the trends identified here.

Turning to the formal exposition, assume that a local housing market—a metropolitan area—contains a finite number of jurisdictions, *J*, and a population of *N* families, $N >> J$. Each jurisdiction, indexed by *j*, contains *n* identical houses and is endowed with an exogenous effectiveness parameter, μ_j. No two jurisdictions have identical effectiveness.

Each family must rent a house. I assume that there are enough houses to go around but not so many that there can be empty communities: $n(J-1) < N < nJ$.[21] All homes are owned by absentee landlords, perhaps a previous generation of parents, who have no current use for them. These owners will rent for any nonnegative price, although they will charge positive prices if the market will support them. There is no possibility for collusion among landlords. Housing supply in each community is thus perfectly inelastic: In quantity-price space, it is a vertical line extending upward from $(n, 0)$.

Family i's exogenous income is $x_i > 0$; the income distribution has density function *f*, which is bounded and has connected support, and distribution function *F*.[22] Families choose the district that offers them the highest utility, taking community composition and housing prices as given. Let $\bar{x}_j$ denote the mean income of families in community *j*, and let h_j be the rental price of local housing. The utility that family *i* would obtain in jurisdiction *j* is $U_{ij} = U(x_i - h_j, \bar{x}_j\delta + \mu_j)$,

[20] Amenities might draw wealthy families to low-peer-group districts, improving those districts' peer groups and reducing the costs borne by other families living there. This could increase effectiveness sorting, although the effect would be weakened if there were a private school sector. Offsetting this, amenities might also prevent families from exiting localities with ineffective schools, reducing effectiveness sorting just as does concern for peer group.

[21] The model is a "musical chairs" game, and the upper constraint serves to tie prices down, while the lower constraint avoids the need to define the mean income of a community with no residents.

[22] Of course, the income distribution cannot truly be continuous for finite *N*. The discussion nevertheless treats it as such, and can be seen as an approximation for large *N*. Relaxing the treatment to allow a discrete distribution would add notational complexity and introduce some indeterminacy in equilibrium housing prices, but would not change the basic sorting results.

where U is twice differentiable everywhere with U_1 and U_2 both positive.[23] I make the usual assumption about the utility function:

Single Crossing Property: $U_{12}U_1 - U_{11}U_2 > 0$ everywhere.

Single crossing ensures that if any family prefers one school quality-price combination to another with lower quality—where quality is $q_j \equiv \bar{x}_j\delta + \mu_j$—all higher-income families do as well; if any family prefers a district to another offering higher quality education, all lower-income families do also. (This is proved in Appendix D.) Put slightly differently, it means that 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 district than will the lower-income family. As in other multicommunity models, the single crossing assumption drives the stratification results outlined below.

Market equilibrium is defined as a set of housing prices and a rule assigning families to districts on the basis of their income that is consistent with individual family preferences, taking all other families' decisions as fixed:

Definition: An equilibrium for a market defined by $\delta; J; \{\mu_1, \ldots, \mu_J\}$; and f consists of a set of nonnegative housing prices $\{h_1, \ldots, h_J\}$ and an allocation rule $G : \text{support}(f) \mapsto \mathbf{Z}_J$ that satisfy the following conditions (where $\bar{x}_j \equiv \int \mathbf{1}(G(x) = j) x dF(x) \Big/ \int \mathbf{1}(G(x) = j) dF(x)$):

EQ1 *No district is over-full.* For each j, $\int \mathbf{1}(G(x) = j) dF(x) \leq n/N$.

EQ2 *Nash equilibrium.* At the specified prices and with the current distribution of peer groups, no family would prefer a district other than the one to which it is assigned: $U(x_i - h_{G(x_i)}, \bar{x}_{G(x_i)}\delta + \mu_{G(x_i)}) \geq U(x_i - h_k, \bar{x}_k\delta + \mu_k)$ for all i and all k.

EQ3 *Normalization of housing prices.* $h_j = 0$ whenever $\int \mathbf{1}(G(x) = j) dF(x) < n/N$

EQ4 *No ties in realized quality.* For any j, k in the range of G (i.e. for any non-empty communities), $\bar{x}_j\delta + \mu_j \neq \bar{x}_k\delta + \mu_k$.[24]

---

[23] I might allow $U_{ij} = U(x_i - h_j, Q(\bar{x}_j, \mu_j))$, with $Q_1 \geq 0$ and $Q_2 > 0$, without changing the basic results; δ then corresponds to Q_1 / Q_2. The key assumption is that all families share the same U and Q functions, with all differences in their behavior resulting from differences in their budget constraints (i.e. from x_i).

[24] Condition EQ4 corresponds to the "stability" notion of Fernandez and Rogerson (1996; 1997). Arrangements that satisfy EQ1 through EQ3 but not EQ4 are unstable, and perturbations in one of the tied communities' effectiveness

An allocation rule that is an equilibrium with some housing price vector is called an *equilibrium allocation*. The following results are proved in Appendix D:

Theorem 1. Equilibrium exists.
Theorem 2. Any equilibrium is perfectly stratified, in the sense that no family lives in a higher-quality, higher-price, or higher-peer-group district than does any wealthier family. The following statements are thus equivalent in equilibrium:

i. $\bar{x}_j\delta + \mu_j > \bar{x}_k\delta + \mu_k$;

ii. $h_j > h_k$;

iii. $\bar{x}_j > \bar{x}_k$;

iv. $\inf\{x \mid G(x) = j\} \geq \sup\{x \mid G(x) = k\}$.

Theorem 2 suggests that equilibrium divides the income distribution into bins of size *n*, and that the only differences among equilibria concern the permutation of these bins across the available districts. Corollaries 2.1 and 2.2 formalize this notion and draw out an important implication:

Corollary 2.1. In any equilibrium, the *n* families with incomes greater than $F^{-1}\left(1-\frac{n}{N}\right)$ live in the same community, which has higher quality $(\bar{x}\delta + \mu)$ than any other. The next *n* families, with incomes in $\left(F^{-1}\left(1-\frac{2n}{N}\right), F^{-1}\left(1-\frac{n}{N}\right)\right)$, live in the community ranked second in quality. This continues down the distribution: For each $j \leq J$, the families with incomes in $\left(F^{-1}\left(\max\left\{1-\frac{jn}{N}, 0\right\}\right), F^{-1}\left(1-\frac{(j-1)n}{N}\right)\right)$ live in the community with the j^{th} ranked schools.[25]
Corollary 2.2. If $\delta = 0$, equilibrium is unique.

Before turning to the next result, it is helpful to introduce some additional notation. For any allocation rule *G*, let $r(j)$ be the index number of the *j*-th ranked district, where ranking is by $\bar{x}\delta + \mu$. That is, $r(\cdot)$ is the permutation that satisfies $\bar{x}_{r(1)}\delta + \mu_{r(1)} \geq \bar{x}_{r(2)}\delta + \mu_{r(2)} \geq \ldots \geq$

or peer group would lead to non-negligible differences between the communities as families adjust. With EQ4, equilibria are locally stable.

[25] I neglect families precisely at the boundary between income bins (i.e. those with incomes satisfying $F(x) = 1 - \frac{jn}{N}$ for some *j*). I demonstrate in the Appendix that families at boundary points are indifferent between the two communities, and the must be apportioned to conform to communities' size constraints. As the income distribution approaches continuity, the potential importance of boundary families declines to zero.

$\geq \bar{x}_{r(J)}\delta + \mu_{r(J)}$.[26] Also, let $\breve{x}_k$ be the lower bound of the k^{th} $\left(\frac{n}{N}\right)$ income bin:

$\breve{x}_k \equiv F^{-1}\left(1 - \frac{kn}{N}\right)$, $k = 1, \ldots, J-1$.

Theorem 2 and its corollaries tell us that any equilibrium must permute the income distribution in a specific way, with all families in the same $\left(\frac{n}{N}\right)$ income bin assigned to the same district. In the new notation, the j^{th} ranked district will have index number $r(j)$; for $1 < j < J$, its residents will be the families with incomes in $\left(\breve{x}_j, \breve{x}_{j-1}\right)$. The converse also holds: Any allocation rule that permutes the income bins in this way and that assigns wealthier bins to higher-ranked districts is an equilibrium allocation.

Theorem 3. Let *G* be an assignment rule satisfying:

i. *G* assigns x_1 and x_2 to the same district whenever x_1 and x_2 are in the same $\left(\frac{n}{N}\right)$ income bin, as defined above, and

ii. *G* assigns x_1 to a more preferred district than that where x_2 is assigned whenever $x_1 > x_2$ and the two are in different $\left(\frac{n}{N}\right)$ income bins.[27]

Then there exist housing prices with which *G* is an equilibrium.

The proof of Theorem 1 constructs an equilibrium that always exists, in which communities are perfectly sorted by district effectiveness: Higher quality/housing price/income communities also have more effective schools. By Corollary 2.2, this is the only equilibrium when $\delta = 0$. When $\delta > 0$, however, Theorem 3 provides conditions under which there are also equilibria that are unsorted on μ. That is, any assignment rule over $\left(\frac{n}{N}\right)$-bins that satisfies $\bar{x}_j\delta + \mu_j > \bar{x}_k\delta + \mu_k \Leftrightarrow \bar{x}_j > \bar{x}_k$ is an equilibrium; it is not required that $\mu_j > \mu_k$.

I refer to an assignment rule that satisfies the requirements of Theorem 3 but that assigns some families to lower-μ districts than those to which some higher-x families are assigned as an

[26] In general, $r(\cdot)$ and $\mu_{r(\cdot)}$ depend on the assignment rule *G*. If the market is in equilibrium, however, Corollary 2.1 indicates that $\bar{x}_{r(\cdot)}$ does not: $\bar{x}_{r(1)}$ is the average income of the richest *n* families; $\bar{x}_{r(2)}$ that of the next richest *n* families; and so on.

[27] Formally, these conditions are:

i. $G(x_1) = G(x_2)$ whenever $\text{int}\left\{\left(1 - F(x_1)\right)\frac{N}{n}\right\} = \text{int}\left\{\left(1 - F(x_2)\right)\frac{N}{n}\right\}$, and

ii. $\bar{x}_{G(x_1)}\delta + \mu_{G(x_1)} > \bar{x}_{G(x_2)}\delta + \mu_{G(x_2)}$ whenever $\text{int}\left\{\left(1 - F(x_1)\right)\frac{N}{n}\right\} < \text{int}\left\{\left(1 - F(x_2)\right)\frac{N}{n}\right\}$.

unsorted (or *imperfectly sorted*) *equilibrium*. Unsorted equilibria arise when the peer group advantage of high-income communities over low-income communities is large enough to overcome deficits in school effectiveness.[28] For fixed income and effectiveness distributions, unsorted equilibria become harder to maintain as the weight that families place on peer group relative to school quality falls:

Corollary 3.1. Let G be an assignment rule which satisfies assumption (i) of Theorem 3 and under which there exist communities j and k satisfying $\mu_j < \mu_k$ but $\bar{x}_j > \bar{x}_k$. Then for

$$C \equiv \max_{\bar{x}_k < \bar{x}_j} \frac{\mu_k - \mu_j}{\bar{x}_j - \bar{x}_k} > 0,$$

i. Whenever $\delta > C$, G is an equilibrium assignment (i.e. there exist housing prices with which G is an equilibrium).
ii. Whenever $\delta < C$, G is not an equilibrium assignment.
iii. If $\delta = C$, G can satisfy requirements EQ1-EQ3 for equilibrium, but violates EQ4.

I do not present formal results on the implications of increases in J for effectiveness sorting, as much depends on the μ_j's assigned to the new districts. Informally, however, Corollary 3.1 suggests that for a stable μ distribution, increasing the number of districts constrains the possibility of unsorted equilibria: With more districts, the distance between the average incomes of districts that are adjacent in the quality distribution is smaller. As C depends on this distance, a higher J reduces the amount by which a low-income district's effectiveness parameter can exceed that of the next-wealthier district before the wealthier families will bid away houses in the more effective district.

This tendency is at the core of my empirical strategy. To clarify it, I present next to a simulation exercise that demonstrates the effect of market structure on effectiveness sorting under different assumptions about the importance of peer group to parental preferences, and thus about the "stickiness" of residential assignments. I begin by describing the allocation of effectiveness in illustrative equilibria, then describe the simulation and its results. Finally, at the end of this section I

[28] One might imagine that unsorted equilibria are less efficient than the perfectly sorted equilibrium, but this need not be true: If the marginal utility of school quality declines quickly enough, it can be more efficient to assign effective schools to low-income bins than to the wealthiest students. In any case, concern for peer group amounts to an externality, and there is no assurance that the efficient assignment of families to districts is an equilibrium at all. It may be efficient to have heterogeneous income distributions at each school, for example, but this is never an equilibrium.

return to the basic model to discuss its allocative implications and the likely effects of endogenizing school effectiveness.

3.1. Graphical illustration of market equilibrium

From Theorem 2 and its corollaries, the income distribution in any equilibrium is divided into *J* quantiles, with wealthier quantiles living in more preferred—higher $\bar{x}_j\delta + \mu_j$—districts. From Theorem 3, this necessary condition is also sufficient for equilibrium. I use these results to construct possible equilibria under different (δ, J) combinations. These sample equilibria make concrete the intuitions about effectiveness sorting discussed in Section 2.

It is helpful to begin by considering a Tiebout market that approximates perfect competition. Assume that there are as many districts as there are families, with only a single house in each district, and suppose that both family income and school effectiveness are uniformly distributed on [0, 1]. There is no peer group externality, as families that move to another house-district take their "peer group" with them. Regardless of parental valuations, then, families always prefer a high-μ house to one with lower μ. Because willingness-to-pay for a preferred school is increasing in *x*, equilibrium is unique, with the ranking of districts by effectiveness is identical to that by the income of the resident family. Panels A and B of Figure 1 graph the allocations of effectiveness (μ) and district desirability ($\bar{x}_j\delta + \mu_j$) as functions of family income when parents have no concern for peer group ($\delta = 0$, Panel A) and when concern for peer group is moderate ($\delta = 1.5$, Panel B).

The competitive case serves as a baseline, but it is not a realistic description of choice in the presence of externalities. We next consider a market with ten equally-sized districts, a degree of Tiebout choice that, as is discussed below in Section 4, corresponds roughly to the 80$^{\text{th}}$ percentile U.S. metropolitan area. Assume that $\mu_j = {}^{j}\!/_{10}$, $j = 1, \ldots, 10$. Panel C of Figure 1 displays the unique, perfectly sorted equilibrium when $\delta = 0$. Families in the j^{th} decile of the income distribution live in the district with the j^{th} most effective schools.

When parental concern for peer group is introduced, the perfectly sorted equilibrium is no longer unique. It is now possible for ineffective districts to retain wealthy peer groups in equilibrium, as long as they are not so ineffective that families would prefer a lower-$\bar{x}$, higher-μ district. One imperfectly sorted equilibrium is displayed in Panel D. Note that district desirability is monotonically increasing in district average income, as Theorem 2 requires that the desirability and income rankings be identical in equilibrium. Effectiveness is not monotonic in family income, however: Some families live in districts that are less effective than those where some poorer families live. Effectiveness sorting nevertheless remains substantial, and effectiveness is highly correlated with peer group average income.

Finally, we consider the case where the housing market gives parents few options, with only three equally-sized districts. This corresponds roughly to the 38th percentile of the U.S. distribution. Suppose here that $\mu_j = j/3$, $j = 1, 2, 3$. When there are no peer effects (Panel E), equilibrium is again unique and is perfectly sorted on effectiveness.

When we add concern for peer group to the three-district market, there is substantially more potential for mis-sortings than even in the ten-district case. The gap in peer quality between adjacent districts has grown substantially, and families therefore require a much larger μ return to justify a move from one district to another whose current residents are lower in the x distribution. Indeed, with the parameter values used here, there is *no* allocation of x terciles to districts in which any family would willingly move to a lower-$\bar{x}$ district; all six of the possible allocations are equilibria. Panel F illustrates one possibility. Here, the most effective district is rewarded with the wealthiest students, but the two remaining districts are mis-sorted.

Recall equation (3), which suggested that a naïve estimate of the peer effect is magnified by effectiveness sorting, with the degree of magnification being $\theta^* \equiv \text{cov}(\bar{x}_j, \mu_j)/\text{var}(\bar{x}_j)$, the coefficient from a regression of μ_j on $\bar{x}_j$ across all districts in the market. $\theta^* = 1$ in the perfectly sorted markets displayed in Panels A, B, C, and E of Figure 1, indicating that the slope of school-level average test scores with respect to student characteristics in these markets will overstate the contribution of individual and peer characteristics to student performance by one. In the imperfectly sorted markets displayed in Panels D and F, however, the magnification effect is smaller: $\theta^* = 0.9$ in D and 0.5 in F. The simulations below suggest that this tendency for effectiveness sorting and magnification to depend on the number of districts when parents care about both peer group and effectiveness holds generally, as long as concern for peer group (δ) is moderate. When δ is large, however, even markets with many districts can have unsorted equilibria, and there is no tendency for $E[\theta^* | \delta, J]$ to increase with *J*, at least in the ranges considered here.[29]

3.2. Simulation of expanding choice

In this subsection, I describe simulations of a hypothetical regional economy under several combinations of (δ, J). As δ grows, the relative importance of school effectiveness diminishes and the likelihood of unsorted equilibria expands. By the logic above, for any fixed δ we might expect unsorted equilibria to be less prominent with many districts than with few.

Where Figure 1 used uniform, nonstochastic distributions for both income and effectiveness, here I adopt the slightly more realistic assumption that income has a normal distribution and I draw random effectiveness parameters from the same distribution.[30] For each market type, I conducted 5,000 draws, first choosing effectiveness parameters for each district and

---

[29] For any δ, there is *some J* for which effectiveness sorting will increase: The perfectly competitive case in Panels A and B would be perfectly sorted for any δ. I simulate only markets with $J \leq 10$—the computational burden increases with $J!$—though this is easily enough to reveal the general trend.

[30] Analysis of varying δ subsumes the variance of the μ_j's: Increased variation in school effectiveness is equivalent, for the purpose of the sorting process, to increased parental valuation of a district with high effectiveness relative to one with a desirable peer group (i.e. to a reduction in δ). A normal (rather than log-normal) income distribution was chosen to avoid complications from the log-normal distribution's skew, and because the *x* index that I use in the empirical analysis is approximately normally distributed.

then permuting the assignment of income bins to districts until I obtained an equilibrium assignment (i.e. one in which no low-income district was preferable to any high-income district).[31]

Figure 2 displays the average allocation of school effectiveness in markets with three and ten equally-sized districts. Panel A depicts the case where parents are unconcerned about the peer group, as in the left-hand panels of Figure 1. Here, families must be perfectly sorted on school effectiveness in equilibrium, and the average μ's depicted in the figure are simply order statistics from the standard normal distribution. The remaining panels show progressively higher valuations for the peer group: $\delta = 0.5$, 1.5, and 3. As δ grows, progressively less complete sortings become equilibria and average μ_j values collapse toward the overall mean.[32] Moreover, the collapse happens more quickly for three-district markets than for those with ten districts. This means that when δ is moderate in Panel C, the gradient of school effectiveness with respect to family income is steeper for $J = 10$ than for $J = 3$. As δ grows, however, Panel D indicates that the differences between the two sorts of markets shrink toward zero.

It is clear from Figure 2 that effectiveness sorting tends to increase with δ and, for moderate values like that shown in Panel C, with *J*. The simulation results can be used to understand the magnification bias in naïve estimates of the peer effect like (3). For each (δ, J) combination, I estimated a regression of μ_j on $\bar{x}_j$, pooling all 5,000 simulated markets and including a fixed effect for each. The resulting estimates of $\theta(\delta, J) = \text{cov}(\bar{x}_j, \mu_j)/\text{var}(\bar{x}_j)$ are displayed in Figure 3. The trends identified in Figures 1 and 2 are again clear. First, θ is well above zero when δ is small, indicating that the residential housing market mechanism rewards administrators of effective schools with the wealthiest students when parents primarily assess

[31] This strategy treats all possible equilibria as equally likely. It might be more realistic to attach higher probability to equilibria that are attracting points for larger ranges of initial assignments under some adjustment process, but this is left for future work.

[32] The nonmonotonicity of the $\delta = 3, J = 10$ case is likely a result of the equal weighting of all possible equilibria, and would probably not be robust to a more realistic adjustment process.

schools by their effectiveness. When δ is large, θ is close to zero for all *J*, as no district structure creates the desired rewards when parents are largely unconcerned with school effectiveness.

The moderate δ case is the most interesting. Here, we observe more perfect sorting on μ—and therefore larger slopes of μ with respect to $\bar{x}_j$—when there are many districts than when there are few. That is, $\partial\theta / \partial J > 0$ for moderate δ.[33] If both peer group and school effectiveness are important to parents, then, the Tiebout mechanism rewards effective administrators only when there are many districts. Model (3) suggests that in this case the test score gap between high- and low-income schools will tend to be larger in markets with a great deal of interdistrict competition than in those with less Tiebout choice. I test for this in the empirical analysis below.

3.3. Allocative implications and endogenous school effectiveness

In the model presented above, Tiebout choice hurts low-income students in two ways. First, it permits increased stratification of students. Because total peer group is in fixed supply, stratification necessarily offers better peers to wealthy students and worse peers to low-income students. Second, if the market mechanism functions and families sort on effectiveness, it assigns low-income students to schools that are below-average in their effectiveness. This is an unavoidable side effect of the Tiebout mechanism, as the flip side of rewarding effective schools with wealthy students is punishing poor students with relatively ineffective schools.

The model stacks the deck, however, by holding the distribution of effectiveness fixed. If school administrators respond to incentives, effectiveness sorting will also induce higher effort and greater effectiveness. This will tend to raise scores for everyone, and the productivity benefits may offset the allocative costs that Tiebout choice imposes on poor students.[34]

---

[33] Figure 3 reveals a small effect of Tiebout choice on the effectiveness gradient even when $\delta = 0$, but this is sensitive to the simulation assumptions (in particular, to the distribution of effectiveness as the number of districts grows). The simulations for positive δ—in which equilibrium need not be unique, so that averages are determined both by the distribution of effectiveness and by the set of equilibria—are much less sensitive.

[34] There is great need for a model of the supply side of Tiebout choice markets that describes the distribution of administrators' responses to incentives. Does competition force the worst districts to catch up to the average, induce the best districts to pull away from the average, or lead all districts to improve effectiveness equally? A Mirrlees-type

My empirical analysis thus has two components. In Section 5, I look for evidence that effectiveness sorting is more complete in high-choice than in low-choice markets, as the simulations above suggest it should be if parental valuations attach substantial weight to school effectiveness. In that section, I identify effectiveness sorting from the distribution of student performance within markets, using fixed effects to absorb any differences in average effectiveness across markets. Then, in Section 6, I examine the distribution of average test scores across markets, looking for evidence that interdistrict competition leads to increases in the average effectiveness of local administrators.

4. 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. I describe first my measure of market structure, defined over district-level enrollment. I then present evidence that this measure represents a binding constraint on parents' ability to exercise Tiebout choice. Finally, I discuss the SAT data that are the primary source of information on student outcomes across schools.

4.1. Measuring market concentration

I define local housing markets as Metropolitan Statistical Areas (MSAs), Census Bureau approximations of local housing markets defined by observed commuting patterns.[35] The SAT data that I use to measure student outcomes are taken from the early 1990s. Consequently, I use 1990 MSA definitions and draw demographic characteristics of each MSA from the 1990 Census.

---

argument suggests that the first is unlikely without market entry, as a district that enrolls the lowest-income students faces little sanction for further reductions in effectiveness. If this intuition holds, administrative responses would not offset the inequality-increasing effects of Tiebout choice identified here.

[35] The Census Bureau classifies the largest urbanizations as Consolidated MSAs (CMSAs), and subdivides them 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 across 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 slightly larger—New England County Metropolitan Areas. For reasons of data availability and comparability, the Honolulu and Anchorage MSAs are excluded from all analyses.

MSAs differ substantially in their educational governance structures. While the median MSA has 9 school districts, there are 25 markets with only a single district each. (Thirteen of these—including Miami and Fort Lauderdale, by far the largest—are in Florida, which has large counties and only one district per county.) Boston, with 132 districts, represents the other extreme; seventeen additional markets have fifty districts or more.[36]

The raw count of districts is a crude measure of market concentration, as it does not distinguish between the New York PMSA, where the three largest districts have 87 percent of enrollment and the remaining 53 districts combine for 13 percent, and the Dallas PMSA, with the same number of districts but only 44 percent of enrollment in the three largest. Following Hoxby (2000a), I calculate a more appropriate index of Tiebout choice as one minus the Herfindahl Index, a concentration measure used by the Federal Trade Commission (FTC) in antitrust deliberations and defined as the sum of firms' squared market shares. Districts' "market shares" are their enrollments in grades 9-12 divided by the total over all public school districts in the MSA, calculated using data from the 1990 Common Core of Data (CCD), an annual census of public schools and districts.

Figure 4 displays the histogram of the resulting choice indices. Nearly all metropolitan markets are highly concentrated by private market standards: Vertical lines on the figure indicate the FTC's thresholds for "concentrated" and "highly concentrated" markets (choice indices below 0.9 and 0.82, respectively); four-fifths of MSAs are concentrated and three-fifths highly concentrated by these definitions.

Table 1 displays summary statistics for several metropolitan-level demographic measures, calculated from county-level tabulations of the 1990 Decennial Census (from the STF-3C file) aggregated to the MSA level. Means of each variable are presented both for the full sample of 318 MSAs and within each quartile of the choice distribution. There are substantial differences across quartiles: Low-choice markets tend to be located in the South, to be smaller, and to have more

[36] All district counts and enrollment figures are calculated for grades 9-12 only (Urquiola, 1999).

Blacks and Hispanics. They are also more likely to be located in states with "Minimum Foundation Plan" financing schemes, a mechanism used by 37 states to reduce inequality in school resources.[37]

4.2. Does district structure matter to school-level choice?

Most of the existing literature, while recognizing that there is heterogeneity across schools within any given school district, has assumed that public school *districts* are the relevant units that compete for students in a Tiebout choice framework (Borland and Howsen, 1992; Hoxby, 2000a). There are two main reasons for this. First, any local tax and spending decisions are made at the district level, and this is also where many key education policies (curriculum, teacher pay scales, etc.) are set. Second, for reasons relating to the jurisprudence of school desegregation and to mechanisms like "open enrollment" and magnet schools, there are not always stable, well-defined catchment areas within districts that link neighborhoods to individual schools, so residential location may not be an important determinant of within-district school assignment.[38] Nevertheless, many districts limit the ability of parents to choose from among the schools in the district except by their location decisions, and even when parents can choose distance is often a major factor. Thus, Tiebout choice may operate across neighborhood schools within a large district as well as across districts. To the extent that peer groups and school-level policies, rather than policies set at the district level, are the primary objects of parental choice, neighborhood sorting within school districts may be a relatively effective form of choice.

In view of this possibility, it is important to ask whether inter-district competition matters to the way that students are assigned to neighborhoods and schools in Tiebout equilibrium. A first answer is provided by the data in Panel B of Table 1. MSAs with more district-level choice have more schools, on average, than do low-choice MSAs, but this is largely a function of population; average school size is only weakly correlated with the district-level choice index. Nevertheless, a choice index calculated across schools is strongly positively correlated with choice across school

[37] Categorizations of state finance plans as of the early 1990s are drawn from Card and Payne (2002).
[38] On desegregation remedies, see Welch and Light (1987), Orfield (1983), and *Milliken v. Bradley* 418 U.S. 717, 1974.

districts: In MSAs in the lowest quartile of district choice, the average school-level choice index is 0.82, versus 0.96 in MSAs in the highest district-level quartile. Moreover, although I do not report regression models for the school-level choice index, the relationship is robust to controls for the demographic characteristics shown in Panel A of Table 1.

The multicommunity model developed above, in which families stratify across jurisdictions, suggests another useful test of the hypothesis that district boundaries are important constraints on the Tiebout choice process. If school districts are a unit over which parents exercise residential choice, we should expect greater stratification by family income in markets with high district choice indices than in markets with more concentrated school governance, and this effect should be robust to the inclusion of a choice index calculated over schools.[39] Table 2 presents evidence on the relationship between the district-level choice index described above and two measures of within-MSA stratification, based on the distribution of household income across districts and the racial composition of schools.

The first three columns present regression models for the across-district share of variance of household income, calculated separately for each MSA with at least two districts.[40] All three models include as explanatory variables the district-level choice index, fixed effects for nine Census-defined geographic divisions, and controls for several MSA-level variables that might have independent effects on the degree of sorting. The second column adds to these a control for the school-level choice index, while the third column also controls for several measures of census-tract-level stratification.[41] All three estimates indicate a strong relationship between district-level choice and income stratification across districts.

---

[39] Eberts and Gronberg (1981) and Epple and Sieg (1999) propose similar stratification tests of Tiebout-style models.

[40] District-level income distributions are drawn from the School District Data Book (SDDB), a tabulation of 1990 Census data at the school district level. I am grateful to Cecilia Rouse for providing access to the SDDB data.

[41] Tract-level data come from the 1990 Census STF-3A files. Census tracts are much smaller than school districts, with 4,000 residents on average. Tiebout models do not speak to within-jurisdiction sorting, and invariance of the choice coefficient to tract-level controls offers reassurance that the relationships observed in Table 2 do not derive from a spurious correlation between district structure and MSA residents' tastes for micro-neighborhood segregation.

There may be a mechanical relationship, however, between measures of across-district sorting and the district structure. To see this, note that areas with more districts—conditional on market size—necessarily have smaller districts, and random distribution of populations would produce higher measures of segregation across these smaller areas. To avoid the bias that this produces, one would ideally estimate the same regressions for measures of across-*school* stratification. Unfortunately, representative income data are not available at the school level. Instead, I use data on the racial composition of each school, collected in both the CCD and the Private School Survey (PSS; National Center for Education Statistics, 2000), a census of private schools. I compute from these data a dissimilarity index (Cutler, Glaeser, and Vigdor, 1999) based on the distribution of white and non-white students across both public and private schools in each MSA.[42] Columns D, E, and F of Table 2 report models using this dissimilarity index as the dependent variable. Again, the coefficient on the district-level choice index is large, significant, and not much changed by the inclusion of the school-level choice index and the tract-level segregation measures.

The estimates in Table 2 are repeated using several additional stratification measures and alternative specifications in Appendix A. The basic result is clear: There is a strong, robust relationship between the structure of an MSA's educational governance (at the district level) and the degree of student stratification across schools and districts within that MSA. District-level market concentration evidently captures real variation in parents' ability to sort themselves across schools.

4.3. SAT data

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 of the distribution of student outcomes across schools within each MSA. The NELS uses a multi-stage sampling procedure and draws data from only three

[42] The earliest year for which I have been able to obtain electronic PSS data is 1997-1998, so they do not line up perfectly with the CCD data. Both the CCD and PSS datasets describe the racial composition of the entire school; when schools include both elementary and secondary grades, I assume that the racial composition of students in grades 9-12 is the same as that for the school as a whole. The 29 MSAs in which the CCD is missing racial composition for schools with more than 20% of MSA enrollment are excluded from the calculations.

schools in the average MSA.[43] 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.[44] As students in this sample generally entered high school in 1990, the MSA demographic data and choice measures discussed above should accurately describe the environment in which students' parents made their locational decisions.

The SAT data are rich, but have a serious limitation: 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 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. To partially remedy this, I discard all observations from the 27 states with SAT-taking rates below one third.[45] The remaining sample consists of 329,205 SAT-takers from 177 MSAs in "SAT states." This sample is likely representative of the college-bound population within the areas under consideration, and I do not further adjust for sample selection.[46] All analyses of the SAT data, however, control for the MSA SAT-taking rate.

---

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

[44] 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.

[45] 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: The marginal states, Colorado and Oregon, have rates of 23% and 38%, respectively. Within the set of states above the cutoff, there is very little evidence of positive selection into SAT-taking in high-SAT-taking states; see Appendix C.

[46] 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).

Exploratory analyses with more involved selection corrections—reported in Appendix C—suggest that the resulting estimates are not seriously biased by within-school selection into SAT-taking.

The size of the SAT database permits precise estimation of school-level measures: I have at least ten observations per school from schools with 77 percent of enrollment in the MSAs studied. 78 percent of schools (enrolling 90 percent of sample students) in the SAT data are public.

It is helpful to have a one-dimensional index of peer group quality at each school. To construct such an index, 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 by ten for father's education, each including a nonresponse category) and the interactions of six ethnicity indicators with two gender categories (eleven parameters) and with twelve family income bins (66 additional parameters).[47] 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. An index of peer quality was constructed by averaging the fitted values (excluding the estimated school effect) of this regression over all students at each school.[48] This index has the interpretation of the peer group's predicted average SAT performance at a nationally representative school.

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

[47] The model explained 33 percent of the cross-sectional variance in individual SAT scores (as compared with 22 percent explained by school effects alone).

[48] The individual characteristics coefficients may be biased by endogenous selection into schools. This is not 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. Tests reported in Appendix C indicate that the school-level index is quite reliable, and in any case specification checks reported in Section 5 indicate that the results are not particularly sensitive to the particular peer group measure used.

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.

Table 3 lists summary statistics for the SAT sample and for that portion of the sample in MSAs in each of the four choice quartiles. High-choice MSAs have substantially higher SAT-taking rates and scores than do low-choice MSAs. The differences in average scores, however, are entirely accounted for by differences in students' background characteristics.[49]

Figure 5 displays the scatterplot of school average SAT scores against the peer group index for a one-quarter subsample of the schools in the data. Circle sizes indicate the number of (weighted) observations entering the school-level averages. The line in the figure indicates the regression of average SAT score on the peer group, controlling for MSA fixed effects. This line has slope 1.74; the construction of the peer index is such that the effect of individual characteristics on own scores (i.e. β in equation 3) accounts for exactly 1 of this, with the remaining 0.74 deriving from the slope of school effects with respect to peer group (i.e. from $\gamma + \theta$, the combination of reduced-form peer effects and effectiveness sorting). In the next section, I look for evidence that the slope of this line is steeper in high-choice than in low-choice MSAs; under the assumption that β and γ do not vary systematically with choice, variation in the overall slope is informative about $\partial\theta / \partial c$, the effect of choice on effectiveness sorting. In Section 6, I estimate a different potential effect of Tiebout choice on the line in Figure 5. There, I look for evidence that choice affects its intercept, as it might if choice is correlated with average effectiveness (i.e. if $\partial E[\mu | c] / \partial c \neq 0$.

5. Empirical Results: Choice and Effectiveness Sorting

The sorting model in Section 3 predicts 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

---

49 Recall that low-choice MSAs are disproportionately Black, Hispanic, and in the South.

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 should be larger in high-choice markets than in low-choice markets.

My first test of this prediction in the SAT data uses nonparametric techniques to allow for a nonlinear educational production function. These offer no evidence of substantial nonlinearity, and I next turn to regression estimates of several 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 data sets or specifications studied here supports the hypothesis that effective schools are more likely to attract advantaged students in markets where the Tiebout choice index is high.

5.1. Nonparametric estimates

If neither the effect of individual characteristics on own scores (β) 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 5 which included data only from markets with high choice indices should exhibit a steeper slope than that shown, while a version estimated only from low-choice markets should be less steep. To avoid possible confounding from nonlinearities in the peer effect, I test this by estimating nonparametric versions of the regression line displayed in Figure 5 separately for high- and low-choice markets.

The median MSA contains only 19 high schools, not nearly enough to permit separate nonparametric estimation for each market. As an alternative, I grouped MSAs into quartiles by the choice index and estimated separate school-level kernel regressions of test scores on student characteristics for each quartile. Figure 6 displays the estimated functions, which use an Epanechnikov kernel and a bandwidth of five, about one-tenth of a school-level standard deviation. The figure offers little evidence of any differences in reduced-form educational production functions between the high-choice and low-choice quartiles, as the quartile functions are quite similar in both their intercepts and slopes.

To better display differences across quartiles, Figure 7 removes the underlying trend by subtracting from each quartile's estimated response function the average over all four quartiles. It is possible to discern here a slight upward slope in the second quartile relative to the third, but there seems to be no trend at all in the first quartile relative to the fourth.

5.2. Regression estimates of linear models

The quartile analysis in Figures 6 and 7 offers no natural way to control for MSA variables that might have independent effects on the housing market or on the causal importance of the peer group. Here, I develop and estimate a more parametric version of the hypothesis of interest. Drawing on the indication in Figure 6 that there is no substantial nonlinearity in the peer effect, I revert to the earlier linear model, letting *m* index housing markets:

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

$$E[\mu_{jm} | \bar{x}_{jm}] = \psi_m + \bar{x}_{jm}\theta_m^{*}. \tag{5}$$

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:

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

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 $\varphi_1 > 0$. Combining (4), (5), and (6), we obtain an estimable equation:

$$\begin{aligned} E[\bar{t}_{jm} | \bar{x}_{jm}] = (\alpha + \psi_m) + \bar{x}_{jm}(\beta+\gamma+\varphi_0) + \bar{x}_{jm}c_m\varphi_1 + \bar{x}_{jm}Z_m\varphi_2 + \\ + \left(\bar{x}_{jm}\omega_m + (\mu_{jm} - E[\mu_{jm} | \bar{x}_{jm}]) + \bar{\varepsilon}_{jm}\right), \end{aligned} \tag{7}$$

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

The effect of choice on the extent of 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 regression for school average test scores. The terms on the second line of (7) are unobserved residuals, and standard errors must be adjusted to account for their nonclassical structure.

Basic results

Table 4 contains the main empirical results of the paper. It presents OLS estimates of model (7), using MSA fixed effects to absorb the effect of variations in ψ_m. Standard errors permit arbitrary heteroskedasticity and are clustered at the MSA level to accommodate the within-MSA autocorrelation implied by the random coefficient ω_m. 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.

Column A displays a very restricted version of model (7) 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 5.) 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 a 0.88 standard deviation (84 points) change in average SAT scores. This, of course, reflects the combined influence of individual characteristics (β), peer effects (γ) and an average of the θ^*'s, the within-MSA gradients of school effectiveness with respect to peer group. The final rows of the table indicate that the peer group index explains 74 percent of the within-MSA variation in school average SAT scores, and that the index plus MSA fixed effects account for 77 percent of the total variation in the sample.

Column B adds a single interaction of the peer group with a choice index. The estimate of φ_1 is small and indistinguishable from zero. The remaining columns add additional interactions of $\bar{x}_{jm}$ with several metropolitan-level controls 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 a control for the MSA-level SAT-taking rate and indicators for six census divisions; the log of the MSA population; and two combinations of additional demographic, income distribution, and institutional controls. In each specification, the φ_1 point estimate is negative, although it is only significantly different from zero in columns C and D. By the rightmost column, the model is somewhat overfit: there are 18 interactions of MSA variables with the $\bar{x}_{jm}$ index and only 177 MSAs in the sample.

All of the models in Table 4 are based on a particular specification of the educational production function, (7), which may not be correct. Table 5 reports the results of several alternative specifications, each using the control variables from Column E of Table 4. Column A repeats the relevant coefficients from that specification. In column B, 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 Column A.

Column C 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.

Column D 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 again negative and insignificant.

Columns E and F explore the impact of varying the sample definition. In Column E, the basic model is estimated on public schools only, while in Column F the 18 MSAs that have only a single district are excluded. The choice-peer group interaction coefficient is again negative in each of these specifications, significantly so (and with a substantially larger point estimate than in the basic specifications) in the latter case.

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. In addition, Appendix B presents several instrumental variables estimates of (7); there is no indication that endogeneity of the choice index biases the estimates presented here.

Evidence from the NELS and from high school completion rates

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 an important determinant of families' evaluations of schools. Nevertheless, it remains possible that selection into SAT-taking biases the above results. To assess their validity, I turn to other data sets describing student outcomes across schools and MSAs. I use test score data from the National Educational Longitudinal Survey (NELS) and high school completion rates from the Common Core of Data for this purpose. Neither of these has nearly the breadth of the SAT data, so the estimates presented here are not as precise as those above, 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 700 schools (534 public and 166 private) in the NELS sample to the 205 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 8^{th} grade. (I continue to use the high school choice index in this analysis; it correlates 0.98 with an elementary-level index.) 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.[50] As in the SAT data, peer effects and effectiveness sorting are together substantial, inflating the school-level background index coefficient by 90 percent 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.

Panel B of the table repeats this analysis, this time with the score earned by students in the 1992 wave, when they were in the 12th grade.[51] Again, estimates of the choice effect are imprecise but are—with one statistically insignificant exception—of the opposite sign from that predicted by the economic model.

One non-test outcome that parents may care about is high school completion. The remaining panels present models for measures relating to school continuation rates. (The continuation rate is one minus the cumulative dropout rate.) In Panel C, the dependent variable is 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. The background index used is the same as that used in Panel B; 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.

The final panel leaves the NELS data, reporting models for high school completion rates, again defined as one minus the dropout rate but here extending through the end of the 12th grade, of the cohort entering 9th grade in the fall of 1993. Data on this outcome come from a district-level tabulation of the Common Core of Data distributed by the National Center for Education Statistics. There are several limitations to the CCD completion rate variable: It is measured at the district level rather than the school; it covers only public schools; it is missing for a great many districts who failed to report one of the component variables; and it may be unreliable if districts cannot distin-

[50] 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.

[51] School averages are still for the 8th grade school, as once students transfer to high schools the NELS sample is no longer representative of the schools attended. Only about half of the original sample is available in the 1992 wave.

guish mobility from dropout. Moreover, the CCD contains very little information about student background, and I therefore use the SAT data student quality index, aggregated to the district, to measure student characteristics. I drop MSAs that are not in SAT states or where available completion rate data cover less than two thirds of public enrollment. This leaves a sample of 931 districts from 50 MSAs. In spite of the serious limitations in the CCD data, the pattern of results in Panel D is quite similar to that in Panel C. Again, the student quality index is a strong predictor of completion rates, but its coefficient is (insignificantly) smaller in high-choice than in low-choice MSAs.

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

Possible biases in estimates of (7)

Several identifiable factors may bias the coefficient on the peer group-Tiebout choice interaction in specifications like (7). I discuss two here; I argue that each has the potential to produce an upward bias in φ_1.

The first source of bias is statistical. There are several reasons to suspect 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 estimated gradient of school average student outcomes with respect to peer group characteristics

The reliability of $\bar{x}_{jm}$ is likely to be higher, however, in markets where schools are more stratified. One reason is that stratification implies a higher true variance of the peer group, and therefore a larger signal component of the signal-to-noise ratio. A second reason is that schools in more stratified markets are likely to be more internally homogenous; as the sampling variance of the school average depends linearly on the within-school variance of individual characteristics, more internally homogenous schools imply more reliable school-level averages. A final reason to suspect a stratification-reliability relationship is that unobserved peer group characteristics are likely to be more strongly associated with observed characteristics in markets that are more heavily stratified.

In MSA-by-MSA regressions, the above arguments imply greater attenuation of the peer group coefficient in MSAs with less stratified schools. As choice is positively correlated with stratification, this produces 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 produces an upward bias in the effect of choice on the peer group gradient, and therefore in the interaction coefficient φ_1.

A second possible source of bias in φ_1 is economic. It seems likely 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) and (4)—which will appear as a positive contribution to φ_1.[52]

Either of these effects would imply upward bias in estimates of φ_1 relative to the effect of interest. To the extent that they are thought to be important, the results presented in Table 4, 5, and 6 should be seen as upper bounds on the effect of Tiebout choice on parental sorting into effective schools.

Calibration of results: Can we reject meaningful effects?

None of the estimates presented in this section supports the hypothesis that effective schools are more likely to attract the best peer groups in markets with fragmented school

[52] 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 districts with good teachers.

governance than in those where Tiebout choice is more difficult to exercise. Point estimates of the choice-peer group interaction are almost uniformly negative, suggesting that effectiveness sorting is *less* complete in high-choice than in low-choice markets. These estimates are imprecise, however, and most cannot reject a zero effect. It is worth considering whether the confidence regions exclude the sorts of effects that we would expect if school effectiveness were a prime determinant of parental location decisions.

Consider the specification in Column E of Table 4. Would a true effect of +0.20—the upper bound of a 95% confidence region for φ_1—be consistent with a Tiebout choice process driven in substantial part by parental pursuit of effective schools? The answer appears to be no. Note that the within-MSA gradient of school average SAT scores with respect to student characteristics is 1.74 (from Column A of the same table). Even at the upper limit of the confidence interval, a move from unified governance to complete decentralization accounts for just over ten percent of this gradient.

We can imagine as a thought experiment fully decentralizing school governance in Miami-Dade County, which is served by a single district.[53] Using the coefficients from Table 4, Column E, the fitted slope of school average SAT scores with respect to student characteristics in Miami is 1.63, implying a 0.82 standard deviation (78 points) average difference in SAT scores between schools that differ by one school-level standard deviation (48 points) in their student composition. At the upper limit of the confidence interval for the effect of choice, completely decentralizing Miami's school governance would increase the fitted slope to 1.83 and increase the SAT gap to 0.93 standard deviations (88 points).

Figure 8 displays the actual distribution of peer groups and school average SAT scores in Miami, as well as the counterfactual distribution that might be observed if the Miami choice index

[53] The district's web site indicates that the county is partitioned into school attendance areas. These can be changed easily, however, and indeed were under the supervision of federal judges for desegregation purposes from 1970 through 2001 (Welch and Light, 1987).

were changed to one and if the effect of choice were at the upper limit of its confidence interval.[54] The actual and counterfactual distributions of school averages are nearly identical. If the counterfactual reflects a substantial increase in sorting on school effectiveness, it must be that effectiveness is responsible for a very small share of the across-school variation in SAT scores.[55]

Recall, moreover, that this thought experiment assumes a choice coefficient at the upper limit of the confidence interval. At the point estimate, choice *reduces* the gradient of SAT scores with respect to student quality. The models in Table 4 reject a sizable—by any reasonable standard—effect of choice on the test score gradient. The estimated effects are difficult to reconcile with a sorting process in which school effectiveness is an important part of both location decisions and educational production.

6. Empirical Results: Choice and Average SAT Scores

The results presented in Section 5 offer no evidence that the allocation of effective schools is systematically different in high-choice than in low-choice markets. If Tiebout choice does not increase the probability that effective schools attract students from advantaged backgrounds, it is not clear how it can provide incentives that will lead administrators to exert greater effort. The above results thus suggest that the argument (Brennan and Buchanan, 1980; Hoxby, 2000a) that average school performance should be higher in markets with decentralized governance may not hold. The SAT data permit a direct test of this prediction, however.

Table 7 presents regression models for the average level of SAT scores across MSAs.[56] Column A includes only the choice index as a regressor. It enters with a positive coefficient,

---

[54] Note that decentralization of Miami's schools would probably change the allocation of peer group as well as the distribution of groups across schools. If, as Table 2 indicates, choice causes increased stratification, the counterfactual Miami market would exhibit more dispersion along the horizontal axis in Figure 8. The figure ignores any such effect, and simply considers whether decentralization would lead to increased dispersion of SAT scores conditional on the observed peer group allocation.

[55] If effectiveness is unimportant, of course, it would be irrational for parents to attach much weight to it, and we would not expect choice to promote effectiveness sorting.

[56] The effect of choice on average scores can only be interpreted as its effect on average productivity if peer effects are linear. If they are not, the stratification effect of choice could raise or lower average scores, depending on the

implying that fully decentralized MSAs produce average SAT scores about forty points higher than those with only a single district. Recall, however, that there are large differences between high-choice and low-choice MSAs in both SAT-taking rates and student characteristics (from Tables 1 and 3). Columns B, C, and D add controls for the SAT-taking rate and the average background index of SAT-takers.[57] As indicated by Table 3, the positive effect of choice on performance seems to result entirely from the omission of students' background characteristics; when they are included in Column C, the coefficient becomes negative and significant. Moreover, the MSA SAT-taking rate, which should have a negative effect on average scores if SAT-takers are positively selected, only has the correct sign when the background index is included as a regressor, and even here it is close to zero. The remaining columns add additional MSA-level regressors. Their inclusion does not substantially alter the estimated effect of choice: It remains negative and significant.[58]

The negative effect of Tiebout choice on average SAT scores indicated by Table 7 is not very large: A one standard deviation (0.28) increase in the choice index corresponds with a reduction in mean scores of only about four points, about one-eighth of an MSA-level standard deviation. Moreover, in some alternative specifications not reported here, the coefficient estimate is statistically insignificant, though still negative. When MSAs are weighted equally, for example, rather than by the number of SAT takers or by the 17-year-old population (not shown, but similar to the SAT-taker weighting in Table 7), the choice effect is about one third as large as that shown here and confidence intervals do not reject zero. Nevertheless, there is no indication that Tiebout choice is associated with higher SAT scores once student background is controlled.[59] Moreover, the coefficient on the background index across MSAs—1.58 in Column C, and slightly higher in later

convexity or concavity of peer effects, independent of any effect on productivity. Although Figure 6 does not indicate substantial nonlinearity, this possibility should be kept in mind when interpreting the results in Table 7.

[57] The background index is the same one described in Section 4, estimated from within-school variation in SAT scores.

[58] Note that in Column F, which controls for several MSA demographic characteristics, the coefficient on the SAT-taking rate finally takes on its expected sign.

[59] Hoxby (2000a), 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 high-income than for low-income students. The SAT sample might be thought analogous to her "not-low-income" group. Hoxby's positive effect is not seen here, either in the OLS results in Table 7 or in instrumental variables specifications (in Appendix B) similar to hers.

columns—is nearly identical to that found within MSAs (Table 4, Column A). This is consistent with the claim that both coefficients measure primarily the peer effect (γ), 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 on SAT scores *across* MSAs thus support those on the distribution of scores *within* MSAs: The evidence does not indicate that Tiebout choice provides incentives to school administrators to improve productivity, as productive administrators appear no more likely to be rewarded for it in high-choice than in low-choice MSAs.

7. Conclusion

This paper has used the Tiebout choice process—the choice of school characteristics via housing decisions—as a lens through which to study whether parental preferences for effective schools are strong enough to outweigh the other neighborhood or school characteristics that affect the desirability of particular residential locations. Earlier work on Tiebout choice markets presumes that parents use their location decisions to choose effective schools; one lesson of the analysis here is that the potential importance of peer group externalities to community desirability can create coordination failures in which ineffective schools are preferred to more effective competitors.

The motivation for the empirical approach is a model of the Tiebout marketplace in which housing prices ration access to desirable schools. As is common in multicommunity models, equilibrium is characterized by maximum stratification of families across school districts, with the wealthiest families residing in the most-preferred communities. Preferred districts need not have particularly effective schools, however, when peer group enters into parental valuations, as wealthy families can be "stuck" in ineffective schools by their unwillingness to abandon the peer group offered. For parental valuations that place substantial weight on school effectiveness, this becomes less likely as Tiebout choice increases parents' exit options.

In so far as student test scores depend on school effectiveness, effectiveness sorting is observable as an increase in the slope of school average scores with respect to student characteristics. I find no evidence that the gradient of school-level SAT scores with respect to student characteristics varies systematically with Tiebout choice, as would be expected if effectiveness allocations were more stratified in high-choice markets. 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 specification tests and alternative data sets fail to reveal important biases in the basic models. Consistent with the results on within-market sorting, I also find no evidence that Tiebout choice increases average SAT scores across markets, as would be expected if choice increases competitive pressure for administrators to run effective schools.

I see four possible explanations for the pattern of results. First, it may be that I have mis-measured the extent of Tiebout choice by focusing on a district-level choice index where in fact the relevant measure of parents' exit options is at the school level. Second, parents may have no concern whatever for the peer group, and may choose schools purely for their effectiveness. (Recall that there is no necessary connection between market structure and effectiveness sorting in this case.) Third, parents' concern for the peer group may be so large that it dominates effectiveness in their choices, so that again there is no effect of choice on effectiveness sorting. Finally, it may be that the sorts of policies that I call "school effectiveness," those not dependent on the peer group, are relatively unimportant determinants of student outcomes (or that they do not vary substantially across schools), and thus that effectiveness sorting and differences in average effectiveness across markets are not observable in the pattern of average SAT scores.

The first two of these are not particularly plausible. I present strong evidence, in Table 2 and in Appendix A, that the district-level choice index is an important determinant of student stratification, even when possible confounding factors are controlled. It seems that parents are sorting on *some* characteristics of school districts, though not on anything that serves to increase student performance conditional on individual and peer characteristics.

It similarly seems unlikely that parents have zero concern for peer group. In the presence of direct or indirect peer effects on student learning, parents would be irrational to ignore peer group in their evaluations of schools, and anecdotal evidence suggests that they do not do so. The likelihood that parents have imperfect information only reinforces this judgment, as the most widely available indicator of school quality, the average test score, loads heavily on the peer group, while value added is much more difficult to observe.

The alternative hypotheses that are consistent with the above results, that parental valuations place a great deal of weight on peer group relative to effectiveness or that administrative and instructional effectiveness is simply unimportant to the distribution of educational outcomes, seem more plausible. I interpret the paper's results as cautious support for the first of these, though the second would equally well explain the results and in any case their implications for the productivity benefits of Tiebout choice are the same.

In the absence of parental sorting on school effectiveness, 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. Caution is required, however, 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 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 thus seems likely that the character of equilibrium will depend crucially on the particular institutions of any choice program. Further research with large-scale voucher programs will be needed to determine whether administrators of effective schools are rewarded by increased demand in the choice regimes that these policies create.

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, Robert McMillan, and Kim Rueben (2002). "An Equilibrium Model of Sorting in an Urban Housing Market: A Study of the Causes and Consequences of Residential Segregation." Unpublished manuscript, August.

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.

Bowen, William G., and Derek Bok (1998). *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton, NJ: Princeton University Press.

Brennan, Geoffrey, and James Buchanan (1980). *The Power to Tax: Analytical Foundation of a Fiscal Constitution*, Cambridge, New York: 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.

Coleman, J., et al. (1966). *Equality of Educational Opportunity*. Washington, DC: United States Office of Education.

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.

Dale, Stacey Berg, and Alan Krueger (1998). "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." Princeton University Industrial Relations Section Working Paper #409, December.

Dee, Thomas S. (1998). "Competition and the Quality of Public Schools," *Economics of Education Review* 17 (4), 419-427.

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, and Thomas Romer (1991). "Mobility and Redistribution," *Journal of Political Economy* 99, August, 828-58.

Epple, Dennis N., and Glenn J. Platt (1998). "Equilibrium and Local Redistribution in an Urban Economy When Households Differ in Both Preferences and Incomes," *Journal of Urban Economics* 43 (1), January, 23–51.

Epple, Dennis N., Radu Filimon and Thomas Romer (1993). "Existence of Voting and Housing Equilibrium in a System of Communities with Property Taxes," *Regional Science and Urban Economics* 23 (5), November, 585– 610.

Epple, Dennis, and Richard E. Romano (1996). "Ends Against the Middle: Determining Public Service Provision When There Are Private Alternatives." *Journal of Public Economics* 62, 297-325.

Epple, Dennis, and Holger Sieg (1999). "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107 (4), August, 645-681.

Epple, Dennis, and Allen Zelenitz (1981). "The Implications of Competition Among Jurisdictions: Does Tiebout Need Politics?" *Journal of Political Economy* 89 (6), 1197-1217.

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.

Fernandez, Raquel, and Richard Rogerson (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.

Hamilton, B.W. (1975) "Zoning and Property Taxation in a System of Local Governments." *Urban Studies* 12, 205-211.

Hanushek, Eric A. (1981). "Throwing Money at Schools," *Journal of Public Policy Analysis and Management* 1 (1), 19-41.

Hanushek, Eric A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24 (3), September, 282-294.

Hanushek, Eric A., and Margaret E. Raymond (2001), "The Confusing World of Educational Accountability," *National Tax Journal* 54 (2), 365-384.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin (2001). "New Evidence about *Brown v. Board of Education*: The Complex Effects of School Racial Composition on Achievement," Working Paper #8741, National Bureau of Economic Research, December.

Heckman, James (1979). "Sample Selection Bias as a Specification Error," *Econometrica* 47, 153-161.

Hirschman, Albert O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Cambridge, Mass.: Harvard University Press.

Hoxby, Caroline Minter (1994). "Do Private Schools Provide Competition for Public Schools?" Working Paper #4978, National Bureau of Economic Research.

Hoxby, Caroline M. (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.

Hoxby, Caroline M. (1999b). "The Productivity of Schools and Other Local Public Goods Producers," *Journal of Public Economics* 74 (1), October, 1–30.

Hoxby, Caroline M. (2000a). "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90 (5), December, 1209-38.

Hoxby, Caroline M. (2000b). "Peer Effects in the Classroom: Learning From Gender and Race Variation," Working Paper #7867, National Bureau of Economic Research.

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., Douglas O. Staiger, and Gavin Samms (2002). "School Accountability Ratings and Housing Values," unpublished manuscript, October 20.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman (2001). "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics* 116 (2), May, 607-654.

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.

Luizer, James, and Robert Thornton (1986). "Concentration in the Labor Market for Public School Teachers," *Industrial and Labor Relations Review* 39 (4), 573-584.

National Center for Education Statistics (1999). *Digest of Education Statistics 1998*. U.S. Department of Education: Washington D.C.

National Center for Education Statistics (2000). *Private School Online Directory, v. 1.00e*. U.S. Department of Education: Washington D.C.

Nechyba, Thomas J. (1997). "Existence of Equilibrium and Stratification in Local and Hierarchical Tiebout Economies with Property Taxes and Voting." *Economic Theory* 10, 277-307.

Oates, Wallace E. (1985). "Searching for Leviathan: An Empirical Study," *American Economic Review* 75 (4), 748-757.

Orfield, Gary (1983). *Public School Desegregation in the United States, 1968-1980*. Washington: Joint Center for Political Studies.

Phillips, Meredith, Jeanne Brooks-Gunn, Greg J. Duncan, et al. (1998). "Family Background, Parenting Practices, and the Black-White Test Score Gap." In Jencks, Christopher, and Meredith Phillips, *The Black-White Test Score Gap*, Washington, D.C.: Brookings Institution Press.

Raudenbush, Stephen W., and Anthony S. Bryk (2002). *Hierarchical linear models : applications and data analysis methods.* Thousand Oaks, California: Sage Publications.

Raudenbush, Stephen W., and J. Douglas Willms (1985). "The Estimation of School Effects," *Journal of Educational and Behavioral Statistics* 20 (4), Winter, 307-335.

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. North-Holland: New York.

Sheppard, Stephen (1999). "Hedonic Analysis of Housing Markets," in Cheshire, Paul, and Edwin S. Mills, eds, *Handbook of Regional and Urban Economics* vol. 3. North-Holland: New York.

Technical Design Group (2000). Construction of California's 1999 School Characteristics Index and Similar Schools Ranks, PSAA Technical Report 00-1. California Department of Education: Sacramento.

Theil, Henri (1972). *Statistical Decomposition Analysis.* London: North-Holland.

Tiebout, Charles M. (1956). "A Pure Theory of Local Public Expenditures." Journal of Political Economy, 64 (5), October, 416–24.

Tracy, Joseph, and Joel Waldfogel (1997). "The Best Business Schools: A Market-Based Approach." *The Journal of Business* 70 (1), January, 1-31.

Urquiola, Miguel (1999). "Demand Matters: School District Concentration, Composition, and Educational Expenditure." Working Paper #14, University of California, Berkeley, Center for Labor Economics, April.

U.S. Geological Survey (2002). *Geographic Names Information System*, State Gazetteer electronic data.

Welch, Finis, and Audrey Light (1987). *New Evidence on School Desegregation.* United States Commission on Civil Rights Clearinghouse Publication 92, June.

Wells, Amy Stuart (1993). "The Sociology of School Choice," in Rasell, E., and R. Rothstein, eds., *School Choice: Examining the Evidence.* Washington, DC: Economic Policy Institute.

Willms, J. Douglas, and Frank H. Echols (1992). "Alert and Inert Clients: The Scottish Experience of Parental Choice of Schools" *Economics of Education Review* 11 (4), 339-350.

Willms, J. Douglas, and Frank H. Echols (1993). "The Scottish Experience of Parental School Choice," in Rasell, E., and R. Rothstein, eds., *School Choice: Examining the Evidence.* Washington, DC: Economic Policy Institute.

Appendices

Appendix A: Choice and School-Level Stratification

Table 2 presents evidence for an effect of Tiebout choice over school districts on the segregation of students across both districts and schools. There are several reasons to be interested in this relationship. First, an effect of interjurisdictional competition on residential stratification is a clear implication of Tiebout-style multicommunity models, and indeed has been proposed as a test of the Tiebout hypothesis (Eberts and Gronberg, 1981; Epple and Sieg, 1990). Second, the determinants of residential sorting are interesting in their own right (Cutler, Glaeser, and Vigdor, 1999). Finally, the allocation effects of school choice policies must be weighed in any evaluation of their social benefits; many authors have argued that choice's first-order effect will be to increase the inequality of peer group allocations (Wells, 1993).

In this Appendix I present additional evidence that district-level choice has a strong, robust effect on student stratification across schools and districts. Interestingly, this result is not affected by the inclusion of controls for micro-neighborhood segregation, and indeed there seems to be no relationship between district structure and stratification measured at levels finer than the school.

Table A1 presents models similar to those in Table 2, but for several alternative measures of student stratification. In Columns A and B, the dependent variable is the across-district share of variance of adult education. The district-level choice index has a strong effect on this variable, one that is robust to the inclusion of controls for the school-level choice index and census tract-level segregation measures in Columns B. As noted in the main text, however, there may be a mechanical relationship between district size and across-district sorting. Columns C and D present estimates for the school-level white/non-white isolation index. The isolation index has the interpretation of the fraction non-white at the average non-white student's school, normalized by the overall fraction non-white in the metropolitan area (Cutler, Glaeser, and Vigdor, 1999). Again, we see a strong effect of district structure that is unchanged by the inclusion of additional controls.

The remaining columns of Table A1 present models for an alternative measure of racial segregation, the Theil segregation measure (Theil, 1972). This measure has two advantages: It can accommodate multiple racial groups, which the two-group dissimilarity and isolation indices cannot, and the school-level measure has a natural decomposition into across-district and across-school, within-district components. Column E presents estimates of the effect of choice on the school-level Theil index; Column F the effect on the across-district component; and Column G the effect on the across-school, within-district component.[60] District structure has a weak, but significant, effect on the school-level measure and a much stronger effect on the across-district component. It also has a large, negative effect on the within-district measure, perhaps resulting from the mechanical relationship discussed earlier or from district-level desegregation policies.

Hoxby (2000a, p. 1236) argues that even at the school level there is a mechanical relationship between the district-level choice measure and her measure of student stratification, as the two are constructed similarly and "the similarity of construction creates spurious correlation." This should not be a problem for the results presented here, as I use measures of stratification that do not share the choice index's construction. Nevertheless, following Hoxby, I present in Table A2 estimates of

[60] I use 5 racial groups to calculate the Theil index: Asian, Black, Hispanic, Native American, and White. The school-level Theil measure in Column E is defined over both public and private schools, so cannot be literally decomposed into across- and within-district components. The decompositions in Columns F and G use an alternate measure defined over public schools only; this correlates 0.998 with the public-plus-private measure.

the choice-stratification relationship that use two alternative measures of district-level choice, the number of districts and the number of districts per 17-year-old resident. Each specification indicates a strong relationship, though the alternative measures of district competition produce somewhat less precise estimates than does the choice index. These results are somewhat at odds with Hoxby's, as she finds no relationship between number of districts and racial segregation.

Finally, Table A3 presents estimates that use measures of segregation at the census tract level as the dependent variable. The choice index coefficient is uniformly small, and is insignificant for three of the four measures. The basic estimates are unchanged in models that include a school-level choice index (not shown). Family sorting across schools and districts evidently does not carry over into substantial sorting at the micro-neighborhood level.

Appendix B: Potential Endogeneity of Market Structure

There has been a broad trend in the U.S. in the post-war era toward consolidation of school districts, and there were only 19% as many districts in the country in 1990 as there were in 1950 (Kenny and Schmidt, 1994). Hoxby (2000a) argues that one factor that may have influenced the degree of consolidation is the quality of local school administration: In markets with particularly badly-run central city schools, suburban districts may have resisted pressures to consolidate, producing a negative effect of past school quality on current Tiebout choice. If quality is sufficiently stable over time, current choice may be endogenous to current school quality, biasing estimates of choice's effect on school performance (Table 7) downward.

Figure B1 graphs the evolution of the number of districts over time in the U.S. as a whole, and in the counties comprising the 1990 MSAs. These data are drawn from a variety of sources, using varying definitions, but the overall trends are clear. The number of districts fell dramatically in the decades before 1970, but has been essentially unchanged in the three decades since. The recent stability provides a reason to be suspicious of the claim that district structure is endogenous to current quality: Unless quality is extraordinarily stable, any quality considerations that influenced consolidation during the 1960s are unlikely to be reflected in 1994 test scores.

Nevertheless, I present here instrumental variables estimates of the effect of choice on student stratification, effectiveness sorting, and average SAT scores across MSAs. I use several instruments for current market structure that are likely to predate possible sources of endogeneity. One candidate is the number of rivers and streams in the area, which Hoxby (2000a) argues predicts current district structure because initial district boundaries are likely to have been drawn along natural barriers to travel and to have persisted in current institutions.

A second strategy takes advantage of the fact that, even where districts are independent entities, school districts rarely cross county lines. Some MSAs have larger counties, and consequently looser upper bounds for district size, than do others. I calculate for each market a choice index for the distribution of population across counties, analogous to the choice index for students across school districts. As county boundaries are effectively unchangeable in nearly all states, differences across MSAs in county choice should predate any endogenous influences.[61]

Yet another strategy is to derive instruments from institutional characteristics governing the size of districts that are stable over time. One such characteristic is an indicator for whether the market is in one of the six states where districts have always been functions of county government: As counties are relatively large, MSAs in these states have always had fewer districts than other

[61] The county choice index depends on the intra-MSA distribution of growth as well as on the original county boundaries. If MSA settlement patterns are endogenous to average school quality, the county choice index is an invalid instrument. As it happens, results are largely insensitive to the particular choice of instruments.

MSAs. Another instrument is derived from a Census Bureau tabulation of independent school districts by county in 1942 (Gray, 1944).[62] I calculate for each 1990 metropolitan area an estimate of the area's choice index in 1942, treating all districts as equally sized and ignoring the primary-secondary distinction.

A final strategy takes more direct account of the differences in state-level institutions: I instrument for each MSAs choice index with the average choice index of other MSAs in the same state, with the idea that institutions determining market structure in Boston (choice index=0.98) and Miami (choice index=0) are likely to be similar to those operative in Worcester, MA (0.96) and Fort Lauderdale, FL (0), respectively.

Table B1 displays first stages for several combinations of the candidate instruments. Each of the proposed instruments is highly significant in the first stage, and each has the expected sign. When all of the instruments are included together, in columns F and G, all but the streams variable remain significant.[63]

Table B2 reports two-stage least squares estimates of several specifications from the main text, using the same combinations of instruments as in Table B1. Each column reports OLS estimates of the choice coefficient in the first row and different 2SLS estimates of the same coefficient in each succeeding row.[64]

Columns A and B present 2SLS estimates of the segregation models from Table 2. Although several 2SLS specifications yield imprecise estimates, the point estimates are uniformly of the same sign as the OLS coefficients and generally are similar in magnitude. Eight of the twelve estimates in these columns reject zero effect of choice, while Hausman tests fail to reject equality of any of the 2SLS choice coefficients with their OLS counterparts. Additional IV estimates of the models in Appendix A, not shown here, similarly support the OLS results. There would seem to be little evidence that the choice-segregation relationship is biased by endogeneity of the choice index.

Column C of Table B2 presents 2SLS estimates of the choice-peer group interaction coefficient in a model for the school average SAT score. Here, the instruments are the interactions between the variables listed and the school-level peer group background index. The estimates are quite noisy, and most are positive. None reject zero, however, and none of the Hausman tests reject the OLS estimate.

Finally, Column D reports 2SLS estimates of the effect of choice on MSA mean SAT scores, as discussed in Section 6. Once again, the estimates are somewhat noisy, but there is again no indication that high-choice MSAs produce higher SAT scores than do low-choice markets, once student background is controlled. One Hausman test—for the model using streams as the sole instrument—rejects the equality of OLS and 2SLS estimates, suggesting perhaps a larger negative effect of choice on average scores than is indicated by OLS.

---

[62] In several states where districts are traditionally dependent upon town or county governments, the 1942 data list zero districts. I assign to counties in these areas either a single district (where schooling is traditionally a county function) or as many districts as there are townships. These imputation rules are quite accurate for later years in which counts of dependent districts are available.

[63] Hoxby actually uses both a "smaller streams" and a "larger streams" variable, but has declined to make them available for this analysis. My streams measure is constructed from the Geographic Names Information System (U.S. Geological Survey, 2002) database following Hoxby's description of her "smaller streams" variable, which she reports as by far the more powerful of her two instruments. The coefficient in column B, 0.32, is nearly identical to that which Hoxby reports for the "smaller streams" variable in her first stage.

[64] Of course, if all the instruments are indeed exogenous, the estimates in the final row—which optimally weight the entire set of instruments—are the most efficient. Other rows are presented to demonstrate that the basic similarity between OLS and IV specifications does not rely on any single exogeneity assumption.

Taking the instrumental variables estimates as a whole, there appears to be no reason to suspect serious endogeneity of the 1990 district-level choice index to any of the dependent variables considered here. I read this pattern of results as justification for my focus in the main text on the somewhat more precise OLS results.

Appendix C: Selection into SAT-taking

The great limitation of the SAT data used in this paper is that students self-select into taking the SAT. Because SAT-taking rates vary considerably across states, estimates based only on SAT-takers' performance may not accurately describe patterns of student performance in the entire population of students. Figure C1 displays the relationship between SAT-taking rates and average SAT scores across MSAs. There is a clear negative relationship, indicating that at this macro level there is probably positive selection into SAT-taking (Dynarski, 1987, and Card and Payne, 2002, present similar graphs).

The picture is very different, however, when one distinguishes between MSAs in "SAT states," indicated by solid diamonds, and MSAs not in SAT states, indicated by pluses. Within the SAT state sample, the correlation disappears: Markets with high participation rates have average scores no lower than do those with relatively low rates. All analyses of the SAT data in this paper use only observations from SAT-state MSAs, and moreover control for the MSA SAT-taking rate. In this appendix, I describe several additional tests that have been performed to gauge the degree to which selection into SAT-taking, and particularly within-MSA selection, may bias the results above.

The first form of analysis involves explicit models for the selection process. Ideally, one would use a variable that predicts a student's probability of taking the SAT but does not predict the student's score conditional on test-taking. It is difficult to think of an instrument for this selection margin, however. Instead, I attempted to use the school SAT-taking rate as a summary of the factors that might determine sample selection. Specifically, I estimated models of the form

$$E\left[t_{ij} \mid X_{ij}, Z_j, \pi_j, i \text{ takes the SAT}\right] = a + X_{ij}b + Z_jc + \lambda\left(\pi_j\right)d, \qquad \text{(C1)}$$

where X_{ij} is a vector of individual characteristics for student i at school j; Z_j is a vector of school-level measures, and π_j is the SAT-taking rate at school j. $\lambda(\cdot)$ is a "control function," which was specified as the inverse-Mills ratio, $\lambda\left(\pi_j\right) \equiv \varphi\left(\Phi^{-1}(\pi_j)\right) / \pi_j$. This specification is appropriate for a conventional Heckman-style model of sample selection in which the factors determining SAT-taking are constant for all students at school *j* and residuals in selection and SAT-score equations are jointly normal (Heckman, 1979; Card and Payne, 2002).

If students are positively selected into SAT-taking, we expect $d > 0$, as increases in a school's SAT-taking rate should reduce average scores. Using a variety of peer group measures in *Z*, OLS estimates of *d* were all large and negative, most likely indicating that this cross-school comparison does not adequately control for the determinants of SAT scores. In an effort to obtain a more reasonable selection model, I also estimated versions of (C1) with school fixed effects, using data from the 1994 through 1998 SAT-taking cohorts and identifying the selection parameter from within-school, across-year variation in SAT-taking. This produced an estimated *d* with the correct sign, although the implied correlation between test-taking propensity and the latent test score was almost implausibly small: $\hat{\rho} = 0.02$.[65]

It is difficult to have much faith in estimates of selection models like (C1) without an adequate instrument for selection. To further explore the potential impact of selection, individual SAT scores were adjusted according to model (C1) under several assumed ρ (and therefore *d*) values. Table C1 reports the correlation of individual and school mean SAT scores and student background indices across different choices of ρ. These correlations are all quite large, indicating that school-level selection adjustments (at least using models like (C1)) are unlikely to affect results greatly. Based on these correlations, the basic analyses in the main text were conducted using unadjusted SAT scores for the sample of 177 high-SAT-participation MSAs. Exploratory analyses with adjusted scores (for moderate assumed ρ) produced substantially similar results to those obtained from raw scores.

Table C2 offers further suggestive evidence that selection bias is not a major problem for the school-level analyses conducted here. It displays the correlation across years in school-level average SAT scores and peer group background indices.[66] The smallest correlation coefficient here is 0.899, indicating that both measures are quite reliable: To the extent that selection into SAT-taking biases the school-level averages that are the focus of the analysis here, there is apparently very little variation in this selection across years. Moreover, the correlations decay quite slowly over time, indicating that schools do not change rapidly and that much of the across-year variation in school averages likely derives from transitory sampling error.

In a final attempt to test the robustness of the basic results to selection into SAT-taking, I made use of a variable in the SAT data describing students' self-reported rank (by grade point average) within their high school classes. Response categories correspond to top decile, second decile, and second through fifth quintile, although the bottom categories are very rarely reported. I used the class rank variable to "re-weight" the SAT data so that one-sixth of the weighted SAT observations at each school come from each of the top two deciles and one-third come from each of the second and third quintiles (observations from the bottom two quintiles are dropped). Under the assumption that sample selection is random within each school-decile cell, these weights produce consistent estimates of average SAT scores and student characteristics for the 60 percent highest-ranked students at each school, and in particular produce averages that are comparable across schools. Table C3 presents estimates of the SAT score-peer group gradient model—equation (7)—from the reweighted data. The estimated models are nearly identical to those in Table 4.

Appendix D: Proofs of Results in Section 3

It is useful to begin with a Lemma that follows directly from the single crossing property:

Lemma 1. Suppose that $\bar{x}_j\delta + \mu_j > \bar{x}_k\delta + \mu_k$ and $h_j > h_k$ and assume the single-crossing property:

i. If a family with income x_0 (weakly) prefers community *j* to community *k*, then all families with $x > x_0$ strictly prefer district *j* to district *k*.

[65] There is almost certainly measurement error in school enrollment, and therefore in the school-level SAT-taking rate. One explanation for the small selection coefficient is attenuation from unreliability of within-school changes in SAT-taking rates, which may contain very little signal but a good deal of noise.

[66] The background index was estimated separately for each year, with a new set of weights for individual characteristics derived from a year-specific regression of SAT scores on individual characteristics with high school fixed effects.

ii. If a family with income x_0 (weakly) prefers community *k* to community *j*, then all families with $x < x_0$ strictly prefer district *k* to district *j*.

Proof of Lemma 1.

I prove part i; the remainder follows directly by a similar argument. Define $q_j \equiv \bar{x}_j\delta + \mu_j$. Suppose first that the two districts' quality and housing prices are "close" to each other, so that first-order Taylor expansions are accurate. Consider an expansion of the utility function around the utility that family x_0 obtains in district *k*, evaluated at $(x_0 - h_j, q_j)$:

$$U(x_0 - h_j,\, q_j) - U(x_0 - h_k,\, q_k) \approx -(h_j - h_k)U_1(x_0 - h_k,\, q_k) + (q_j - q_k)U_2(x_0 - h_k,\, q_k). \quad \text{(D1)}$$

By the assumption that family x_0 weakly prefers district *j*, the left-hand side must be non-negative. Rearranging terms, this implies that

$$\frac{U_2(x_0 - h_k,\, q_k)}{U_1(x_0 - h_k,\, q_k)} \geq \frac{h_j - h_k}{q_j - q_k} > 0\,. \quad \text{(D2)}$$

Note that the derivative of $U_2(x - h_k,\, q_k)/U_1(x - h_k,\, q_k)$ with respect to *x* is $(U_{21}U_1 - U_{11}U_2)/U_1^2$. As the denominator is always positive, the single crossing property says that $U_2(x - h_k,\, q_k)/U_1(x - h_k,\, q_k)$ is strictly increasing in *x*. If $x > x_0$, then,

$$\frac{U_2(x - h_k,\, q_k)}{U_1(x - h_k,\, q_k)} > \frac{U_2(x_0 - h_k,\, q_k)}{U_1(x_0 - h_k,\, q_k)} \geq \frac{h_j - h_k}{q_j - q_k} > 0. \quad \text{(D3)}$$

An expansion similar to (D1) for family *x* easily establishes that $U(x - h_j,\, q_j) > U(x - h_k,\, q_k)$.

Now suppose that districts *j* and *k* are discretely different. The single-crossing property holds everywhere. Consider family *x*'s indifference curve ($x > x_0$)through (q_j, h_j) in *q*-*h* space. (Refer to Figure D1.) We have shown that this curve passes below $(q_j - \varepsilon,\, h_j - \nu)$ for small ε and ν such that $U(x_0 - h_j + \nu, q_j - \varepsilon) = U(x_0 - h_j, q_j)$. Because it crosses family x_0's indifference curve at (q_j, h_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, h_j). As (q_k, h_k) is one such point by assumption, and as family x_0's curve passes no higher than (q_k, h_k), family *x* must prefer district *j* to *k*.

Proof of Theorem 1.

We 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:

$$\tilde{G}(y) = \begin{cases} j & \text{whenever} \quad F(x) \in \left[1 - {}^{jn}\!/_{N},\, 1 - {}^{(j-1)n}\!/_{N}\right),\ j = 1, \ldots, J; \\ 1 & \text{when} \quad F(x) = 1. \end{cases} \quad \text{(D4)}$$

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 $\tilde{h}_J = 0$. For $j < J$, let

$$\tilde{h}_j = \tilde{h}_{j+1} + (q_j - q_{j+1})\frac{U_2(\breve{x}_j - h_{j+1},\, q_{j+1})}{U_1(\breve{x}_j - h_{j+1},\, q_{j+1})}, \quad \text{(D5)}$$

where $\breve{x}_j \equiv F^{-1}\left(1-\frac{jn}{N}\right)$. (Note that $\breve{x}_j = \inf\left\{x \mid \widetilde{G}(x)=j\right\} = \sup\left\{x \mid \widetilde{G}(x)=j+1\right\}$, by the construction of $\widetilde{G}$.)

I demonstrate that $\widetilde{G}(\cdot)$ and $\left\{\widetilde{h}_1,\ldots,\widetilde{h}_J\right\}$ are an equilibrium. To begin, 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 $\int \mathbf{1}(G(x)=J)f(x)dx < \frac{n}{N}$, the latter a direct result of $1-\frac{Jn}{N}<0$. EQ1 and EQ3 are thus clearly satisfied. What about EQ2? It suffices to show that for each district *j*, the "boundary" family—the family with income $\breve{x}_j$—is indifferent between districts j and j+1. If this is true, Lemma 1 provides that all families in districts $k>j$—who under $\widetilde{G}(\cdot)$ have incomes $x<\breve{x}_j$—will strictly prefer district $j+1$ to *j* under $\widetilde{h}$, while all families in districts $k<j+1$—other than the boundary family—will strictly prefer district j to $j+1$. Since this will be true for all *j*, there cannot be any family who prefers another district to the one to which it is assigned by $\widetilde{G}(\cdot)$.

To demonstrate boundary indifference, plug the housing price equation (D5) into the first-order Taylor expansion of the utility function around $\left(q_j,\widetilde{h}_j\right)$, evaluated at $\left(q_{j+1},\widetilde{h}_{j+1}\right)$:

$$\begin{aligned} U\left(\breve{x}_j-\widetilde{h}_j,\, q_j\right) &\approx U\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right) - \left(\widetilde{h}_j-\widetilde{h}_{j+1}\right)U_1\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right) \\ &\quad + \left(q_j-q_{j+1}\right)U_2\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right) \\ &= U\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right) - \left(q_j-q_{j+1}\right)\frac{U_2\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right)}{U_1\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right)}U_1\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right) \\ &\quad + \left(q_j-q_{j+1}\right)U_2\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right) \\ &= U\left(\breve{x}_j-\widetilde{h}_{j+1},\, q_{j+1}\right). \end{aligned} \tag{D6}$$

All that remains is to demonstrate that EQ4 is satisfied. By definition of $\widetilde{G}(\cdot)$, $\bar{x}_j > \bar{x}_k$ whenever $j<k$, which also implies that $\mu_j > \mu_k$. For any $\delta \ge 0$, then, $\bar{x}_j\delta+\mu_j > \bar{x}_k\delta+\mu_k$, so in particular $\bar{x}_j\delta+\mu_j \ne \bar{x}_k\delta+\mu_k$.

Proof of Theorem 2.

Given EQ1-EQ4, I show that (i) holds if and only if (ii) does; that (i) and (ii) imply (iii) and (iv), and that either (iii) or (iv) implies (i).

By assumption, all families prefer a high-quality community to a low-quality community if there is no extra cost associated with it, and a low-priced community to a high-priced community if there is no loss of quality. Thus, (i) must imply (ii) and vice versa, as no one would live in a low-quality community if houses were no more expensive in a higher-quality community.

Lemma 1 tells us that if any family prefers community *j* to *k* when (i) and (ii) hold, all higher-income families must as well. There cannot, therefore, be any residents of community *k* who have incomes higher than any residents of district *j*, establishing both (iv) and, trivially, (iii). This argument can be reversed: Let x_j be the income of some family in district *j* and x_k the income of some family in *k*, with $x_j > x_k$. If either (iii) or (iv) holds, there must be such a pair. Now suppose that $q_j < q_k$. Then it must be that $h_j < h_k$, else x_j would strictly prefer district *k*. By Lemma 1, however, x_k would also prefer district *j* in this situation. Thus, $q_j > q_k$; equality is ruled out by EQ4.

Proof of Corollary 2.1.

For finite *J*, in any equilibrium there must be one community which has higher quality than any other. Theorem 2 provides that every resident of this community has higher income than any resident of any other community. As Theorem 2 also establishes that the high-quality community has higher housing prices than any other, and as this can only occur when all homes are occupied, the community must contain the *n* highest-income families. By definition of *F*, these are precisely those families with incomes above $F^{-1}\left(1-{}^{n}\!/_{N}\right)$. (As in the main text, I neglect families precisely at the boundary point.)

Now consider the second-ranked district by quality. Again, it has positive prices and higher income families than any district save the highest-ranked district, so must have families with incomes in $\left(F^{-1}\left(1-{}^{2n}\!/_{N}\right), F^{-1}\left(1-{}^{n}\!/_{N}\right)\right)$. The argument proceeds identically for the next-ranked district, and so on to the one of lowest quality.

Proof of Corollary 2.2.

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. (When $\delta > 0$, a high-income population can allow an ineffective school to outrank an effective one.) Corollary 2.1 thus describes the only possible allocation function: The highest-income families must live in the district with the highest μ; the next highest in the next-most effective district; and so on. Moreover, in order to maintain this allocation as an equilibrium, housing prices must keep boundary families indifferent. The price vector described in the proof of Theorem 1 accomplishes this; because $U_1 > 0$, no other price vector can do so.[67] As an equilibrium is completely described by the allocation rule and price vector, it must be unique.

Proof of Theorem 3.

Let $r(j)$ be the quality rank-community index correspondence under an assignment function *G* satisfying the assumptions of the theorem, and define the following housing prices:

$$h_{r(j)} = \begin{cases} 0 & \text{for} \quad j = J \\ h_{r(j+1)} + \dfrac{U_2\left(\breve{x}_j - h_{r(j+1)}, q_{r(j+1)}\right)}{U_1\left(\breve{x}_j - h_{r(j+1)}, q_{r(j+1)}\right)}\left(q_{r(j)} - q_{r(j+1)}\right) & \text{for} \quad j < J \end{cases} \tag{D7}$$

These housing prices, together with *G*, form an equilibrium.

EQ1, EQ3, and EQ4 follow directly from the assumptions. To demonstrate that EQ2 is satisfied by the stated housing prices, it suffices to show boundary indifference. That is, the family with income $\breve{x}_j$ must be indifferent between district $r(j)$ and $r(j+1)$ given *G*'s allocation of peer group and $h_{r(j)}$ and $h_{r(j+1)}$. This is the same result shown in (D6), above; it follows from a direct Taylor expansion of the utility function around family $\breve{x}_j$'s consumption and quality in district $r(j+1)$. Lemma 1 then guarantees that no family in the districts $\{r(1), r(2), \ldots, r(j)\}$ prefers any of

[67] This is where the assumption of extra houses comes in; without it, the lowest-quality district could have positive prices, with a corresponding (but not necessarily identical) shift in each higher-quality district's prices.

the districts $\{r(j+1), r(j+2), \ldots, r(J)\}$ and vice versa. As this must hold for each *j*, EQ2 must be satisfied.

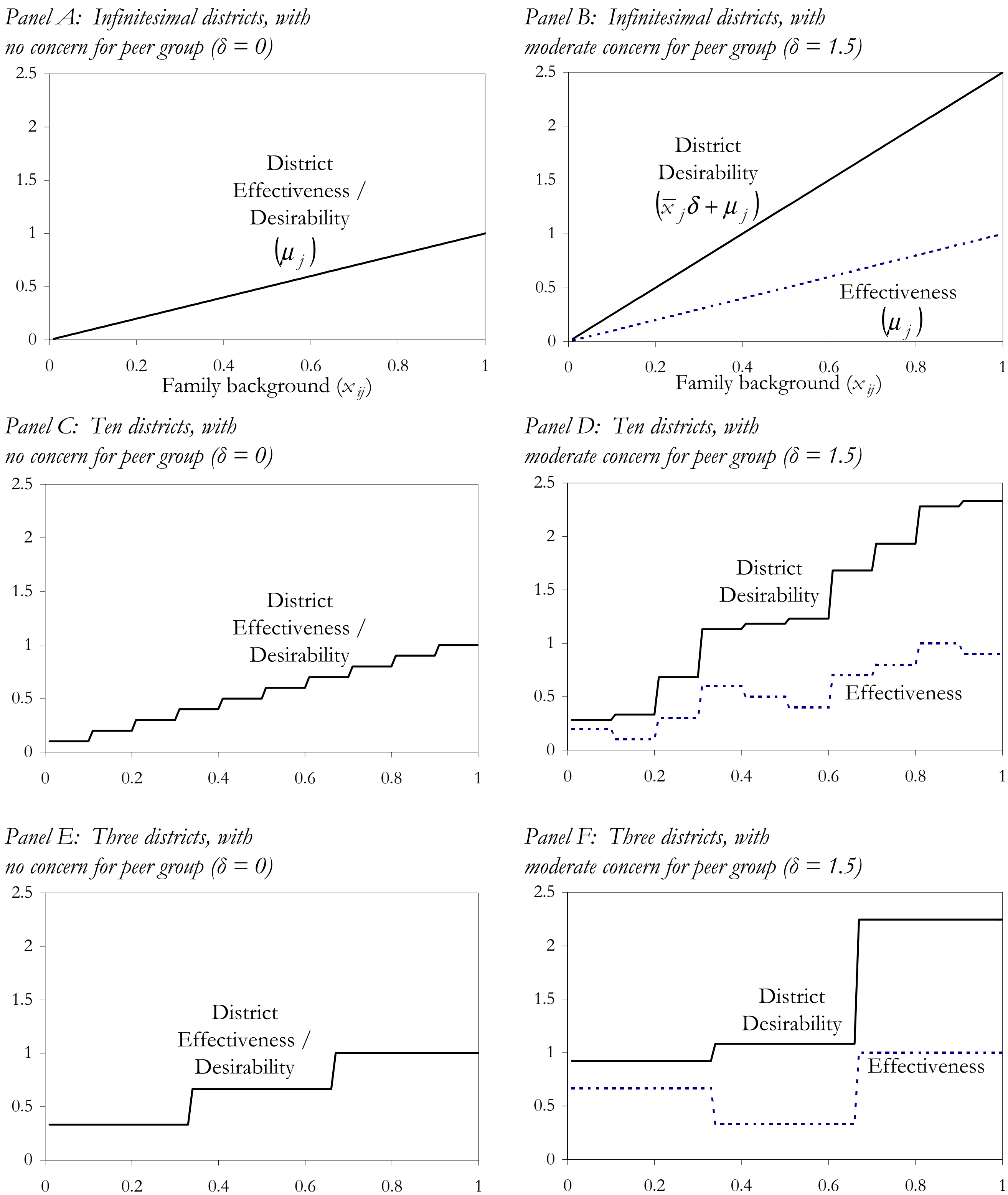
Proof of Corollary 3.1.

Let $\bar{x}_{(j)}$ denote the mean income of the j^{th} bin, and let $\mu_{(j)}$ be the effectiveness of the community to which *G* assigns that income bin. By Theorem 3, *G* is an equilibrium assignment if and only if it assigns higher-income bins to higher-quality communities; that is, if and only if $\bar{x}_{(j)}\delta + \mu_{(j)} > \bar{x}_{(k)}\delta + \mu_{(k)}$ for all *j* and all $k > j$. Note that the latter is equivalent to $\delta > \dfrac{\mu_{(k)} - \mu_{(j)}}{\bar{x}_{(j)} - \bar{x}_{(k)}}$ for all *j* and all $k > j$. Recall that

$$C \equiv \max_{j,k>j} \frac{\mu_{(k)} - \mu_{(j)}}{\bar{x}_{(j)} - \bar{x}_{(k)}} \tag{D8}$$

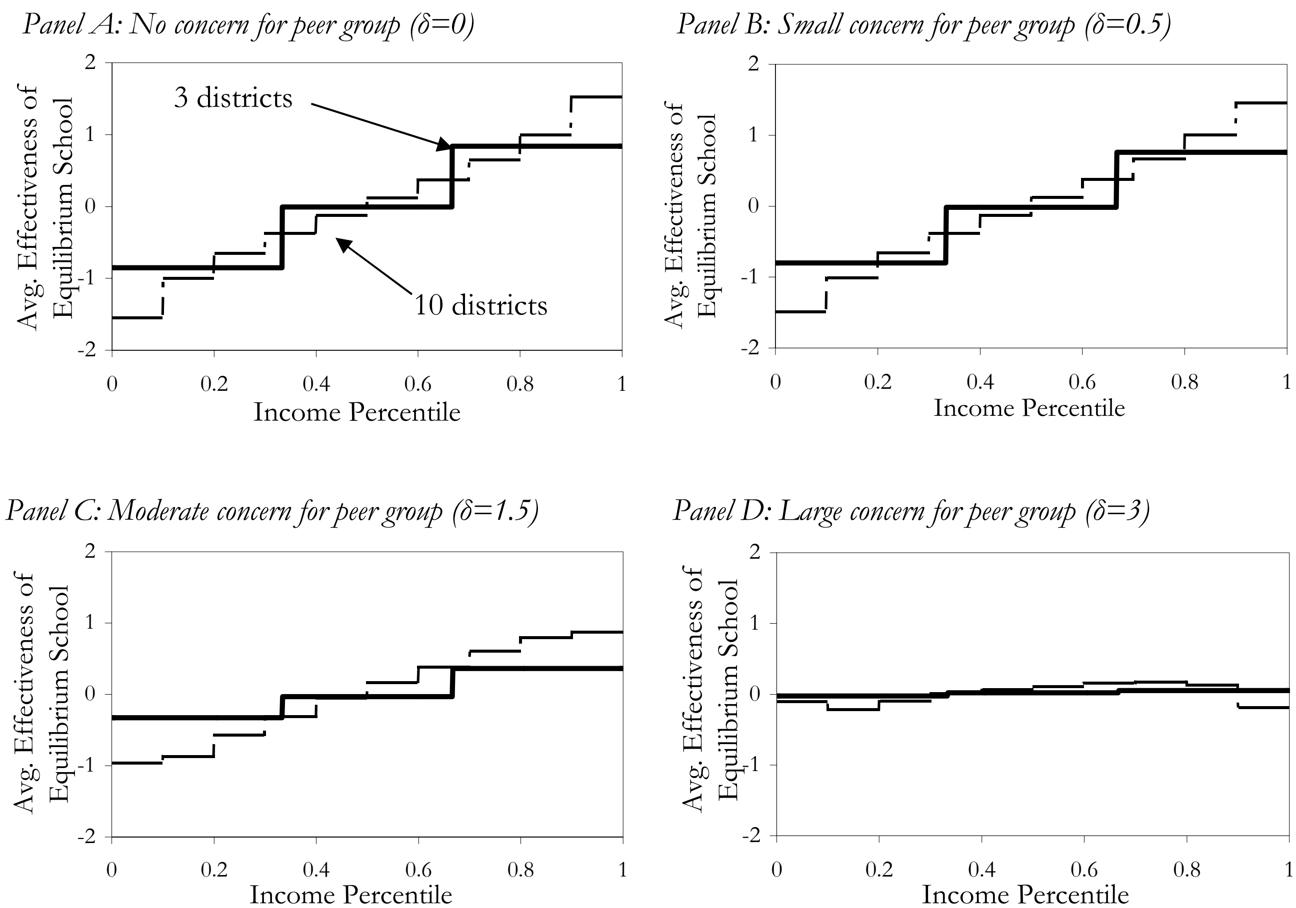
It is immediately clear that when $\delta > C$, assumption (ii) of Theorem 3 is satisfied, so *G* is an equilibrium. Similarly, when $\delta < C$, there exist some *j* and some *k* that violate the equivalence of Theorem 2's (i) and (iii), so *G* cannot be an equilibrium. When $\delta = C$, there are at least two districts for which $\bar{x}_{(j)}\delta + \mu_{(j)} = \bar{x}_{(k)}\delta + \mu_{(k)}$, violating EQ4, but otherwise the argument for Theorem 3 could proceed.

Figure 1.
Schematic: Illustrative allocations of effective schools in Tiebout equilibrium, by size of peer effect and number of districts



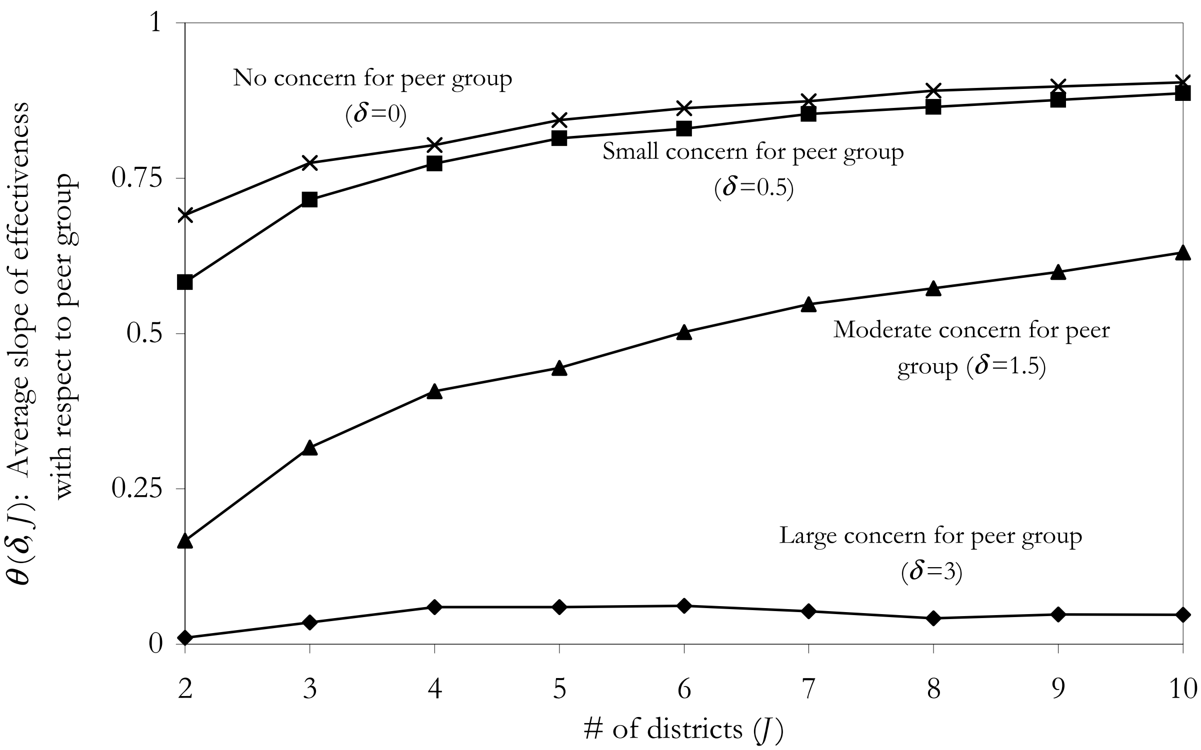
Notes: Each panel illustrates one possible equilibrium in a market characterized by the listed market structure and parental valuations. In each panel, income is uniformly distributed and effectiveness parameters are equally spaced on the [0, 1] interval. See text for details.

Figure 2.
Simulations: Average effectiveness of equilibrium schools in 3- and 10-district markets, by income and importance of peer group



Notes: Each horizontal segment in each figure represents the average of 5,000 draws, where income has a standard normal distribution and effectiveness parameters for each income bin are drawn from the same distribution, then permuted to find an equilibrium assignment. See text for details.

Figure 3.
Simulations: Slope of effectiveness with respect to average income in Tiebout equilibrium, by market structure and importance of peer group



Notes: Each point is the coefficient of a separate "within" regression of school effectiveness (μ) on average income, estimated on 5,000 simulated markets with a fixed effect for each market. See text for details.

Figure 4.
Distribution of district-level choice indices across 318 U.S. metropolitan areas

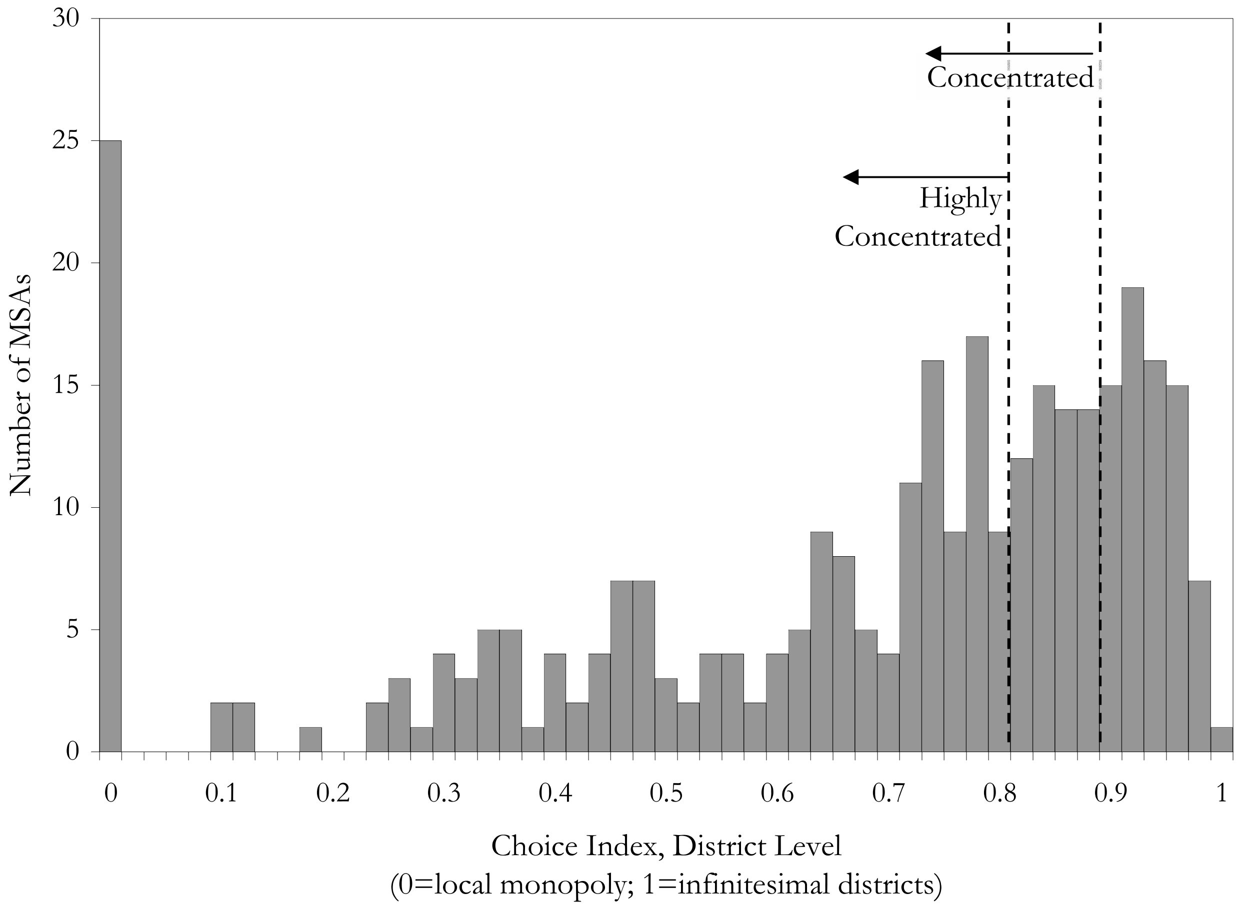
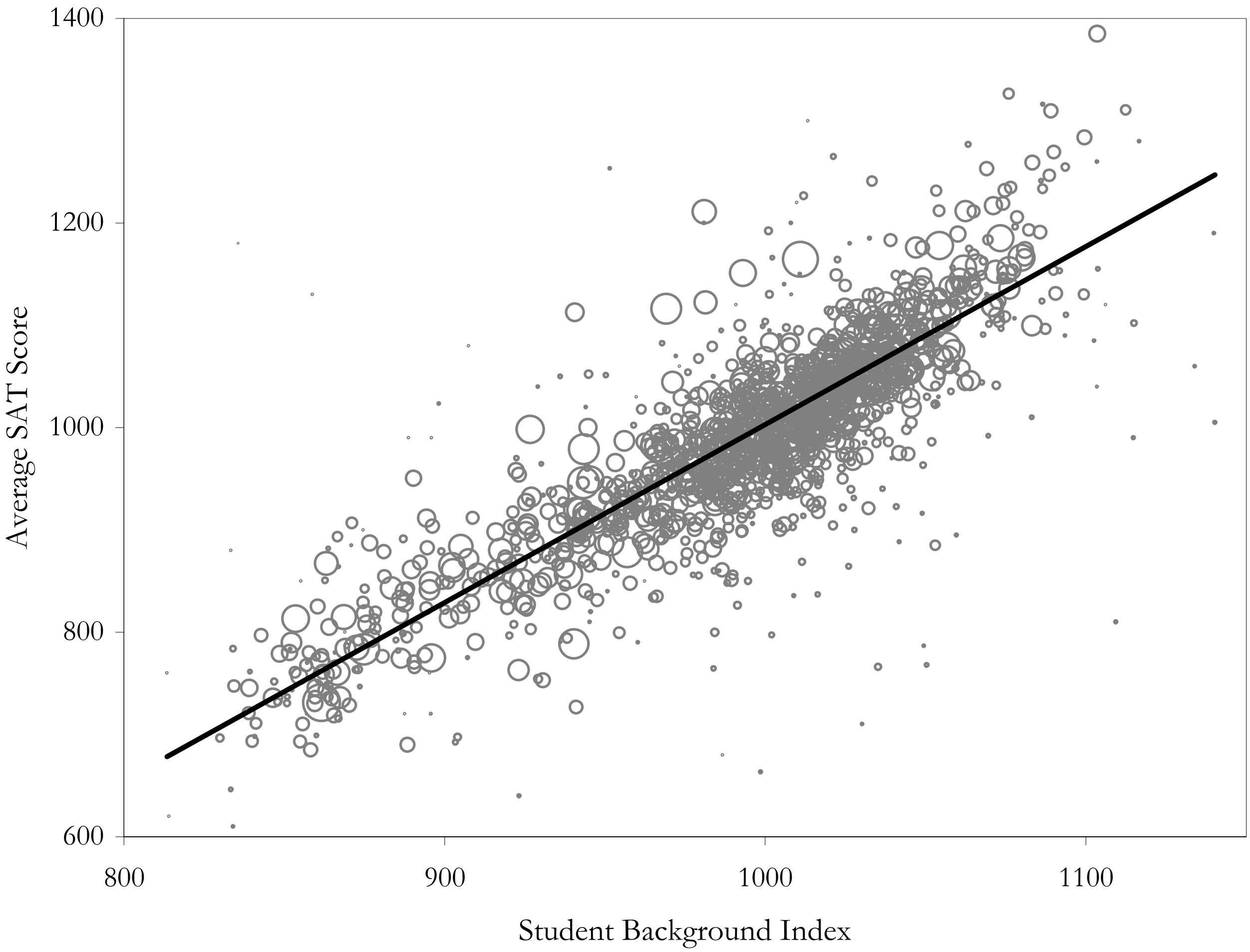
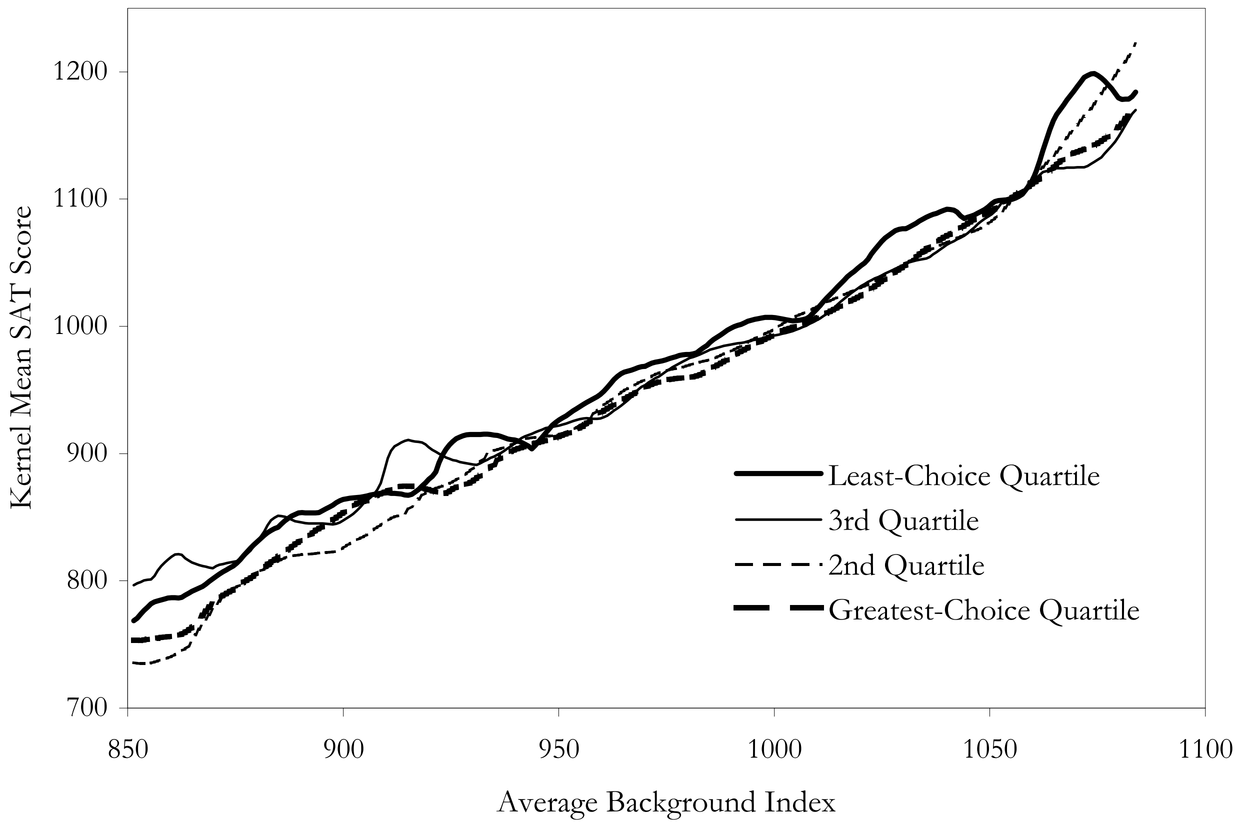


Figure 5.
Student characteristics and average SAT scores, school level



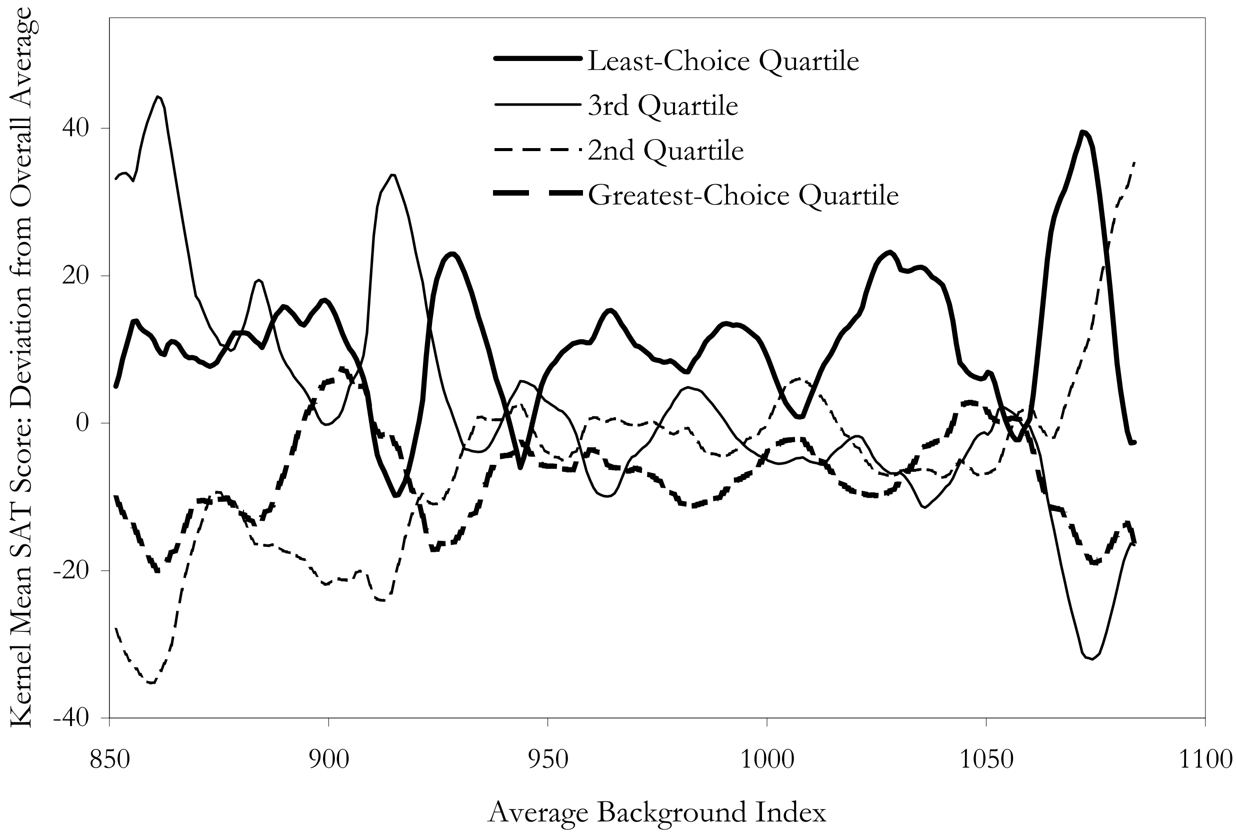
Notes: Each point represents a single school; a randomly selected 25% subsample of schools is shown here. Circle areas are proportional to the sum of SAT-taker weights at the school. The dark line represents a weighted regression on the full sample with fixed effects for 177 MSAs; the line has slope 1.74.

Figure 6.
Nonparametric estimates of the school-level SAT score-peer group relationship, by choice quartile



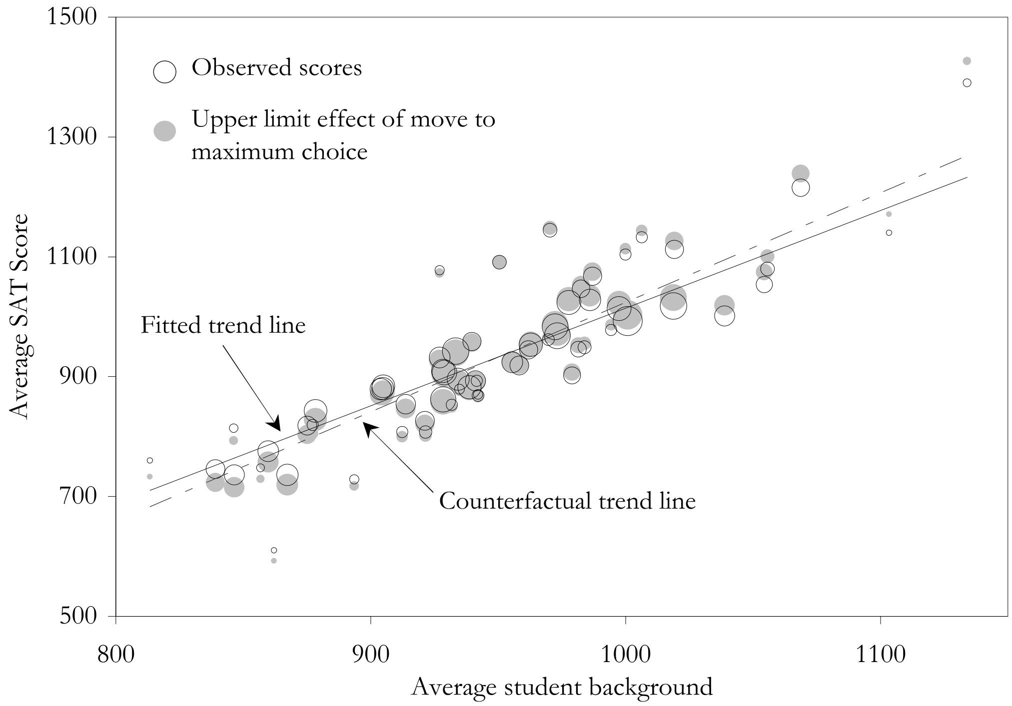
Notes: Figure displays kernel estimates (using an Epanechnikov kernel and a bandwidth of 5 points) of the school-level conditional mean SAT score as a function of the school average background index in each of 4 quartiles of the district-level Tiebout choice index. Schools are weighted by the number of SAT-takers, with weights adjusted so that MSA-level total weights are proportional to 17-year-old populations. Estimates are not displayed for background index values below the first percentile or above the 99th percentile of the school-level distribution.

Figure 7.
Detrended display of nonparametric SAT score-peer group relationship, by choice quartile



Notes: Figure displays kernel estimates from Figure 6, adjusted by subtracting from each point the simple average of the four quartile estimates at that background index value.

Figure 8.
"Upper limit" effect of fully decentralizing Miami's school governance on the across-school distribution of SAT scores



Notes: Hollow circles are observed average SATs at schools in the Miami PMSA; circle areas are proportional to the square root of the number of SAT-takers at the school. "Fitted trend line" represents fitted values from the model in Table 4, Column E. "Counterfactual trend line" represents the fitted values after complete decentralization of Miami school governance (i.e. after the choice index goes from 0 to 1), if the choice-background index interaction effect is assumed to be at the upper limit of the estimated 95% confidence region from that model. Shaded circles represent counterfactual SAT averages for the schools that observed Miami peer groups might attend under these assumptions.

Table 1.
Summary statistics for U.S. MSAs

	All MSAs		Mean by Choice Quartile			
	Mean	S.D.	Least Choice	Q3	Q2	Most Choice
	(A)	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**
Panel A: Basic Descriptive Statistics						
N	318		81	76	83	78
Choice index (district level)	0.66	0.28	0.25	0.66	0.81	0.92
ln(Population)	12.7	1.0	12.1	12.3	12.8	13.5
ln(Mean HH inc.)	10.5	0.2	10.4	10.4	10.5	10.6
Gini coeff., HH inc.	0.36	0.02	0.37	0.36	0.35	0.35
Fraction Black	10%	10%	12%	11%	9%	8%
Fraction Hispanic	7%	13%	9%	7%	7%	5%
Fraction college grads	20%	6%	20%	19%	20%	20%
In foundation plan state	74%		89%	82%	75%	50%
South	38%		65%	47%	30%	12%
Private enrollment share	8%	5%	7%	8%	8%	9%
Panel B: Districts and schools (public, grades 9-12)						
# of districts	14.7	17.9	3.3	7.7	13.5	34.5
# of schools	33.5	40.6	16.5	20.1	38.1	59.1
# of students (thousands)	25.8	40.1	13.4	14.1	30.2	45.3
Average district enrollment	3,053	6,103	6,557	2,247	2,015	1,303
Average school enrollment	709	242	754	692	680	710
Choice index (school level)	0.89	0.08	0.82	0.88	0.92	0.96

Sources: Common Core of Data, 1990; 1990 Decennial Census STF-3C; Card and Payne (1998). Choice quartiles are index values 0-0.5 (Q4); 0.5-0.75 (Q3); 0.75-0.875 (Q2); and 0.875-1 (Q1).

Table 2.
Effect of district-level choice index on income and racial stratification

Dependent Variable:	Across-District Share of Variance, HH Income			White/Non-White Dissimilarity Index (School Level)		
	(A)	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**
Choice	0.07	0.09	0.10	0.15	0.10	0.10
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)
ln(Population) / 100	0.77	1.29	0.22	4.39	2.21	-0.64
	(0.22)	(0.29)	(0.27)	(0.83)	(1.13)	(0.90)
Pop: Frac. Black	0.09	0.09	-0.06	0.33	0.30	0.12
	(0.022)	(0.022)	(0.026)	(0.09)	(0.09)	(0.08)
Pop: Frac. Hispanic	0.02	0.01	0.01	0.03	0.04	0.11
	(0.01)	(0.01)	(0.01)	(0.06)	(0.06)	(0.05)
ln(mean HH income)	0.07	0.07	0.03	0.15	0.17	-0.06
	(0.018)	(0.018)	(0.016)	(0.07)	(0.06)	(0.05)
Gini coeff., HH income	0.36	0.33	0.15	1.88	2.03	0.20
	(0.11)	(0.11)	(0.10)	(0.42)	(0.42)	(0.33)
Pop: Frac. BA+	-0.06	-0.07	-0.05	-0.47	-0.40	0.32
	(0.033)	(0.033)	(0.036)	(0.13)	(0.13)	(0.12)
Foundation plan state / 100	-0.04	-0.04	0.24	-2.95	-3.00	0.24
	(0.40)	(0.39)	(0.35)	(1.60)	(1.58)	(1.15)
School-level choice index		0.10	0.10		-0.37	-0.38
		(0.036)	(0.031)		(0.13)	(0.10)
Census tract- level segregation measures:						
Isolation index (white/non-white)			0.10			-0.27
			(0.03)			(0.09)
Dissimilarity index (white/non-white)			-0.07			1.06
			(0.03)			(0.10)
Across share of variance, education			-0.15			-0.38
			(0.046)			(0.15)
Across share of variance, HH inc.			0.37			0.40
			(0.05)			(0.16)
N	293	293	293	289	289	289
R^2	0.63	0.64	0.74	0.63	0.64	0.82

Notes: Observations are MSAs/PMSAs. Regressions are unweighted. Dependent variable has mean (S.D.) 0.041 (0.038) in columns A-C; 0.413 (0.151) in columns D-F. Columns A through C exclude 25 one-district MSAs. Dissimilarity index is calculated over public and private schools; 29 MSAs in which racial composition is missing for schools with more than 20% of public enrollment are excluded. All columns include fixed effects for nine census divisions.

Table 3.
Summary statistics for SAT sample

	All MSAs		By Choice Quartile			
			Least Choice	Q3	Q2	Most Choice
	(A)	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**
# of observations	329,025		42,286	30,298	117,274	139,167
# of schools	5,727		755	619	1,648	2,705
# of MSAs	177		42	37	47	51
	Mean	S.D.	Mean			
MSA SAT-taking rate	39.5%	10.6%	32.8%	37.8%	38.7%	47.0%
Individual-level						
SAT score	997	201	973	995	997	1004
Black	12%		19%	11%	13%	9%
Hispanic	10%		17%	12%	13%	6%
Asian	9%		8%	4%	14%	7%
Female	54%		56%	55%	55%	54%
Father's education	14.3	2.7	13.9	14.3	14.3	14.3
Mother's education	13.8	2.5	13.6	13.9	13.8	13.9
Family income ($1,000s)	47.0	25.9	40.7	45.6	47.3	48.7
Student background index	997	81	974	997	996	1004
School-level						
# of SAT observations	57	67	56	49	71	51
Sum of SAT weights	102	90	101	89	101	105
S.D., SAT score	171	31	168	170	172	171
S.D., student background	62	16	65	65	65	59

Notes: See text for description of SAT sample. Individual-level measures weight observations by inverse sampling probability. Schools are unweighted for school-level measures. Individual- and school-level standard deviations in Column B are computed over individuals and schools, not over MSA means. Choice quartiles are index values 0-0.5 (Q4); 0.5-0.75 (Q3); 0.75-0.875 (Q2); and 0.875-1 (Q1).

Table 4.
Effect of Tiebout choice on the school-level SAT score-peer group gradient

	(A)	(B)	(C)	(D)	(E)	(F)
Avg. student background index	1.74	1.72	1.49	0.09	-2.35	0.76
	(0.04)	(0.17)	(0.15)	(0.27)	(2.34)	(2.45)
Interaction of student background average with:						
* Choice index		0.02	-0.41	-0.34	-0.09	-0.11
		(0.20)	(0.13)	(0.12)	(0.15)	(0.17)
* MSA SAT-taking rate			2.08	1.94	0.99	1.16
			(0.51)	(0.46)	(0.44)	(0.49)
* ln(Population)				0.09	0.04	0.03
				(0.02)	(0.02)	(0.03)
* Pop: Frac. Black					-0.33	-2.33
					(0.36)	(1.09)
* Pop: Frac. Hispanic					0.03	-1.60
					(0.18)	(0.80)
* ln(mean HH inc.)					0.18	-0.02
					(0.21)	(0.20)
* Gini, HH inc.					3.02	1.89
					(1.71)	(1.90)
* Pop: Frac. BA+					1.56	2.31
					(0.52)	(0.64)
* Foundation plan state					-0.02	-0.03
					(0.06)	(0.05)
* Pop: Frac. White^2						-1.16
						(0.65)
* ln(Density)						0.02
						(0.03)
* Pop: Frac. LTHS						1.14
						(0.82)
* Census division FEs	n	n	y	y	y	y
R^2	0.77	0.77	0.78	0.78	0.78	0.78
R^2, within MSAs	0.74	0.74	0.74	0.75	0.75	0.75

Notes: Sample in each column is 5,727 schools in 177 MSAs. Dependent variable is the weighted mean SAT score at the school. Within MSAs, observations are weighted by the estimated number of SAT-takers at the school (i.e. by the sum of individual sampling weights); these are adjusted at the MSA level to make total MSA weights proportional to the 17-yr-old population. All models include 177 MSA fixed effects, and standard errors are clustered at the MSA level.

Table 5.
Effect of Tiebout choice on the school-level SAT score-peer group gradient: Alternative specifications

	Full Sample				Public Schools Only	Multi-District Markets Only
	(A)	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**
Mean peer quality	-2.35	-2.77	-2.80	1.61	-2.55	-3.31
	(2.34)	(2.25)	(2.27)	(0.07)	(2.58)	(2.32)
Mean peer quality * choice	-0.09	-0.15	-0.11		-0.10	-0.46
	(0.15)	(0.15)	(0.13)		(0.17)	(0.17)
S.D.(peer quality)		0.53				
		(0.08)				
Mean family inc. ($1,000s)				-19.44		
				(7.80)		
Mean family inc. ($1,000s) * choice				-0.16		
				(0.44)		
Peers: Fr. Black			126.5			
			(21.9)			
Peers: Fr. Hispanic			74.1			
			(14.5)			
Peers: Fr. Asian			82.1			
			(22.7)			
Peers: Fr. other race			38.9			
			(20.3)			
N	5,727	5,139	5,727	5,690	4,453	5,476
R^2	0.77	0.78	0.80	0.78	0.80	0.78
R^2, within MSAs	0.74	0.75	0.77	0.75	0.75	0.75

Notes: Dependent variable in all columns is school mean SAT score. All models include 177 MSA fixed effects and controls for interaction of "MSA Characteristics" used in Table 4, Column E with average background index (or, in Column D, with average income). Observations are schools, weighted within MSAs by the sum of individual weights and across MSAs by the 17-year-old population; see text. Standard errors are clustered at the MSA level. Sample size varies due to availability of regressors: S.D.(peer quality) is set to missing when there are 5 or fewer observations; mean family income is calculated over students who report non-missing values. Column E excludes private schools, while Column F excludes 18 MSAs with only a single district.

Table 6.
Effect of Tiebout choice on the school-level SAT score-peer group gradient: Evidence from the NELS and the CCD

	Base	No Controls	Basic Controls	Preferred Controls	Full Controls
	(A)	**(B)**	**(C)**	**(D)**	**(E)**
Panel A: NELS 8th grade score (205 MSAs; 707 schools; 23.3 students per school)					
Avg. student background index	1.90	1.97	2.62	7.76	7.24
	(0.08)	(0.15)	(0.30)	(12.75)	(13.04)
* choice		-0.09	-0.42	-0.57	-0.56
		(0.20)	(0.25)	(0.40)	(0.56)
Panel B: NELS 12th grade score (202 MSAs; 682 schools; 12.1 students per school)					
Avg. student background index	1.47	1.62	2.26	10.65	3.94
	(0.14)	(0.19)	(0.42)	(12.13)	(13.82)
* choice		-0.19	-0.37	0.14	-0.04
		(0.32)	(0.44)	(0.53)	(0.57)
Panel C: NELS 8th-12th grade continuation rate (202 MSAs; 682 schools; 12.1 students per school)					
Avg. student background index / 100	2.53	3.19	3.03	26.46	55.22
	(0.49)	(2.55)	(1.73)	(44.64)	(49.82)
* choice		-0.84	-1.28	-0.19	-0.69
		(3.04)	(1.99)	(2.27)	(2.34)
Panel D: CCD 9th-12th grade completion rate (50 MSAs; 931 school districts)					
Avg. student background index / 1,000	1.99	2.79	5.34	-28.06	-33.35
	(0.21)	(2.45)	(2.00)	(10.59)	(14.37)
* choice		-0.90	-5.43	-7.08	-6.34
		(2.56)	(2.47)	(4.29)	(4.25)

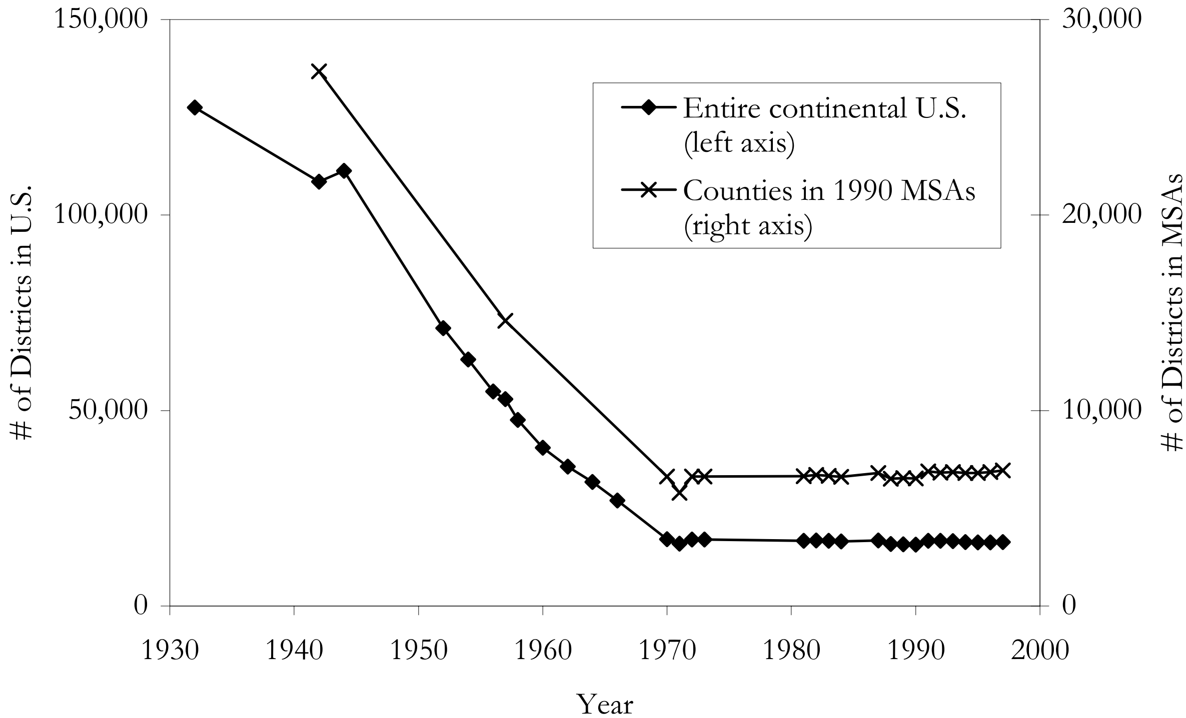
Notes: Specifications are similar to those in Table 4, columns A, B, C, E, and F, although the MSA SAT-taking rate is excluded from all models. All models control for MSA fixed effects and all standard errors are clustered at the MSA level. Sample for Panel A is schools in the original NELS 8th grade sample; Panels B and C restrict sample to those schools with students in the 1988-1992 NELS panel. Student Background Index in Panels A-C is fitted value from a within-school regression of composite test scores (8th grade in A; 12th in B and C) on student race, gender, and parental education measures, averaged to the school level and dropping the school fixed effects. Sample in Panel D is public school districts in SAT-sample MSAs with non-missing completion data (from the Common Core of Data) for at least two thirds of metropolitan enrollment. Student quality in this Panel is the index constructed from the SAT data, averaged over schools in the district.

Table 7.
Effect of Tiebout choice on average SAT scores across MSAs

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Choice index	40.7	36.5	-16.6	-16.1	-26.3	-16.1	-14.1	-13.7
	(9.2)	(10.1)	(5.0)	(5.2)	(5.1)	(5.0)	(5.1)	(5.8)
MSA SAT-taking rate		28.7		-3.5	3.8	-88.4	39.1	25.2
		(27.4)		(13.1)	(19.1)	(17.8)	(68.5)	(72.9)
MSA SAT-taking rate^2							-157.2	-141.9
							(81.6)	(86.3)
Avg. bkgd. index, SAT-takers			1.58	1.58	1.79	1.75	1.78	1.80
			(0.06)	(0.07)	(0.06)	(0.12)	(0.12)	(0.13)
ln(Population)					5.9	0.6	0.1	-0.2
					(0.9)	(1.1)	(1.1)	(1.3)
Pop: Frac. Black						39.9	41.5	50.3
						(22.2)	(22.0)	(41.8)
Pop: Frac. Hispanic						51.0	58.1	63.2
						(14.1)	(14.5)	(29.7)
ln(mean HH inc.)						-4.3	-2.8	-5.0
						(8.6)	(8.6)	(9.8)
Gini, HH inc.						-180.9	-170.2	-178.5
						(61.8)	(61.6)	(70.7)
Pop: Frac. BA+						164.6	162.6	163.8
						(27.0)	(26.8)	(33.1)
Foundation plan state						-3.2	-2.9	-2.8
						(2.4)	(2.3)	(2.4)
Pop: Frac. White^2								4.7
								(24.1)
ln(Density)								1.0
								(1.6)
Pop: Frac. LTHS								2.4
								(33.0)
Census division FEs	n	n	n	n	y	y	y	y
R^2	0.10	0.11	0.80	0.80	0.87	0.93	0.93	0.93

Notes: Dependent variable is the weighted mean SAT score at the MSA level; there are 177 MSAs in the sample. MSAs are weighted by the sum of SAT-taker weights.

Appendix Figure B1.
Number of School Districts Over Time



Sources:

Statistics of state school systems , 1966: 1932, 1944, 1952, 1954, 1956, 1958, 1962, 1964, 1966

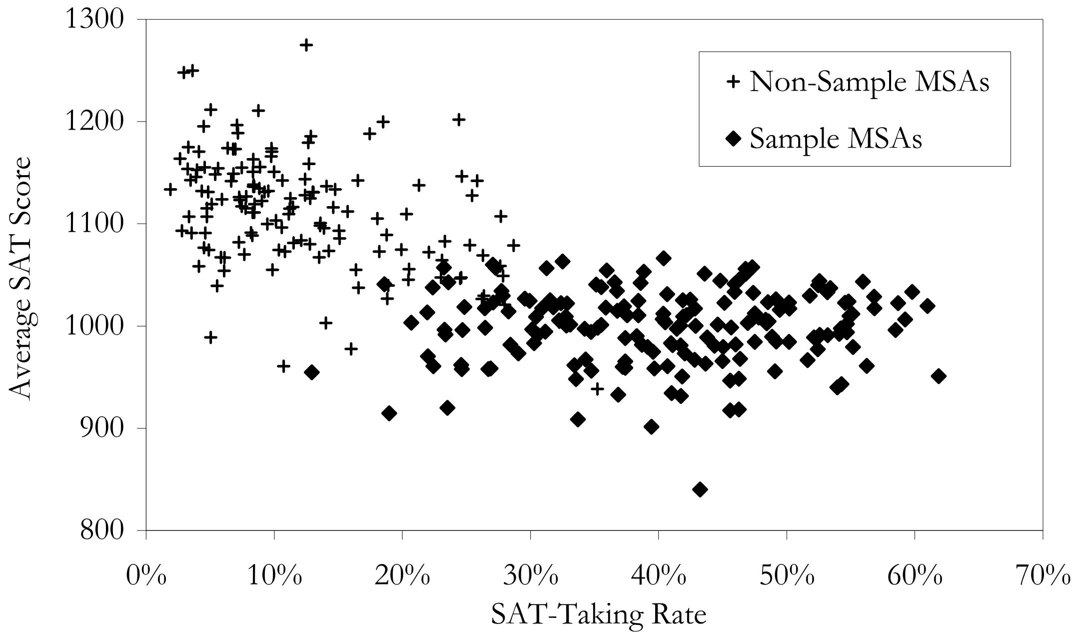
Gray, E.R., 1944, *Governmental Units in the United States 1942:* 1942

Governments in the United States 1957: 1957

Elsegis electronic file, ICPSR #2238: 1969, 1970, 1971, 1972

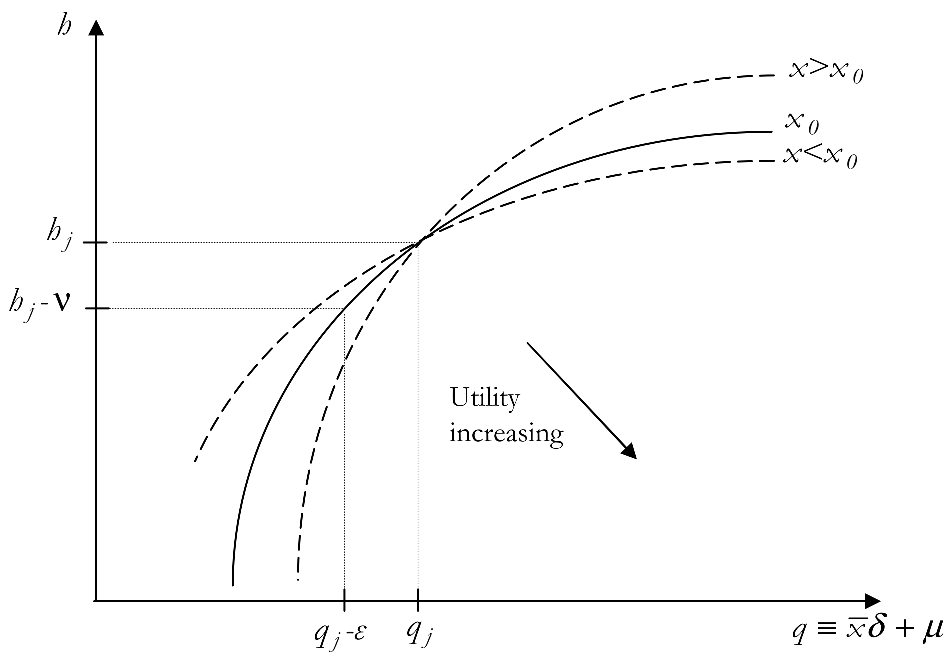
Common Core of Data: 1981 forward

Appendix Figure C1.
SAT-taking rates and average SAT scores across MSAs



Notes: Sample MSAs are those used in main analysis (i.e. those in states with SAT-taking rates above one third). Honolulu and Anchorage MSAs are excluded.

Appendix Figure D1.
Illustration of single-crossing: Indifference curves in q-h space



Appendix Table A1.
Evidence on choice-stratification relationship: Additional measures

Dependent Variable:	Across-District Share of Variance: Adult Educ.		School-Level White/Non-White Isolation Index		Theil Segregation Measure: Across Schools	Theil Segregation Measure: Across Districts	Theil Segregation Measure: Across Schools, Within Districts
	(A)	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**	**(G)**
Choice	0.08	0.10	0.07	0.06	0.06	0.26	-0.14
	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
ln(Population) / 100	0.05	0.53	4.27	3.56	0.60	-0.18	1.73
	(0.25)	(0.34)	(0.80)	(1.10)	(0.76)	(0.74)	(0.50)
Pop: Frac. Black	0.03	0.03	0.81	0.80	0.19	-0.07	0.07
	(0.03)	(0.03)	(0.09)	(0.09)	(0.07)	(0.07)	(0.05)
Pop: Frac. Hispanic	0.04	0.03	0.07	0.08	0.06	0.05	-0.01
	(0.02)	(0.02)	(0.06)	(0.06)	(0.04)	(0.04)	(0.03)
ln(mean HH income)	0.02	0.02	0.29	0.29	-0.05	-0.12	-0.01
	(0.02)	(0.02)	(0.06)	(0.06)	(0.04)	(0.04)	(0.03)
Gini coeff., HH income	0.50	0.46	1.74	1.79	0.35	-0.15	0.28
	(0.13)	(0.13)	(0.41)	(0.41)	(0.28)	(0.28)	(0.19)
Pop: Frac. BA+	0.22	0.21	-0.47	-0.44	0.26	0.42	-0.03
	(0.04)	(0.04)	(0.12)	(0.12)	(0.10)	(0.10)	(0.07)
Foundation plan state / 100	0.17	0.17	-3.27	-3.28	0.46	1.10	0.40
	(0.47)	(0.46)	(1.53)	(1.53)	(0.96)	(0.93)	(0.63)
School-level choice index		-0.07		0.17	0.24	-0.15	0.24
		(0.04)		(0.10)	(0.08)	(0.08)	(0.06)
Census tract- level segregation measures:							
Isolation index (white/non-white)		0.06		0.50	0.13	0.45	-0.07
		(0.03)		(0.10)	(0.08)	(0.08)	(0.05)
Dissimilarity index (white/non-white)		-0.04		0.27	0.47	0.25	0.15
		(0.04)		(0.11)	(0.08)	(0.08)	(0.06)
Across share of variance, education		0.52		-0.44	-0.36	-0.51	0.04
		(0.05)		(0.16)	(0.13)	(0.12)	(0.08)
Across share of variance, HH inc.		-0.16		0.08	0.07	0.05	0.06
		(0.05)		(0.16)	(0.13)	(0.13)	(0.09)
N	293	293	289	289	289	264	264
R^2	0.48	0.62	0.65	0.79	0.78	0.81	0.62

Notes: Observations are unweighted MSAs/PMSAs. Columns C-G exclude MSAs missing racial composition data for more than 20% of public enrollment. Columns A, B, F, and G exclude MSAs with only one district. See Theil (1972) for description of the Theil segregation measure, which is calculated over all schools in column E and over public districts and schools in F and G. All columns include fixed effects for 9 census divisions.

Appendix Table A2.
Alternate measures of Tiebout choice: Effects on segregation and stratification

	School-Level Racial Segregation			Across-District Share of Variance	
	Isol. Index	Dissim. Index	Theil Measure	Income	Education
	(A)	**(B)**	**(C)**	**(D)**	**(E)**
Tiebout Choice Measure					
District-level choice index	0.10	0.16	0.11	0.08	0.08
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Number of districts (00s)	0.15	0.15	0.16	0.09	0.06
	(0.04)	(0.04)	(0.03)	(0.01)	(0.01)
Districts per 17-yr-old population (* 10)	0.59	0.91	0.77	0.25	0.34
	(0.30)	(0.31)	(0.25)	(0.10)	(0.11)

Notes: Each entry is the coefficient on a single choice measure from a distinct MSA-level regression, with control variables as in Table 2, column C (except that the school-level choice index is excluded and population is entered here in levels rather th

Appendix Table A3.
Effect of district-level choice on tract-level income and racial stratification

Dependent Variable:	Tract-Level Racial Segregation		Across-Tract Share of Variance	
	Dissimilarity	Isolation	Income	Education
	(A)	**(B)**	**(C)**	**(D)**
Choice	0.00	-0.03	-0.02	0.01
	(0.02)	(0.02)	(0.01)	(0.01)
ln(Population) / 100	3.51	4.39	2.51	1.24
	(0.65)	(0.70)	(0.28)	(0.26)
Pop: Frac. Black	0.32	0.75	0.27	0.11
	(0.07)	(0.07)	(0.03)	(0.03)
Pop: Frac. Hispanic	-0.03	0.00	0.05	0.12
	(0.04)	(0.05)	(0.02)	(0.02)
ln(mean HH income)	0.31	0.41	0.08	0.02
	(0.05)	(0.06)	(0.02)	(0.02)
Gini coeff., HH income	2.25	2.36	0.66	0.71
	(0.33)	(0.36)	(0.14)	(0.13)
Pop: Frac. BA+	-0.75	-0.77	0.15	0.37
	(0.10)	(0.11)	(0.04)	(0.04)
Foundation plan state / 100	-4.03	-3.15	-0.50	0.31
	(1.27)	(1.36)	(0.54)	(0.50)
N	318	318	318	318
R^2	0.66	0.71	0.70	0.68

Notes: Observations are MSAs/PMSAs, unweighted. Each model includes fixed effects for 9 census divisions.

Appendix Table B1.
First stage models for MSA choice index

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Instruments							
# of streams/1000		0.32					0.01
		(0.08)					(0.06)
County choice index			0.41			0.19	0.18
			(0.05)			(0.04)	(0.05)
Est. 1942 choice index				0.62		0.50	0.50
				(0.05)		(0.05)	(0.05)
County-district state indic.				-0.08		-0.05	-0.05
				(0.04)		(0.04)	(0.04)
Avg. choice index,					0.49	0.17	0.17
rest of state					(0.07)	(0.06)	(0.06)
Controls							
ln(Population)	0.13	0.09	0.05	0.09	0.13	0.06	0.06
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Pop: Frac. Black	0.07	0.23	0.10	-0.14	-0.12	-0.14	-0.14
	(0.17)	(0.17)	(0.16)	(0.13)	(0.16)	(0.13)	(0.13)
Pop: Frac. Hispanic	-0.16	0.01	0.08	-0.19	-0.22	-0.10	-0.09
	(0.11)	(0.12)	(0.11)	(0.09)	(0.11)	(0.09)	(0.09)
ln(mean HH inc.)	-0.40	-0.28	-0.25	-0.13	-0.30	-0.08	-0.08
	(0.13)	(0.13)	(0.12)	(0.10)	(0.12)	(0.10)	(0.10)
Gini, HH inc.	-2.88	-3.16	-2.80	-1.29	-2.36	-1.38	-1.38
	(0.84)	(0.82)	(0.76)	(0.64)	(0.79)	(0.62)	(0.63)
Pop: Frac. BA+	0.28	0.22	0.27	-0.18	0.14	-0.15	-0.15
	(0.26)	(0.25)	(0.23)	(0.19)	(0.24)	(0.19)	(0.19)
Foundation plan state	0.01	0.01	-0.01	0.00	0.02	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
N	318	318	318	318	315	315	315
R^2	0.51	0.54	0.60	0.73	0.58	0.75	0.75
F statistic, exclusion of instruments		17.7	64.3	122.0	54.1	72.2	57.6

Sources: Electronic Geographic Names Information System (Streams); 1990 Census STF-3C (County choice); Gray, 1944 (1942 choice index); Kenny and Schmidt, 1994 (County Districts); author's calculations.

Notes: Dependent variable is the district-level choice index. Observations are MSAs. All columns include fixed effects for 9 census divisions. Columns E, F, and G exclude 3 MSAs for which there are no other MSAs in the same state.

Appendix Table B2.
2SLS Estimates of Effect of Tiebout Choice

Model:	Across-District Share of Variance, HH Income	Dissimilarity Index	SAT Score-Peer Group Gradient	Avg. SAT Score
Source Table, Specification	Table 2 , Col. C	Table 2 , Col F.	Table 4, Col. E	Table 7, Col. G
	(A)	**(B)**	**(C)**	**(D)**
OLS	0.10	0.10	-0.09	-14.1
	(0.01)	(0.02)	(0.15)	(5.1)
2SLS				
Streams	0.13	0.17	-0.27	**-55.9**
	(0.10)	(0.14)	(0.36)	(21.3)
County choice	0.08	0.02	0.14	-18.7
	(0.06)	(0.08)	(0.40)	(15.1)
Historical (1942 choice	0.06	0.08	0.17	-6.1
+ county districts)	(0.03)	(0.03)	(0.25)	(7.3)
Rest of state	0.16	0.16	1.27	-35.0
	(0.06)	(0.08)	(1.30)	(36.7)
All but streams	0.07	0.07	0.12	-5.7
	(0.02)	(0.03)	(0.25)	(7.2)
All	0.07	0.07	0.02	-9.9
	(0.02)	(0.03)	(0.23)	(7.0)

Notes: Each entry represents the coefficient on the district-level choice index (or, in Column C, on the interaction between that index and the peer group background index) from a separate regression. Specifications are the same as the OLS specification listed at top, but are estimated by instrumental variables. Bold coefficient indicates that a Hausman test rejects equality of the 2SLS and OLS choice coefficients at the 5% level.

Appendix Table C1.

Sensitivity of individual and school average SAT variation to assumed selection parameter

	Correlation between actual and selection-adjusted value	
	Individual SAT score	School average SAT score
	(A)	**(B)**
Assumed selection parameter		
$\rho = 0.05$	1.000	0.999
$\rho = 0.1$	0.999	0.998
$\rho = 0.25$	0.996	0.987
$\rho = 0.5$	0.983	0.956
$\rho = 0.75$	0.956	0.910
$\rho = 0.9$	0.930	0.873

Notes: Entries in table represent cross-sectional correlation between observed score (or average score) and that obtained by adjusting scores using the school-average SAT-taking rate and within-school selectivity described by the listed parameter. Obser

Appendix Table C2.

Stability of school mean SAT score and peer group background characteristics over time

	1994	1995	1996	1997	1998
1994		0.906	0.908	0.902	0.899
1995	0.957		0.912	0.908	0.909
1996	0.957	0.961		0.918	0.915
1997	0.955	0.959	0.963		0.921
1998	0.952	0.957	0.961	0.963	

Notes: Entries above diagonal represent correlations across years in schools' average SAT scores. Entries below diagonal are correlations of school peer group background index values.

Appendix Table C3.
Effect of Tiebout choice on the school-level SAT score-peer group gradient: Estimates from class rank-reweighted sample

	(A)	(B)	(C)	(D)	(E)	(F)
Avg. student background index	1.79	1.70	1.40	-0.14	-5.18	-2.66
	(0.04)	(0.19)	(0.16)	(0.24)	(2.51)	(2.79)
Interaction of student background average with:						
* Choice index		0.11	-0.40	-0.32	-0.09	-0.07
		(0.22)	(0.15)	(0.12)	(0.17)	(0.18)
* MSA SAT-taking rate			2.19	2.03	1.18	1.25
			(0.52)	(0.45)	(0.46)	(0.49)
* ln(Population)				0.10	0.05	0.05
				(0.02)	(0.02)	(0.03)
* Pop: Frac. Black					-0.45	-2.37
					(0.37)	(1.28)
* Pop: Frac. Hispanic					0.02	-1.47
					(0.20)	(0.94)
* ln(mean HH inc.)					0.42	0.28
					(0.23)	(0.23)
* Gini, HH inc.					3.20	2.88
					(1.56)	(1.77)
* Pop: Frac. BA+					0.77	1.12
					(0.56)	(0.69)
* Foundation plan state					0.02	0.01
					(0.07)	(0.06)
* Pop: Frac. White^2						-1.17
						(0.76)
* ln(Density)						0.01
						(0.03)
* Pop: Frac. LTHS						0.39
						(0.88)
* Census division FEs	n	n	y	y	y	y
R^2	0.78	0.78	0.79	0.79	0.80	0.80
R^2, within MSAs	0.75	0.75	0.76	0.76	0.76	0.76

Notes: Sample in each column is 5,671 schools in 177 MSAs. Dependent variable is the weighted mean SAT score at the school, with weights adjusted using students' self-reported rank in class to balance the first and second deciles and second and third quintiles within the school; students not reporting a class rank or reporting a rank in the bottom 40% are dropped. Within MSAs, schools are weighted by the number of twelfth grade students; these are adjusted at the MSA level to make total MSA weights proportional to the 17-yr-old population. All models include 177 MSA fixed effects, and standard errors are clustered at the MSA level.