Does Competition Among Public Schools Benefit Students and Taxpayers?

By Caroline M. Hoxby*

**Tierbuck choice among districts is the most powerful market force in American public education. Naive estimates of its effects are biased by endogenous district formation. I derive instruments from the natural boundaries in a metropolitan area. My results suggest that metropolitan areas with greater Tiebout choice have more productive public schools and less private schooling. Little of the effect of Tiebout choice works through its effect on household sorting. This finding may be explained by another finding: students are equally segregated by school in metropolitan areas with greater and lesser degrees of Tiebout choice among districts. (JEL H70, I20)**

Many proposed reforms for elementary and secondary schooling in the United States share a common driving force: increased parental choice. These reforms include intradistrict choice, interdistrict choice, vouchers for private schools, and charter schools. It might seem that such reforms would propel American schools into wholly unknown territory, where proponents hope that competition would improve schools and opponents fear that students would sort themselves among schools in a way that would impair the educational prospects of some students. In fact, this territory is not wholly unknown. The reforms extend the traditional method of school choice in the United States—that which takes place when households make residential choices among local school districts. This choice process has long been considered the primary example of the Tiebout process, whereby residential choices determine the quality of, and expenditures on, local public goods. Although this consolidation and states’ school finance programs increasingly limit the Tiebout process, it is still the most powerful force in American schooling. In this paper, I attempt to show the effects of this type of school choice. My goals are to shed light on the system we have and to demonstrate general properties of school choice that are helpful for thinking about reforms.

With any form of school choice, there are potential trade-offs. Choice may allow students to sort in a manner that impedes the learning of at least some children. On the other hand, choice may intensify the competitive mechanism that rewards schools with high productivity (high student achievement per dollar spent). Choice may also allow students to sort in a manner that facilitates learning—for instance, a disabled child may be able to attend a school that has an especially good program for disabled children.

In order to analyze the effects of choice on school productivity and sorting of students, I explore some related questions. Are parents less likely to send their children to private schools when they have more choice among public schools? Do states’ school finance programs that weaken the financial implications of the Tiebout process lessen the effects of choice? The last question is important because we need to know whether the effects of Tiebout choice depend on its financial implications or just on parents being better able to match their children with schools.

Empirical work on Tiebout choice is important precisely because theory—which is discussed in Section II—does not provide us with much guidance. The theoretical predictions can be briefly summarized as follows. The incentives that schools have to be productive are generally increased by Tiebout choice because it gives households more information and leverages in the principal-agent problem that exists between them and the people who run their local schools. Self-sorting of students is generally increased by Tiebout choice, and people sort themselves so as to maximize private allocative efficiency (their own welfare). Self-sorting may produce poor social allocative efficiency, however. Each school may be more productive given its student body but students may be sorted so that good parents are not in contact with the students who would benefit from them most. Theory is ambiguous as to whether Tiebout choice will increase or decrease total spending on schools.

The empirical focus of this paper is not only used to illustrate these points, but also to document the empirical challenges. The first challenge is creating satisfactory measures of Tiebout choice. The second is identifying variation in Tiebout choice that is driven by exogenous factors that affect the supply of school districts. I attempt to exclude variation that is endogenous to observed student achievement or that is driven by the demand for school districts. I use instrumental variables based on topographic (specifically, streams) to identify natural differences in areas’ propensity to have numerous school districts. The final challenge is distinguishing between competition and sorting. For instance, a school district may be highly productive either because it has strong incentives to be efficient or because its students are self-sorted so that only one instructional method is required. To meet this challenge, I employ data not only on the average characteristics of people in each district, but also on the heterogeneity of each district.

### I. The Importance of Analyzing Tiebout Choice

Analysis of Tiebout choice is important because its long history and widespread application allow us to understand the general equilibrium effects of choice among schools. It will be years before any reform could have the pervasive effects that Tiebout choice has had on American schooling. Moreover, the short-term effects of reforms are misleading because they depend unduly on the students who actively make a choice in the years immediately following reform. Even if reforms did not require these students to be few and atypical, their experience would be unrepresentative because the supply response to a reform—the entry or expansion of successful schools and the shrinking or exit of unsuccessful schools—may take a decade or more to fully evolve itself. Until many students experience an increased degree of choice, reforms are unlikely to affect public schools much, either through competitive pressure or through sorting. We are mainly interested in the new general equilibrium that would exist if choice-based reforms were widely enacted, not in the partial effect on the students who are the first to take advantage of a reform.

The second reason it is important to understand Tiebout choice is that reforms are layered on top of the existing system. Most reforms would not substitute, not improve, choice. If one ignores Tiebout choice, one neglects the fact that some of the predicted effects of reforms are attained by Tiebout choice already. For instance, one is likely to miscalculate the distributional consequences of reforms if one neglects the fact that some people, such as the wealthy, already have a high degree of choice so that reforms would hard affect their behavior. Other people would have their choice sets greatly expanded by reforms. Oddly enough, many analyses of reforms ignore Tiebout choice. For instance, some influential analyses of vouchers assume that there is only one large public school in which all households participate in the absence of vouchers.

The third reason to analyze Tiebout choice is that it is a practical one. Identifying the effects of choice, as opposed to the causes and correlates of choice programs, is very difficult except in the case of Tiebout choice. For example, when a

---

* Department of Economics, Harvard University, Cambridge, MA 02138. The author gratefully acknowledges generous help from the anonymous referee, Henry S. Farber, Jonathan Gruber, Lawrence F. Katz, D. Daniele Faerber, James M. Poterba, and seminar participants at several universities.

1 The seminal article is by Charles M. Tiebout (1956). Daniel L. Rubinfeld (1987) surveys the extensive theoretical and empirical literature related to Tiebout's model. In the text, I cite a number of articles in this literature that are particularly relevant to choice among public-school districts, but Melvin V. Borland and Roy M. Howsen (1992) and Charles F. Manski (1992) initiated recent commentary on this topic.

2 Lawrence W. Kenay and Amy B. Schmidt (1994) describe the decline in the number of school districts in the United States.

school district enacts a policy of intradistrict school choice, the policy grows out of the district's circumstances and often is part of a package of policies enacted simultaneously. As a result, such policies are difficult to evaluate. With Tiebout choice, the identification problem is more manageable because the Tiebout choice in an area largely depends on historical circumstances that are arbitrary with respect to modern schooling. Nevertheless, I try to remedy potential identification problems by using variation in school districting that is literally natural.

II. What Theory Predicts About Tiebout Choice

Tiebout choice can affect private allocative efficiency, social allocative efficiency, and schools' productivity. Informal and formal versions of the Tiebout model demonstrate that private allocative efficiency tends to be increased by Tiebout choice, even when it is combined with political shortcomings, such as voting on local property tax rates. Intuitively, when there are more school districts, it is easier for households to sort themselves into groups that are relatively homogeneous in terms of their preferences with regard to schooling and property. As a result, an equilibrium in which households get schools close to what they privately prefer is more likely to exist. The more school districts there are, the less troublesome are free-rider problems, which tend to make Tiebout equilibrium break down.

What does the private allocative efficiency result imply? First, to the extent that greater sorting reduces the degree to which households pay for school programs they do not value, Tiebout choice raises the amount of school quality that households want to buy. Also, if households achieve greater private allocative efficiency in the public schools, they will be less inclined to send their children to private schools. Under greater Tiebout choice, households will be more sorted among districts on the basis of their preferred type of schooling and level of school spending. In part, this implies more sorting on the basis of income. But, it also implies more sorting on the basis of tastes for (or ability to benefit from) education. For instance, a school district might end up with a combination of households—some of whom are richer and want to spend a small share of their incomes on education, and others of whom are poorer but want to spend a large share of their incomes on education. Also, a school district might end up with a group of households who share a taste for progressive curricula. To the extent that greater sorting improves match quality between students' needs and schools' offerings, Tiebout choice will raise average student achievement.

The increase in private allocative efficiency generated by Tiebout choice does not necessarily correspond to an increase in social allocative efficiency if there are human-capital spillovers among students or neighbors. Tiebout choice takes no account of such spillovers, since it depends on people making choices that are privately optimal. However, an equilibrium in which more learned students are self-segregated may be socially inefficient if forcing more learned students into contact with less learned students would raise social welfare. Unfortunately, empirical evidence on the nature of human-capital spillovers is very poor, so that we do not even know whether contact between better and worse students raises the achievement of all students, degrades the achievement of all students, reduces the initial achievement differences among students, or exacerbates the initial achievement differences among students. Thus, we do not know whether social allocative efficiency always increases, sometimes increases, or never increases as private allocative efficiency rises.

Theory generally predicts that Tiebout choice raises schools' productivity. Schooling producers can earn rent if households have difficulty observing producers' effort, verifying the quality of schooling inputs (especially student ability), and verifying schooling outcomes. This principal-agent problem is alleviated by Tiebout choice. Intuitively, school budgets based on property taxes form a mechanism that naturally incorporates many households' private observations of schooling outcomes and schooling inputs (including their children's abilities). Although only a minority of households are on the move at any given time, their observations about schools determine property prices in all school districts, and schooling producers who take excessive rent are penalized by reductions in their school budgets. This system works better when there are more Tiebout choice because changes in property prices depend more on information about schools and less on other factors that are essentially noise.

Theory is ambiguous as to whether Tiebout choice will increase or decrease total spending on schools. The productivity prediction discussed above means that a unit of achievement should cost less where there is greater choice. As a result, one expects households to purchase more achievement, but one cannot predict whether they will spend more or less in total. Moreover, in areas with little Tiebout choice, asset-rich households live in districts with asset-poor households. On the one hand, this tends to depress the school spending preferred by asset-rich house- holds because they pay a disproportionately large share of every dollar that is spent. On the other hand, this tends to increase the school spending preferred by asset-poor households. The net effect depends on the political mechanism and on the preferences of asset-rich versus asset-poor households. For instance, if asset-poor households have a demand for school spending that does not elastically increase when their cost of a dollar of school spending falls, then raising the cost of a dollar of school spending for asset-rich households while lowering the cost for asset-poor households may decrease the amount spent on education.

Whether Tiebout choice affects productivity by making school districts fall below minimum efficient scale is, in practice, a nonissue. Empirical evidence suggests that inefficiencies in cities close to a school district is so small that most metropolitan districts easily exceed it. Also, temporary or partial consolidation between mutually agreeable districts is easy, while breaking up too-large is hard. For instance, districts often retain separate elementary systems while sharing a high school. These arrangements are reversible, and it is not uncommon to see districts dispense with a sharing arrangement when their populations are large enough to allow each to achieve minimum efficient scale on its own.

III. An Empirical Version of the Theoretical Predictions

Theory suggests that Tiebout choice indirectly affects sorting and the incentives to be productive that schools face. Tiebout choice indirectly affects productivity, achievement, school spending, and private schooling. In this section, I present an empirical model that summarizes the theoretical predictions. In subsequent sections, I consider problems like identification, aggregation, and how to measure Tiebout choice. For now, assume that choice is measured accurately and that all variation in choice is exogenous. Let I index individual students and their associated households. Let k index school districts, and let m index educational markets. (An educational market is the set of school districts in which a household could reside, given its employment situation. Below, I argue that metropolitan areas are reasonable concrete versions of educational markets.) Let \( C_m \) measure how much Tiebout choice exists in educational market \( m \), where larger \( C_m \) means greater choice. Let \( r_m \) be the reward the market gives to administrators who improve productivity. The prediction that choice creates greater rewards for schooling producers who run schools efficiently can be written:

\[
\frac{\partial r_m}{\partial C_m} > 0.
\]

(1)

In Hoxby (1999a) offers a formal principal-agent model of the productivity of schooling producers. This is a well-known result that has recently been explored but may also be interpreted as evidence for increased school finance equalization. See Falbo Silva and Jon Sonstebl (1995) and Raúl Fernández and Richard Rogerson (1998).

(8) Randell Bénabou (1996) has recently drawn attention to the macroeconomic implications of this familiar point.
This prediction can be tested only indirectly—by looking at the relationship between choice and productivity and trying to eliminate the effect of sorting.

The second prediction is that choice induces self-sorting so that each school district contains households that are more homogeneous in their preferences for a type of schooling and amount of school spending. Theory does not predict that any one household characteristic, such as income, becomes more homogeneous in each district as choice increases. It is the combined effect of household characteristics on education preferences that becomes more homogeneous. Thus, we can only assert that, as choice increases in an educational market, the homogeneity of household characteristics in its districts will change and some characteristics are likely to become more homogeneous in every district. Let $X_{km} = (X_{1km}, X_{2km}, \ldots)$ be a vector of measures of the heterogeneity of household characteristics $X_1, X_2, \ldots$ in school district $k$ in metropolitan area $m$. Let $X_m = (X_{1m}, X_{2m}, \ldots)$ be a vector of measures of the heterogeneity of household characteristics over the entire population of the educational market $m$. Then, the second prediction can be summarized as:

$$
\frac{\partial x_{km}(C_m, X_m)}{\partial C_m} \geq 0 \text{ for } k = 1, 2, \ldots, K,
$$

and so on for characteristics $X_2, X_3, \ldots$. The inequality states that Tiebout choice affects (in an unknown direction) the heterogeneity of characteristic $X_1$ in each school district, relative to the heterogeneity that would exist if educational market $m$ were one district.

Schools' productivity can be affected by choice-driven incentives, sorting, demographic, and numerous environmental factors:

$$
\frac{A_{km}}{\ln(E_{km})} = f(C_m, \ldots, X_{1km}, X_{2km}, X_{3km}, X_m, X_n),
$$

where $A$ represents student achievement and $\ln(E)$ represents the log of per-pupil expenditure. $X_{km}$ and $X_m$ are defined above. $X_{km}$ is a vector of characteristics of household $i$ in district $k$ in market $m; X_m$ is a vector of mean characteristics of households in district $k$ in market $m; and X_n$ is a vector of mean characteristics of households in market $m.$

Since we do not directly observe incentives for productivity, empirical tests must be based on a reduced form where the first term in $C_m$ is not $\tau(C_m, \ldots).$ Also, the amount of variation in choice that we observe is limited by the number of educational markets in the United States, so it is necessary to impose a simple functional form such as:

$$
A_{km} = \beta_1 C_m + X_{1km} \beta_2 + X_{2km} \beta_3 + X_{3km} \beta_4 + X_{4km} + \epsilon_{km} + \epsilon_n + \epsilon_m.
$$

Equation (4) is the basic specification in this paper, but I use versions of it in which the dependent variable is just achievement or just per-pupil spending. It is useful to look at achievement and per-pupil spending separately because their relationships with choice may suggest how social allocative efficiency is affected by choice. For instance, one may have prior beliefs that a dynamic drop in average achievement would not be part of an increase in social allocative efficiency.

Equation (4) should help us differentiate the impact of competition from the impact of sorting. Many of the effects of choice on sorting will be captured by $X_m$ if it includes all available measures of heterogeneity that are likely to have a significant effect on productivity. I test whether the estimate of $\beta_1$ changes significantly when measures of district heterogeneity are excluded from the equation: this is a partial test because it assumes that the observable measures of heterogeneity are correlated with the unobservable measures of heterogeneity.

### IV. The Identification Problem

There are two identification problems likely to affect the analysis of Tiebout choice. The first is the potential for omitted variables bias. The second is the potential for observed choice to be endogenous.

The degree of choice that one observes in an educational market is the result of factors that affect the supply of school districts and factors that affect the population's demand for school districts. For studying the effects of choice, one wants to rely solely on variation in choice that comes from the supply side. In particular, one might worry that the factors that affect demand for school districts have a direct effect on productivity or achievement. If they do, then naive estimates of the effect of choice would be biased by omitted variables. For instance, Alberto Alesina et al. (1999) show that areas with greater ethnic heterogeneity demand more jurisdictions. Thus, if ethnic heterogeneity has an independent effect on productivity, and if ethnic heterogeneity is not fully controlled, then an ordinary least-squares (OLS) estimate of the effect of choice would be biased. The magnitude of the bias would shrink as one added measures of ethnic heterogeneity to equation (4), but it would be impossible to say when all bias had been eliminated. The sign of such omitted variables bias is not predictable.

The degree of choice that we observe in a market can also be, in part, a response to schools' observed productivity. This is a strict endogeneity problem that is best explained by an example. Consider an educational market that contains a district that has, for ideosyncratic reasons, a highly productive administration. Other districts will want to consolidate with the productive district so that its talented administrators can serve more students. But, such consolidation will lessen the degree of observed choice. Similarly, households with school-aged children will want to move into the highly productive district, excluding places with household that do have any school-aged children. But, such moves will lessen the degree of observed choice for any measure of choice that is sensitive to how many children each district serves. Endogeneity will negatively bias estimates of the effect of choice on productivity.

Intuitively, areas with more observed choice will be areas in which no districts were idiosyncratically good enough to attract consolidation or a disproportionate share of students. The best response to the identification problem is a set of valid instruments—that is, variables that affect the supply of jurisdictions but are uncorrelated with factors that affect the demand for jurisdictions.

### V. Measuring the Degree of Tiebout-Style Choice

The key to measuring the degree of Tiebout choice is to think about how households make residential decisions. One needs to consider, first, the boundaries of the educational market over which households exercise choice and, second, the costs associated with exercising choice. In some previous studies, insufficient attention to these matters has resulted in confused evidence about the effects of choice. For instance, researchers have sometimes assumed that the Tiebout process applies equally in rural and metropolitan areas.

If we take households' endowments as given, then each household faces two principal constraints on its residential choice: income and job location. The educational market over which it exercises Tiebout choice includes all school districts within a feasible commuting distance of its job(s). Such markets tend to correspond to Census-defined metropolitan areas of the United States because Census definitions are based, in

---

11 The measures of achievement are standardized so that they are approximately in percentile terms. It is natural, therefore, to use the natural logarithm of per-pupil spending in the denominator. Results based on a productivity measure that has per-pupil spending in the denominator are roughly similar and available from the author.

12 It is a partial test because it assumes that the observable measures of heterogeneity are correlated with the unobservable measures of heterogeneity.
part, on actual commuting behavior. Many rural districts do not belong to any educational market, so in far as most of their residents could feasibly commute to only a few (if any) other districts. I confine the empirical work, therefore, to analysis of metropolitan schools and students.

Next, consider the costs of exercising Tiebout choice within a metropolitan area. The important costs are not the costs associated with moving from one residence to another. The important costs are the costs of choosing a residence for its associated schools rather than for its other characteristics. Costs of the second type vary more across metropolitan areas and are incurred daily (unlike moving costs). For example, suppose that a household's elder has a job located at the center of a metropolitan area, and suppose that the household cares about only two characteristics of residences: commuting distance and local per-pupil spending. The household can choose different levels of school spending only by choosing among different school districts. In a metropolitan area where one school district contains the vast majority of jobs and residences and the commute to the nearest alternative district is long, the cost of being able to exercise choice is high. Conversely, the cost is low in a metropolitan area where many school districts are within a few minutes of most jobs. To generalize, for example, one only need add other characteristics of residences that matter to households: house prices, house sizes, police services, recreational opportunities, and so on. If households can choose among many school districts that offer comparable residences, where comparability is based on all these characteristics, then the cost of exercising choice is low. If households can only achieve their schooling desires by deviating far from their (otherwise) preferred residence characteristics, then the cost of exercising choice is high. Cost will naturally be a function of the number, size, geographic location, and housing stock of school districts within the metropolitan area. Miami, for instance, has high costs of exercising choice because the Dade County school district covers virtually all of the metropolitan area. At the other extreme, Boston has low costs of exercising choice: there are 70 school districts within a

30-minute commute of the downtown area, with several districts in each range of house prices.14

There are a variety of measures that correspond to the notion of choice just described. They include:

(a) the number of districts per student in the metropolitan area;

(b) a district-level choice index based on a Herfindahl index of school districts' shares of the metropolitan area's total land area:

\[ 1 - H_n = 1 - \sum_{k=1}^{K} S_{0k} \]

\[ S_{0k} = \frac{\text{land area}_{0k}}{\text{land area}_{\text{total}}} \]

(c) a district-level choice index based on a Herfindahl index of school districts' shares of the metropolitan area's total enrollment:

\[ 1 - H_n = 1 - \sum_{k=1}^{K} S_{1k} \]

\[ S_{1k} = \frac{\text{enrollment}_{0k}}{\text{enrollment}_{\text{total}}} \]

and so on.

Measure (c) has a particularly nice interpretation. It is the probability that a student would find himself in another district if he were to switch places with a another, randomly selected, student in his metropolitan area. Measures (b) and (c) vary between zero and one, where a value of zero indicates that one school district monopolizes the entire metropolitan area and a value close to one indicates that there are many, relatively equal-sized districts in the metropolitan area.

There is a conceptual difference between measures like (c) and measures like (a) and (b). Measures like (c) use enrollment to summarize a variety of characteristics that make some residences more desirable than others to households with school-aged children. Thus, measures like (c) condense more information relevant to choice

(housing stock, recreational opportunities for children, etc.), but they are also more prone to be endogenous to schools' observed productivity. Measures like (a) and (b) are unaffected by endogenous residential choices of households with school-aged children, but they are still affected by endogenous consolidation. In short, OLS estimates based on measures like (a) and (b) are likely to be biased, whereas OLS estimates based on measure (c), but—if a valid set of instruments is available—one wants to use measures like (c) to benefit from the information they contain.

One can construct a version of measure (c) that is based on each school's share of metropolitan area enrollment, instead of each district's share. Schools do not, however, have financial autonomy in the United States. Theory suggests that incentives for productivity depend on the financial repercussions associated with productivity gains and losses. Such repercussions are felt at the district level because they work through property prices, which affect districts' tax revenue and districts' budgets.15 Therefore, we expect productivity to be affected more by the degree of choice among districts than by the degree of choice among schools. In contrast, we expect sorting to be affected at least as much, if not more, by the degree of choice among schools as by the degree of choice among districts. Households with different endowments and preferences must sort themselves into different districts in order to get different levels of school spending, but households can determine the peers their children experience simply by sorting themselves into school attendance areas.

All of the proposed measures are erroneous measures of true choice. Thus, an additional benefit of instrumenting is that it remedies attenuation bias caused by classical error in the measurement of choice. One might, however, worry about a source of nonclassical measurement error. Even in the metropolitan areas that have the most Tiebout choice, there is a large central city district that dominates the stock of low-income housing. Thus, the variation in measures of choice probably overstates the variation in the degree of choice available to low-income families and represents more accurately the variation in the degree of choice available to middle-income families. If such nonclassical measurement error exists, it will bias the results toward finding that choice has no effect (either positive or negative) on low-income families.

VI. Instruments for Measures of Tiebout Choice

As instruments, I propose variables—specifically, streams—that reflect the number of natural school district boundaries in a metropolitan area.16 The logic is that in the eighteenth and nineteenth centuries, when school-district boundaries were initially set in America, an important consideration was students' travel time to school. In fact, in petitions for school-district boundaries, travel time was usually the primary justification for a set of boundaries. For a given travel distance, as the crow flies, school boundaries could significantly increase travel time. With automobiles, buses, paved roads, bridges, and flood controls, many of the barriers that would have caused students to travel miles out of their way are now hardly noticed. Yet, the vestigial importance of natural barriers is preserved because they determined initial school-district boundaries, which are the key supply-side factor that determines today's boundaries.

Thus, the number of school districts in a given land area at a given time of settlement was an increasing function of the number of natural barriers. I focus on streams, because they are the most common and most easily quantified natural barriers. I reserve discussion of the formation of the stream variables for the data section, but I implied first-stage equation states that the degree of choice in a metropolitan area is a function of its number of streams:

13 In the United States, revenue is raised and per-pupil spending is determined at the district level. Per-pupil spending is tied to the school level because a district cannot be meaningfully allocated among its schools. For instance, districts maintain programs for disabled children that potentially benefit all of their households. The spending on such programs is not fully assignable to individual disabled students or the schools they happen to attend.


15 According to the U.S. Geological Survey (USGS) (1998), streams include brooks, streams, and rivers. In this paper, streams also include the following bodies of water if they are roughly curvilinear in form: inlets, lakes, ponds, marshes, and swamps.
(5)  \[ C_n = S_n \alpha_1 + \tilde{X}_n \alpha_2 + \tilde{X}_n \alpha_3 + v_n. \]

where \( S_n \) is a vector of measures of the number of streams in metropolitan area \( m \). \( \tilde{X}_n \) is included many area characteristics that help to ensure that \( C_n \) is identified by stream topography, not by a metropolitan area's size or region.

In the results section, I demonstrate that streams fulfill the first condition for valid instrumental variables—that is, they are correlated with measures of choice. But, what about the second condition for valid instrumental variables—that streams are exogenous to school productivity? The condition is highly plausible. Such plausibility is important because it is impossible to fully test the second condition, but I do show some partial tests of it, including two overidentification and an examination of the covariances between streams and industrial composition and between streams and modern commuting times.

VII. School-District Characteristics and Aggregation Issues

A final set of econometric issues concerns the level of aggregation at which equations like (4) should be estimated. These issues are tricky, but they become straightforward if broken down into parts.

A. Aggregation When All the Covariates Are Exogenous

First consider the situation that would exist if measures of choice and all other covariates were exogenous. For instance, consider the equation:

(6)  \[ A_{i,m} = \gamma_1 C_n + \gamma_2 X_{1,m} + \gamma_3 X_{2,m} + \gamma_4 X_{1,m}^2 + \gamma_5 X_{2,m}^2 + \gamma_6 X_{1,m} X_{2,m} + v_{i,m}. \]

This is a version of equation (4) in scalar algebra, which makes the exposition that follows more intuitive. The dependent variable is achievement, and I have specified that there are only two variables in the \( X \) vectors: \( X_1 \) and \( X_2 \). These simplifications are without loss of generality. However, the omission of measures of heterogeneity is a loss of generality. I reintroduce these measures below.

To avoid tangential issues related to weighting, consider the case in which all educational markets have the same number of observations and, within each metropolitan area, all districts are of the same size.\(^{16}\) The key question about aggregation is whether the estimated effect of choice depends on whether equation (6) is estimated at the individual level, aggregated up to the district, and estimated at that level, or aggregated up to the educational market and estimated at that level. The answer is that the estimate of \( \gamma_1 \) does not depend on the level of aggregation at which equation (6) is estimated.

Intuitively, why is this so? In a linear regression, the coefficient on \( C_n \) (or any variable at the level of the educational market) is affected by another covariate only through the covariance between \( C_n \) and the market-level mean of that other covariate. Therefore, by the definition of a market-level mean, the inclusion of individual-level variables and district means does not affect the estimated coefficient on \( C_n \) so long as their market-level counterparts are included.

The proof is as follows. Suppose one estimates equation (6) by OLS at the individual level. One takes the usual first-order conditions with respect to the parameters. Even before solving the system of equations, however, one can impose constraints that hold exactly such as:

\[ \sum_{i=1}^{NM} 1 x_{1,m} c_{i,m} = \sum_{i=1}^{NM} x_{1,m} c_{i,m}, \]

and so on.\(^{19}\) Each constraint holds because deviations from a mean sum to zero for the observations over which the mean is taken. For instance, the first constraint listed can be rewritten as:

\[ \sum_{i=1}^{NM} x_{1,m} c_{i,m} = \sum_{i=1}^{NM} \sum_{i=1}^{N_i} (x_{1,m} + u_{1,m}) c_{i,m} \]

where \( u_{1,m} \) is defined so that \( x_{1,m} = x_{1,m} + u_{1,m} \). But, by definition,

\[ \sum_{i=1}^{NM} u_{1,m} = 0 \quad \forall m. \]

Therefore:

\[ \sum_{i=1}^{NM} x_{1,m} c_{i,m} = \sum_{i=1}^{NM} x_{1,m} c_{i,m}. \]

After imposing the full set of constraints on the first-order conditions, one can solve the system of equations to obtain the OLS estimate of \( \gamma_1 \):

\[ \gamma_1 = \frac{\sigma_{x_1,x_2} \sigma_{y_n} - \left[ \sum_{i=1}^{NM} x_{1,i} c_{i,n} \right] \sum_{i=1}^{NM} y_{n,i} c_{i,n}}{\sigma_{x_1,x_2} \sigma_{y_n} - \left[ \sum_{i=1}^{NM} x_{1,i} c_{i,n} \right] \sum_{i=1}^{NM} x_{1,i} c_{i,n}}. \]

where all the sums are from \( i = 1 \) to \( NM \).

\(^{17}\) Aggregation issues, including aggregation issues for the first-stage equation, are discussed in the next section. Equation (5) is the reduced-form of a structural problem in which early settlers tried to maximize the output of local schools, taking into account economies of scale and travel time. Streams enter the problem through equations for travel time, and the number and location of school districts are the solution to the problem.

\(^{18}\) That is, consider the case in which some metropolitan areas are broken into many equal-sized districts and others are broken into few equal-sized districts. For simplicity, I ignore weights in this exposition, but weights are used in the actual estimation so that observations do aggregate up to metropolitan area means. In addition, I use weights for individual students that are provided by the National Education Longitudinal Survey (NELS) (U.S. Department of Education, 1994a) and the National Longitudinal Survey of Youth (NLSY) (U.S. Department of Labor, 1998) to account for their sampling.

\(^{19}\) Variables in lowercase are in deviations-from-means form. The full set of additional constraints includes the three written above plus:

\[ \sum_{i=1}^{NM} x_{2,m} c_{i,m} = \sum_{i=1}^{NM} x_{2,m} c_{i,m}, \]

\[ \sum_{i=1}^{NM} x_{1,m} c_{i,m} = \sum_{i=1}^{NM} x_{1,m} c_{i,m}, \]

\[ \sum_{i=1}^{NM} x_{2,m} c_{i,m} = \sum_{i=1}^{NM} x_{2,m} c_{i,m}. \]
One obtains exactly the same formula for $\hat{\gamma}_l$ if one aggregates equation (6) to the district level (thereby eliminating the terms beginning with $\gamma_l$ and $\gamma_n$), takes the OLS first-order conditions, imposes the constraints listed in footnote 19 that are relevant, and solves them.20

One might wonder why I describe this result, since it seems to be a red herring. It is useful when we come to instrument for $C_m$. Also, although the point estimate of the coefficient on $C_m$ is unaffected, an equation like (6) is most efficiently estimated at the individual level. The gain in efficiency, relative to an aggregate regression, is due to the fact that an individual-level regression generates more precise estimates of the covariances between variables that vary at the individual or district level. Intuitively, the individual-level regression employs more information—such as the fact that black children are more likely to be poor. Of course, the calculation of the standard errors should reflect the fact that the multiple observations of $C_m$ in an educational market are not independent and the fact that the multiple observations of district-level variables in a district are not independent.

This is a common problem solved by Brent R. Moulton (1986), whose correct standard error formulas I use. Intuitively, Moulton’s formulas allow each educational market and district to have a random effect. Random effects are shown in equation (6) as a reminder that the standard errors are calculated using Moulton’s formulas.

B. Should One Worry That District-Level Means Are Endogenous to Choice?

Choosing choice in an educational market affects the mean characteristics of individuals in each school district. That is, district mean variables such as $X_{1am}$ are endogenous to the degree of choice $C_m$ in the educational market. The discussion above, however, showed that the inclusion of district mean variables does not affect the estimated coefficient on $C_m$. Therefore, the endogeneity of district mean variables to choice is irrelevant to the estimated coefficient on $C_m$. Intuitively, when greater choice increases the mean of some characteristic in one district, there is an exactly offsetting decrease in the mean of that characteristic in other districts. Therefore, although mean characteristics of districts are endogenous to choice, the effect of choice through its effect on mean characteristics is mechanically equal to zero.

C. Aggregation When Choice Needs to Be Instrumented

The aggregation results extend naturally to the case where instruments are needed because measures of choice are potentially endogenous. It is easiest to see the extension using a two-stage least-squares (2SLS) formulation of the instrumental variables (IV) procedure. The first-stage equation that corresponds to equation (6) is:

$$ C_m = \beta_0 + \beta_2 X_{1am} + \beta_3 X_{2am} + \omega_m $$

If we substitute the predicted value, $\hat{C}_m$, from this regression into the 2SLS formula for $\gamma_l$:

$$ \gamma_l = \frac{\sum_i x_{1im} \hat{\beta}_2 X_{1am} + \sum_i x_{2im} \hat{\beta}_3 X_{2am} + \omega_l}{\sum_i x_{2im}^2 + \sum_i x_{1im}^2} $$

we get the same reduced form regardless of whether the second stage is estimated at the individual, district, or market level of aggregation. The 2SLS formula and reduced form are identical for all three levels of aggregation because $C_m$ varies only at the level of the educational market.

D. Measures of District Heterogeneity

Finally, reintroduce the measures of district heterogeneity and educational market heterogeneity:

$$ A_{1am} = \beta_1 C_m + \beta_3 X_{1am} + \beta_3 X_{2am} + \beta_4 X_{1at} + \beta_4 X_{2at} $$

21 Assuming that the standard errors are calculated appropriately, the following first-stage regressions are identical because there is no correlation between the dependent variable and the individual-level and district-level independent variables:

$$ C_m = \beta_0 + \beta_2 X_{1am} + \beta_3 X_{2am} + \omega_m $$

$$ C_m = \beta_0 + \beta_2 X_{1am} + \beta_3 X_{2am} + \beta_4 X_{1at} + \beta_4 X_{2at} + \omega_m $$

22 By stating that the coefficients on the district-level mean variables do not have a ready structural interpretation, I mean that they should not be interpreted naively as pure peer effects. The coefficients on the district-level mean variables reflect a mixture of peer and individual effects. For the purposes of this paper, there is no need to dissect the mixture because the variables are added solely to improve precision and, as has been shown, they do not affect the coefficient on the choice variable.
Data on school districts and schools come from two sources. The first is the Census of Governments (COG) (U.S. Department of Commerce, 1984), which contains administrative data on the expenditures, enrollment, and instructional staff of every district in the United States. The second is the National Center for Education Statistics Common Core of Data (CCD) (U.S. Department of Education, 1995), which contains administrative data on enrollment, instructional staff, and student demographic information available at the school level is gender, race, and free-lunch eligibility (a common proxy for poverty).

Demographic information at the district level is, however, much richer. I use the special school-district tabulation of the Census of Population and Housing (SDDB) (U.S. Department of Commerce, 1983a; U.S. Department of Education, 1994b) for data on the percentage of students in private school, mean demographic characteristics each district, and mean demographic heterogeneity of each school district. For example, the equations contain only the number of students between two social groups and educational attainment group, but also indices of racial heterogeneity, ethnic heterogeneity, and educational heterogeneity.21 I derive demographic measures at the metropolitan-area level from the City and County Data Book (CCDB) (U.S. Department of Commerce, 1983b). Finally, I include measures of metropolitan-area size (based on population and land area) to ensure that the measure of choice is not merely picking up larger metropolitan areas, and I include indicator variables for the nine Census regions to ensure that the measure of choice is not picking up regional effects.

The school-district and school samples are drawn from the U.S. Geological Survey's (USGS) 1:24,000 quadrangle maps. It was by using these extremely detailed maps—which allow the viewer to identify even very small streams, buildings, and boundaries—that I initially recognized the relationship between natural barriers and school-district boundaries. The measurement of the streams variable was in two stages. Using the physical maps, I first counted all streams that were at least 3.5 miles long and of a certain width on the map. These data were checked against the Geological Survey's Geographic Names Information System (GNIS) for accuracy. I derived smaller streams directly from GNIS.24 I employ two stream variables: the number of larger streams (measured by hand and often traversing multiple districts, sometimes containing the population that fit into each racial group and educational attainment group, but also indices of racial heterogeneity, ethnic heterogeneity, and educational heterogeneity. One obvious way to combine the two is to use the interaction between the two variables, and the streams variables, to instrument for the interaction between the two variables, and the streams variables, and so on.

21 All three indices are based on the Herfindahl index and therefore vary from close to zero (substantial heterogeneity) to 1 (complete homogeneity). The index of racial homogeneity is a Herfindahl index built upon shares of the population who belong to each of five racial groups: non-Hispanic white, non-Hispanic black, Hispanic, Native American, and Asian. The index of educational homogeneity is an Herfindahl index built upon shares of the population that belong to each of four educational attainment groups: less than high school, high school graduate, some college, and four years or more years of college. The index of ethnic homogeneity starts with the same structure as the index of racial homogeneity, but discounts the homogeneity of the white population if it is heterogeneous in ancestry and discounts the homogeneity of Hispanic population if it is heterogeneous in ancestry. The index thus varies from 0 to 1, and does not differ from the index of racial homogeneity if the white and Hispanic populations have homogeneous ancestry. See Altonji et al. (1999) for a complete description of the index.

24 This two-part strategy was needed because bigger, linear bodies of water are sometimes inlets, bays, lakes, ponds, marshes, or swamps which must be judged visually. Also, larger bodies of water traverse multiple counties and, thus, visual counting prevents double counting. Smaller streams are more accurately measured using GNIS, which provides the longitude and latitude of their origin and destination.

25 Because the NLSY students were born in several years, every regression based on NLSY observations contains indicator variables for year of birth.
Table 1—Measures of Tiebout Choice

<table>
<thead>
<tr>
<th>Measure</th>
<th>Panel A: Descriptive Statistics</th>
<th>Standard deviation, controlling for metropolitan-area size*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index of choice among districts, based on enrollment</td>
<td>0.686</td>
<td>0.271</td>
</tr>
<tr>
<td>Index of choice among districts, based on land area</td>
<td>0.761</td>
<td>0.269</td>
</tr>
<tr>
<td>Districts in metropolitan area</td>
<td>21.132</td>
<td>27.611</td>
</tr>
<tr>
<td>Difference in commuting time (minutes) between the district with the shortest commute and the district with the shortest commute in the third shortest commute in the district with the shortest commute</td>
<td>6.498</td>
<td>8.551</td>
</tr>
<tr>
<td>Index of choice among schools, based on enrollment</td>
<td>0.974</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Panel B: Correlations Among the Residual Measures (Controlling for Metropolitan-Area Size)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index of choice among districts, based on enrollment</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index of choice among districts, based on land area</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index of choice among schools, based on enrollment</td>
<td>0.45</td>
<td>0.42</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference in commuting time (minutes) between the district with the shortest commute and the district with the shortest commute</td>
<td>0.65</td>
<td>0.71</td>
<td>0.40</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Index of choice among schools, based on enrollment</td>
<td>0.14</td>
<td>0.15</td>
<td>0.07</td>
<td>0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: There are 316 observations (metropolitan areas). See text for variable definitions.
Sources: Author’s calculations based on data from SDBD and CDC.
* Residuals are based on OLS regressions containing a constant, the metropolitan area’s population, the square of the population, the metropolitan area’s land area, and the square of the land area.

IX. Results

Table 1 shows several measures of the degree of choice among school districts. All of the measures show that the degree of choice varies widely across metropolitan areas in the United States—and not just because some metropolitan areas are larger than others. For instance, the choice index based on districts’ enrollment shares has a mean of 0.69 and a standard deviation of 0.27. Even after controlling for a metropolitan area’s land area, the square of its land area, its population, and the square of its population, the (residual) index has a standard deviation of 0.25. This standard deviation corresponds to the difference between having, say, four equal-sized districts and a very large number of equal-sized districts (more than 50).

Table 1 also shows there is not much variation across metropolitan areas in the degree of choice among schools. The choice index based on school’s enrollments has a mean of 0.97. The standard deviation of its residual (controlling for metropolitan-area size) is only 0.06.

The bottom panel of Table 1 shows correlations among the residual measures of choice. Residual measures eliminate the correlation caused purely by metropolitan-area size. The table shows that the measures of choice among districts are highly correlated. For instance, the correlation between the residual choice index based on districts’ enrollments and the residual choice index based on districts’ land areas is 0.86. Given the substantial correlation among the district-based measures, the important differences among them are the substantive ones (such as the greater information and greater potential for endogeneity in the enrollment-based measure). A metropolitan area’s degree of choice among districts is not, however, highly correlated with its degree of choice among schools. The correlation between the residual district-based choice index and the residual school-based choice index is 0.14. Metropolitan areas with more choice among districts do not necessarily offer more choice among schools. A metropolitan area that is monopolized by one district may have neighborhood schools to the same extent as a metropolitan area with 50 districts. The lack of correlation is important because the peers whom a student actually encounters depend on the school he attends. Also, the lack of correlation is the key to interpreting some of the results shown below.

A. First-Stage Results

Table 2 contains estimates of the implied first-stage regressions for the district-based and school-based measures of Tiebout choice. The district-based choice index is statistically significantly related to the streams variables. The F-statistic on the joint significance of the two excluded instruments is 24.4 (the p-value is less than 0.001), so the instruments are not weakly correlated. One standard deviation in the number of smaller streams generates about one-fourth of a standard deviation in the choice index. Moreover, the number of smaller streams has, by itself, nearly as much explanatory power as the two streams variables used in combination. If only the smaller streams variable is included in the regression, the F-statistic on the excluded instrument is 20.2.

The streams variables have a weak statistical relationship with the school-based choice index. The F-statistic on the joint significance of the two excluded instruments is 1.86 (the p-value is 0.174).

There are a few ways to test whether the streams variables are effectively exogenous to school productivity—that is, whether they affect productivity only through their effect on school districting. I discuss the statistical tests below, but it is natural to discuss two informal tests here. To address the possibility that streams influence school productivity by affecting a metropolitan area’s industrial composition, one can examine the correlation between measures of industrial composition and the residual streams variables. The correlation between the residual larger streams variable and the percentages of employment in manufacturing, mining, and durable goods manufacturing, wholesale trade, and financial services are, respectively, 0.007, 0.004, −0.074, and −0.029. The corresponding correlations for the smaller streams variables are 0.066, 0.043, −0.068, and −0.007. To address the possibility that streams somehow influence school productivity by increasing travel times in general in the metropolitan area, one can examine the correlation between the average travel time to work in a metropolitan area and the residual streams variables. The correlation between the larger streams variable and average travel time is −0.015. The corresponding correlation for smaller streams is 0.070. The smallness of the “suspect” correlations

26 Two background variables had missing observations because the background variables employed were key variables in each survey. The NELS sample was altered by design with every wave, so there are, for instance, more 8th-grade than 10th-grade test scores available. I use earned income data only for NLSY students who have positive earnings.
Table 2—Selected Coefficients from the Implied First-Stage Regression

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Index of choice among districts (based on enrollment)</th>
<th>Index of choice among schools (based on enrollment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of larger streams in metropolitan area*</td>
<td>0.080</td>
<td>-0.040</td>
</tr>
<tr>
<td>Number of smaller streams in metropolitan area</td>
<td>0.034</td>
<td>0.004</td>
</tr>
<tr>
<td>Population of metropolitan area (thousands)</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>Land area of metropolitan area (thousands of square miles)</td>
<td>0.085</td>
<td>0.033</td>
</tr>
<tr>
<td>Mean of log(gini) of metropolitan area</td>
<td>-0.246</td>
<td>-0.263</td>
</tr>
<tr>
<td>Gini coefficient of metropolitan area</td>
<td>-3.381</td>
<td>-1.300</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is 0 to 19 years of age</td>
<td>(0.811)</td>
<td>(0.349)</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is 65 years of age or older</td>
<td>1.859</td>
<td>0.239</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is Asian</td>
<td>0.799</td>
<td>(0.344)</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is black</td>
<td>0.447</td>
<td>0.583</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is Hispanic</td>
<td>0.605</td>
<td>(0.260)</td>
</tr>
<tr>
<td>Index of racial homogeneity of metropolitan area</td>
<td>1.877</td>
<td>0.894</td>
</tr>
<tr>
<td>Index of ethnic homogeneity of metropolitan area</td>
<td>(1.053)</td>
<td>(0.432)</td>
</tr>
<tr>
<td>Share of adults in metropolitan area the highest grade which is some college</td>
<td>0.827</td>
<td>0.654</td>
</tr>
<tr>
<td>Share of adults in metropolitan area the highest grade which is B.A. or more</td>
<td>(0.414)</td>
<td>(0.178)</td>
</tr>
<tr>
<td>Index of educational homogeneity of metropolitan area</td>
<td>0.089</td>
<td>0.114</td>
</tr>
<tr>
<td>Indicator variables for the nine Census regions of the United States</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>$F_{iv}^{29}$ statistic, joint significance of the excluded instruments</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: The table shows selected coefficients from the first-stage regression that is implied by the IV regressions (which are actually estimated in one stage). There are 316 observations (metropolitan areas). The two stage variables are the excluded instruments. The dependent variables have the following means (standard deviations): 0.686 (0.271), 0.974 (0.089). The two stage variables have the following means (standard deviations): 7.893 (14.782), 182.749 (208.810). Coefficient estimates that are not shown in the table are available from the author. See text for variable definitions. Sources: Author’s calculations based on data from SDBB, CCD, CCDB, GNIS, and USGS maps.

*The variables are measured in hundreds.

suggests that, if streams affect schools, it is because they affect district boundaries, not because they are otherwise important today.

B. The Effect of Tieout Choice on Student Achievement

Ultimately, we are interested in schools’ productivity, but it makes sense to look first at the two components of productivity: student achievement and school spending. Table 3 shows IV estimates of equation (4) for one measure of student achievement: 12th-grade reading test scores from the NELS. The table shows not only the coefficient of interest, but coefficients for other interesting covariates, such as family background and measures of metropolitan-area heterogeneity. I present

Table 3—Effect of Tieout Choice on Achievement: Instrumental Variables Estimates of Selected Coefficients

<table>
<thead>
<tr>
<th>Dependent variable: 12th-grade reading score</th>
<th>Index of choice among districts, based on enrollment</th>
<th>Index of choice among schools, based on enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(household income)</td>
<td>1.536</td>
<td>(0.164)</td>
</tr>
<tr>
<td>Female</td>
<td>1.959</td>
<td>(0.227)</td>
</tr>
<tr>
<td>Asian</td>
<td>0.284</td>
<td>(0.591)</td>
</tr>
<tr>
<td>Black</td>
<td>-5.491</td>
<td>(0.497)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-2.866</td>
<td>(0.518)</td>
</tr>
<tr>
<td>Parents' highest grade is some college</td>
<td>2.306</td>
<td>(0.296)</td>
</tr>
<tr>
<td>Parents' highest grade is B.A. or more</td>
<td>5.453</td>
<td>(0.299)</td>
</tr>
<tr>
<td>Mean of log(gini) of metropolitan area</td>
<td>-5.421</td>
<td>(5.527)</td>
</tr>
<tr>
<td>Gini coefficient of metropolitan area</td>
<td>-12.770</td>
<td>(12.022)</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is Asian</td>
<td>-5.620</td>
<td>(13.067)</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is black</td>
<td>-9.732</td>
<td>(6.063)</td>
</tr>
<tr>
<td>Share of metropolitan-area population that is Hispanic</td>
<td>0.247</td>
<td>(3.516)</td>
</tr>
<tr>
<td>Index of racial homogeneity of metropolitan area</td>
<td>-9.598</td>
<td>(7.840)</td>
</tr>
<tr>
<td>Index of ethnic homogeneity of metropolitan area</td>
<td>16.313</td>
<td>(10.697)</td>
</tr>
<tr>
<td>Share of adults in metropolitan area the highest grade which is some college</td>
<td>5.274</td>
<td>(7.135)</td>
</tr>
<tr>
<td>Share of adults in metropolitan area the highest grade which is B.A. or more</td>
<td>3.163</td>
<td>(5.929)</td>
</tr>
<tr>
<td>Index of educational homogeneity of metropolitan area</td>
<td>-5.447</td>
<td>(12.952)</td>
</tr>
</tbody>
</table>

Notes: IV estimates based on 6,119 students who live in 316 metropolitan areas. Standard errors are in parentheses and use formulas (Moulton, 1986) for data grouped by districts and metropolitan areas. The regression is weighted so that each metropolitan area receives equal weight. The standardized reading scores have a mean of 50 and a standard deviation of 10. The choice index has a mean of 0.765 and a standard deviation of 0.236. Covariates that are not shown are: population of metropolitan area, land area of metropolitan area, mean of log(household income) in district, Gini coefficient for district, share of district population that is Asian, share of district population that is Black, share of district population that is Hispanic, index of racial homogeneity in district, index of ethnic homogeneity in district, share of adults in district the highest level of education which is some college, share of adults in district the highest level of education which is B.A. or more, and index of educational homogeneity of metropolitan area. See text for variable definitions. Sources: Author’s calculations based on data from SDBB, CCD, CCDB, GNIS, and USGS maps.

Table 3 to illustrate the coefficients on these other covariates.

I find conventional effects of individual background characteristics on 12th-grade reading scores. (All of the test scores are standardized to have a mean of 50 and a standard deviation of 10.) Students from households with income that is 10 percent higher have test scores that are about 0.15 standardized points higher. Females have reading scores that are 2 standardized points higher than those of males. Black and Hispanic students have reading scores that are, respectively, 5.5 and 2.9 standardized points lower than those of white non-Hispanic students. Compared to students whose
The OLS results reveal the sign of the bias due to omitted variables and endogeneity. They suggest that successful districts do attract household members with school-aged children and do attract other districts into consolidation. Also, achievement gains may be negatively affected by the presence of students on free/reduced lunch, factors such as disension, that raise the demand for districts in a metropolitan area.

A comparison between the first and third rows of Table 4 shows that if measures of a district's heterogeneity were omitted from the equation, the estimated coefficient on choice does not change by a statistically significant amount. This is not just because the standard errors are too large for a plausible change to be statistically significant: the change in the point estimates is very small. That is, the main effect of choice on student achievement does not appear to be working through the effect of choice on districts' heterogeneity. (We cannot make this comparison for measures of achievement from the NLSY because the students cannot be matched to specific districts.)

The fourth row of Table 4 shows a version of the base specification that has been aggregated up to the metropolitan-area level. The resulting estimates are similar to, but have larger standard errors. As with the base specification, the estimated coefficient on choice is significantly different from zero at the 0.05 level, except for the educational attainment result which is statistically significantly different from zero at the 0.10 level.

Are these positive effects on achievement large or small? They are impressive if one considers an increase from 0 to 1 in the choice index—that is, if one compares metropolitan areas at opposite ends of the choice spectrum, like Miami and Boston. This is the relevant comparison for thinking about the potential of Tiebout choice as a policy. However, a standard deviation in the choice index is 0.27, so only a modest amount of the current variation in American students' achievement is explained by Tiebout choice.

Now consider the OLS results that we obtain if we naively ignore the probability that observed choice is endogenous to achievement and factors that affect the demand for school districts' enrollment. To interpret them naively, the OLS results in the second and third rows of Table 4 would suggest that Tiebout choice has no effect or small positive effects on achievement.

Parents have no college education, students whose parents have some college education (but no baccalaureate degree) have reading scores that are 2.3 standard deviations higher, and students whose parents have at least one baccalaureate degree have scores that are 5.5 standardized points higher. Most of the metropolitan-area characteristics do have statistically significant effects, but these coefficients do not have a ready structural interpretation so they are not shown.

Table 4 shows the estimated effect of choice (that is, just the coefficient of interest) for several measures of student achievement and specifications. The measures of achievement are 8th-grade reading scores, 10th-grade math scores, 12th-grade reading scores, ASVAB math knowledge scores, highest grade completed, and the income a student earns when he is 32 years old. The first row of Table 4 shows the base IV specification, which is the specification shown in detail in Table 3. The results for the base specification suggest that student achievement is higher when there is more choice among districts. An increase from 0 to 1 in the index of Tiebout choice generates 8th-grade reading scores that are 3.8 points higher, 10th-grade math scores that are 3.1 points higher, and 12th-grade reading scores that are 5.8 points higher, and math knowledge scores that are 2.7 points higher. In short, test scores rise by one-quarter to one-half of a standard deviation. In addition, such an increase in choice generates educational attainment that is 1.4 grades higher and income at age 32 that is about 15 percent higher. All of this evidence suggests that Tiebout choice is significantly different from zero at the 0.05 level, except for the educational attainment result which is statistically significantly different from zero at the 0.10 level.

Parents have no college education, students whose parents have some college education (but no baccalaureate degree) have reading scores that are 2.3 standard deviations higher, and students whose parents have at least one baccalaureate degree have scores that are 5.5 standardized points higher. Most of the effects of the coefficients on district characteristics (especially since district-level variables are only crude measures of the environment that a family possesses for a student). An Appendix is available from the author that covers issues such as (1) how the inclusion of individual-level variables improves the precision of the regressions, (2) how the effects of family background are reflected by both the individual-level variables and the district-level variables, or (3) how error in the measurement of family background affects the coefficients on the individual-level and district-level variables. Among other things, the Appendix shows that, in practice, the coefficients on the individual-level variables are only very slightly affected by the inclusion of the district-level variables. Thus, it is safe to say that one should interpret the coefficients on the individual-level variables as the effects directly, and indirect of, an individual's background on his achievement.
The fifth row of Table 4 shows the less preferred measure of choice based on land area. Recall that this measure tends to be less informative about choice because it does not reflect the enrollment structure of metropolitan areas. On the other hand, this measure cannot reflect some types of endogenous behavior, such as parents of school-aged children moving to good school districts. If this type of endogeneity is remedied by the instrumental variables strategy, then the IV estimates for the land-based measure should be similar to those for the enrollment-based measure. This is, in fact, what Table 4 shows. The point estimates for the land-based measure are similar to those in the top row of the table (and are far from being statistically significantly different). The sixth row of Table 4 shows IV estimates of the effect of more choice among schools on student achievement. Recall that there is a weak statistical relationship between the streams variables and the index of choice among schools, so that one is unlikely to get meaningful IV results for this type of choice. Indeed, this is what Table 4 shows. The standard errors of the row six estimates are so large that the estimates (none of which is significantly different from zero) are uninterpretable. Thus, the effect of choice among schools is probably impossible to determine. The instrumental variables procedure is not usable, but OLS estimates would not be credible. The comparison between OLS and IV for district-level choice suggested that endogeneity plagues OLS, and measures of choice among schools are far more likely to be endogenous than measures of choice among districts.

The next two rows of Table 4 allow the effects of Tiebout choice to differ for students who come from low-income and not-low-income families. For the purpose of this table, low-income families are those that have household income less than or equal to 70 percent of the mean household income in their metropolitan area. Not-low-income households are all others. Families are classified relative to their metropolitan area’s income because the main reason for considering heterogeneous effects is the potential for Tiebout choice to affect sorting of families within their metropolitan area. The results shown in the two rows provide little evidence of heterogeneous effects. The estimates are slightly lower for low-income families, but they are in the same range (one-quarter to one-half of a standard deviation in test scores) and not statistically significantly different from the estimates for not-low-income families. The ninth and tenth rows of Table 4 allow the effects of Tiebout choice to vary for minority and nonminority students. The purpose of this table, minority students are black and Hispanic students, and nonminority students are all others. The results suggest that Tiebout choice does not have a statistically significant effect on the 8th-grade reading scores or the 10th-grade math scores of minority students, while the effects on nonminority students are statistically significant and positive. The minority-nonminority difference in the effect on 8th-grade reading scores is not statistically significant, but the minority-nonminority difference in the effect on 10th-grade math scores is statistically significant at the 0.10 level. This weak evidence of heterogeneous effects is not, however, confirmed by the other measures of achievement. The effects on 12th-grade reading scores are in the same range for minority and nonminority students, and the point estimates for math knowledge and educational attainment are higher for minority than for nonminority students, although the difference is not statistically significant. The estimates for income at age 32 are nearly identical.

Finally, the last two rows of Table 4 show partial tests of the exogeneity of the instrument variables. The omnibus test attempts to show whether variations in the streams variables that is not correlated with variation in school districting (or other observable determinants of school productivity) is correlated with school productivity. Intuitively, it is a test of whether, after eliminating its correlation with choice (instrumented) and the other covariates, achievement is still correlated with the streams variables. The test statistics are distributed as $\chi^2$ with one degree of freedom. I use formulas that account for the fact that individuals in the same metropolitan area do not have independent values of metropolitan-level variables (including the choice index and the streams variables). The omnibus test consistently fails to reject the null hypothesis that streams affect student achievement only via their effect on choice. The other tests use a Hausman test of the exogeneity of the larger streams variable. It is based on the premise that one has more a priori confidence in the exogeneity of the smaller streams variable because smaller streams are more consistent with modern life. The test statistics, which are distributed as $\chi^2$ with one degree of freedom, show that the Hausman test consistently fails to reject the null hypothesis that the larger streams variable is a valid instrument. Overall, Table 4 demonstrates that an increase in Tiebout choice has a statistically significant, positive effect on measures of achievement that range from test scores to wages. Naive OLS estimates of the effect of choice on student achievement are likely to be downward biased, and the stream variables appear to be valid instruments. It is possible that the effect on minority and low-income students is smaller than the effect on other students, but the evidence for such a conclusion is only suggestive. Even if the evidence were stronger, it would be hard to interpret. It could be that sorting caused by Tiebout choice has negative effects on disadvantaged students that offset some of the gains they experience from competition. Alternatively, as discussed above, it could be that the choice measure is particularly erroneous for disadvantaged students.

C. The Effects of Tiebout Choice on Per-Pupil Spending

Table 5 shows the effect of Tiebout choice on per-pupil spending and private schooling. Recall that these results are based on nearly the entire population of metropolitan school districts, unlike the achievement results which are based on a sample. I use Table 5 to discuss the covariates other than the coefficients of interest. Metropolitan areas with higher household incomes spend more per pupil, as do metropolitan areas that are more Hispanic and more racially homogeneous. The elasticity of per-pupil spending with respect to the mean of log income in the metropolitan area is estimated to be 0.54. Also, metropolitan areas with larger populations have a higher percentage of students in private schools, as do metropolitan areas whose adults have more heterogeneous educational attainments. The coefficients on district characteristics have statistically significant effects, but these coefficients do not have a ready structural interpretation, so they are not shown.

Table 6 shows the coefficient of interest for several specifications. Consider the results for per-pupil spending, which are in the first column. The results for the base specification suggest that per-pupil spending is lower where there is more choice among districts. An increase from 0 to 1 in the index of Tiebout choice generates a 7.6-percent decrease in per-pupil spending. The OLS estimate is 7.2 percent, which is insufficiently different from the IV estimate to suggest bias. That is, while the effects of choice on achievement appear to be significantly affected by endogeneity and omitted variables bias, the effects of choice on per-pupil spending appear to be only slightly affected. The specification that omits measures of district heterogeneity and the specification that is aggregated to the metropolitan area level produce estimates similar to the base estimates.
The estimate for the land-based choice index is a 10.1-percent decrease in per-pupil spending for an increase in the choice index from 0 to 1. The index of choice among schools generates results that are interpretable and statistically significant. The statistic for the Hausman test of the exogeneity of the instrumental variables fails to reject the null hypothesis that larger streams are a valid instrumental variables. In cases where the choice index among schools is significantly related to student achievement only via their effect on choice, the statistic for the Hausman test of the exogeneity of the larger streams instrument also fails to reject the null hypothesis that larger streams are a valid instrumental variables.

Notes: IV estimates based on data from school districts in metropolitan areas. Standard errors are in parentheses and use formulas (Moulton, 1986) for data gathered by metropolitan areas. The regression parameters are weighted equally, given that each metropolitan area represents an equal number of observations (school districts) in a given metropolitan area. The log of per-pupil spending has a mean of 8.462 and a standard deviation of 0.256. The share of students in private school has a mean of 0.118 and a standard deviation of 0.064. The index of choice has a mean of 0.765 and a standard deviation of 0.256. Covariates that are not shown are: share of district population that is Asian, share of district population that is black, share of district population that is Hispanic, index of racial homogeneity in district, index of ethnic homogeneity in district, mean of log(household income) in district, Gini coefficient for district, share of adults in district that are high school graduates, share of adults in district that are college graduates, share of adults in district that are B.A. or more, and share of educational attainment in district. See text for variable definitions. Sources: Author's calculations based on data from SDDB, CCD, CCDB, GNIS, and USGS maps.

The evidence on per-pupil spending, the evidence on private school using data from nearly all the public-school districts in the United States. For this reason, the private-schooling re-

---

Table 5—Effect of Tiebout Choice on Per-Pupil Spending and Private Schooling: Instrumental Variables Estimates of Selected Coefficients

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Log(per-pupil spending)</th>
<th>Share of students in private school</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index of choice among districts, based on enrollment</td>
<td>-0.076 (0.034)</td>
<td>-0.042 (0.012)</td>
</tr>
<tr>
<td>Population of metropolitan area (thousands)</td>
<td>0.004 (0.003)</td>
<td>0.006 (0.003)</td>
</tr>
<tr>
<td>Land area of metropolitan area (thousands of square miles)</td>
<td>0.003 (0.002)</td>
<td>-0.001 (0.001)</td>
</tr>
<tr>
<td>Mean of log(average) of metropolitan area</td>
<td>0.543 (0.115)</td>
<td>0.044 (0.040)</td>
</tr>
<tr>
<td>Gini coefficient of metropolitan area</td>
<td>0.372 (0.601)</td>
<td>-0.251 (0.247)</td>
</tr>
<tr>
<td>Share of metropolitan population that is Asian</td>
<td>0.620 (0.713)</td>
<td>0.264 (0.225)</td>
</tr>
<tr>
<td>Share of metropolitan population that is black</td>
<td>0.299 (0.146)</td>
<td>0.097 (0.074)</td>
</tr>
<tr>
<td>Share of metropolitan population that is Hispanic</td>
<td>0.463 (0.134)</td>
<td>0.012 (0.027)</td>
</tr>
<tr>
<td>Index of racial homogeneity of metropolitan area</td>
<td>0.762 (0.334)</td>
<td>0.096 (0.087)</td>
</tr>
<tr>
<td>Index of ethnic homogeneity of metropolitan area</td>
<td>-0.581 (0.492)</td>
<td>0.001 (0.127)</td>
</tr>
<tr>
<td>Share of adults in metropolitan area who have a high school degree or more</td>
<td>0.480 (0.348)</td>
<td>-0.327 (0.103)</td>
</tr>
<tr>
<td>Share of adults in metropolitan area who have a B.A. or more</td>
<td>-0.426 (0.502)</td>
<td>-0.081 (0.076)</td>
</tr>
<tr>
<td>Index of educational homogeneity of metropolitan area</td>
<td>-1.231 (0.945)</td>
<td>-0.609 (0.233)</td>
</tr>
<tr>
<td>Indicator variables for the nine Census regions of the United States</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Notes: IV estimates based on school districts in metropolitan areas. Standard errors are in parentheses and use formulas (Moulton, 1986) for data gathered by metropolitan areas. The regression coefficients are weighted equally, given that each metropolitan area represents an equal number of observations (school districts) in a given metropolitan area. The log of per-pupil spending has a mean of 8.462 and a standard deviation of 0.256. The share of students in private school has a mean of 0.118 and a standard deviation of 0.064. The index of choice has a mean of 0.765 and a standard deviation of 0.256. Covariates not shown are: share of district population that is Asian, share of district population that is black, share of district population that is Hispanic, index of racial homogeneity in district, index of ethnic homogeneity in district, mean of log(household income) in district, Gini coefficient for district, share of adults in district that are high school graduates, share of adults in district that are college graduates, share of adults in district that are B.A. or more, and share of educational attainment in district. See text for variable definitions. Sources: Author's calculations based on data from SDDB, CCD, CCDB, GNIS, and USGS maps.

The evidence on per-pupil spending, the evidence on private school using data from nearly all the public-school districts in the United States. For this reason, the private-schooling re-

D. The Effect of Tiebout Choice on Private-School Enrollment

Like the evidence on per-pupil spending, the evidence on private school using data from nearly all the public-school districts in the United States. For this reason, the private-schooling re-
The Effect of Tiebout Choice on Productivity

Table 7 shows results for productivity measures that are created by dividing the six measures of achievement by the log of per-pupil spending. Given the achievement and per-pupil spending results already discussed, one expects choice to have a positive effect on productivity.

To get an accurate estimate, however, one should explicitly examine productivity because the data on achievement are available for only a sample of the districts for which spending data are available. The results are best summarized in terms of standard deviations of the productivity measures (the dependent variables), since the units of the productivity measures are not intuitive. In Table 7, the numbers in square brackets show each coefficient estimate as a share of a standard deviation of the dependent variable.

The results for the base specification indicate that an increase from 0 to 1 in the index of Tiebout choice among districts raises productivity by between one-fourth and six-tenths of a standard deviation. Greater productivity effects are found for productivity measures that are based on measures of achievement that are recorded later in life (for instance, income as opposed to 8th-grade test scores). The next two rows of Table 7 show no statistical evidence that productivity effects are different for students from poor and nonpoor families, although the estimates suggest that productivity effects may be slightly greater for students from nonpoor families. The following two rows show conflicting evidence on whether productivity effects are different for minority and nonminority students. On the other hand, the productivity measures based on 10th-grade math scores indicates that productivity effects are statistically significantly smaller (at the 0.10 level) for minority students than nonminority students. On the other hand, three productivity measures suggest that the productivity effects are greater (by a statistically insignificant amount) for minority students. In short, there is weak evidence that poor or minority students experience a smaller productivity boost than other students. Even if one were to find strong evidence, it would be unclear whether the smaller effect is due to choice being measured poorly for disadvantaged students.

Choice among districts appears to raise productivity, but how essential to the result are the financial incentives for districts that work through property tax revenues? I attempt to answer this question with the next two rows of

<table>
<thead>
<tr>
<th>8th-grade math score</th>
<th>10th-grade math score</th>
<th>12th-grade math score</th>
<th>ASVAB math knowledge</th>
<th>Highest grade attained</th>
<th>Incomes at age 32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base IV specification</td>
<td>0.290</td>
<td>0.308</td>
<td>0.579</td>
<td>0.516</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>(0.140)</td>
<td>(0.158)</td>
<td>(0.226)</td>
<td>(0.202)</td>
<td>(0.056)</td>
</tr>
<tr>
<td></td>
<td>[0.246]</td>
<td>[0.262]</td>
<td>[0.495]</td>
<td>[0.410]</td>
<td>[0.494]</td>
</tr>
<tr>
<td>Base IV with choice index interacted with family income: &lt; low-income households</td>
<td>0.227</td>
<td>0.268</td>
<td>0.406</td>
<td>0.513</td>
<td>0.213</td>
</tr>
<tr>
<td>Effect for low-income households</td>
<td>(0.173)</td>
<td>(0.182)</td>
<td>(0.263)</td>
<td>(0.193)</td>
<td>(0.053)</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.228)</td>
<td>(0.347)</td>
<td>(0.408)</td>
<td>(0.489)</td>
</tr>
<tr>
<td>Effect for non-low-income households</td>
<td>0.312</td>
<td>0.298</td>
<td>0.572</td>
<td>0.619</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(0.169)</td>
<td>(0.264)</td>
<td>(0.056)</td>
<td>(0.020)</td>
</tr>
<tr>
<td></td>
<td>(0.265)</td>
<td>(0.234)</td>
<td>(0.489)</td>
<td>(0.492)</td>
<td>(0.514)</td>
</tr>
<tr>
<td>Base IV with choice index interacted with minority status: &lt; minority households</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect for minority households</td>
<td>-0.141</td>
<td>-0.157</td>
<td>0.428</td>
<td>0.695</td>
<td>0.277</td>
</tr>
<tr>
<td></td>
<td>(0.348)</td>
<td>(0.385)</td>
<td>(0.419)</td>
<td>(0.310)</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.134)</td>
<td>(0.366)</td>
<td>(0.552)</td>
<td>(0.636)</td>
</tr>
<tr>
<td>Effect for nonminority households</td>
<td>0.374</td>
<td>0.556</td>
<td>0.595</td>
<td>0.417</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>(0.155)</td>
<td>(0.182)</td>
<td>(0.289)</td>
<td>(0.346)</td>
<td>(0.333)</td>
</tr>
<tr>
<td></td>
<td>(0.318)</td>
<td>(0.474)</td>
<td>(0.509)</td>
<td>(0.331)</td>
<td>(0.377)</td>
</tr>
<tr>
<td>Base IV with choice index interacted with state control of school revenue: &lt; mostly state controlled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect for mostly state controlled</td>
<td>0.110</td>
<td>0.196</td>
<td>0.469</td>
<td>0.304</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.173)</td>
<td>(0.245)</td>
<td>(0.217)</td>
<td>(0.053)</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.276)</td>
<td>(0.401)</td>
<td>(0.242)</td>
<td>(0.584)</td>
</tr>
<tr>
<td>Effect for mostly locally controlled</td>
<td>0.290</td>
<td>0.357</td>
<td>0.315</td>
<td>0.401</td>
<td>0.302</td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.160)</td>
<td>(0.245)</td>
<td>(0.203)</td>
<td>(0.056)</td>
</tr>
<tr>
<td></td>
<td>(0.246)</td>
<td>(0.304)</td>
<td>(0.512)</td>
<td>(0.330)</td>
<td>(0.694)</td>
</tr>
</tbody>
</table>

Mean of dependent variable: 0.917 1.937 5.937 5.973 5.973 1.642 1.138

Standard deviation of dependent variable: 0.177 1.172 1.170 1.258 0.435 0.136

Notes: Each dependent variable is formed by dividing a measure of achievement by the log of per-pupil spending. The numbers in square brackets show each coefficient estimate as a share of a standard deviation of the dependent variable. The notes for Table 4 apply to this table. States are mostly state controlled if state revenue accounts for more than 50 percent of school spending. The remaining states are mostly locally controlled.

Sources: Author's calculations based on data from NELS, NLSS, NSD, SCD, CCDB, GNIS, and USGS maps.

Table 7, where the specification allows the effect of Tiebout choice to depend on the degree to school revenues in a state come from the state government. I divide the states into those in which state government accounts for more than 50 percent of school spending ("mostly state controlled") and less than 50 percent of school spending ("mostly locally controlled"). This is a crude measure of state control, but it is sufficient for the present purpose. Conveniently, half of the states fall into each category. The estimated productivity effects for mostly state-controlled districts are consistently smaller than those for mostly locally controlled districts, but none of the differences is statistically significant. The point estimates do suggest, however, that Tiebout choice has stronger effects in states where districts have more financial independence. An increase from 0 to 1 in the choice index is estimated to raise productivity by at least one-quarter of a standard deviation in states with mostly local control, but only by at least one-tenth of a standard devia-
Table 8—Homogeneity of Students Within Schools

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>OLS</th>
<th>IV</th>
<th>OLS</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of districts in metropolitan area (in hundreds)</td>
<td>-0.066</td>
<td>-0.071</td>
<td>-0.047</td>
<td>-0.060</td>
</tr>
<tr>
<td>Number of schools in metropolitan area (in hundreds)</td>
<td>(0.059)</td>
<td>(0.158)</td>
<td>(0.007)</td>
<td>(0.030)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>OLS</th>
<th>IV</th>
<th>OLS</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of districts in metropolitan area (in hundreds)</td>
<td>0.082</td>
<td>0.118</td>
<td>-0.012</td>
<td>-0.010</td>
</tr>
<tr>
<td>Number of schools in metropolitan area (in hundreds)</td>
<td>(0.050)</td>
<td>(0.158)</td>
<td>(0.005)</td>
<td>(0.021)</td>
</tr>
</tbody>
</table>

Notes: OLS and IV estimates based on schools in metropolitan areas. Standard errors are in parentheses and use formulas (Moulton, 1986) for data grouped by metropolitan areas. The regression is weighted so that each metropolitan area receives equal weight. There are 30,901 observations (schools) in 316 metropolitan areas. The full set of covariates is the same as those listed in Table 2. See text for variable definitions. The racial homogeneity indices are Herfindahl indices based on student shares of five racial/ethnic groups (Asian, Native American, Hispanic, non-Hispanic white, and non-Hispanic black). The poverty homogeneity indices are Herfindahl indices based on student shares of two groups: eligible to receive free lunch, and not eligible to receive free lunch.

Sources: Author’s calculations based on data from SDDB, CCD, CCDB, GNIS, and USGS maps.

The ratio of racial heterogeneity has no statistically significant relationship with the number of districts in a metropolitan area. In other words, students are just as segregated in schools in metropolitan areas that contain few districts as they are in metropolitan areas that contain many districts. Households sort themselves into school attendance areas inside districts so that district boundaries have little effect on the racial heterogeneity experienced by students.

The results in the bottom panel of Table 8, which focus on poverty heterogeneity, exhibit a roughly similar pattern but have large standard errors. That is, the point estimates hint that the ratio of poverty heterogeneity has a negative relationship with the number of schools in a metropolitan area. The point estimates do not hint at a similar negative relationship with the number of districts in a metropolitan area (in fact, the estimates are of the wrong sign). The large standard errors are probably caused by measurement error in free-lunch eligibility as an indicator of household income. Unfortunately, a school-level measure of household income is available. In summary, choice among districts may not have much effect on the peers a student actually experiences because households sort themselves into school attendance areas regardless of whether they have much choice among districts. Therefore, the effect of choice on productivity is more likely to be caused by competitive pressure among districts than by student sorting.

**F. Further Evidence on Tiebout Choice and Student Sorting**

In the results discussed thus far, there has been little evidence that Tiebout choice affects productivity through sorting. That is, although Tiebout choice certainly affects how households are sorted across districts, the resulting sorting appears to have little effect on the average level of achievement or per-pupil spending in a metropolitan area. This may indicate that peer effects are small or that there are offsetting benefits and losses when students experience heterogeneous peers. Beneficial peer influences, for instance, may be offset by the difficulties that teachers have in communicating material to heterogeneous classes.

School-level demographic data, however, suggest yet another possibility. It may be that peers have important net effects, but that even metropolitan areas with little Tiebout choice among districts have substantial peer sorting among schools. After all, the peers whom a student actually encounters depend on sorting among schools, and such sorting can occur in areas where there is little Tiebout choice among districts. Table 8 shows how the probability that a student experiences a heterogeneous student body is affected by Tiebout choice.

The dependent variable in the top panel of Table 8 is a ratio. In the numerator is the probability that, in a random encounter with another student in his school, a student from metropolitan area $m$ meets a student of a different racial group. In the denominator is the probability that, in a random encounter with another student in his metropolitan area, a student from metropolitan area $m$ meets a student of a different racial group. If all the schools in a metropolitan area have the same racial composition as one another, then the ratio is equal to one. On the other extreme, if each school in a metropolitan area is racially homogeneous despite the existence of racial heterogeneity in the metropolitan area, then the ratio is equal to zero. Thus, an increase in the ratio indicates that schools are more heterogeneous racially, given the underlying racial heterogeneity of the metropolitan area in which they are located. The dependent variable in the bottom panel of Table 8 is defined analogously, except that—instead of being divided into racial groups—students are divided into a group that is eligible to receive free lunch and another that is not eligible to receive free lunch. The ratio in the bottom panel of Table 8 attempts to measure the poverty heterogeneity of schools.

Apart from the dependent variable, the specification estimated in Table 8 is similar to the specification estimated in the first-stage equations. The only other difference is that choice is measured by the number of districts and schools in the metropolitan area, not by choice indices. It is necessary to avoid choice indices because choice indices have the same basic construction as the dependent variable and the similarity of construction creates spurious correlation. The top panel of Table 8 shows that the racial heterogeneity of a student’s peers is related to the number of schools, but not to the number of districts, in his metropolitan area. This statement holds for both the OLS and IV estimates, although the IV estimates are preferable for all the reasons described above. The ratio of racial heterogeneity has a statistically significant, negative relationship with the number of schools in a metropolitan area: the more schools there are (for a metropolitan area of a given size), the less likely a student is to experience the racial heterogeneity that exists in his metropolitan area.

---

25 That is, the dependent variable and the choice indices are both built upon Herfindahl indices. Using a dependent variable and independent variable with the same structure generates correlated measurement error between the two sides of the equation. This problem is common in the labor-supply literature where the construction of hours of work affects both the dependent variable and an independent variable (the wage). The example in this paper is somewhat less severe, but the scale of the problem can be gauged by the finding that, if randomly generated data on racial composition is assigned to each school, there is statistically significant correlation between the dependent variable (the ratio) and the choice indices. This correlation is generated purely by the parallel construction of the dependent variable and the choice indices.

---

The American Economic Review December 2000
and lowering spending. The effects on productivity, student achievement, and per-pupil spending are substantial. In size (generally one-quarter to one-half of standard deviation) if one considers the potential of Tiebout choice as a policy. That is, the effects are substantial if one considers moving from one end of the Tiebout choice spectrum (a metropolitan area like Miami) to the other (a metropolitan area like Boston). Naturally, most metropolitan areas are between the two ends of the spectrum, and the current variation in Tiebout choice among metropolitan areas explains a modest amount of their differences in school productivity.

There is suggestive evidence that Tiebout choice needs to have financial consequences if it is to produce the productivity effects described. Tiebout choice appears to have larger productivity effects in areas where school districts have greater financial independence.

Where households have more Tiebout-style choice, they are less likely to choose private schools. The fact that households with more Tiebout-style choice are more likely to stay in the public-school system is important for policy planning. For instance, policies that reduce choice among districts (district consolidation) or the benefits of choice (more state control of school expenditures) are likely to increase the share of students in private schools and reduce the share of voters who are interested in the general well-being of public education.

Although Tiebout choice among districts allows more sorting of households by district, the resulting increases in district homogeneity have little net effect on achievement, per-pupil spending, or productivity. The estimated effects of Tiebout choice are not significantly affected by controlling for district-level measures of demographic heterogeneity. This may be due to offsetting effects of having heterogeneous peers (benefits from students experiencing heterogeneous peers may be countered by the difficulties that teachers have in teaching heterogeneous groups of students). However, the results suggest another explanation: much of the sorting of students by racial and income groups is at the school level, not the district. Whether a student experiences peers of different racial groups or different poverty status is not significantly affected by the degree of choice among school districts. Metropolitan areas that have little Tiebout choice among districts have about the same level of sorting among schools as metropolitan areas with a lot of Tiebout choice among districts. This suggests that metropolitan areas with little Tiebout choice among districts could experience the productivity benefits of choice with little change in the nature of student sorting among their schools.

The effects of Tiebout choice are not significantly different for lower-income and higher-income families. Also, the effects of Tiebout choice are generally not significantly different for minority and nonminority families. There is suggestive evidence (rather than statistically significant evidence) that the effects of Tiebout choice may be somewhat larger for better-off, nonminority families. Even if this evidence were strong, however, it would be hard to interpret. One explanation could be that, although sorting caused by Tiebout choice has little overall effect, it has negative effects on disadvantaged students that offset some of the gains they experience from competition. Another explanation could be that most of the variation in the choice measure reflects true variation in choice experienced by nondisadvantaged students.

REFERENCES


