Does Competition Among Public Schools Benefit Students and Taxpayers?

By CAROLINE M. HOXBY*

Tiebout 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.[1] Although district consolidation and states' school finance programs increasingly limit the Tiebout process, it is still the most powerful force in American schooling.[2] 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 self-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 self-sort among schools 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 addition to analyzing the effects of choice on schools' 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

* Department of Economics, Harvard University, Cambridge, MA 02138. The author gratefully acknowledges generous help from the anonymous referees, Henry S. Farber, Jonathan Gruber, Lawrence F. Katz, M. Daniele Paserman, 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. Kenny and Amy B. Schmidt (1994) describe the decline in the number of school districts in the United States.

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 leverage 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 peers 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 useful because theory is indecisive; it is necessary because the empirical challenges are formidable. 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 topographics (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 schools. 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 evince 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 *extend*, not introduce, 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 hardly affect their behavior. Other people would have their choice sets greatly expanded by reform. 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.[3]

The third reason to analyze Tiebout choice 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

[3] See, for instance, Dennis N. Epple and Richard E. Romano (1998). An exception is Thomas Nechyba (1996), who describes a Tiebout equilibrium with multiple public-school districts and uses computable general-equilibrium techniques to predict the effects of vouchers.

school district enacts a policy of *intra*district 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.[4] 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 mechanisms, such as voting on local property tax rates.[5] 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 equilibria 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 taste 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.[6] 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.[7] Tiebout choice takes no account of such spillovers, since it depends on people making choices that are privately optimal. Therefore, 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.[8] 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 is 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 in total, given the fall in price. 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 households 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.[9]

Whether Tiebout choice affects productivity by

---

making school districts fall below minimum efficient scale is, in practice, a nonissue. Empirical evidence suggests that minimum efficient scale for a school district is so small that most metropolitan districts easily exceed it.[10] Also, temporary or partial consolidation between mutually agreeable districts is easy, while breaking up a too-large district 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 directly 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 possibilities. 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 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:

$$(1) \qquad r_m = r(C_m, \cdots), \qquad \frac{\partial r_m}{\partial C_m} > 0.$$

---

[4] In Hoxby (1999b), I show that naive estimates of the effects of intradistrict choice are misleading.

[5] Informal versions of Tiebout's model rely on the analogy to private goods. Formal versions of Tiebout's model that include median-voter politics are offered by Epple et al. (1984, 1993). Past empirical research has often tested whether Tiebout choice actually attains allocative efficiency. Such tests are misguided, given that Tiebout choice operates imperfectly even in the markets where it is most prevalent. It is more reasonable to test whether an increase in Tiebout choice raises private allocative efficiency.

[6] This result is derived by Epple and Glenn J. Platt (1998), who solve a Tiebout model in which households can differ both in their incomes and their tastes for education.

[7] Roland Bénabou (1996) has recently drawn attention to the macroeconomic implications of this familiar point.

[8] Hoxby (1999a) offers a formal principal–agent model of the productivity of schooling producers.

[9] This is a well-known result that has recently been explored in the context of school finance equalization. See Fabio Silva and Jon Sonstelie (1995) and Raquel Fernández and Richard Rogerson (1998).

[10] Randall W. Eberts et al. (1990) present evidence on minimum efficient scale of schools.

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 $\tilde{X}_{km} = (\widetilde{X1}_{km}, \widetilde{X2}_{km}, \cdots)$ be a vector of measures of the heterogeneity of household characteristics $X1$, $X2$, $\cdots$ in school district k in metropolitan area m. Let $\tilde{X}_m = (\widetilde{X1}_m, \widetilde{X2}_m, \cdots)$ 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:

$$(2) \quad \frac{\partial \widetilde{X1}_{km}(C_m, X1_m)}{\partial C_m} \gtreqless ?$$

$$0 \text{ for } k = 1, \ldots, K,$$

and so on for characteristics $X2$, $X3$, $\ldots$. The inequality states that Tiebout choice affects (in an unknown direction) the heterogeneity of characteristic $X1$ 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, demographics, and numerous environmental factors:

$$(3) \quad \frac{A_{ikm}}{\ln(E_{km})}$$

$$= f(r(C_m, \cdots), X_{ikm}, \bar{X}_{km}, \tilde{X}_{km}, \bar{X}_m, \tilde{X}_m),$$

where A represents student achievement and $\ln(E)$ represents the log of per-pupil expenditure.[11] $\bar{X}_{km}$ and $\tilde{X}_m$ are defined above. X_{ikm} is a vector of characteristics of household i in district k in market m; $\bar{X}_{km}$ is a vector of mean characteristics of households in district k in market m; and $\bar{X}_m$ 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 is C_m, not $r(C_m, \cdots)$. 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:

$$(4) \quad \frac{A_{ikm}}{\ln(E_{km})}$$

$$= \beta_1 C_m + X_{ikm}\beta_2 + \bar{X}_{km}\beta_3 + \tilde{X}_{km}\beta_4$$

$$+ \bar{X}_m\beta_5 + \tilde{X}_m\beta_6 + \varepsilon_{ikm} + \varepsilon_{km} + \varepsilon_m.$$

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 dramatic 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 $\tilde{X}_{km}$ if it includes all available measures of heterogeneity that are likely to have a significant effect on productivity. I test whether the estimate of β_1 changes significantly when measures of district heterogeneity are excluded from the

---

[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.

equation: this is a partial test of whether β_1 mainly evinces the effects of competition (as opposed to sorting) on productivity.[12]

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 idiosyncratic 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, exchanging places with households 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.[13]

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

---

[13] As demonstrated by Kenny and Schmidt (1994), the twentieth-century history of American public education is a history of *consolidation*. Although most of the consolidation has affected rural districts, the number of metropolitan districts has also decreased by nearly 40 percent since 1950. Thus, a metropolitan area that has little observed Tiebout choice is likely to be either an area that has always had large districts or an area that has experienced significant district consolidation. The latter type of metropolitan area is likely to introduce endogeneity.

---

[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, in so 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 earner 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 the 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:

$$1 - H_m = 1 - \sum_{k-1}^{K} s_{km}^2,$$

$$s_{km} = \frac{land\ area_{km}}{land\ area_m};$$

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

$$1 - H_m = 1 - \sum_{k=1}^{K} s_{km}^2,$$

$$s_{km} = \frac{enrollment_{km}}{enrollment_m};$$

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

[14] See Universal Publishing Company (1998).

(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 less biased than 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

[15] In the United States, revenue is raised and per-pupil spending is determined at the district level. Per-pupil spending is undefined at the school level because a district's costs cannot be meaningful 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.

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," natural barriers 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 streams variables for the data section, but the implied first-stage equation states that the degree of choice in a metropolitan area is a function of its number of streams:

[16] 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.

$$(5) \quad C_m = S_m \alpha_1 + \bar{X}_m \alpha_2 + \tilde{X}_m \alpha_3 + v_m,$$

where S_m is a vector of measures of the number of streams in metropolitan area m.[17] $\bar{X}_m$ and $\tilde{X}_m$ include many area characteristics that help to ensure that C_m 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 tests 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:

$$(6) \quad A_{ikm} = \gamma_1 C_m + \gamma_2 X1_{ikm} + \gamma_3 X2_{ikm}$$
$$+ \gamma_4 \overline{X1}_{km} + \gamma_5 \overline{X2}_{km} + \gamma_6 \overline{X1}_m$$
$$+ \gamma_7 \overline{X2}_m + v_{ikm} + v_{km} + v_m.$$

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: $X1$ and $X2$. 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.[18] 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 γ_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_m (or any variable at the level of the educational market) is affected by another covariate only through the covariance between C_m 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_m, 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

---

[18] That is, consider the case in which some metropolitan areas are broken into many equal-sized districts and others are broken into a 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.

---

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:

$$(7) \quad \sum_{i=1}^{NM} x1_{ikm} c_m = \sum_{i=1}^{NM} \overline{x1}_m c_m,$$

$$\sum_{i=1}^{NM} x1_{ikm} \overline{x1}_m = \sum_{i=1}^{NM} \overline{x1}_m^2,$$

$$\sum_{i=1}^{NM} x1_{ikm} \overline{x2}_m = \sum_{i=1}^{NM} \overline{x1}_m \overline{x2}_m,$$

and so on.[19] Each constraint holds because deviations from a mean sum to zero for the observa-

---

[19] Variables in lowercase are in deviations-from-means form. The full set of additional constraints includes the three written above plus:

$$\sum_{i=1}^{NM} x2_{ikm} c_m = \sum_{i=1}^{NM} \overline{x2}_m c_m, \quad \sum_{i=1}^{NM} x2_{ikm} \overline{x1}_m = \sum_{i=1}^{NM} \overline{x2}_m \overline{x2}_m,$$

$$\sum_{i=1}^{NM} x2_{ikm} \overline{x2}_m = \sum_{i=1}^{NM} \overline{x2}_m^2,$$

$$\sum_{i=1}^{NM} \overline{x1}_{km} c_m = \sum_{i=1}^{NM} \overline{x1}_m c_m, \quad \sum_{i=1}^{NM} \overline{x1}_{km} \overline{x1}_m = \sum_{i=1}^{NM} \overline{x1}_m^2,$$

$$\sum_{i=1}^{NM} \overline{x1}_{km} \overline{x2}_m = \sum_{i=1}^{NM} \overline{x1}_m \overline{x2}_m,$$

$$\sum_{i=1}^{NM} \overline{x2}_{km} c_m = \sum_{i=1}^{NM} \overline{x2}_m c_m, \quad \sum_{i=1}^{NM} \overline{x2}_{km} \overline{x1}_m = \sum_{i=1}^{NM} \overline{x2}_m \overline{x2}_m,$$

$$\sum_{i=1}^{NM} \overline{x2}_{km} \overline{x2}_m = \sum_{i=1}^{NM} \overline{x2}_m^2,$$

$$\sum_{i=1}^{NM} \overline{x1}_{ikm} \overline{x1}_{km} = \sum_{i=1}^{NM} \overline{x1}_{km}^2, \quad \sum_{i=1}^{NM} x1_{ikm} \overline{x2}_{km} = \sum_{i=1}^{NM} \overline{x1}_{km} \overline{x2}_{km},$$

$$\sum_{i=1}^{NM} x2_{ikm} \overline{x1}_{km} = \sum_{i=1}^{NM} \overline{x1}_{km} \overline{x2}_{km}, \quad \sum_{i=1}^{NM} x2_{ikm} \overline{x2}_{km} = \sum_{i=1}^{NM} \overline{x2}_{km}^2.$$

---

tions over which the mean is taken. For instance, the first constraint listed can be rewritten as:

$$(8) \quad \sum_{i=1}^{NM} x1_{ikm} c_m = \sum_{m=1}^{M} \sum_{\substack{i=1 \\ in\ m}}^{i=N\ in\ m} (\overline{x1}_m + u1_{ikm}) c_m$$

$$= \sum_{i=1}^{NM} \overline{x1}_m c_m$$

$$+ \sum_{m=1}^{M} c_m \sum_{\substack{i=1 \\ in\ m}}^{i=N\ in\ m} u1_{ikm},$$

where $u1_{ikm}$ is defined so that $x1_{ikm} = \overline{x1}_m + u1_{ikm}$. But, by definition,

$$(9) \quad \sum_{\substack{i=1 \\ in\ m}}^{i=N\ in\ m} u1_{ikm} = 0 \quad \forall m.$$

Therefore:

$$(10) \quad \sum_{i=1}^{NM} x1_{ikm} c_m = \sum_{i=1}^{NM} \overline{x1}_m c_m.$$

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 γ_1:

$$(11) \quad \hat{\gamma}_1 = \frac{\begin{matrix} \sum x_m^2 \sum z_m^2 \sum y_m c_m - [\sum x_m z_m]^2 \sum y_m c_m \\ + \sum x_m z_m \sum x_m c_m \sum y_m z_m \\ + \sum x_m z_m \sum z_m c_m \sum y_m x_m \\ - \sum x_m^2 \sum z_m c_m \sum y_m z_m \\ - \sum z_m^2 \sum x_m c_m \sum y_m x_m \end{matrix}}{\begin{matrix} \sum x_m^2 \sum z_m^2 \sum c_m^2 \\ - 2 \sum x_m z_m \sum x_m c_m \sum z_m c_m \\ - \sum c_m^2 [\sum x_m z_m]^2 - \sum x_m^2 [\sum z_m c_m]^2 \\ - \sum z_m^2 [\sum x_m c_m]^2 \end{matrix}},$$

where all the sums are from $i = 1$ to NM.

One obtains exactly the same formula for $\hat{\gamma}_1$ if one aggregates equation (6) to the district level (thereby eliminating the terms beginning with γ_2 and γ_3), takes the OLS first-order conditions, imposes the constraints listed in footnote 19 that are relevant, and solves. Moreover, one obtains exactly the same formula for $\hat{\gamma}_1$ if one aggregates equation (6) up to the district level, takes the OLS first-order conditions, 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?

Tiebout choice in an educational market affects the mean characteristics of individuals in each school district. That is, district mean

variables such as $X1_{km}$ 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 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:

$$(12) \quad C_m = S_m\delta_1 + \delta_2\overline{X1}_m + \delta_3\overline{X2}_m + \omega_m{}^{21}$$

If we substitute the predicted value, $\hat{C}_m$, from this regression into the 2SLS formula for γ_1—

[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:

$$C_m = S_m\delta_1 + \delta_2\overline{X1}_m + \delta_3\overline{X2}_m + \omega_m,$$

$$C_m = S_m\delta_1 + \delta_2\overline{X1}_m + \delta_3\overline{X2}_m + \delta_4\overline{X1}_{km} + \delta_5\overline{X2}_{km}$$

$$+ \delta_6\overline{X1}_{lkm} + \delta_7 X2_{ikm} + \omega_m + \omega_{km} + \omega_{lkm}.$$

$$(13) \quad \hat{\gamma}_1^{2SLS} = \frac{\begin{array}{c}\sum x_m^2 \sum z_m^2 \sum y_m\hat{c}_m - [\sum x_m z_m]^2 \sum y_m\hat{c}_m + \sum x_m z_m \sum x_m\hat{c}_m \sum y_m z_m + \sum x_m z_m \sum z_m\hat{c}_m \sum y_m x_m \\ - \sum x_m^2 \sum z_m\hat{c}_m \sum y_m z_m - \sum z_m^2 \sum x_m\hat{c}_m \sum y_m x_m\end{array}}{\sum x_m^2 \sum z_m^2 \sum \hat{c}_m^2 - 2 \sum x_m z_m \sum x_m\hat{c}_m \sum z_m\hat{c}_m - \sum \hat{c}_m^2[\sum x_m z_m]^2 - \sum x_m^2[\sum z_m\hat{c}_m]^2 - \sum z_m^2[\sum x_m\hat{c}_m]^2},$$

—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 $\hat{C}_m$ varies only at the level of the educational market. Furthermore, the endogeneity of the district-level means to C_m (and the question of whether one instruments for them) is not relevant to the estimate of $\hat{\gamma}_1^{2SLS}$. That is, if one has a valid instrument for C_m, one can get an unbiased estimate of γ_1 without instrumenting for the district-level means. Again, the intuition is that the effect of choice *through its effect on district mean characteristics* is mechanically equal to zero.

It would be incorrect, however, to interpret the coefficients on district mean characteristics as though they were exogenous. Therefore, I include the district mean characteristics purely to improve the fit of the equation. Since I cannot give them a ready structural interpretation, I do not interpret them at all in the results section.[22]

D. Measures of District Heterogeneity

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

$$(14) \quad A_{ikm} = \beta_1 C_m + \beta_{2a}X1_{ikm} + \beta_{2b}X2_{ikm}$$

$$+ \beta_{3a}\overline{X1}_{km} + \beta_{3b}\overline{X2}_{km}$$

[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.

$$+ \beta_{4a}\overline{X1}_{km} + \beta_{4b}\overline{X2}_{km}$$

$$+ \beta_{5a}\overline{X1}_m + \beta_{5b}\overline{X2}_m$$

$$+ \beta_{6a}\overline{X1}_m + \beta_{6b}\overline{X2}_m$$

$$+ \varepsilon_{ikm} + \varepsilon_{km} + \varepsilon_m.$$

This is a scalar version of equation (4) that is without loss of generality. I demonstrated above that the inclusion of district-level means does not affect the estimated coefficient on C_m. The same cannot be said for the measures of district heterogeneity. District heterogeneity is endogenous to the degree of choice, and the endogenous changes in heterogeneity will not cancel out across districts in an educational market. For example, suppose that $X1$ is family income. If we compare two educational markets that have the same heterogeneity of family income in their populations, the market with greater choice is likely to have districts that are less heterogeneous. That is, the mean level of district heterogeneity in an educational market is likely to be correlated with choice. (Recall that choice need not increase the within-district homogeneity of all characteristics.)

In short, the mean level of district heterogeneity in an educational market depends on the *interaction* between choice and the heterogeneity of the market's population. If an educational market is homogeneous initially, then no increase in choice can increase the homogeneity of its average district. However, if a educational market is heterogeneous initially, then an increase in choice is likely to change the homogeneity of its average district. Represent this interaction between choice and the heterogeneity of a market population by functions such as:

$$(15) \quad \sum_{k=1 \text{ in } m}^{K \text{ in } m} \overline{X1}_{km} = g^1(C_m, \overline{X1}_m).$$

If we define $w1_{km}$ so that

$$(16) \quad \widetilde{X1}_{km} = \left(\sum_{k=1 \ in \ m}^{K \ in \ m} \widetilde{X1}_{km} \right) + w1_{km}$$

$$= g^1(C_m, \widetilde{X1}_m) + w1_{km},$$

and write similar definitions for $X2$, then we can rewrite equation (15) as follows:

$$(18)$$

$$A_{ikm} = \beta_1 C_m + \beta_{2a} X1_{ikm} + \beta_{2b} X2_{ikm}$$

$$+ \beta_{3a} \overline{X1}_{km} + \beta_{3b} \overline{X2}_{km}$$

$$+ \beta_{4a} g^1(C_m, \widetilde{X1}_m)$$

$$+ \beta_{4b} g^2(C_m, \widetilde{X2}_m) + \beta_{4a} w1_{km}$$

$$+ \beta_{4b} w2_{km} + \beta_{5a} \overline{X1}_m + \beta_{5b} \overline{X2}_m$$

$$+ \beta_{6a} \widetilde{X1}_m + \beta_{6b} \widetilde{X2}_m$$

$$+ \varepsilon_{ikm} + \varepsilon_{km} + \varepsilon_m.$$

The sixth through ninth right-hand terms of equation (18) are of interest. The terms in $w1$ and $w2$ will be irrelevant to the estimated coefficient on C_m since they sum to zero over each educational market. The $g^1()$ and $g^2()$ terms, however, need to be instrumented (for the same reasons C_m needs to be instrumented) if they are not to cause bias. The $g^1()$ and $g^2()$ terms embody the way that choice interacts with an educational market's overall heterogeneity to produce more or less heterogeneous districts. Thus, it is natural to instrument for $g^1()$ with the interaction between $\widetilde{X1}_m$ and the streams variables, to instrument for $g^2()$ with the interaction between $\widetilde{X2}_m$ and the streams variables, and so on.

VIII. Data

I use several sources of data, all matched geographically at the school-district or metropolitan-area level.

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 of Education Statistics *Common Core of Data* (CCD) (U.S. Department of Education, 1995), which contains administrative data on enrollment, instructional staff, and student demographics for every school in the United States. The only 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 of each district, and measures of the demographic heterogeneity of each school district. For example, the equations contain not only mean household income but also the Gini coefficient based on household income. The equations contain not only the percentages of the population that fit into each racial group and educational attainment group, but also indices of racial heterogeneity, ethnic heterogeneity, and educational heterogeneity.[23] I derive demographic measures at the metropolitan-area level

---

[23] All three indices are based on Herfindhal indices 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 still 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 Alesina et al. (1999) for a complete description of the index.

---

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 streams variables are derived 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 multiple counties) and the number of smaller streams (from GNIS). There are practical reasons for creating two stream variables, but the division is also useful for testing the second instrumental variables condition. Smaller streams are frequently associated with district boundaries, which suggests that they were once natural barriers. They are far too small, however, to affect present-day commuting times or to have determined local industrial history. Thus, smaller streams do not fit the few stories that suggest how streams might affect student achievement through routes other than district boundaries.

The bottleneck in this and similar studies is getting data on achievement. The school and

demographic data are all *population* data, but *sample* data must be used to get many interesting measures of achievement such as test scores, educational attainment, and income. It is best to have student data that is matched to individual school districts, so I use the restricted-access version of the *National Education Longitudinal Survey* (NELS) (U.S. Department of Education, 1994a) for 8th-, 10th-, and 12th-grade test scores. Equation (4) can be consistently estimated at the metropolitan-area level, however, so I also use measures of student achievement from the restricted-access version of the *National Longitudinal Survey of Youth* (NLSY) (U.S. Department of Labor, 1998), which I match at the metropolitan-area level. The NLSY sample is older (ages 32–40 in the 1997 data) so it is a better source for measures of achievement like college completion and income. The three measures of achievement taken from the NLSY are the math knowledge score from the Armed Services Vocational Aptitude Battery (ASVAB), highest grade completed, and earned income at age 32.[25] None of the achievement measures is definitive, but together the measures form a picture. Students from both samples are matched with the geographic area where they attended high school and with the most appropriate year of *Census of Population and Housing* (U.S. Department of Commerce, 1983a; U.S. Department of Education, 1994b) and *Census of Governments* data (1990/92 data for NELS students, 1980/82 data for NLSY students) (U.S. Department of Commerce, 1984; U.S. Department of Education, 1994b).

The school and district data have virtually no missing observations. All 6,523 regularly functioning metropolitan districts are included in the regressions. Among NELS and NLSY students who attended high schools in metropolitan areas, fewer than 100 had to be dropped for missing background data. Variation in the availability of the achievement measures accounts for the variation in the number of observations among regressions, and most of the variation in availability is not due to missing observations but to the survey structure or the nature of the achievement mea-

---

[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 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

Panel A: Descriptive Statistics

Measure	Mean	Standard deviation	Standard deviation, controlling for metropolitan-area size[a]
Index of choice among districts, based on enrollment	0.686	0.271	0.250
Index of choice among districts, based on land area	0.761	0.269	0.252
Districts in metropolitan area	21.132	27.611	18.751
Difference in commuting time (minutes) between the district with the third shortest commute and the district with the shortest commute	6.498	8.551	
Index of choice among *schools*, based on enrollment	0.974	0.069	0.062

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

Measure	(1)	(2)	(3)	(4)	(5)
(1) Index of choice among districts, based on enrollment	1.00				
(2) Index of choice among districts, based on land area	0.86	1.00			
(3) Districts in metropolitan area	0.45	0.42	1.00		
(4) Difference in commuting time (minutes) between the district with the third shortest commute and the district with the shortest commute	0.65	0.71	0.40	1.00	
(5) Index of choice among *schools*, based on enrollment	0.14	0.15	0.07	0.01	1.00

Note: There are 316 observations (metropolitan areas). See text for variable definitions.
Sources: Author's calculations based on data from SDDB and CCD.
[a] 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.

sure.[26] Every table has notes that describe the variables and observations included.

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 *schools'* enrollments has a mean of 0.97. The standard deviation of its residual (controlling for metropolitan-area size) is only is 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 necessary 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.[27] 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 variable 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 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

TABLE 2—SELECTED COEFFICIENTS FROM THE IMPLIED FIRST-STAGE REGRESSION

	Dependent variable	
	Index of choice among districts (based on enrollment)	Index of choice among schools (based on enrollment)
Number of larger streams in metropolitan area[a]	0.080 (0.040)	−0.040 (0.045)
Number of smaller streams in metropolitan area[a]	0.034 (0.007)	0.004 (0.004)
Population of metropolitan area (thousands)	0.015 (0.013)	0.001 (0.006)
Land area of metropolitan area (thousands of square miles)	0.005 (0.005)	0.003 (0.002)
Mean of log(income) of metropolitan area	−0.246 (0.155)	−0.263 (0.067)
Gini coefficient of metropolitan area	−3.581 (0.811)	−1.500 (0.349)
Share of metropolitan-area population that is 0 to 19 years of age	1.859 (0.799)	0.239 (0.344)
Share of metropolitan-area population that is 65 years of age or older	0.447 (0.605)	0.558 (0.260)
Share of metropolitan-area population that is Asian	1.877 (1.003)	0.894 (0.432)
Share of metropolitan-area population that is black	0.827 (0.414)	0.054 (0.178)
Share of metropolitan-area population that is Hispanic	0.089 (0.181)	0.114 (0.078)
Index of racial homogeneity of metropolitan area	−0.183 (0.491)	0.014 (0.212)
Index of ethnic homogeneity of metropolitan area	0.698 (0.697)	−0.078 (0.300)
Share of adults in metropolitan area the highest grade which is some college	−0.998 (0.407)	−0.215 (0.176)
Share of adults in metropolitan area the highest grade which is B.A. or more	0.953 (0.419)	0.520 (0.181)
Index of educational homogeneity of metropolitan area	−2.948 (1.025)	−0.831 (0.441)
Indicator variables for the nine Census regions of the United States	yes	yes
$F_{2,291}$-statistic, joint significance of the excluded instruments	24.370	0.860

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 stream variables are the excluded instruments. The dependent variables have the following means (standard deviations): 0.686 (0.271), 0.974 (0.089). The two streams variables have the following means (standard deviations): 7.893 (14.782), 182.748 (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 SDDB, CCD, CCDB, GNIS, and USGS maps.

[a] The variables are measured in hundreds.

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

B. *The Effect of Tiebout 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 TIEBOUT CHOICE ON ACHIEVEMENT: INSTRUMENTAL VARIABLES ESTIMATES OF SELECTED COEFFICIENTS

	Dependent variable: 12th-grade reading score
Index of choice among districts, based on enrollment	5.770 (2.208)
Log(household income)	1.536 (0.164)
Female	1.959 (0.227)
Asian	0.284 (0.591)
Black	−5.491 (0.497)
Hispanic	−2.866 (0.518)
Parents' highest grade is some college	2.306 (0.296)
Parents' highest grade is B.A. or more	5.453 (0.299)
Mean of log(income) of metropolitan area	−5.421 (5.527)
Gini coefficient of metropolitan area	−12.770 (12.022)
Share of metropolitan-area population that is Asian	−5.620 (13.067)
Share of metropolitan-area population that is black	−0.732 (6.063)
Share of metropolitan-area population that is Hispanic	0.247 (3.516)
Index of racial homogeneity of metropolitan area	−9.598 (7.840)
Index of ethnic homogeneity of metropolitan area	16.313 (10.697)
Share of adults in metropolitan area the highest grade which is some college	5.274 (7.135)
Share of adults in metropolitan area the highest grade which is B.A. or more	3.163 (5.929)
Index of educational homogeneity of metropolitan area	−5.447 (12.952)
Indicator variables for the nine Census regions of the United States	yes

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 NELS, SDDB, 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

TABLE 4—EFFECT OF TIEBOUT CHOICE ON ACHIEVEMENT:
COEFFICIENT ON INDEX OF CHOICE FOR VARIOUS SPECIFICATIONS

Specification:	Dependent variable:					
	8th-grade reading score	10th-grade math score	12th-grade reading score	ASVAB math knowledge	Highest grade attained	ln(income) at age 32
Base IV specification (see previous table)	3.818	3.061	5.770	2.747	1.381	0.151
	(1.591)	(1.494)	(2.208)	(1.570)	(0.469)	(0.072)
Base specification estimated by OLS	−0.236	−0.733	−1.434	2.024	0.323	0.055
	(0.493)	(0.564)	(0.650)	(0.561)	(0.150)	(0.029)
Base IV without measures of district heterogeneity	4.649	2.573	6.084	does not	does not	does not
	(1.598)	(1.478)	(2.276)	apply	apply	apply
Base IV aggregated to metropolitan-area level	5.137	2.663	7.149	2.860	1.285	0.170
	(3.428)	(3.419)	(4.844)	(4.587)	(1.229)	(0.239)
Base IV with choice index based on district land area	4.761	2.875	5.803	2.855	1.516	0.159
	(1.429)	(1.486)	(2.179)	(1.597)	(0.517)	(0.073)
Base IV with choice index based on *schools'* enrollment	61.357	−57.414	−130.577	−18.832	8.031	1.436
	(44.128)	(52.959)	(95.960)	(23.835)	(12.013)	(2.341)
Base IV with choice index interacted with family income:						
Effect for low-income students	3.364	2.825	4.350	4.148	1.564	0.189
	(1.776)	(1.767)	(2.297)	(1.633)	(0.447)	(0.094)
Effect for not-low-income students	4.028	3.043	5.810	5.639	1.708	0.193
	(1.802)	(1.747)	(2.303)	(1.735)	(0.473)	(0.091)
Base IV with choice index interacted with minority status:						
Effect for minority students	−0.376	−2.830	4.234	5.485	1.835	0.188
	(2.761)	(3.604)	(4.218)	(2.629)	(0.730)	(0.098)
Effect for nonminority students	4.589	5.116	6.096	2.907	1.267	0.187
	(1.685)	(1.769)	(2.205)	(1.708)	(0.572)	(0.083)
Test statistic, omnibus overidentification test (distributed $\chi^2_{d.f.-1}$)	0.404	0.001	0.001	0.118	0.237	0.020
Test statistic, exogeneity of larger streams variable (distributed $\chi^2_{d.f.-1}$)	0.003	0.001	0.001	0.116	0.251	0.021

Notes: The base specification is shown in the previous table. The notes for that table apply to this table. The test scores have means of approximately 50 and standard deviations of approximately 10. Highest grade completed has a mean of 13.928 and a standard deviation of 2.855. The log of income at age 32 has a mean of 9.655 and a standard deviation of 1.152. Observations are metropolitan-area students from the NELS (three left-hand columns) and the NLSY (three right-hand columns). The number of observations in each column are: 10,790 (from 211 metropolitan areas), 7,776 (from 211 metropolitan areas), 6,119 (from 209 metropolitan areas), 7,112 (from 218 metropolitan areas), 7,538 (from 221 metropolitan areas), and 5,944 (from 209 metropolitan areas). The number of observations varies due to the availability of the dependent variable (see text). Low-income families are those whose household income is less than or equal to 70 percent of mean household income in their metropolitan area. Minority students are black, Hispanic, or Native American.
Sources: Author's calculations based on data from NELS, NLSY, SDDB, CCD, CCDB, GNIS, and USGS maps.

parents have no college education, students whose parents have some college education (but no baccalaureate degree) have reading scores that are 2.3 standardized points higher, and students whose parents have at least one baccalaureate degree between them have reading scores that are 5.5 standardized points higher.[28] Most of the

[28] One should interpret the coefficients on the individual-level variables as the effects of an individual's background on his achievement. Some of the effects of an individual background are indirect, in the sense that they work through parents' decisions to put their children into a particular environment. For instance, one of the effects of coming from a well-off family works through living, in all probability, in a safer neighborhood. It is normal and inevitable that the individual-level variables pick up both direct and indirect effects of family background. They do this to some extent even though there are district-level variables included in the regression

metropolitan-area characteristics do not have a statistically significant effects on achievement. Several of the coefficients on district characteris-

(especially since district-level variables are only crude measures of the environment that a family provides to 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 neighborhood 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, direct and indirect, of an individual's background on his achievement.

tics 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, 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 the above results are statistically 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. If we were to interpret them naively, the OLS results in the second row of Table 4 would suggest that Tiebout choice has no effect or small positive effects on achievement.

The OLS results reveal the sign of the bias due to omitted variables and endogeneity. They suggest that successful districts do attract households with school-aged children and do attract other districts into consolidation. Also, achievement may be negatively affected by unobserved factors, such as dissension, 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 heterogeneity are 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 than, the estimates from the base specification. Given the foregoing discussion of aggregation issues, this is what one would expect.[29]

[29] If the weights were perfect and there were no measurement or sampling errors, then the coefficient estimates on metropolitan-level variables would be the same in aggregate specifications and specifications that include district-level mean variables but not district-level heterogeneity variables. The standard errors would, however, be larger in the aggregate specifications. For instance, in Table 4, the third and fourth rows would contain the same estimates, but the fourth row would have larger standard errors. In practice, however, measurement error, sampling error, and imperfect weighting prevent the estimates in the third and fourth rows from being *exactly* the same. First, sampling error occurs because the NELS and NLSY are samples from the population. Second, some of the variables from the *Census of Population and Housing* (U.S. Department of Commerce, 1983a) (regardless of whether they are drawn from the SDDB or CCDB) are based on a 20-percent sample of the population, not the entire Census population. Third, all of the variables are potentially measured with error and the measurement error in the NELS and NLSY variables need not be identical to that in the Census variables. Fourth, there is measurement error in the creation of district-level statistics because some Census blocks straddle district

The fifth row of Table 4 shows the less preferred measure of choice based on land area. Recall that this measure is likely 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).[30]

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 in row six 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.[31]

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 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 differ for minority and nonminority students. For 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

---

boundaries. Fifth, even when used with the best available methods, the survey weights supplied by the NELS and NLSY generate statistics that are not *exactly* representative of the underlying population. Sixth, different district weights are ideal for different district variables—some ideal weights would be based on the population, a few would be based on the adult population, and a couple would be based on the number of households. The ideal district weight that is modal is based on the population, so that is what I use. The alternative choices generate similar results. Not using the ideal district weight for every variable is the sixth and final reason why the estimates in the third and fourth rows are not exactly the same.

[30] The difference between the land-based OLS and IV results is smaller than the difference between the enrollment-based OLS and IV results. This suggests that the land-based measure does, in fact, eliminate some of the endogeneity.

[31] That is, the cost of moving a student between schools within a district is much smaller than the cost of moving a student between districts. The data required to compute a land-based measure of choice among schools would make such a measure nearly impossible to compute, even if it were likely to be useful without instrumenting.

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 instrumental variables. The omnibus test attempts to show whether variation 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 χ^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).[32] The omnibus test consistently fails to reject the null hypothesis that streams affect student achievement only via their effect on choice.

The other test shown is 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 too small to affect modern life. The test statistics, which are distributed as χ^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 attainment. Several of 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.

---

[32] Both the omnibus test and the Hausman test are described by Jerry Hausman (1983). Conventional formulas for the tests are *incorrect* whenever Moulton standard errors are appropriate. Hoxby and M. Daniele Paserman (1998) describe the problem and provide a method for calculating correct test statistics.

TABLE 5—EFFECT OF TIEBOUT CHOICE ON PER-PUPIL SPENDING AND PRIVATE SCHOOLING: INSTRUMENTAL VARIABLES ESTIMATES OF SELECTED COEFFICIENTS

	Dependent variable	
	Log(per-pupil spending)	Share of students in private school
Index of choice among districts, based on enrollment	−0.076 (0.034)	−0.042 (0.012)
Population of metropolitan area (thousands)	0.004 (0.003)	0.006 (0.003)
Land area of metropolitan area (thousands of square miles)	0.003 (0.002)	−0.001 (0.001)
Mean of log(income) of metropolitan area	0.543 (0.115)	0.044 (0.040)
Gini coefficient of metropolitan area	0.372 (0.601)	−0.251 (0.247)
Share of metropolitan area population that is Asian	−0.620 (0.715)	0.264 (0.225)
Share of metropolitan area population that is black	−0.070 (0.299)	0.097 (0.074)
Share of metropolitan area population that is Hispanic	0.463 (0.146)	−0.012 (0.027)
Index of racial homogeneity of metropolitan area	0.762 (0.331)	0.096 (0.087)
Index of ethnic homogeneity of metropolitan area	−0.581 (0.492)	0.001 (0.127)
Share of adults in metropolitan area the highest grade which is some college	0.480 (0.348)	−0.327 (0.103)
Share of adults in metropolitan area the highest grade which is B.A. or more	−0.426 (0.302)	−0.081 (0.076)
Index of educational homogeneity of metropolitan area	−1.231 (0.945)	−0.609 (0.233)
Indicator variables for the nine Census regions of the United States	yes	yes

Notes: IV estimates based on school districts in metropolitan areas. Standard errors are in parentheses and use formulas (Moulton, 1986) for data grouped by metropolitan areas. The regressions are weighted so that each metropolitan area receives equal weight. There are 6,523 observations (school districts) in 316 metropolitan areas. 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 choice index has a mean of 0.765 and a standard deviation of 0.236. 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 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 district. See text for variable definitions.
Sources: Author's calculations based on data from SDDB, CCD, CCDB, GNIS, and USGS maps.

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 uninterpretable because of their extremely large standard errors. The statistic for the omnibus test of the exogeneity of the instrumental variables fails to reject the null hypothesis that streams affect 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.

Are these effects on per-pupil spending large or small? Again, it depends on the question one wants to answer. A 7.6- to 10.1-percent decrease in per-pupil spending is substantial. This decrease is associated with an increase from 0 to 1 in the choice index and thus reflects the potential of Tiebout choice as a policy. However, a standard deviation in the choice index is only 0.27, so differences in Tiebout choice account for only a modest amount of the variation in

TABLE 6—EFFECT OF TIEBOUT CHOICE ON SCHOOL INPUTS AND PRIVATE SCHOOLING: COEFFICIENT ON INDEX OF CHOICE FOR VARIOUS SPECIFICATIONS

	Dependent variable:		
Specification:	ln(per-pupil spending)	Student-teacher ratio	Share of students in private school
Base IV specification (see previous table)	−0.076 (0.034)	−2.669 (1.084)	−0.042 (0.018)
Base specification estimated by OLS	−0.072 (0.022)	0.375 (0.268)	0.006 (0.006)
Base IV without measures of district heterogeneity	−0.058 (0.033)	−2.493 (0.994)	−0.067 (0.022)
Base IV aggregated to metropolitan-area level	−0.064 (0.049)	−2.448 (1.463)	−0.069 (0.031)
Base IV with choice index based on district land area	−0.101 (0.043)	−2.582 (1.122)	−0.043 (0.020)
Base IV with choice index based on *schools'* enrollment	−0.803 (0.934)	−3.828 (5.372)	−0.180 (0.159)
Test statistic, omnibus overidentification test (distributed $\chi^2_{d.f.=1}$)	1.222	0.085	0.144
Test statistic, exogeneity of larger streams variable (distributed $\chi^2_{d.f.=1}$)	1.021	0.047	0.171

Notes: The base specification is shown in the previous table. The notes for that table apply to this table.
Sources: Author's calculations based on data from SDDB, CCD, CCDB, GNIS, and USGS maps.

per-pupil spending in the United States. Undoubtedly, the most striking result is not the positive effect of choice on student achievement or the negative effect of choice on per-pupil spending but the opposite direction of the achievement and spending results. An increase in choice among districts lowers per-pupil spending with no loss—in fact, a gain—in student achievement. Of course, this has powerful implications for productivity.

The middle column shows that, interestingly, the estimated decrease in per-pupil spending is associated with a *decrease* in the student-teacher ratio (an increase in teaching resources per student). An increase from 0 to 1 in the index of Tiebout choice generates a decrease of 2.7 students in the student-teacher ratio. The estimated decrease in the student-teacher ratio is similar for the specification without measures of district heterogeneity. The version of the specification that is aggregated to the metropolitan-area level produces estimates that are, as expected, similar to those of the base specification but less precise. The OLS point estimates are not statistically significantly different from zero.

Given that the per-pupil spending results and student-teacher ratio results are based on ex-

actly the same data, the results imply that districts that face more Tiebout choice allocate their lower levels of per-pupil spending in such a way that they actually have smaller student-teacher ratios. This suggests that Tiebout choice makes districts allocate money away from other inputs and towards reducing the student-teacher ratio. This suggests that reducing the student-teacher ratio is either an unusually productive use of funds or a policy that enjoys unusual popularity with parents.[33]

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

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

[33] In Hoxby (1999b), I present empirical evidence that choice tends to make schools fit parents' stated preferences. In addition to studying student-teacher ratios, one can examine the effect of Tiebout choice on teacher salaries and teacher quality. Such an examination is beyond the scope of this paper because there are several important channels—including unionization—by which salaries and quality might be affected by choice.

sults are included in Tables 5 and 6, and it will be convenient, at this point, to digress briefly and discuss them.

The first row of Table 6 shows that choice among public schools is a substitute for choice of private schools. The IV estimates indicate that an increase in the choice index from 0 to 1 causes private-school enrollment to fall by 4.2 percentage *points*. To understand the significance of this decrease, recall that most metropolitan areas have private-school enrollment rates between 9 and 14 percent. That is, if exercised to its full potential, Tiebout choice can have a dramatic effect on the percentage of children who attend private schools. Because a standard deviation in the choice index is 0.27, differences in Tiebout choice account for a more modest, but still substantial, amount of the variation in private schooling in the United States. The trade-off between parents choosing among public-school districts and parents choosing outside the public sector altogether is important for policy decisions. As shown by Thomas Nechyba (1996), Epple and Romano (1998), and Gerhard Glomm and B. Ravikumar (1998), when families with a strong taste for education leave the public sector by shifting their children into private schools, their voting behavior changes radically and their decisions have a disproportionate effect on support for public education.

Naive OLS estimation suggests that Tiebout choice has no effect on private schooling. The significant difference between the OLS and IV estimates gives us important evidence on the nature of endogeneity. An unsuccessful public-school district tends to drive its students into private schools. Because this phenomenon increases the concentration of public-school students in a few districts, it endogenously lowers the choice index based on district enrollment. (This is not to suggest that the private schooling is the major source of endogeneity in observed public-school choice. The major sources of endogeneity are probably district consolidation and the residential decisions of households with school-aged children.)

E. *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 hint 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 one hand, the productivity measure 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 was due to sorting or 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 7—EFFECT OF TIEBOUT CHOICE ON PRODUCTIVITY:
COEFFICIENT ON INDEX OF CHOICE FOR VARIOUS SPECIFICATIONS

Specification:	Dependent variable: (measure of achievement listed below)/ (log of per-pupil spending)					
	8th-grade reading score	10th-grade math score	12th-grade reading score	ASVAB math knowledge	Highest grade attained	ln(income) at age 32
Base IV specification	0.290	0.308	0.579	0.516	0.215	0.077
	(0.140)	(0.158)	(0.236)	(0.202)	(0.056)	(0.028)
	[0.246]	[0.262]	[0.495]	[0.410]	[0.494]	[0.566]
Base IV with choice index interacted with family income:						
Effect for low-income households	0.227	0.268	0.406	0.513	0.213	0.053
	(0.173)	(0.182)	(0.263)	(0.193)	(0.053)	(0.027)
	[0.192]	[0.228]	[0.347]	[0.408]	[0.489]	[0.389]
Effect for not-low-income households	0.312	0.298	0.572	0.619	0.224	0.092
	(0.161)	(0.169)	(0.264)	(0.204)	(0.056)	(0.030)
	[0.265]	[0.254]	[0.489]	[0.492]	[0.514]	[0.676]
Base IV with choice index interacted with minority status:						
Effect for minority households	−0.141	−0.157	0.428	0.695	0.277	0.096
	(0.348)	(0.385)	(0.419)	(0.310)	(0.087)	(0.045)
	[−0.120]	[−0.134]	[0.366]	[0.552]	[0.636]	[0.705]
Effect for nonminority households	0.374	0.556	0.595	0.417	0.164	0.059
	(0.155)	(0.182)	(0.289)	(0.246)	(0.068)	(0.033)
	[0.318]	[0.474]	[0.509]	[0.331]	[0.377]	[0.434]
Base IV with choice index interacted with state control of school revenue:						
Effect for mostly state controlled	0.110	0.323	0.469	0.304	0.254	0.085
	(0.179)	(0.173)	(0.245)	(0.217)	(0.053)	(0.028)
	[0.093]	[0.276]	[0.401]	[0.242]	[0.584]	[0.625]
Effect for mostly locally controlled	0.290	0.357	0.600	0.415	0.302	0.099
	(0.163)	(0.160)	(0.245)	(0.203)	(0.056)	(0.026)
	[0.246]	[0.304]	[0.512]	[0.330]	[0.694]	[0.728]
Mean of dependent variable	5.917	5.937	5.957	5.973	1.642	1.138
Standard deviation of dependent variable	1.177	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, NLSY, SDDB, CCD, 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.[34] 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

---

[34] In other words, an individual district is categorized based on its state, not based on how much revenue it receives from its state government. The latter procedure would introduce omitted variables bias.

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

	Dependent variable: Ratio of racial heterogeneity experienced in metropolitan-area schools to racial heterogeneity that would be experienced in metropolitan area if students were uniformly distributed			
	OLS	IV	OLS	IV
Number of districts in metropolitan area (in hundreds)	−0.066 (0.039)	−0.071 (0.158)		
Number of schools in metropolitan area (in hundreds)			−0.047 (0.007)	−0.060 (0.030)

	Dependent variable: Ratio of poverty heterogeneity experienced in metropolitan-area schools to poverty heterogeneity that would be experienced in metropolitan area if students were uniformly distributed			
	OLS	IV	OLS	IV
Number of districts in metropolitan area (in hundreds)	0.082 (0.050)	0.138 (0.158)		
Number of schools in metropolitan area (in hundreds)			−0.012 (0.005)	−0.010 (0.021)

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.

tion in states with mostly state control. It may be that where competition does not translate into much financial pressure, districts have a smaller productivity response to competition.

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 heterogenous 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 a group that is ineligible. 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.[35]

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.

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, no other 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.

X. Conclusions

Let me conclude by collecting the results into a picture of how Tiebout choice affects American schools. The first conclusion is a practical one: naive estimates (like OLS) that do not account for the endogeneity of school districts are biased towards finding no effects. This is probably mainly due to the tendency of successful school districts to attract households with school-aged children, thereby increasing their market share and reducing the observed degree of choice.

The key result of the paper is that Tiebout choice among public-school districts raises school productivity. The most dramatic finding is the opposite sign of the achievement and spending results: Tiebout choice raises productivity by simultaneously raising achievement

---

[35] 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 obvious, 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.

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 states 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 suggest 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

Alesina, Alberto; Baqir, Reza and Hoxby, Caroline M. "Ethnic Diversity and the Number of Jurisdictions." Working paper, Harvard University, August 1999.

Bénabou, Roland. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance." *American Economic Review*, June 1996, *86*(3), pp. 584–609.

Borland, Melvin V. and Howsen, Roy M. "Student Academic Achievement and the Degree of Market Concentration in Education." *Economics of Education Review*, March 1992, *2*(1), pp. 31–39.

Eberts, Randall W.; Schwartz, E. K. and Stone, Joe A. "School Reform, School Size, and Student Achievement." *Economic Review, Federal Reserve Bank of Cleveland*, Summer 1990, *26*(2), pp. 2–26.

Epple, Dennis N.; Filimon, Radu and Romer, Thomas. "Equilibrium among Local Jurisdictions: Toward an Integrated Treatment of Voting and Residential Choice." *Journal of Public Economics*, August 1984, *24*(3), pp. 281–308.

_____. "Existence of Voting and Housing Equilibrium in a System of Communities with Property Taxes." *Regional Science and Urban Economics*, November 1993, *23*(5), pp. 585–610.

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

Epple, Dennis N. and Romano, Richard E. "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review*, March 1998, *88*(1), pp. 33–62.

Fernández, Raquel, and Rogerson, Richard. "Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education-Finance Reform." *American Economic Review*, September 1998, *88*(4), pp. 813–33.

Glomm, Gerhard, and Ravikumar, B. "Opting Out of Publicly Provided Services: A Majority Voting Result." *Social Choice and Welfare*, February 1998, *15*(2), pp. 187–99.

Hausman, Jerry. "Specification and Estimation of Simultaneous Equation Models," in Zvi Griliches and Michael Intriligator, eds., *Handbook of econometrics*, vol. 1. Amsterdam: North-Holland, 1983, pp. 391–448.

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

_____. "When Parents Can Choose, What Do They Choose? The Effects of School Choice on Curriculum and Atmosphere," in Paul E. Peterson and Susan Mayer, eds., *Earning and learning: How schools matter*. Washington, DC: Brookings Institution Press, 1999b, pp. 281–316.

Hoxby, Caroline M. and Paserman, M. Daniele. "Overidentification Tests with Grouped Data." National Bureau of Economic Research (Cambridge, MA) Working Paper No. T0223, February 1998.

Kenny, Lawrence W. and Schmidt, Amy B. "The Decline in the Number of School Districts in the U.S.: 1950–1980." *Public Choice*, April 1994, *79*(1–2), pp. 1–18.

Manski, Charles F. "Educational Choice (Vouchers) and Social Mobility." *Economics of Education Review*, December 1992, *11*(4), pp. 351–69.

Moulton, Brent R. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics*, August 1986, *32*(3), pp. 385–97.

Nechyba, Thomas. "Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects and Private School Vouchers." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5642, June 1996.

Rubinfeld, Daniel L. "The Economics of the Local Public Sector," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of public economics*, vol. 2. Amsterdam: North-Holland, 1987, pp. 571–645.

Silva, Fabio, and Sonstelie, Jon. "Did Serrano Cause a Decline in School Spending?" *National Tax Journal*, June 1995, *48*(2), pp. 199–215.

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

U.S. Department of Commerce, Bureau of the Census. *Census of population and housing 1980, summary tape file 3F*. Washington, DC: U.S. Bureau of the Census, 1983a.

_____. *City and county data book*. Washington, DC: Bureau of the Census, 1983b.

_____. *Census of governments, 1982: Finance summary statistics*. Computer file, 1984.

U.S. Department of Education, National Center for Education Statistics. *National education longitudinal study, 1988: Third follow-up* [restricted access computer file]. Washington, DC: National Center for Education Statistics, 1994a.

_____. *School district data book: 1990 census school district special tabulation* [computer file]. Washington, DC: National Center for Education Statistics, 1994b.

_____. *Common core of data: School years 1987–88 through 1992–93* [computer file]. Washington, DC: National Center for Education Statistics, 1995.

U.S. Department of Labor, Bureau of Labor Statistics. *National longitudinal survey of youth, 1979–1997*, Release 10.0 [restricted access computer file]. Columbus, OH: Center for Human Resource Research, Ohio State University, 1998.

U.S. Geological Survey. *Geographic names information system: Digital gazetteer* [computer file]. Reston, VA: U.S. Geological Survey, National Mapping Division, 1998.

Universal Publishing Co. *Universal atlas of metropolitan Boston & eastern Massachusetts*, 31st Ed. Stoughton, MA: Universal Publishing Co., 1998.