DOCUMENT RESUME

ED 267 499 EA 018 296

AUTHOR Bania, Neil; And Others
TITLE Effective Environments for Secondary Schooling: Modelling the Process of Choosing between Public and Private Schools. Final Report.
INSTITUTION Oregon Univ., Eugene. Center for Educational Policy and Management.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Feb 86
NOTE 38p.
AVAILABLE FROM Publication Sales, Center for Educational Policy and Management, University of Oregon, Eugene, OR 97403 ($2.50 prepaid; quantity discounts; postage will be added to billed orders).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Quality; Models; Private Schools; Public Schools; *Research Methodology; *School Choice; *School Effectiveness; Secondary Education; Statistical Analysis; *Statistical Bias
IDENTIFIERS High School and Beyond (NCES)

ABSTRACT

To assess the relative effectiveness of public and private school environments, researchers must distinguish between the effects of the schools' programs and the students' innate abilities. Student background variables do not appear to account for all the important differences among students attending public and private schools. This document proposes and tests a model of school selection that relates school choice to a family's assessment of public and private school quality. The assumption is that unless private schools appear to offer distinct advantages they would not be selected over tax-supported public schools. The model was tested using data from the High School and Beyond studies of1980 and 1982. The researchers determined that when estimates of the educational production function for public and private schools were corrected for sample selection bias using the school choice model, private schools no longer displayed a significant advantage over public schools in educating the average student. This report describes the model, reviews the data used to test the model, presents the equation that expresses the concepts in the model, discusses the application of educational production functions, and explains how these elements can be manipulated to reveal relative school effectiveness. (PGD)

Effective Environments for Secondary Schooling: Modelling the Process of Choosing between Public and Private Schools

by

Neil Bania, Randall W. Eberts, and
Joe A. Stone

February 1986

Center for Educational Policy and Management
College of Education
University of Oregon
Eugene, Oregon 97403
(503) 686-5173

The preparation of this final report was made possible through an Institutional Grant awarded by the National Institute of Education to the Center for Educational Policy and Management. The opinions expressed in this report do not necessarily reflect the positions or policies of the NIE or the U.S. Department of Education.

I. Introduction

A lively debate about the relative effectiveness of private and public schools has ensued over the last five years. In April 1981, Coleman, Hoffer, and Kilgore (henceforth CHK) completed a highly publicized study of high schools in the U.S. Using the High School and Beyond (HSB) longitudinal study of sophomore and senior students, CHK concluded among other things that Catholic schools are more effective than private schools in enhancing the cognitive skills of secondary students.

This analysis has received a fair amount of criticism, and one of the most common criticisms was the methodology used to distinguish between the effectiveness of program and the ability of students. CHK attempted to control for student abilities by including 17 student background variables in the student achievement equations. Barnow, Cain, and Goldberger (1980) questioned whether the inclusion of student characteristics is sufficient to eliminate selectivity bias. Noell (1982) applied the Heckman approach of correcting for sample selection bias to the same database as used by CHK and found no statistically significant difference between the effectiveness of Catholic and public schools. In response to Noell's reestimation of the HSB dataset, CHK also used the Heckman technique and once again found, contrary to Noell's results, a statistically significant difference between the effectiveness of public and Catholic schools.

The discrepancy in the results of two apparently similar analyses applied to the same set of data was puzzling. However, much of the mystery was unraveled by Murnane, Newstead, and Olsen (1984). They reported that in fact there was no selectivity bias among black or hispanic student samples in either public or Catholic schools. However, they did find selectivity bias among white students in the sample.

Why would selectivity bias in the whites but not in the minorities cause the results to differ? Murnane, Newstead, and Olsen stress two basic reasons. First, minority students were more predominate in CHK's analysis than in Noell's estimation. This difference in representation occurred because of the nature of the HSB database. The HSB database is a stratified random sample with an oversampling of black and hispanic students. Design weights have been calculated that permit the creation of a weighted sample that reflects the U.S. high school student population. CHK did not use the design weights; Noell did use them. Thus, the oversampled blacks and hispanics were much more predominate in CHK's analysis than in Noell's. In addition, not all records in HSB contain complete information, and a greater percentage of incomplete records came from minority students than from white students. Since Noell included more student characteristics in his estimation than CHK, Noell further undersampled minority students as compared with CHK. Consequently, the estimated comparative advantage of Catholic schools over public schools for minority students was not picked up by Noell's estimation to the degree it was by CHK.

Second, the estimation of sample selection bias and its importance as an explanatory variable in the achievement equation was hampered by an unidentified system of equations. "Unidentified" is a statistical term referring to the technical inability to distinguish between two structural relationships. In this case, it is difficult to distinguish between the educational process and the selection process because similar variables are used to explain each relationship. Murnane and others found that selectivity bias among white students did not show up because the achievement equation was unidentified and the effect of selectivity bias was confounded with the impact of religious status on student achievement. Thus, they concluded that an improper exclusion restriction can lead either to the conclusion of selectivity bias when there is in fact none, or to the conclusion of no selection when in fact selection is present.

The purpose of this paper is to explore in more depth the selection process used to choose between public and private schools. In doing so, we do not intend to address directly the controversy over the difference in effectiveness between Catholic and public high schools. Rather, we intend to pursue an aspect of the analysis that Murnane and others concluded was important in properly comparing the effectiveness of these two school environments: identification of the selection process.

Most studies have used only student and family characteristics along with regional identifiers to explain the selection process. However, we believe that the quality of local public school programs may be an important factor in the selection process. This is because families will tend to choose public schools, already paid for through local taxes, unless the quality is low.

To investigate this proposition, we organize the paper in the following manner. First, we model the process of choosing between public schools and private schools. Next we describe the relevant portions of the High School and Beyond dataset which has been used by CHK, Noell, and Murnane and others. We should mention, however, that since we are interested in the selection process between public and private schools, we consider both Catholic schools and non-Catholic private schools, whereas the earlier papers considered only Catholic schools as the private entity. Finally, we estimate the selection process using logit analysis and estimate separate achievement equations for public and private schools. Estimates of the logit analysis are then used to correct for selectivity bias in the two achievement equations. Differences in predicted student math test scores are calculated with and without adjusting for selectivity bias and the results are compared.

II. Model of Choosing between Private and Public Schools

The model of the demand for public school quality and private school enrollments is based on the median voter model development by Bergstrom and Goodman (1973) and its extension by Sonstelie (1982). In this model, each family participates in the collective choice mechanism that determines the common level of average public school quality within a specific local school district. Once the level of public school quality is determined, families choose between the collectively-determined public school quality and various levels of private school quality available in or around the district.

Since public schools are available to all residents of a school district, the public school option involves no cost to the individual family beyond the mandatory tax payment. Thus a family that chooses private schools incurs an additional cost but cannot avoid its tax burden. Families will send their children to private schools only if the net-of-tuition benefit of attending private schools exceeds the "gross" benefits of attending public school.

To model the process more formally, let each family value educational services and the consumption of noneducational goods according to the following utility function: $U(z,q)$, where z is a composite noneducation good (measured in dollars spent on all other noneducation goods) and q is a measure of educational quality. Let q_{pub} be the quality of education in the local public schools and y be the family's income after paying local property taxes. We assume for now that the unit cost of educational quality is the same in both the public and private sector. Given that private schools are assumed to be perfectly competitive, they will earn no economic profits. Therefore, the unit cost of quality will be the same as the unit price of quality and may be denoted by P. We also assume that the cost of local education is paid entirely out of local property taxes.

If the child attends public schools, then the utility level is: $U(y,q_{pub})$. If the child attends private schools, then the utility level is: $U^{*}=U(y-Pq^{*},q^{*})$, where q^{*} is the level of quality that the family would choose in the absence of the public school alternative. That is, q^{*} represents the level of quality preferred by a family if its only contraint were income and relative prices. Under this formulation, a family will choose public education if and only if $U(y,q_{pub})$ is greater than U^{*}; otherwise it will choose a private school. For each family there will be some unique public school reservation quality, that is, the quality level for which the family will be indifferent between private and public schools.

More formally, each family's maximization problem can be viewed as a two-step procedure. In the first step each family determines its preference for public school quality and private school quality. The private school preference is the quality level that would be chosen in the absence of any governmental support for education. This is the level of quality that would be chosen if there were only private schools, and a sufficiently large number of schools to supply a broad distribution of levels of quality. In addition, each family must express its preferences via the collective choice process. In general, each family's preference for private educational quality will not coincide with its preference for public education quality since the demand for public quality depends upon the tax price. Thus, even if the unit cost of educational quality is the same for both private and public schools, families may prefer more or less public school quality according to whether their tax cost of supporting public schools is less than or greater than the price of private educational quality.

Consequently, in the first step of the process, each family participates in the political process which determines q_{pub}. Even though each family may participate in the collective choice mechanism, q_{pub} is still viewed as exogenous by the individual family. In addition, each family determines the level of private school quality that would be purchased in the absence of the government provision of "free" public schools.

The second step of the maximization process involves a comparison of the relative benefits of choosing a public versus a private school. That is, public schools will be chosen if $U(y,q_{pub})$ is greater than $U^*(y-Pq^*,q^*)$; otherwise private schools will be chosen.

The choice between public and private schools can also be viewed in a somewhat different manner. The reservation quality is defined to be the quality of public schools that would make the family indifferent between public schools (of quality q_R) and private schools (of quality q^*). The reservation quality (q_R) then, is defined implicitly by the following relationship:

$$U(y,q_R) = U^* = U(y-pq^*,q^*).$$

If public schools are of lower quality than q_R, the family will choose private schools; if public schools are of higher quality than q_R, the family will choose public schools.

Under certain standard properties imposed on the family's preferences for education and noneducation goods, a family will never choose a private school of lower quality than the public school alternative. This implies that the public school quality level will form a lower bound for the range of educational quality outputs, with private schools offering various quality levels above the public school quality level. Obviously, no private school would offer a quality level below that of the public school because no one would attend it, since they could receive higher quality at a public school without paying the tuition.

When the model is generalized to allow for different preferences among families for various components of educational ouput, it will be possible for some private schools to be of lower quality than public schools. For example, Catholic families may prefer the religious content offered in Catholic schools to the secular orientation of public schools.

Catholic families with these preferences may select private (Catholic) schools where the quality level, as measured by achievement test scores, is lower than in public schools. Of couse, this does not contradict the model since quality could, conceptually at least, be redefined to include a vector of outputs such as test scores and religious education. Viewed in this way, Catholic families would simply weigh the religious output more heavily than other families and, from the Catholic family's perspective, private Catholic schools would be of higher quality than public schools.

The public-private choice can be illustrated with the family's income-compensated demand curves. In Figure 1, AD is the income-compensated demand curve for educational quality, and OP measures the price of educational quality in both the public and private sector. The quality of public schools is viewed by an individual family as fixed at OE units of quality. The gross benefit of attending public schools is equal to the area OABE, and the net-of-tuition benefit of attending private schools is equal to the area of PAC. The family will choose private schools if the area of PAC exceeds the area of OABE; otherwise, it will choose public schools.

The public school quality level that makes the two areas exactly equal is referred to as the family's reservation quality. If public school quality is below this level, then the family chooses private schools. Alternatively, if public school quality is above this level the family will choose public schools.

The relationship between the reservation quality and the family's income level can also be seen in Fig. 1. Assuming that demand for education increases with income, a rise in income will increase the maximum utility that can be achieved. Since the level of utility has increased, the income-compensated demand curve will shift to the right, increasing both the gross benefit of attending public school and the net-of-tuition benefit of attending private school. The increase in gross benefit of public school attendance is equal to the area BB'A'A, while the increase in net-of-tuition benefit of private school attendance is equal to the area A'C'CA, which is clearly larger. Thus, the reservation quality rises with the level of household income as long as the income elasticity of education is positive.

This implies that for a given quality of public schools, there is some income level such that the family with this income level is indifferent between public and private schools. Families with income greater than this level will choose private schools and families below this income level will choose public schools.

If the income distribution is known in the district, then the proportion of students in private schools is defined as the proportion of students coming from families with incomes greater than the reservation level. Turning this around, the proportion of students in private schools is one minus the proportion of students coming from families with incomes less than the reservation income level.

In summary, the proportion of all students who choose private schools depends upon two factors: the quality of public schools and the dispersion in the demand for educational quality as reflected in student and family background and community characteristics. The next step in the analysis is to use these factors to explain the observed choice between public and private schools. First, however, a brief description of the data is provided in the next section.

III. Data

The student level and school level data for this study come from the High School and Beyond study conducted by the National Center for Education Statistics (NCES) in 1980 with a follow-up survey in 1982. Additional data employed in the cross-sectional analysis of the choice between public and private schools was drawn from well-known census sources.

The High School and Beyond (HSB) study was designed as a two-stage stratified probability sample. In the first stage of the sampling process, approximately 1,000 public and private high schools were chosen for inclusion in the sample. In the second stage, 36 sophomore and 36 senior students in each school were randomly selected. Since only the sophomore data (which includes their responses as seniors two years later) were used in this study, the data description will be limited to that part of the sample. The HSB data used in this research included a student questionnaire, the results of student exams on cognitive tests, and a school questionnaire. The student questionnaire included information on student and family background characteristics, the test covered a wide range of subjects including math and reading, and the school questionnaire covered school resources and programs.

In the initial selection of schools, two general strata were identified. The regular strata were not oversampled and included public and Catholic schools. Catholic schools were further stratified by four census regions. In the case of public schools, the sample was stratified according to the nine census regions, racial composition, enrollment, and the degree of urbanization (central city, suburb, rural). For schools in the regular strata, the probability of selection was proportional to the number of students in the school. The special strata included public schools with "alternative" programs, public schools with a high percentage of Cuban students, Catholic schools with a high percentage of Cuban students, public schools with a high percentage of non-Cuban Hispanic students, "high performance" private schools, other non-Catholic private schools, and Catholic schools with a high percentage of Black students. Within this stratification, the other non-Catholic private schools were further sub-stratified by the four Census regions. These schools were all oversampled to ensure adequate representation in the sample to conduct separate analysis.

The initial drawing of the schools came from the universe of schools in the United States that had either tenth- or twelfth-grade students. The list of schools was compiled from a merged list of schools provided by NCES and the Curriculum Information Center, a private firm. Of the initial sample of 1,122 schools, 811 agreed to participate in the survey. For the schools that refused to participate, substitution was carried out within strata, and 204 schools were added, which brought the total to 1,015.

Within each school, 36 sophomore students were randomly selected. Students who refused to participate or who were absent on the day of the test were not replaced. If the school contained fewer than 36 students, then all students were selected.

For the 1982 follow-up sample, 40 schools were dropped for various reasons, bringing the total number of schools to 975. Of the total 70,704 senior and sophomore students initially selected to participate, 58,270 eventually completed the survey in 1980--this included a sample of 29,737 sophomore students. Of this total, 25,150 sophomore students participated in the follow-up survey in 1982.

Since the sample was highly stratified, the data were weighted to ensure that the analysis would reflect the outcomes for the entire population. The general approach to weighting involved two steps. First, a weight which reflected the unequal probabilities of selection was calculated. At the school level, the weight for each school is equal to the number of schools in the population represented by the sampled school. These school-level weights ranged from 1.00 to 169.00, and summed to 21,174. In this way, schools with weights of 1.00 are a 100 percent sample of their sub-strata where (1/169th) of the schools in the population were sampled. The sum of the weights indicates that the 1,015 schools were sampled from a population of 21,174 schools.

To form weights for the student-level data, the school level weight was multiplied by the probability of each student being selected for the sample. This probability was calculated as the number of students selected for the sampled divided by the actual number of students in the selected school. To form weights for the follow-up analysis, the weights for both the schools and the students were multiplied by the inverse of the probability of selection in the follow-up sample. For most schools, this probability was equal to one since most all schools were included in the follow-up survey. For students, the probability of being included in the follow-up analysis was equal to one for students still in high school. Students who transferred, graduated early, or dropped out of school were not included in the analysis.

The second step in computing the weights for students was to adjust for non-response. Although the procedure was somewhat complex, the idea was to reapportion the weights of non-respondents among those who participated in the survey. For this analysis, a sample of students that completed the cognitive tests in both years of the sample was employed. Among these 22,436 students, the mean weight was 168.00, the minimum was 2.13, the maximum was 2,774, and the sum of the weights was 3,769,248. For the purpose of data analysis, these weights were always divided by a constant such that the sum of the weights equaled the actual number of observations in the analysis.

Murnane, Newstead, and Olsen recommend not using the design weights to avoid the problems mentioned earlier with using the weighting scheme faced by CHK and Noell. However, Catholic students are oversampled in the HSB database, and not using the design weights would bias the estimates of the relationship in the population between the student background variables and school choice. Thus, we take a random sample of the HSB database to form an appropriate sample. This reduces the number of usable observations to a little under 3,000.

Student Level Data

Student-level data are used to estimate the educational production function for students in private and public schools. In addition, student-level data were aggregated to the state-level for use in estimating the choice process between public and private schools.

Table 1 describes the coding of the variables. Most are self-explanatory, but a few need additional comment. The variable Mother

Worked has a value of one if the student's mother worked either parttime or fulltime; otherwise it was zero. Parental involvement is a composite of three questions regarding the degree of parent participation in school PTA, frequency of classroom visits, and involvement with school projects. The scale ranged from a low of 1 (never involved) to a high of 3 (often involved). Parent Talk measured the amount of time the student spent talking with either parent. The scale ranged from a high of 3 (every day, or almost every day), to a low of 1 (rarely or never). Parent Reading measured the amount of time parents spent reading to the student before he or she started school. This variable ranged from a low of 1 (the parent never read to the student) to a high of 5 (the parent read every day). SES Status is a composite variable that was constructed from five components: the father's occupation, the father's education, the mother's education, the family's income, and a composite variable that is an index of household possessions.

Two school level variables were included in the educational production functions. The first was the average level of expenditures per student. This variable was equal to the district average for public schools and to the school average for private schools. In the case of private schools, if the value of expenditures per student was missing, then it was constructed by multiplying the reported tuition level by the inverse of the percent of school funds derived from tuition. The percentage of high achievers in each school was derived by aggregation from the student level to the school level. A high achiever was defined as a student who scored in the top 25 percent of all students in the HSB sample on the sophomore year composite test. This test was a composite of all eight tests that the students took in their sophomore year (1980).

State-Level Data

State-level data are used to estimate the school sector choice equation. Although we would have preferred district-level data, the lowest level of geographic disaggregation available in the HSB data was the nine census regions. The state in which each school was located had to be identified in the following way. First, a subset of students who indicated that they would attend college within their home state was identified. Next a list of the students' college choices, and the state in which these colleges were located, was compiled by school. Each high school was assumed to be located in the state which contained the most colleges on the list of in-state student choices for each school. When ambiguous results were obtained, then the school was omitted from the analysis. The hypothesized location of each school was then checked against the census region reported in the HSB data for consistency. Variables such as percentage of students from Catholic families and percentage of nonwhite students were constructed at the state level.

IV. Derivation and Estimation of the Choice Equation

Since the quality of public schools is determined through the political process of allocating local funds and setting local district policy, an individual family will consider the average quality of public schools to be outside its control. Thus, a family will enroll its child (children) in public schools only if the gross benefit of public school attendance exceeds the net-of-tuition benefit of private school attendance. Using Sonstelie's (1979) terminology, the gross benefit of public school attendance less the net benefit of private school attendance is denoted as the public school surplus. Thus, families will choose public schools only when the public school surplus is positive. In Figure 1 the public school surplus is the area OABE (the gross benefit of private school attendance) less the area of PAC (the net benefit of private school attendance). If relative prices remained the same, an increase in income would result in a higher utility level. Thus, the income-compensated demand curve, as shown in Section II, will shift to the right (to D' in Figure 1) and the public school surplus will decrease by an amount equal to the area BB'C'C. An increase in public school quality will increase the gross benefit of public attendance but leave the net benefit of private schools unchanged.

The size of the public school surplus will also depend upon the family's preferences for educational surpluses. As mentioned in an earlier section, some private schools, such as Catholic schools, provide a religious perspective to education that is not available in public schools. Thus, some families may prefer private schools of lower academic quality than found in public schools simply because of the religious dimension to the school's curricula. Other things being equal, then, families who have this religious preference will have a lower public school surplus.

The number of children in a family will also affect the family's public school surplus. First, the cost to the family of educating children in public schools (paid in taxes) is independent of the number of children enrolled whereas the cost of sending children to private schools rises proportionately with each additional child. The additional cost will decrease the net-of-tuition benefit derived from private school attendance, while leaving the gross benefit of public school attendance unchanged. The result is an increase in the public school surplus. Moreover, as family size increases, holding income constant, it is likely that the fraction of income spent on education will decline (or at least remain constant). This implies that spending per child will decrease with family size. Both of these factors work to decrease the demand for education, and thus to increase the public school surplus. As one would expect then, as the number of children in each family rises, the family is more likely to select public schools.

The public school surplus is assumed to be a random variable distributed according to the logistic probability distribution. Under these conditions, the probability that any individual family will choose public schools is given by:

$$(1) \qquad \Pr(PUBLIC_i)=1/(1+\exp(-S(q_{pub},y,CHILDREN,CATHOLIC)))$$

where $\Pr(PUBLIC_i)$ is the probability of choosing public schools (PUBLIC=1 if attends public school), q_{pub} is the quality of public schools, y is family income, CHILDREN is the number of children in the family, and CATHOLIC is a measure of the family's preference for religious content in schools. The notation exp denotes the exponential function.

We can also express equation (1) as the log of the odds of choosing public schools:

$$(2) \qquad \log(PUBLIC_i/(1-PUBLIC_i)) =$$

$$a_1 q_{pub}+a_2 y+a_3 CHILDREN_i+a_4 CATHOLIC_i+a_5 X_i+e_i$$

where X represents variables describing district and regional characteristics. Using this specification, the coefficients can be estimated using maximum likelihood techniques.

We estimated this equation using the HSB dataset. Two measures of school quality are used. The first variable is the average gain in the students' math achievement test scores between their sophomore and senior years for each state. Math test scores are used instead of reading scores because it is generally believed that school-based inputs play a larger role in determining mathematics achievement gains than reading gains, in part because family background variables are relatively more important for reading and language arts (Madaus, 1979). The second measure of public school quality was the average expenditures per student of districts in the state. Since we also control for regional and urban-nonurban variations in costs through the use of dummy variables (WEST, CENTRAL, SOUTH, URBAN, SUBURB, with rural districts the excluded variable), variations in expenditures per student reflect differences in the amount and quality of resources available to students.

In addition to measures of public school quality, we also included information regarding each student's family characteristics such as income (INCOME), religious preference (CATHOLIC), race (MINORITY), and number of children in the household (CHILDREN).

The results are shown in Table 2 for each measure of public school quality. In the first set of results, expenditures per student is positive and statistically significant at the 5 percent level. This means that the odds of choosing a public school over a private school increases with an increase in public school quality, which supports our hypothesis.

The signs of the other coefficients indicate that children from non-Catholic, white families with lower income and more siblings prefer public schools to private schools. The same results hold when average math test score gains are used as the measure of public school quality, with one exception. The coefficient on the math score variable is not statistically different from zero. Several things could explain this result. One reason, of course, may be that math score gains are a poor proxy of public school quality since they may not pick up the cumulative nature of the educational process. Another reason may be related to the information available to families in selecting schools. One statistic that is easily obtainable is per pupil district expenditures. Test scores, on the other hand, are not as readily available and may be more difficult for families to interpret. Both sets of estimated coefficients will be used to calculate the predicted probability of choosing public schools which will then be used to correct for selectivity bias.

V. Educational Production Functions

Educational production functions relate differences in quality of student outcomes to differences in innate student ability and school resources received by students. Because specifications of educational production functions differ among studies, it is impossible to capture with one specification all the features of all the models constructed to date. However, most studies share the features described in equation (3), which is borrowed from Hanushek (1979).

$$A_{it} = f(B_{it}, P_{it}, S_{it}, I_i), \tag{3}$$

where A_{it} = student outcomes of ith students at time t

B_{it} = vector of family background influences of ith student cumulative to time t

F_{it} = vector of influences of peers of ith student cumulative to time t

S_{it} = vector of school inputs of ith student cumulative to time t

I_i = vector of innate abilities of ith student.

The model incorporates a number of essential aspects of the educational process. First, inputs are those that are relevant to the individual student. Second, the inputs are cumulative, which reflects the fact that schooling and other experiences in past years have a bearing on student outcomes in the present period. Third, school inputs include purchased inputs (e.g., teachers) as well as nonpurchased inputs (e.g., peer groups). Fourth, the allocation of resources is predetermined from the perspective of the production function.

A somewhat popular variant of the model and one that requires substantially less data collection is the value-added model. Instead of considering the contribution of past inputs on student outcomes, this specification considers the changes in student outcomes between two time periods, in this case sophomore year and senior year. This formulation reduces the data requirements, since inputs are only collected for two years and not from the beginning of the child's schooling (e.g., kindergarten).

The value-added model results from simply subtracting equation (3) for period t^* from equation (3) for period t.

$$(4)\ A_{it} = f^*(B_i(t-t^*), P_i(t-t^*), S_i(t-t^*), I_i, A^*_{it})$$

Student outcomes in the earlier period (A^*_{it}) may be reflected in scores from tests taken by students in the first year. These scores are then compared with scores of tests taken during the last year. In this way the gains in student outcomes attributed to a flow of educational services within a given time period can be assessed.

Equation (4) is estimated at the student level for each school type: public and private. This allows for the possibility that the parameters of the production function differ between school types. The dependent variable is the student's score on an objective math test taken early in the senior year. The explanatory variables fall into three basic groups: student characteristics, school characteristics, and peer group characteristics. To measure the student characteristics, sixteen variables are employed. The object is to measure the student's innate ability and motivation as well as aspects of each student's socio-economic background which might be related to his or her performance on achievement tests.

The student background variables can be further sub-divided into three categories. First, to measure past achievement, the student's score on an objective math test taken early in the sophomore year is used (SOPHMATH). The sophomore score is expected to be related to the student's innate ability and motivation as well as to school resources received prior to entering high school. Using a value-added model of this type allows the explicit consideration of the relationship between gains in the student's score between the senior and sophomore years and the flow of school resources during this time period. Since the raw score of the test is used, students who received high scores in the sophomore year have relatively little room for improvement on subsequent tests covering similar material compared to students who earned lower scores. This ceiling effect suggests that the relationship between the senior year score and the sophomore year score may be non-linear. To allow for this possibility, the square of the sophomore year score was also used as an explanatory variable (SOPHMSQ).

A second major group of student background variables is those that measure the characteristics of the student's family. These variables include dummy variables for the student's sex and race (BLACK and HISPANIC). In addition, the number of siblings in the household (CHILDREN) and an index of socio-economic status (SOCIO-ECON) are used. Finally, dummy variables are used to indicate whether the student's mother worked before the student entered first grade (MWBS), and whether there are currently two parents present in the household (TWOPAR).

The last group of student background variables is designed to measure the motivation level of the student and of the student's family with regard to educational achievement. Three school-related variables provide a direct measure of student and parent motivation. A composite variable was constructed by averaging the response to three questions regarding the degree of parent involvement in school activities (PARENT). The possible responses on this question ranged from a low of 1 (never involved) to a high of 3 (often involved). To measure the motivation of the student, the number of days absent without an excuse (ABSENT) and the number of hours per week spent on homework (HWHOURS) were also used.

Finally, four additional variables complete the explanatory student background variables. These include three variables to measure student time spent on various activities: the number of hours each student worked for pay each week (WORKHRS), the number of hours spent watching television each week (TVHOURS), and the amount of time spent talking with parents (PARTALK). A final variable measures the amount of time parents spent reading to the student prior to first grade (READTO).

Two variables are employed to measure the flow of school resources and the quality of student peers. School resources are measured by expenditures per student, a single variable intended to summarize the overall amount of purchased productive school inputs (EXPEN). The peer group effect (PEERS) is measured by computing the percentage of high achievers in each school. High achievers are defined as students who scored in the 25th percentile of a composite exam given to all students in the sample, public and private, in their sophomore year. The composite exam included the results from the math test as well as the other subject areas.

Estimates of the educational production function using ordinary least squares regression are presented in Table 3 for public and private schools. The signs of the coefficients are in the anticipated direction. In public schools for instance, students who spend more time on homework score higher on math tests than students who neglect homework. Time spent watching TV and working at a job results in lower test scores. Parents also play an important role in achievement gains. Students whose parents are involved in school-related projects and who maintain a dialogue with their children at home perform much better on tests than those whose parents have little involvement at school or interaction at home. The purchased resources of the school also have a positive effect on student achievement gains. Peer groups, although exhibiting a positive effect, do not significantly affect senior test scores. The socioeconomic background of the student's family has a major effect on achievement as does the student's ethnic background.

Results for students attending private schools were of similar signs and magnitudes in many cases but in general had larger standard errors. One reason for this lack of statistical significance of the individual coefficients could be the much smaller sample of private school students than public school students. In addition, students in private schools may be much more homogeneous than students in public schools, which would result in multicollinearity. Thus, it would be impossible to separate out the individual effects of the explanatory variables used to explain test score gains. Another reason may be the distinct difference between Catholic private schools and other private schools. We divided these two types of private schools into two samples and estimated the production function separately. This approach yielded coefficients with roughly the same signs but with much lower standard errors. Nonetheless, we choose to stay with the public/private distinction in comparing the coefficients of the two educational sectors.

VI. Differences in Effectiveness of Private and Public Schools

Differences in the educational technologies, or environments, of public and private schools can be estimated by taking the difference in the coefficients for each explanatory variable in the production function, including the intercept, multiplying the difference by the public school mean and then adding up these weighted differences. Table 4 shows these calculations for the estimates reported in Table 3. We find that the total advantage of public schools is calculated to be -.64. This is interpreted to mean that private school students on average score .64 points higher than public school students.

The advantage of private schools can be placed in perspective in two ways. First, we can express this difference in test score gain relative to the average score of a public school sophomore. The result would be a 2.9 percent advantage to private schools. We could also express the difference between the two school environments relative to the average test score gain of public school students. In this case the result would be a 45 percent advantage for private schools. Regardless of the basis of comparison, private schools are more effective under this estimation. Our results support the conclusions drawn by CHK who found Catholic schools to be more effective.

Of course, the central issue is whether the student background variables included in the educational production function are sufficient to control for differences in the students' innate abilities within each of the two schooling environments. As mentioned earlier, Barnow and others (1980) and Murnane and others (1984) claim that additional ways of accounting for student effects need to be implemented.

We used the logit estimates of school choice to construct the predicted probability that a student will choose public schools over private schools. Heckman (1978) has shown that these predicted probabilities can be used to correct for a biased sample selection. Sample selection bias results when higher innate achievers naturally select private schools instead of public schools. If this is the case, then private schools will appear to be more effective even though their advantage is not in the effectiveness of their program but in the ability of their students, which is indepeadent of program. By correcting for the truncated distribution of students across the private and public sectors, we can also correct for sample selection bias.

To implement this scheme, the predicted probabilities are used to construct an inverse Mill's ratio. An inverse Mill's ratio is comprised of the cumulative distribution of students ($F(\theta)$) and the density function ($f(\theta)$) where θ is the estimated probability of choosing public schools. The inverse Mill's ratio is then defined as

$h_i = -f(\theta_i)/F(\theta_i)$ for the ith public school student and

$h_j = f(\theta_j)/(1-F(\theta_j))$ for the jth private school student.

Including these two statistics into the achievement equation adjusts the mean of the math test score distribution for the fact that either the upper tail or lower tail of the distribution may be missing due to the decision of families to send their children to either public or private school. If the coefficient on the inverse Mill's ratio variable for the public school achievement equation is negative, this indicates a loss of top achievers from the public school sample. A negative sign on the inverse Mill's ratio variable in the private school achievement equation indicates that the private school sample has gained top achievers.

The achievement equations are reestimated with the inverse Mill's ratio included as an explanatory variable. The results are shown in Table 5 with expenditures per student used as a measure of school quality in the school choice equation and in Table 6 with average gain in math test scores used as school quality in the choice equation. In both cases, the coefficient on the inverse Mill's ratio is negative and it is statistically significant at the 10 percent level for public schools. The signs on these coefficients indicate that the public school sample has lost high achievers while the private school sample has gained these high achievers. Thus, it would appear that the advantage estimated for private schools may be reduced when these corrections for sample selectivity bias are included.

This is indeed the case. We find that when test scores are used in the prediction equation public schools gain an advantage over private schools with a difference of .90. When expenditures per student are used in the prediction equation, the public school advantage disappears and the effectiveness of public and private schools becomes very similar with a difference of only -.04. Thus, it appears that when sample selection bias is taken into account, the estimated advantage of private schools in educating the average student disappears. There may be certain advantages in private schools for educating students from various ethnic groups as found by other studies. However, our conclusion is much more general and considers what happens to estimated differences in effectiveness when the selection process is modelled in a way that allows both the selection equation and the achievement equation to be properly identified.

VII. Conclusion

Assessing the relative effectiveness of public and private school environments requires distinguishing between the effectiveness of school program and the innate ability of students. Including student background variables in a student achievement equation does not appear sufficient to control for important differences in students who attend public and private school. We have proposed and estimated a model of school selection that is based upon the family's evaluation of the relative quality of public and private schools. Since all students are entitled to attend public school free of tuition, the relevant point of reference is to compare the relative quality of private schoool alternatives to public schools. Thus, the quality of public schools was entered into the choice equation, and it was found that the higher the quality of public schools the less likely a student will choose private schools.

Estimates of this selection process were used to correct for the truncated distribution of students found in the public school sample of students and the private school sample. Differences in the estimated educational production technologies, when sample selection bias was not corrected, revealed that private schools are more effective than public schools. When estimates of the educational production function were corrected for sample selection bias using the model of school choice presented here, private schools no longer displayed a significant advantage over public schools in educating the average student.

References

Bergstrom, Theodore C., and Robert P. Goodman. "Private Demands for Public Goods." American Economic Review 63 (1973): 280-96.

Barnow, S.N., G.G. Cain, and A.S. Goldberger. "Issues in the Analysis of Selectivity Bias," in Evaluation Studies, Vol. 5, (1980), E.W. Stromsdorfer and G. Farkas (eds.). Beverly Hills, CA: Sage Publications.

Coleman, J., S.N. Hoffer, and S. Kilgore. "Questions and Answers: Our Response." Harvard Educational Review 51 (1981): 526-545.

________________. High School Achievement: Public, Catholic and Private Schools Compared. New York: Basic Books, 1982.

Eberts, Randall W., and Joe A. Stone. Unions and Public Schools: The Effect of Collective Bargaining on American Education. Lexington, MA: Lexington Books, 1984.

Hanushek, Eric. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions." Journal of Human Resources 14 (1979): 351-358.

Heckman, James J. "Dummy Endogenous Variables in a Simultaneous Equations System." Econometrica 46 (1978): 931-960.

Madaus, G.F. "The Sensitivity of Measures of School Effectiveness." Harvard Educational Review 2 (1979): 207-30.

Murnane, Richard, Stuart Newstead, and Randall J. Olsen. "Comparing Public and Private Schools: The Puzzling Role of Selectivity Bias." Institution for Social and Policy Studies, Yale University, 1984.

National Center for Education Statistics. 1982 High School and Beyond: Sample Design Report. Washington, D.C., 1982.

Noell, J. "Public and Catholic Schools: A Reanalysis of 'Public and Private Schools'." Sociology of Education 55 (April/July 1982).

Sonstelie, Jon. "Public School Quality and Private School Enrollments." National Tax Journal 32 (1979): 343-53.

________________. "The Welfare Cost of Free Public Schools." Journal of Political Economy 90 (1982): 794-808.

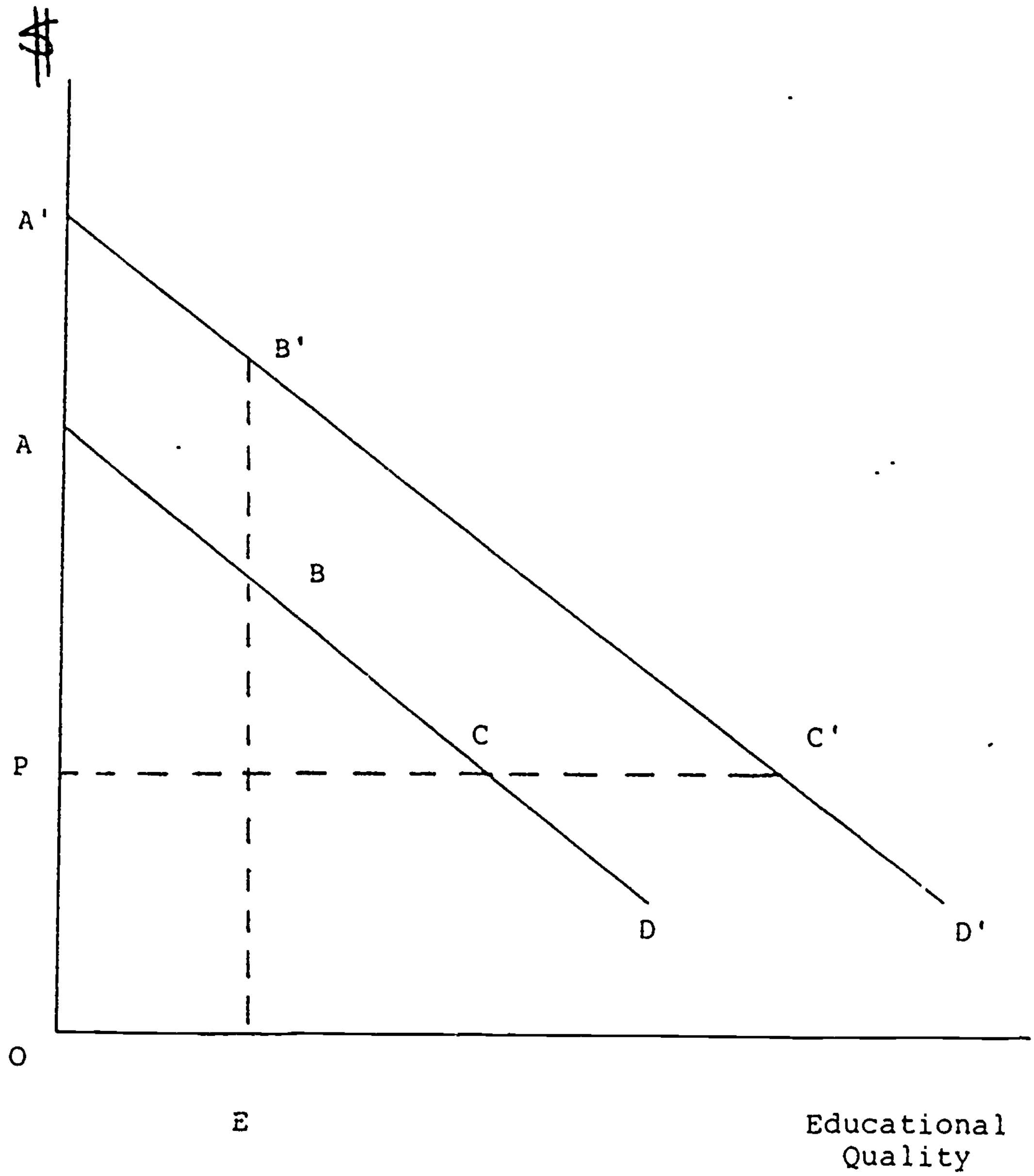

Figure 1. The Choice Between Public and Private Schools

Table 1. Recoding of Student Variables

Variable	HSB Name	Recode/Transformation*
Sex	Flag21	Male=1, Female=0
Black	Flag22	Black=1, Otherwise=0
TV Hours	FY61	0 to 1 hours = 0.5 hours
		1 to 2 hours = 1.5 hours
		2 to 3 hours = 2.5 hours
		3 to 4 hours = 3.5 hours
		4 to 5 hours = 4.5 hours
		5 + hours = 6 hours
Work Hours	FY25	1 to 4 hours = 2.5 hours
		5 to 14 hours = 9.5 hours
		15 to 21 hours = 18 hours
		22 to 29 hours = 25.5 hours
		30 to 34 hours = 32 hours
		35 to 40 hours = 37.5 hours
		41 + hours = 42.5 hours
SES	Flag29	
Siblings	BB096A-E	
Two Parent	BB036B-E	1=Two Parents, 0=Otherwise
Mother Worked	BB037C	1=Mother Worked full/part time
Parents Read	BB095	
Parent Talk	FY60F	
Hispanic	Flag22	1=Hispanic,0=otherwise
Homework Hrs	FY15	0 to 1 hours = 0.5 hours
		1 to 3 hours = 2.0 hours
		3 to 5 hours = 4.0 hours
		5 to 10 hours = 7.5 hours
		10 to 15 hours = 12.5 hours
		15 + hours = 17.5 hours
Days Absent	FY16	1 to 2 days = 1.5 days
		3 to 4 days = 3.5 days
		5 to 10 days = 7.5 days
		11 to 15 days = 13 days
		16 to 20 days = 18 days
		21 + days = 25 days
Parent Involvement	FY58A,C,E	
Expen/Student	School Level	
% of High Ach	School Level	
Senior Score	FYMTH1RT/2RT	FYMTH1RT + FYMTH2RT
Soph Score	BBMTH1RT/2RT	BBMTH1RT + BBMTH2RT

*If blank, then no recode/transformation was made.

Table 2: Estimation of Odds of Choosing Public Schools over Private Schools

PUBLIC SCHOOL QUALITY = EXPENDITURES/STUDENT

Variable	Mean	Coefficient	Coeff/std. Error
INCOME	27379.12	-.00002*	-8.66
EXPEND/STUDENT	2278.67	.00018*	2.39
CHILDREN	1.79	.061*	2.85
CATHOLIC	.34	-.975*	-13.70
MINORITY	.18	-.011*	-7.46
SOUTH	.24	.442*	3.74
CENTRAL	.41	.135*	1.63
WEST	.15	.566*	4.98
CITY	.30	.218*	2.03
SUBURB	.14	-.106	-1.37
INTERCEPT		6.695*	31.41
CHI SQUARE = 3679.96	DF = 3257		

PUBLIC SCHOOL QUALITY = MATH SCORE GAIN

Variable	Mean	Coefficient	Coeff./Std. Error
INCOME	27379.12	-.00002*	-8.99
MATH SCORE	1.91	-.058	-.65
CHILDREN	1.79	.066*	3.06
CATHOLIC	.34	-.97*	13.73
MINORITY	.18	-.012*	8.13
SOUTH	.24	.247*	2.21
CENTRAL	.41	-.004	.04
WEST	.15	.51*	4.28
CITY	.30	.228*	2.14
SUBURB	.14	-.122*	-1.60
INTERCEPT		7.25	38.08
CHI-SQUARE = 3574.1	DF = 3202		

Note: (*) denotes statistical significance at the 5 percent level.

Table 3: Estimates of the Educational Production Function for Public and Private Schools.

Variable	Mean Public	Mean Private	Coefficient (T-statistic) Public	Coefficient (T-statistic) Private
SOPHMATH	20.54	22.62	.62 (8.87)	1.19 (6.12)
SOPHMSQ	477.22	557.78	.005 (2.95)	-.007 (1.75)
SEX	.46	.45	1.15 (5.80)	.74 (1.53)
BLACK	.08	.04	-1.03 (2.73)	.19 (.16)
HWHRS	4.59	6.29	.15 (6.50)	.11 (2.12)
ABSENT	3.11	2.39	-.03 (1.22)	.04 (.62)
TVHOURS	2.54	2.20	-.13 (2.23)	.08 (.51)
WORKHRS	14.83	15.39	-.02 (2.73)	-.02 (.95)
SOCIO-ECON	.28	.39	.90 (5.77)	.55 (1.43)
CHILDREN	1.82	1.69	-.06 (1.04)	-.08 (.49)
TWOPAR	.81	.85	-.25 (1.00)	.14 (.22)
MWBS	.37	.29	-.16 (.82)	-.61 (1.20)
READTO	2.98	3.15	-.04 (.60)	.02 ([illegible]6)
PARTALK	3.46	3.62	.29 (2.56)	.28 (.89)
HISPANIC	.10	.06	-1.64 (4.97)	.39 (.40)
EXPEN	2096.49	1678.28	.00003 (1.94)	.000004 (.13)
PEERS	.28	.42	.51 (.70)	-1.69 (1.04)
PARENT	1.26	1.47	.50 (1.87)	.09 (.19)
INTERCEPT			5.18 (5.49)	.52 (.19)
R-square			.73	.70
Observations			2075	338

Table 4: Advantage (disadvantage) of Public over Private Schools Without Correcting for Selectivity Bias

Variable	Coefficient Public (b1)	Private (b2)	b1-b2	Public Mean (x*)	(b1-b2)x*
SEX	1.15	.74	.407	.463	.188
BLACK	-1.026	.186	-.212	.077	-.016
HWHRS	.147	.106	.042	4.62	.194
ABSENT	-.030	.042	-.071	3.09	-.219
TVHRS	-.125	.076	-.201	2.56	-.515
WORKHRS	-.023	-.019	-.003	14.70	-.044
SOCIO-ECON	.90	.547	.353	.044	.016
CHILDREN	-.062	-.079	.017	1.82	.065
TWOPAR	-.251	.143	- 394	.809	-.319
MWBS	-.163	-.610	.447	.374	.167
READTO	-.038	.023	-.061	3.03	-.185
PARTALK	.291	.279	.012	3.45	.041
HISPANIC	-1.64	.392	-2.03	.104	-.211
EXPEN	2.60 E-4	3.98 E-5	.00022	2106.08	.463
PEERS	.511	-1.69	2.198	.279	.613
PARENT	.501	.093	.408	1.26	.514
SENIOR MATH				22.15	
SOPHMATH	.615	1.19	-.573	20.72	-11.87
SOPHMSQ	.0047	-.0074	.012	484.55	5.82
INTERCEPT	5.18	.518	1	4.66	4.66
Total advantage (disadvantage if negative) of public schools					-.64

Table 5: Advantage (disadvantage) of Public over Private Schools Correcting for Selectivity Bias with Expenditures per Student as Measure of Public School Quality

Variable	Coefficient Public (b1)	Private (b2)	b1-b2	Public Mean (x*)	(b1-b2)x*
SEX	1.18	.442	.738	.463	.342
BLACK	-1.10	-.192	-.912	.077	-.070
HWHRS	.144	.092	.052	4.62	.240
ABSENT	-.049	.040	-.089	3.09	-.275
TVHRS	-.164	.0669	-.231	2.56	-.591
WORKHRS	-.027	-.028	.0012	14.70	.018
SOCIO-ECON	.839	.215	.624	.044	.028
CHILDREN	-.039	-.074	.036	1.82	.065
TWOPAR	-.296	.656	-.952	.809	-.770
MWBS	-.120	-.477	.357	.374	.134
READTO	-.067	-.033	-.034	3.03	-.102
PARTALK	.229	.219	.010	3.45	.035
HISPANIC	-1.78	.448	-2.23	.104	-.232
EXPEN	1.98 E-4	3.28 E-4	-.00013	2106.08	-.273
PEERS	.835	-1.99	2.82	.279	.788
PARENT	.619	.166	.453	1.26	.569
SENIOR MATH				22.35	
SOPHMATH	.649	1.218	-.569	20.72	-11.79
SOPHMSQ	.004	-.008	.012	484.55	5.80
CONSTANT	5.17	-.895	6.07	1	6.07
BIAS	-.886	-.985	.099	-.162	-.016

Total advantage (disadvantage if negative) of public schools	-.04

Table 6: Advantage (disadvantage) of Public over Private Schools Correcting for Selectivity Bias with Average Gain on Math Tests as Measure of Public School Quality

Variable	Coefficient Public (b1)	Private (b2)	b1-b2	Public Mean (x*)	(b1-b2)x*
SEX	1.18	.442	.738	.463	.342
BLACK	-1.10	-.192	-.912	.077	-.070
HWHRS	.144	.092	.052	4.62	.240
ABSENT	-.049	.040	-.089	3.09	-.275
TVHRS	-.164	.067	-.231	2.56	-.591
WORKHRS	-.027	-.028	.0012	14.70	.018
SOCIO-ECON	.839	.215	.624	.044	.028
CHILDREN	-.039	-.074	.036	1.82	.065
TWOPAR	-.296	.656	.-952	.809	-.770
MWBS	-.120	-.477	.357	.374	.134
READTO	-.067	-.033	-.034	3.03	-.102
PARTALK	.229	.219	.010	3.45	.035
HISPANIC	-1.78	.448	-2.23	.104	-.232
EXPEN	1.98 E-4	3.28 E-5	-.00013	2106.08	-.453
PEERS	.803	-2.09	2.89	.279	.807
PARENT	.634	.159	.475	1.26	.597
SENIOR MATH				22.35	
SOPHMATH	.654	1.21	-.551	20.721	-11.42
SOPHMSQ	.004	-.008	.012	484.55	5.62
CONSTANT	5.14	-1.86	7.01	1	7.01
BIAS	-1.107	-1.98	.873	-.167	-.164

Total advantage (disadvantage if negative) of public schools .904