Discusssion Paper No. 18

Williams Project on the Economics of Higher Education
Williams College
Denison Gatehouse
Williamstown, MA 01267

The College Investment Decision: Direct and Indirect Effects of Family Background on Choice of Postsecondary Enrollment and Quality

Jere R. Behrman, Lori G. Kletzer,
Michael S. McPherson and Morton Owen Schapiro

July 1992
DP-18

THE COLLEGE INVESTMENT DECISION:

DIRECT AND INDIRECT EFFECTS OF FAMILY BACKGROUND ON CHOICE OF POSTSECONDARY ENROLLMENT AND QUALITY

Jere R. Behrman, Lori G. Kletzer, Michael S. McPherson, and Morton Owen Schapiro

July, 1992

Behrman, University of Pennsylvania; Kletzer and McPherson, Williams College; Schapiro, University of Southern California. Scott Blasdell and Anne Karabinos provided valuable research assistance. The authors wish to acknowledge the helpful comments of Richard Murnane and participants in the National Bureau of Economic Research conference on economics of higher education in April, 1992. The Andrew Mellon Foundation and the National Science Foundation (grant number SED-9115440) provided financial support.

The College Investment Decision: Direct and Indirect Effects of Family Background on Choice of Postsecondary Enrollment and Quality

by Jere R. Behrman, Lori G. Kletzer, Michael S. McPherson, and Morton Owen Schapiro

This paper focuses on two components of the postsecondary schooling decision process. The first concern is to investigate the implications of the endogeneity of high school scholastic achievement in analyzing postsecondary school choice. The second concern is to incorporate an explicit analysis of choice of institutional quality into the investigation of postsecondary enrollment behavior. Our theme in both these components is the role of family background. Our basic data source is the National Longitudinal Study of the High School Class of 1972 (NLSHS72). Family background characteristics show the expected effect on scholastic achievement: students from families with high income, better educated parents, or parents with higher socioeconomic status generally do better on tests of scholastic achievement. Scholastic achievement is positively related to attending a four-year postsecondary school, but not to attending a two-year school. The estimated effects of achievement are larger with achievement treated as endogenous than when achievement is treated as exogenous. We find similar influences of family background directly on postsecondary enrollment and quality.

It is well known that young people's decisions about whether and where to attend college are importantly influenced by such family background characteristics as parental education and income, family size and birth order. Some of these variables no doubt have at least part of their influence directly, at the time of college choice - as, for example, differences in family income influence the range of postsecondary alternatives families find affordable. But many of these variables operate as well by influencing the environment in which children grow up. Thus, the positive correlation generally observed between higher family income and scholastic achievement in high school is presumably partly due to the broader range of educational opportunities more affluent families can provide for their children.

Sorting out these direct and indirect effects of family background on postsecondary educational choices is interesting in its own right, and also carries implications for correctly estimating econometric equations describing patterns of college choice. For if such personal characteristics of young people are in fact endogenously determined, reflecting the influence of family background variables, then studies that treat variables like high school achievement as exogenous or predetermined will result in biased estimates.[1] As we report below, exogeneity tests performed on our data lead us to reject the hypothesis that high school academic achievement is exogenous with respect to the college attendance decision. Modeling achievement as endogenous rather than exogenous turns out to have substantial effects on the estimated magnitude of coefficients of interest. We therefore proceed in our estimates both to account for this endogeneity in estimating the effect of high school achievement on postsecondary attainment and quality decisions, and to distinguish the part of the influence of family background on college attainment and quality decisions that operates through affecting high school achievement from the part that operates directly.

---

1 Most studies of college attainment and choice behavior treat test performance or other achievement measures in high school as exogenous. See, for example, Manski and Wise (1983), Altonji (1990) and Grubb (1992). In treating achievement as endogenous we are not attempting to distinguish between family effects through heredity and family effect s through home environment.

As we have noted, students differ not only in whether but where they choose to attend college - that is, these choices have an important qualitative dimension. Here we capture these dimensions of choice in two steps: by distinguishing between the choice of enrollment at a two-year versus four-year college, and by measuring the quality of four-year college by the level of an institution's instructional expenditures per student.

Section 1 presents our model. Section 2 presents and discusses the data. We discuss our estimates of the determinants of post-secondary school attendance and present simulations of the effects found in Sections 3 and 4, respectively. Section 5 presents an analysis of choice of quality of four-year institution. Section 6 presents concluding comments.

Section 1. Model of Post-Secondary School Determination

We follow the basic approach to modelling choice of educational attainment level set out in Manski and Wise (1983). Assume that individuals choose a level of educational attainment that maximizes lifetime utility. Utility (V) for individual i in educational attainment level j can be expressed as:

$$V_{ij} = X_{ij}\beta + u_{ij}$$

where X is a vector of family background characteristics, labor market characteristics, and achievement, and u is a random error term. Individual i chooses attainment level j if $V\{ij\} > V\{ik\}$ for all k not equal to j.

For our educational attainment model, there are 3 attainment levels ([1] high school only, [2] two-year college, [3] four-year college/university). Let x(i) be the vector of characteristics for individual i. Then the probability that an individual with characteristics x(i) will choose the two-year college attainment level is:

$$P_{i2} = \exp(\beta_2' x_i) / \sum \exp(\beta_k' x_i),$$

with a normalization that $\beta[1] = 0$. With this normalization, the number of parameters to be

estimated is equal to the number of individual characteristics multiplied by two (the number of educational attainment levels minus one).

We adopt a fairly standard general approach to the determination of post-secondary school attendance through conditional reduced-form demand relations.[2] The following theoretical discussion focuses on variables for which we have measures in our data set. We posit that post-secondary schooling (S) for high school seniors depends on: family background characteristics (F) such as parental education and income[3],father's Duncan socioeconomic index (which we interpret as a longer-run income measure[4]), number of siblings, and birth order;[5] individual demographic characteristics (I) such as sex and race; individual "achievement", measured as a score on a test administered during the high school senior year (A[6]); labor market characteristics (LM) such as the opportunity costs of time spent in post-secondary school (the wage rate for those who have completed high school but not continued with postsecondary schooling in the same geographic region and of the same sex and race) and the unemployment rate (for the sex/race group) and the expected gains from college education (i.e., the present discounted value of the lifetime wage gain expected from completing a four-year college course for those in the same region and of the same

---

2 The reduced form relations should not be interpreted as pure demand measures. Where admissions are rationed, the reduced forms include admissions effects. Admissions rationing is likely to be important only for four-year colleges and universities, and even for these institutions, rationing is far from pervasive. In 1989, 69% of freshmen reported that the college they were attending was their first choice (Chronicle of Higher Education, January 24, 1990).

31ncome usually is included in such relations if data on it are available. However, if schooling is a pure investment decision and if access to credit is not correlated with income, a priori income should not be a factor. If, on the other hand, schooling is partly a consumption good for the parents or for the children, income should be relevant.

4The original basis for this index was the association between occupational prestige scores and mean occupational income and education, with the association between the index and income particularly strong. Others also have used this index as an indicator of permanent income (e.g., Chamberlain and Griliches 1977). Zimmerman (1992) presents empirical evidence indicating that the Duncan index is a preferred measure of permanent status.

5The number of siblings could be viewed as an endogenous variable within a quality-quantity framework. We discuss this point further below.

6We discuss these tests in more detail below.

sex and race); prices broadly defined (P) of local post-secondary school options as represented by state and local expenditures on post-secondary schooling and distance to closest two-year and closest four-year post-secondary schools; and a stochastic disturbance term (U):

(1) S = S(F, I, A, LM, P, U).

In order to explore the hypothesis that high school achievement (A) is endogenously determined, we include in the model a second equation explaining high school achievement. We expect achievement levels to depend on family background characteristics (F) that influence the quantity of home-produced human capital, individual demographic characteristics (I) including race and sex, and a set of variables (C) reflecting characteristics of the student's high school, the local labor market, and (through expectations effects) the quality and price of postsecondary education in the student's home state. (The variables in C include some of those in LM and P above.) The equation also includes a stochastic disturbance term V.

(2) A = A(F, I, C, V).

We make three observations about our specification. First one basic difference between ours and the usual specification is that we explicitly realize that S is not a scalar for years of post-secondary schooling but a vector that includes both the years of post-secondary schooling (Y) and the quality of that schooling (Q). That is, we are explicit about the constrained choice element regarding the quality of post-secondary schooling in addition to the quantity of such schooling. Second, while our model focuses on the possible endogeneity of high school achievement, other variables included in equation 1 might also be endogenous. Parents might make simultaneous decisions about the number and the quality of their children, their own time use (and therefore their income), and their location (and therefore local labor market conditions, high school quality

and post-secondary schooling options). Indeed parental schooling may be associated through genetic and home environmental links with unobserved ability and motivation of their children.[7] Ultimately, from a long enough dynastic perspective, all of the variables in our and other data sets can be viewed as possibly endogenous. However our prior is that the variable of most concern with regard to the possible endogeneity question on the right-side of relation (1) is individual achievement (A) and that is where we concentrate. Third while the right side of our specification in a general sense is similar to the standard ones in the literature, as Mies van der Rohe stated, "God is in the details." Our estimates of this model, in addition to the inclusion of quality as a choice variable, are distinguished from other studies in the details of the empirical specification in regard to the empirical representation of various variables. In Section 2 we discuss our linkage of our basic data (the National Longitudinal Study of the High School Class of 1972) with other data sets (e.g., CPS, HEGIS) in order to obtain better representations of variables related to labor market conditions, post-secondary school quality, and state and local post-secondary school options. We also explore interactions between individual characteristics such as sex and race and labor market conditions by defining our labor market variables for such groups.

Section 2. Data

Our basic data source is the National Longitudinal Study of the High School Class of 1972 (NLSHS72). The NLSHS72 is a Department of Education survey of 22,652 individuals who were high school seniors during the 1971-72 school year. The Department recovered further information on 13,274 of these individuals through its Postsecondary Education Transcript Survey (PETS). The initial interview was conducted in the Spring of 1972, with follow-up surveys in 1973, 1974, 1976, 1979, and 1986. We restricted the sample in this paper to the 8,484 individuals in the Base Year (1972) Survey who were included in PETS and about whom there is information

7See, for example, Behrman and Taubman (1989).

for all of the variables in which we are interested.[8] In Appendix I, we provide a detailed description of NLSHS72 and the other data sets we use.

The NLSHS72 data provide considerable detail on family background and measures of achievement. As a longitudinal data set, NLSHS72 provides a number of rounds of data on individuals so that the data that we use are not subject to recall problems over long durations. The sample statistics are presented in Table 1.[9]

As explained more fully in Appendix 1, we use data from the Current Population Survey of the U. S. Census to develop measures of labor market conditions. First is the opportunity cost of student time, represented by the median wage for individuals who have completed high school but not gone on to post-secondary school and who are 18-21 years of age. Second is the expected return to investment in higher education, proxied by the discounted value of the wage differential between persons who have four years of college and persons who stopped their formal education with completion of high school. Of course for both of these variables the individuals concerned may make adjustments given their knowledge of their own characteristics and of the selection processes involved in the decisions about who continues in post-secondary schooling that underlie the CPS data. We control for such adjustments by constructing these two labor market variables for individuals who are similar with respect to race, sex and region and by controlling in our

---

8 Several variables important to our analysis cause reductions in sample size: 1972 family income (lose 3384 observations); Duncan index for the father (lose 1633 observations); distance to postsecondary school (lose 117 1 observations); birth order (lose 614 observations); and race (lose 432 observations). Assorted smaller reductions in sample size occur due to missing information on weeks in the school year, average family income for the high school, and father's education. A comparison of simple variable distributions on the full sample versus the smaller analysis sample shows that they are on the whole quite similar. The analysis sample has a smaller representation of Blacks (10.2% vs. 14.3%) and is a bit more "male" (52.4% vs. 49.6%). Differences in family income, parental education and father's Duncan index are quite small.

9NLSHS72 is a multistage cluster sample with nonrandom selection probabilities. In the original design schools with high percentages of minority students and schools from low-income areas were overrepresented. Sampling weights also vary by status (public, private), region, and city size. The results reported here come from an unweighted estimation. We did not weight the sample because, after observation losses due to missing data, it is difficult to relate our final sample to the underlying population.

estimates for the observed individual, family background, and test score/achievement characteristics.

We use data from the federal government's Higher Education General Information Survey (HEGIS) to construct our measure of the quality of higher education -- instructional expenditures per student over the 1975-76 period.[10] The use of a continuous variable to measure educational quality is a distinctive feature of this study. This measure exhibits substantial variance across four year colleges and universities. However there is relatively little variance across two-year institutions. Moreover, the vast majority of students who attend two-year institutions attend one to which they can commute, with little apparent effort to choose on the basis of quality. Therefore we limit our exploration of quality choices below in Section 5 to choices among four-year institutions (with control, of course, for who selects to attend four-year institutions).

HEGIS also provides us with data on local and state appropriations for higher education in 1974-5 for the state in which the individual attended high school. These are indicators of local resources devoted to public higher education -- an important component of the local options for post-secondary education. We also employ a measure of tuition for state residents at public four-year institutions averaged over the 1972- 1976 period as an additional measure of price.[11]

NLSHS72 provides various test scores that can be used as measures of individual "achievement." As part of the Base Year survey in 1972, each student was asked to complete a 69-minute Test Book measuring verbal and nonverbal ability. The Test Book consisted of six tests, which we list here and describe in more detail in Appendix 2. The tests were: Vocabulary, Reading, Letter groups, Mathematics, Picture number, and Mosaic Comparisons. The first four

---

[10]It is likely that there is some downward bias in estimates of instructional spending per student at public institutions, because in many states significant parts of employment costs (pensions and insurance) may be paid directly by the state government rather than by the institution. We have no systematic means for correcting this possible bias, which presumably varies across states.

[11]We are grateful to Tom Kane for making us aware of these data and providing us with a copy of them.

components of the Test Book test for skills akin to skills tested in the SAT[12]. Our ability measure is a simple sum of scaled scores from all six parts of the Test Book.[13]

Section 3. Estimates of Post-Secondary Education Attendance

Section 3.1. Determinants of high school achievement

Because our estimates of post-secondary education decisions are conditioned on estimates of determinants of high school achievement, we turn first to our estimates of that latter equation (equation 2 above). These results are reported separately for Whites and for Blacks and Hispanics in Table 2. Family background characteristics show the expected effects: students from families with higher income, better educated parents, or parents with higher socioeconomic status generally do better on the test battery we use to measure achievement, although the effect of father's education is of the wrong sign (but not significant) for the Black and Hispanic subsample. Family size (number of siblings) has no significant effect, but among Whites, older children do significantly better than younger ones, controlling for number of children. It is of some interest that a variable indicating whether the student's mother was in the labor force during the student's pre-school years has a significant effect: students with non-working mothers have higher achievement. This may represent a human capital effect - as non-working mothers may devote more time to their children's early education - but it may also proxy for other variables (like family affluence at that point in the child's life cycle) which we capture only imperfectly with other variables. The demographic variables, race and sex, tell us that among Whites women score significantly higher than men. Among Blacks and Hispanics, there is no significant gender difference, but Hispanics score higher than Blacks.

12Manski and Wise (1983) use these four components to predict an individual's SAT score.

13Duc to the proprietary nature of the Test Book, a copy of the test is not available to data users. Only formula and scaled scores exist on the data file.

We turn next to our measures of variables external to the family. Our measure of labor market conditions for high school students (median wage for high school students), which is intended to proxy the opportunity cost of studying, is not significant. Length of the school year is not significant, but students from schools with more affluent clienteles do better on the tests. This may reflect peer effects in learning (given that more affluent students generally have higher achievement), differences in the resources available to schools with wealthier and poorer students, or other effects associated with the neighborhoods where students reside. We had expected our measure of the quality and availability of post-secondary options (state and local higher education expenditures per capita) to have a positive sign, reflecting incentives to invest in high school achievement, but the variable is insignificant for Blacks and Hispanics, and significant and negative for Whites.

Our general conclusion from Table 2 is that our model explains a small but significant part of the variation in high school achievement as measured by test scores, with coefficients that generally conform to theoretical expectations. This does not, of course, in itself tell us whether high school achievement should be treated as endogenous in explaining post-secondary attainment, a question which turns on the correlation between the error terms U and V in equations 1 and 2.

Our exploration of this issue involves a Hausman (1978) test. The test involves estimating the model for post-secondary enrollment summarized in equation 1 above, including in the estimating equation both observed test scores and the residuals from the equation explaining test scores. If the estimated coefficient on the residual is significantly nonzero, then a null hypothesis of exogeneity can be rejected. As explained further below, we estimate post-secondary enrollment in a multinomial logit framework. When this equation is estimated with the residuals from the test score equation included, we find that the residuals are significantly different from zero for explaining four-year college, although they are not significant for two-year college enrollment. (These results are not reported here.) Based on the results for four-year colleges, we conclude that the hypothesis of endogeneity should be maintained. Hence, our focus in the

ensuing is on equations in which predicted values of achievement from the equation reported in table 2 are used in place of the measured values.[14] Results treating achievement as exogenous (and therefore using the measured values of achievement) are reported in Tables 6 and 7 and are briefly discussed in the following subsection, when the test scores variable is considered.

Section 3.2. Summary of Multinomial Logit Estimates for Postsecondary Attainment[15]

We turn to our estimates of a multinomial logit for three outcomes: no post-secondary school, post-secondary education attendance at two-year schools only (whether or not completed), and attendance at four-year schools (whether or not completed or continued on).[16] Table 3a gives these estimates for 7281 Whites and table 3b for 1203 Blacks and Hispanics. We separate Whites from Blacks and Hispanics because of differences in the relations between the two groups. Within each group we include dichotomous variables for females versus males. Among the Whites, 3218 (44.2%) had no post-secondary schooling, 1252 (17.2%) attended two-year schools, and 2796 (38.4%) attended four-year schools. Among the Blacks and Hispanics, 634 (52.7%) had no post-secondary schooling, 222 (18.5%) attended two-year schools, and 345 (28.7%) attended four-year schools.

Family background **(F):** For Whites, parental schooling has significantly positive effects on attendance at both two-year and four-year schools. For father's schooling, but not for mother's schooling, the four-year effect is significantly larger than the two-year effect. The

---

14The following variables, included in the equation for achievement but omitted from the multinomial logit equation explaining post-secondary enrollment, permit the impact of high school achievement on post-secondary choice to be identified: average family income for students in the same high school, high school opportunity wage, number of weeks in the school year, and mother not in labor force before elementary school. The first and last of these variables are statistically significant in the achievement equation.

15Throughout this section, unless otherwise noted, we use "significant" to refer to two-tailed asymptotic t tests at the standard 5% level.

16The estimated standard errors in Tables 3a and 3b must be adjusted to account for the use of a predicted regressor (achievement). We use the adjustment procedure outlined in Murphy and Topel (1985).

coefficient estimate for father's schooling is larger than that for mother's schooling for four-year schools and vice versa for two-year schools, but in neither case is the difference significant. For Blacks and Hispanics, parental schooling has a significant positive effect at the four-year level.

Family income higher than $15,000 and father's socioeconomic index both have significantly positive coefficient estimates for four-year schools for Whites.[17] Neither of these variables has significant coefficient estimates for two-year schools for Whites, nor for either two-year or four-year schools for Blacks and Hispanics. These estimates suggest that neither current nor more permanent indicators of family income affected two-year post-secondary school attendance of Whites, Blacks, and Hispanics. Such a result is consistent with investment motives dominating in these post-secondary school decisions, though it is not clear how to reconcile the interpretation of these results as supporting the dominance of investments motives for these decisions with the results for four-year schools for Whites. It may be the case that the two-year results reflect the relative lack of liquidity constraints at that educational level.

Overall, these estimates must be qualified since the available representations of both current and more permanent income may have measurement problems. The family income variable is an estimate by the sample members when they were seniors in high school in the first NLSHS72 round and probably is contaminated by considerable measurement error (quite possibly with a systematic tendency to over-report low income and under-report high income due to community norms regarding the prevalence of middle income families in the U.S.). Father's socioeconomic index is not designed to measure permanent income, though it is highly correlated with longer-run income in samples that include both and does effectively average out transitory within occupation fluctuations in income.

The number of siblings has significantly negative coefficient estimates for both subsamples in the four-year equation. This suggests a greater constraint on family resources of having more

17Median family income was $10,880 in 1971 and $11,629 in 1973 (United States, Economic Report of the President, 1992).

children whether or not the family size-child quality decisions were simultaneous as in the Becker and Lewis (1973) and Willis (1973) models. As noted, these estimated effects are significant only at the four-year level. Interestingly, this is also the only case in which the socioeconomic status variable (and income for Whites) is significant. This may reinforce the resource constraint interpretation since only if income is an important constraint would it seem relevant among how many individuals the income had to be shared.

For Blacks and Hispanics, higher birth order (being a younger sibling) is positively and significantly related to four-year post-secondary schooling attendance. For Whites, the four-year estimate is also positive, but small and not statistically significant. At the two-year level for Whites, birth order has a negative and statistically significant estimated coefficient. For Blacks and Hispanics, these results are consistent with results in Steelman and Powell (1989), who found higher birth order children to be favored with respect to financial aid. The failure to find a positive effect of birth order on attendance at community colleges may be explained by the low levels of tuition and student aid at these institutions.

Individual characteristics **(I):** There are significant differences between the two racial/ethnic groups, as we note at the start of this section and in the discussion of the all of the variable groups. These are explored further in the simulations of Section 4. Within the Black and Hispanic group, however, the dichotomous additive variable for being Hispanic is negative, but not significant. The absence of a significant difference is somewhat comforting given our aggregation of Blacks and Hispanics (impelled by sample size considerations) into one group.

The dichotomous variable for being female has negative and significant coefficients for Blacks and Hispanics at two-year schools and for Whites for four-year schools. In both cases, the effect is relatively large. The absolute magnitude of the White female effect on estimated four-year school attendance is three times larger than the effect of increasing family income from below $6,000 to over $15,000 per year. For the Black and Hispanics subsample, the female effect

is more than three times greater than the effect of increasing father's education from less than high school to four years or more of college.

Test Score/"High school Achievement" (A): For both racial/ethnic groups, test scores have significantly positive estimates for attending four-year (but not two-year) institutions. These results use predicted values for test scores to account for the endogeneity of this variable. Tables 6 and 7 report results which use observed test scores in place of these predicted values, as would be appropriate if achievement were exogenous. These equations yield the same pattern of sign and significance as the results with test scores endogenous. However, the coefficient on achievement in the four-year equation for Blacks and Hispanics is more than twice as high when achievement is treated as endogenous (.021 exogenous; .059 endogenous). For Whites, the coefficient value in the four-year equation is also higher when the variable is treated as endogenous, although the difference is smaller and not significant (.029 exogenous; .038 endogenous). Apparently, when achievement is treated as exogenous - a familiar assumption, as we noted earlier, in studies of college enrollment decisions - part of the effect of achievement is erroneously attributed to the influence of background variables that influence achievement levels.

Labor market conditions (LM) The variable representing the opportunity cost of time to attend post-secondary school in terms of foregone labor market earnings has significant negative effects on enrollment for Whites at four-year colleges and for Blacks and Hispanics at two-year schools.[18]

The representation of the expected returns from college is significant only for Whites at two-year institutions, and here it has the expected positive sign. Notice that in none of the four

---

[18] In contrast to our results, Grubb (1990, p. 16) finds the signs on his labor market opportunity cost variables to be the "opposite to what one might expect" and conjectures that such results may reflect that "students rely on earnings to finance a substantial part of the postsecondary expenses, and therefore . . . high earnings may increase their ability to enroll, rather than reducing it because of increasing the opportunity cost....".

cases (two-year vs. four-year, White vs. Black and Hispanic) are the coefficients on both the opportunity cost and the returns variables significant. This may well reflect the substantial collinearity between these measures. On the whole, however, our results seem more consistent with expectations regarding both opportunity costs and returns than those Grubb (1990) developed using the same NLSHS72 data. This may result from our efforts, discussed above, to develop measures that better capture the underlying processes at work.[19]

Prices of post-secondary education, broadly defined (**P**): Local and state expenditures per capita on higher education have significantly positive effects on two-year school attendance for Whites, Blacks and Hispanics. This result suggests that at the margin, such resources were directed more at two-year institutions in the early and mid 1970s. Differences in public post-secondary school availability and quality, thus, apparently induce substantial post-secondary school attendance, at least at two-year institutions.

Higher public tuition at four-year institutions is associated with a higher probability of enrollment at two-year institutions for Hispanics and Blacks. For Whites there is an unexpected positive relationship between public four-year tuition and four-year enrollment. It is possible that price is capturing a quality effect here for which we are unable to control.

The distances to the closest two-year and four-year colleges have exactly the sign pattern of estimates that would be predicted a priori: for each distance variable negative for attendance at that type of school and positive for attendance at the other type of school. Three of the four own effects are significantly nonzero (the exception being White attendance at four-year schools). The own distance effects for two-year schools are somewhat larger than the own effects for four-year schools for both groups, which seems plausible a priori since students attending two-year schools are much more likely to live at home and to commute than students attending four-year schools.

---

19 Grubb (1990, p. 15) summarizes his results by stating that "[t]he measure of rates of return . . . are not generally significant except that enrollments by women in community colleges . . . respond positively to higher returns to one to three years of college."

Finally, for Whites the greater the distance to a four-year college, the greater the probability of attending a two-year school.

Section 4. Simulations of Various Effects

The multinomial logit estimates are somewhat difficult to interpret directly because of their interactive nature and the dependence of their effects on the probabilities of school attendance on the values of all of the right-side variables. Therefore we now present some simulations based on the estimates that are discussed in Section 3.1 and 3.2 of a number of different sets of assumptions that are summarized in Table 4. For those variables that appear in the high school achievement equation of Table 2, the simulations account for both the direct effect of the variable on attendance probabilities and the indirect effect operating through achievement. In several cases, we report the direct effect separately in parentheses to illustrate the impact of these direct and indirect effects.

To set a benchmark for later comparisons, we start with two base cases, which set the right hand side variables to their mean (or for categorical variables, their median) values. Base case A uses the mean and median values for the White subsample, while Base case B uses means and medians for the Black and Hispanic subsample. These two cases allow us to consider how college attendance patterns would differ if Whites had the family background and environmental characteristics of Blacks and Hispanics, and conversely. From this starting point, we proceed to simulate the effect on these base values of changes in the values of specific variables from their base levels, allowing us to provide a more tangible sense of the influence of different variables.

1. Base Case A: Predicted enrollments for Whites. Blacks. and Hispanics. setting right hand side variables at means for White sample. The model reports that, for a White male with mean characteristics, the probability of attendance at a two-year college is .142, while the probability of

four-year college attendance is .476.[20] For White females, the attendance probability at two-year colleges is .214; at four-year colleges it is .331. Note that the indirect effects of background characteristics on high school achievement are important in explaining women's college enrollment: if those indirect effects are omitted, the predicted enrollment probability for women drops from .331 to .278.

More interesting, however, are the effects on Black and Hispanic college attendance rates that would be observed if their parents' education levels and incomes (as well as levels of other variables included in the model) matched those of Whites. If we insert the mean values of right hand side variables observed for Whites into the equation used for estimating the enrollment behavior of Blacks and Hispanics, we observe striking effects. Four-year college enrollment probabilities for Blacks and Hispanics would be very close to, and in fact slightly higher than, those for Whites with the same (white mean) characteristics. Interestingly, two-year college enrollment probabilities would be substantially lower than those for Whites. To put the point differently: were it not for differences in family background and other conditioning variables, Hispanics and Blacks would be substantially more likely than Whites to attend four-year colleges. This is the reverse of the pattern actually observed.[21]

2. Base Case B: Predicted enrollments for Whites, Blacks, and Hispanics setting right hand side variables at means for Black and Hispanic sample. These points can be further clarified through examining Base Case B, which shows enrollment probabilities for all race and sex groups on the assumption that all students have the mean characteristics of the Black and Hispanic

---

20 As it happens, this value is fairly close to actual observed attendance probabilities (these are .17 and .46). However, the correspondence between the predicted enrollment probability for students with average characteristics and observed enrollment probability in the sample need not be close in general. Differences may come about because the logit model is non-linear and because categorical variables are set to median rather than mean values.

21 Manski and Wise (1983) report a similar finding: black high school graduates are considerably more likely to apply to four-year colleges than whites with the same characteristics.

subsample. Comparing the two base cases lets us identify more precisely the impact of the difference between Black and White background characteristics. Thus, if White students had the mean characteristics of nonwhite students (including, among other things, less parental education, lower family income, larger family size and higher unemployment) their enrollment probabilities would be lower than those of Blacks and Hispanics with those same characteristics at both two-year and four-year colleges. Comparing Base Case A to Base Case B, we see that for White males, a change from assigning average White characteristics to assigning average nonwhite characteristics results in a drop in the predicted enrollment probability from .142 to .092 at two-year colleges and from .476 to .132 at four-year colleges. A difference of comparable magnitude is predicted for White females (from .214 to .120 at two-year colleges; from .331 to .079 at four-year colleges). Conversely, if a Black female with average characteristics had instead the background characteristics of an average White female, her two-year college enrollment probability would fall from .153 to .038, a decline that would be more than offset by an increase in the four-year enrollment probability from .235 to .459. Comparable changes would occur for other nonwhite groups.

The next two sets of simulations examine the effects of changes in parents' education and family income on enrollment patterns. All variables except those noted remain at sample means.

3. Varying parental education and family income: White students. The first simulation here shows the impact on White enrollments of increasing the parental education level from "high school graduate" to "some college". For White males, two-year college enrollment probabilities drop by .038, compared to Base Case A, while four-year college enrollment probabilities rise by .228. For White females, the effects go in the same direction, but the positive impact on four-year college enrollment is much smaller (5.4 percentage points).

Case (3b) shows the effect of variation in income on enrollment patterns for White

students. For men, the impact of an increase in income from the $9-12,000 interval to the $12-15,000 interval is positive, but strikingly smaller than the impact of an increase in educational level. Women's two-year enrollment rises, but by less than four-year enrollment drops. This somewhat anomalous outcome results from the fact that the coefficient on the $12-15,000 income variable is smaller than that on the $9,-12,000 variable. However, since this difference is not statistically significant, the result shouldn't be given much weight.

The importance of interaction effects in the logit estimates is illustrated by (c) under case 3, which shows the combined effect of the education and income changes shown under (a) and (b). For men, the results are almost identical in case (a) and (c), but for women, the interaction effect is important, so that the combined effect of improved parental income and education produces a large increase in the four-year college enrollment probability.

Finally, (d) shows the effect of further increases in both parental income and education. Highly educated and affluent parents are substantially less likely to have children enrolled in two-year colleges and substantially more likely to have them enrolled in four-year colleges.

4. Varvina parental education and family income: nonwhite students. Increasing parental education from the median value of "less than high school" to "high school" produces drops in two-year college enrollment probabilities that are more than offset by increases in expected four-year enrollment compared to base case (b). The increase in the four-year enrollment probability is substantial - on the order of 0.1, except for Black women, where it is about half as large.

A modest income increase, from between $6,000 and $9,000 to between $9,000 and $12,000, illustrated in case 4b, yields quite a different response pattern. All groups show a substantial drop in two-year enrollment probability with only a very modest increase in four-year enrollment probability. When the education and income effects are interacted in case c, the result is higher four-year and lower two-year enrollment than in either (a) or (b). The effect is primarily to redistribute enrollment between the two sectors, with little increase in overall

enrollment compared to the base case. But as (d) shows, a further increase in both income and education of parents leads to quite substantial increases in both two-year and four-year enrollment relative to c. Here the effect is much more to raise total enrollment than to produce enrollment redistribution, relative to the base case. Nonlinearities and interaction effects appear to be quite important.

Section 5. **Choice of Quality Among Four-Year Schools**

We now turn to the determinants of schooling quality for those who attend four-year postsecondary institutions. Of course those who attend such schools are not a random draw from the population of high school seniors, as is reflected in the estimates of Section 3. Therefore we control for selectivity in our estimates by estimating a probit for whether or not a high senior in the Class of 1972 sample attended a four-year post-secondary institution (first and third sets of estimates in Table 5). The set of right-side variables in this probit is the same as that for the multinomial logits that are discussed in Section 3. The signs and relative magnitudes of the coefficient estimates are basically the same, so we do not discuss them here. The second and fourth columns in Table 5 give our estimates of the choice of quality, as represented by instructional expenditures per student, with control for sample selectivity using the probits, respectively for Whites and for Blacks and Hispanics.[22] For Blacks and Hispanics we find little in the results that is statistically significant. Therefore we focus exclusively on the estimates for Whites in the rest of this section.

Family background variables **(F):** Mother's schooling has a significant positive impact on

22These estimates include the selectivity controls from the probit. The selection control is identified by the nonlinearity of the probit equation. Further selection controls can be provided by adding exclusion restrictions. Decisions about which variables to exclude are, however, somewhat arbitrary. (Perhaps the most plausible candidates for exclusion would be the labor market variables.) Experimenting with alternative sets of excluded variables suggests that the results reported in Table 5 are generally robust with respect to these alternatives. This is not true of the results for females as noted in footnote 23 below.

choice of school quality, with completion of some college or of a college degree having a larger effect than completion of high school. For fathers, the sign on education is also positive, but is significant only for college graduates. Having a family income above $15,000 has a significant positive effect on choice of college quality, as does the Duncan index measuring father's socioeconomic status. Having more siblings significantly reduces the quality of college chosen, suggesting the presence of a quality-quantity trade-off for families in choosing colleges for their children.

Individual characteristics (I): While our earlier estimates showed that females have a lower probability of four-year college attendance, the sign on choice of college quality is negative but not significant.[23]

High school achievement/test scores (A): Like our earlier estimates, the estimates of the impact of test scores treat the test score (or achievement) variable as endogenous. Higher test scores have a positive association with the quality of four-year institution attended. This may reflect a greater return to investments in college quality for higher achieving students. The result may also reflect that higher quality schools (as measured by instructional spending per student) have more selective admissions procedures that are based in part on tests similar to the ones used here (e.g., achievement tests such as the SAT and ACT).

Other variables. The labor market variables (opportunity wage, unemployment rate and expected returns) are not significant. In general, residing in a state with higher public four-year college tuition leads to attendance at an institution with higher per student instructional spending.

[23] In an equation where labor market variables are excluded, the coefficient remains negative and becomes significant.

This may come about because states with higher tuition also spend more on instruction, or because higher tuition induces students to turn to alternative institutions with higher instructional spending. Not surprisingly, students from states with higher state and local higher education spending per capita tend to choose institutions with higher instructional spending per student. The coefficients on the distance variables are not significant.

Section 6. Conclusions

Explaining postsecondary attainment and institutional choice is a complex problem, and this study has focused on a limited part of the overall subject. Our principal concerns have been two: (1) to investigate the implications of the endogeneity of high school achievement in analyzing behavior toward post-secondary education and (2) to incorporate an explicit analysis of choice of institutional quality into the investigation of post-secondary enrollment behavior.

We find evidence that high school achievement levels (as measured by test scores) depend on many of the same background variables that influence college attendance patterns. Further, we are unable to reject the hypothesis that high school achievement is endogenous in our model of post-secondary enrollment behavior. Our evidence indicates that failure to account for endogeneity will lead one mistakenly to attribute part of the effect of high school achievement to family background variables that influence achievement levels. How much does endogeneity matter? One measure of the impact is to ask how much accounting for endogeneity changes the influence of a higher test score on the probability of attending a four-year rather than a two-year school. Table 3a shows that with achievement treated as endogenous, a one point improvement in test scores will increase the probability of attending a four-year school by 0.032 (=0.0382-0.0062). When achievement is treated as exogenous, as in the estimates in tables 6 and 7, the probability increase is only 0.0183 (=0.0294-0.0109). If the true coefficient is the endogenous one, then the treatment of the test score as exogenous results in a downward bias of 43% in this coefficient estimate - a rather large effect.

Our analysis of choice of quality of post-secondary institution relied on instructional expenditures per student as a simple measure of quality of four-year college or university. Although (perhaps owing to sample size) our results for nonwhite students were not satisfactory, we had considerable success in explaining choice of quality among White students attending four-year institutions. The education level and income of parents proved significant in explaining choice of quality, as did high school achievement levels (again treated as endogenous) and some variables measuring the price and availability of public higher education.

These two aspects of our study should, we believe, be taken into account in future work in this area. Acknowledging and accounting for the endogeneity of earlier achievement in explaining later educational choices - and more generally recognizing the interdependence of decisions taken throughout childhood and young adulthood regarding investment in human capital - complicates theoretical formulations and econometric estimation, but is, we believe, important in developing a more adequate understanding of these sequences of decisions. And the fact that these decisions have important qualitative as well as quantitative dimensions is of undoubted significance.

Finally, we would underline the important differences between the determinants of postsecondary school attendance and quality for Whites and for Blacks and Hispanics. If Blacks and Hispanics had the same values for the variables that determine postsecondary school attendance as do Whites, for example, our estimates suggest that they would have higher enrollment rates rather than the lower ones actually observed. Moreover, policy variables appear to have disparate effects on White and nonwhite populations, given the differences in background characteristics. Thus, increases in state and local spending on higher education tend to move Whites from two-year to four-year colleges, while moving Blacks and Hispanics in the reverse direction. There also are some substantial differences in the probabilities of postsecondary school attendance for females compared to males within each racial/ethnic group. That means that if there are reasons to influence the post-secondary education decisions of a particular group, the

policies adopted may vary depending on the target group.

References

Altonji, Joseph G. (1990) "The Effects of High School Curriculum on Education and Labor Market Outcomes," National Center on Education and Employment, Teachers College, Columbia University, Technical Paper No. 16, September.

Becker, Gary S. and H. Gregg Lewis, 1973, "On the Interaction between the Quantity and Quality of Children," Journal of Political Economy, 81:2, part 2 (March/April), S279-288.

Behrman, Jere R., Michael S. McPherson, and Morton Owen Schapiro, 1991, "What Governments Want: Understanding State and Local Higher Education Appropriations," Williamstown, MA: Williams College, mimeo.

Behrman, Jere R. and Paul Taubman, 1989, "Is Schooling 'Mostly in the Genes'? Nature-Nurture Decomposition with Data on Relatives", Journal of Political Economy 97:6 (December), 1425-1446.

Boskin, Michael J., 1974, "A Conditional Logit Model of Occupational Choice," Journal of Political Economy 82:2 (Part 1, March/April), 389-98.

Center for Educational Research and Evaluation, 1981, National Longitudinal Study: Base Year (1972) through Fourth Followup (1979). Data Users Manual. Research Triangle Park, NC.

Chamberlain, Gary and Zvi Griliches, 1977, "More on Brothers," in P. Taubman, ed. Papers in Kinometrics: Determinants of Socioeconomic Success Within and Between Families, Amsterdam: North-Holland Publishing Company.

Chronicle of Higher Education, January 24, 1990.

Grubb, Norton W., 1990, "The Causes of Enrollment in Postsecondary Education: Evidence from the National Longitudinal Study of the Class of 1972," Berkeley, CA: University of California, mimeo.

Hausman, Jerry, 1978, "Specification Tests in Econometrics," Econometrica, 46, 1251-1271.

Manski, Charles F. and David A. Wise, 1983, College Choice in America. Cambridge, MA: Harvard University Press.

Mincer, J.B., 1974, Schooling, Experience, and Earnings, New York: NBER.

Murphy, Kevin M. and Robert Topel, "Estimation and Inference in Two-Step Econometric Models," Journal of Business and Economic Statistics, 3 (1985): 370-379.

Steelman, Lala Carr and Brian Powell, 1989, "Acquiring Capital for College: The Constraints of Family Configuration," American Sociological Review 54, 844-55.

United States. President, 1992, Economic Report of the President. 1992. Washington, D.C.: U. S. Government Printing Office.

Willis, Robert J., 1973, "A New Approach to the Economic Theory of Fertility Behavior," Journal of Political Economy, 81:2 , part 2 (March/April), S14-S64.

Willis, Robert J. and Sherwin Rosen, 1979, "Education and Self-Selection," Journal of Political Economy 85:5, part 2 (October), S7-S36.

Zimmerman, David J., 1992, "Regression Towards Mediocrity in Economic Stature," American Economic Review, 82:3 (June), 409-429.

Appendix 1

Description of Data and Variables

We combine the NLSHS72 data with information on highest degree attained from the Postsecondary Education Transcript Study (PETS). The PETS survey culled the NLSHS72 first through fourth follow-ups for the names of post-secondary schools attended by the 14,700 respondents who reported attending postsecondary school. Transcripts were requested from all schools respondents reported attending. PETS allows us to calculate educational attainment as of 1984, 12 years after high school graduation. These data improve on the self-reported enrollments provided in NLSHS72. Grubb (1990) shows that individuals on average over-report their postsecondary attainment, particularly those not in academic courses in high school and with lower ability, high school grades and family socioeconomic status.[1] Thus, use of the self-reported data would bias downward the estimated effects of high school and family background on post-secondary school enrollment. Our measure of attainment is constructed by finding the highest degree earned or level of schooling attended as reported in the PETS survey. The attainment variable is underestimated for some observations; this might be due to transcripts being lost or destroyed, respondents not listing a school they attended, or respondents missing a follow-up survey. Of the 13,274 individuals for whom we have information on educational attainment, 8484 (63.9%) had valid data on the variables used in the analysis presented below. Of these 8484 observations, 4996 (58.8%) attended some form of postsecondary schooling (not including business or vocational schools).

Our family background and achievement variables are drawn directly from the NLSHS72

---

1 For the same sample he reports that by 1978-9 self-reported enrollments were 68.1% as compared with transcript-reported enrollments of 61.4%. For the bottom quarter by ability the respective percentages were 45.2 and 32.9%, for those with high school grades of D's 42.2 and 31.7%, for those with father's in the bottom quartile of socioeconomic indices 48.1 and 40.7%, for those in general high school programs 57.7 and 47.0%, and for those in vocational high school programs 43.9 and 31.7%.

data file. NLSHS72 also provides information on the distance to the closest two-year and the closest four-year college (which is part of the representation of prices broadly-defined).

With state identifiers, we can link individuals in the NLSHS72 data set with information from the Current Population Survey (CPS), and the Bureau of the Census. We use local and state expenditure on higher education in 1974, from HEGIS, for the state in which the individual attended high school as an indicator of local resources devoted to public higher education. Of course states vary substantially in the absolute magnitudes of such expenditures, in part because of different population sizes. Therefore we normalize by population in the 14-18 age cohort. We use this cohort to lessen the endogeneity that would be involved if we used, for example, the 18 22 age cohort and individuals migrated to states with better public higher education systems and established state residency where they went to school in order to pay in-state tuition rates (as apparently occurred substantially, for example, in California during the sample period). Even with such normalization there is substantial variation in appropriations across states.

We use CPS data for 1972-1976 to obtain two important measures of labor market conditions for individuals of the same sex, race and region. We use six Census regions (as listed in Table 1) with aggregation of contiguous regions in cases in which otherwise the cell sizes would be less than 25. The first variable is the opportunity cost of attending a postsecondary institution, which we represent by the median wage for individuals who have completed high school but not gone on to postsecondary school and who are 18-21 years of age. We use the median instead of the mean so that our representation will not be influenced by outliers in the distribution of wage rates. Our second labor market variable is the present discounted value[2] of the wage rate differential between those who have four years of college and those who terminated their schooling with high school, based on the assumption that the cross-sectional experience for 1972-76 captures expectations regarding future developments over time. We discount these wage

2We use a discount rate of 5%. The previous literatures on college and occupational choice include a range for discount rates from 3 to 15% (e.g., Boskin 1974, Willis and Rosen 1979). Our results are fairly insensitive to a range of discount rates between 5 and 10%.

differentials because the apparent life cycle paths in the cross-sectional data vary significantly for different schooling levels, as is well known (e.g., Mincer 1974).

In our achievement equation, we use variables that measure characteristics of the high school attended. These variables come from a survey of the high schools of the seniors in the NLSHS72 sample. For this reason, all students from one high school will have the same response to any one question. The variables include: average income of the high school community, and the number of weeks in the high school year.

PETS contains FICE codes for postsecondary schools, which allow us to link our data with HEGIS data to measure institutional quality. We use instructional expenditures per FTE student for the academic year 1975-76: the dollar amount of money spent on instruction and self-supported research by the four-year college an undergraduate attended.

Appendix 2

Description of Test Book

There are six tests, described below in sequential order of administration. These descriptions are taken from Center for Education Research and Evaluation (1981, pg. 17- 18).

a. Vocabulary: A brief test using synonym format. The items were selected to avoid academic or collegiate bias and to be of an appropriate level of difficulty for the twelfth grade population.

b. Picture Number: A test of associative memory consisting of a series of drawings of familiar objects each paired with a number. The student, after studying the picture number pairs, is asked to recall the number associated with each object.

c. Reading: A test based on short passages (100-200 words) with several related questions concerning a variety of reading skills (analysis, interpretation) but focusing on straightforward comprehension. In combination with the vocabulary test, it provides a means to derive a verbal score which can allow links to the normative data available for the SAT.

d. Letter Groups: A test of inductive reasoning requiring the student to draw general concepts from sets of data or to form and try out hypotheses in a nonverbal context. The items consist of five groups of letters among which four groups share a common characteristic while the fifth group is different. The student indicates which group differs from the others.

e. Mathematics: Quantitative comparisons in which the student indicates which of two quantities is greater, or asserts their equality or the lack of sufficient data to determine which quantity is greater. This type of item is relatively quickly answered and provides measurement of basic competence in mathematics.

f. Mosaic Comparisons: A test which measures perceptual speed and accuracy through items that require that small differences be detected between pairs of otherwise identical mosaics or tile-like

patterns. A deliberately speeded test, it has three separately timed sections consisting of increasingly more complex mosaic patterns.

Table 1

Descriptive Statistics

	ALL		WHITE		BLACK & HISPANIC	
	frequency	percent	frequency	percent	frequency	percent
N	8484		7281		1203	
MOM'S EDUCATION						
less than high school	2388	28.1%	1732	23.8%	656	54.5%
high school	3717	43.8%	3358	46.1%	359	29.8%
some college	1445	17.0%	1315	18.1%	130	10.8%
4 years college +	934	11.0%	876	12.0%	58	4.8%
DAD'S EDUCATION						
less than high school	2869	33.8%	2117	29.1%	752	62.5%
high school	2587	30.5%	2318	31.8%	269	22.4%
some college	1469	17.3%	1342	18.4%	127	10.6%
4 years college +	1559	18.4%	1504	20.7%	55	4.6%
FAMILY INCOME						
less than $5,999	1300	15.3%	739	10.1%	561	46.6%
$6,000-$8,999	1786	21.1%	1483	20.4%	303	25.2%
$9,000-$11,999	1914	22.6%	1742	23.9%	172	14.3%
$12,000-$14,999	1387	16.3%	1301	17.9%	86	7.1%
over $15,000	2097	24.7%	2016	27.7%	81	6.7%
SEX						
male	4428	52.2%	3895	53.5%	533	44.3%
female	4056	47.8%	3386	46.5%	670	55.7%
REGION						
North Central	240 1	28.3%	2293	31.5%	108	9.0%
South Central	1500	17.7%	1099	15.1%	401	33.3%
Atlantic	2766	32.6%	2328	32.0%	438	36.4%
Mountain	374	4.4%	307	4.2%	67	5.6%
Pacific	1053	12.4%	878	12.1%	175	14.5%
New England	390	4.6%	376	5.2%	14	1.2%
INCOME OF H.S. COMMUNITY						
less than $3,000	518	6.1%	353	4.8%	165	13.7%
$3,000-$4,999	3117	36.7%	2485	34.1%	632	52.5%
$5,000-$6,999	3424	40.4%	3080	42.3%	344	28.6%
$7,000-$8,999	1063	12.5%	1021	14.0%	42	3.5%
over $9,000	362	4.3%	342	4.7%	20	1.7%

	mean	std. dev	mean	std. dev	mean	std. dev.
NUMBER OF SIBLINGS	2.98	2.23	2.76	2.00	4.32	2.95
BIRTH ORDER	2.38	1.57	2.27	1.43	3.01	2.13
OPPORTUNITY WAGE OF 18-21 YR OLDS	8.28	1.06	8.39	1.02	7.60	1.07
OPPORTUNITY WAGE OF 16 YR OLDS	7.87	1.24	7.92	1.21	7.57	1.39
EARNINGS DIFFERENTIAL (1982 $)	111771	2664 1	114439	23077	95625	38464
DUNCAN SE1 SCORE FOR DAD'S OCC	42.46	24.79	45.07	24.48	26.65	20.40
MILES TO CLOSEST TWO-YEAR SCHOOL	19.38	26.29	19.00	25.36	21.64	31.27
MILES TO CLOSEST FOUR-YEAR SCHOOL	20.81	26.79	20.93	26.85	20.14	26.46
STATE AND LOCAL HIGHER EDUCATION EXPENDITURES PER CAPITA	568.65	174.67	564.97	173.47	590.92	180.29
ACHIEVEMENT SCORE (SUM OF SIX COMPONENTS)	306.73	43.08	313.68	39.55	264.67	39.47
WEEKS IN HIGH SCHOOL YEAR	36.83	1.66	36.85	1.68	36.72	1.49

Table 1, continued
Descriptive Statistics

	High School	Two-Year	Four-Year
ACTUAL SAMPLE PROPORTIONS			
White Female	0.430	0.158	0.412
White Male	0.368	0.171	0.461
Black Female	0.512	0.149	0.339
Black Male	0.522	0.135	0.343
Hispanic Female	0.475	0.219	0.306
Hispanic Male	0.435	0.271	0.294

Table 2
OLS Estimates of Achievement:(dependent variable = observed test score)
(standard errors)

Variable	Whites	Blacks & Hispanics
Mom's education:		
High School	8.9943**	7.9919**
	(1.1606)	(2.7244)
Some College	14.4770**	8.0535**
	(1.4881)	(3.9629)
Four years college or more	18.4254**	23.9348**
	(1.7754)	(6.1622)
(Less than high school omitted)		
Dad's education:		
High School	4.3183**	-1.4295
	(1.1786)	(2.9538)
Some College	8.5284**	-2.4287
	(1.4356)	(3.9507)
Four years college or more	12.3339**	-8.7422
	(1.6891)	(6.4791)
(Less than high school omitted)		
Family Income in 1972:		
$6000-$8999	5.1056**	6.6848**
	(1.6667)	(2.7533)
$9000-$11999	8.9578**	5.7948*
	(1.6600)	(3.4427)
$12000-$14999	8.6906**	6.1311
	(1.7781)	(4.6590)
>=$15000	10.2458**	12.2603**
	(1.7729)	(5.0003)
(Less than $6000 omitted)		
Female	7.8381**	3.0231
	(1.0627)	(3.0551)
Hispanic		17.5356**
		(2.8243)
Number of siblings	-.2465	-.1898
	(.2762)	(.5118)
Birth Order	-.7217*	.0103
	(.3883)	(.6955)
High School Opportunity Wage	-.0703	-1.7622
	(.4439)	(1.1546)
Dad's Duncan index	.1264**	.2128**
(socioeconomic index)	(.0226)	(.0625)
State & Local HE expenditures	-.0092**	-.0003
(per capita)	(.0025)	(.0071)
Avg. High School Family Income	2.8912**	2.5222*
(HSFINCOME)	(.4621)	(1.3903)
Number of weeks in the school	.2719	1.0662
year	(.2620)	(.8074)
Mom not in labor **force**	6.9436**	1.0659
before elementary school	(.9066)	(2.3161)
Constant	267.769**	214.116**
	(9.8108)	(27.5786)
Adj. R2	.143	.094
SER	36.61	37.55
Number of observations	7281	1203

Notes: ** **significantly different from zero at a** 5% level of significance
* significantly different from zero at a 10% level of significance

Table 3a
Multinomial Logit Estimates of the Probability of Post-secondary school attendance - White sample (asymptotic standard errors)

Variable	Probability of attending: [a] Two-year school	Four-year school
Mom's education:		
High School	.3187**	.1443
	(.1223)	(.1134)
Some College	.5510**	.5856**
	(.1763)	(.1592)
Four yrs. college or more	.6557**	.6226**
	(.2198)	(.1955)
(Less than high school omitted)		
Dad's education:		
High School	.2329**	.2630**
	(.1008)	(.0955)
Some College	.3227**	.5300**
	(.1405)	(.1289)
Four yrs. college or more	.3490*	.6878**
	(.1873)	(.1671)
(Less than high school omitted)		
Family Income in 1972:		
$6000-$8999	.0688	-.1165
	(.0795)	(.0709)
$9000-$11999	.0827	-.0585
	(.llll)	(.1028)
$12000-$14999	.1709	-.0134
	(.1201)	(.1118)
>=$15000	.2105	.2793**
	(.1303)	(.1177)
(Less than $6000 omitted)		
Female	.1803	-.8398**
	(.2120)	(.1722)
Number of siblings	.0035	-.0998**
	(.0222)	(.0214)
Birth Order	-.0769**	.0347
	(.0329)	(.0308)
Achievement (predicted)	.0062	.0382**
	(.0084)	(.0076)
Opportunity Wage	.1215	-.1382*
	(.lOOO)	(.0784)
College/High School Earnings Differential	.000004**	-.000001
	(.000001)	(.000001)
Unemployment rate for 18-21 year olds	-2.3509	-3.0911
	(3.2036)	(2.5518)
Dad's Duncan index (socioeconomic index)	.0012	.0037*
	(.0022)	(.0020)
Average state resident tuition for 4 yr. college	-.0003	.0006**
	(.0003)	(.0003)
State & Local HE expenditures (per capita)	.0027**	.OOOl
	(.0003)	(.0002)

Table 3a (continued)

Variable	Probability of attending: [a] Two-year school	Four-year school
Miles to closest 2 yr. school (DISTANCEl)	-.0104** (.0021)	.0015 (.0012)
Miles to closest 4 yr. school (DISTANCE2)	.0042* (.0015)	-.0009 (.0012)
Constant	-6.3771** (2.5103)	-10.6895** (2.2474)
Log L -6488.7		
N=7281		
59.2% predicted correctly		

Notes:
a. Coefficients for High School set to zero
** significantly different from zero at a 5% level of significance
* significantly different from zero at a 10% level of significance

Table 3b
Multinomial Logit Estimates of the Probability of Post-secondary school attendance - Nonwhite sample (Blacks & Hispanics)
(asymptotic standard errors)

	Probability of attending:[a]	
Variable	**Two-year school**	**Four-year school**
Mom's education:		
High School	-.3868	-.2117
	(.3785)	(.3729)
Some College	-.0066	.5957*
	(.5313)	(.4535)
Four yrs. college or more	-.5713	-.6188
	(1.002)	(.9678)
(Less than high school omitted)		
Dad's education:		
High School	.1655	.3161
	(.2576)	(.2539)
Some College	.5959*	.6443*
	(.3438)	(.3497)
Four yrs. college or more	.1980	.9389*
	(.7390)	(.6264)
(Less than high school omitted)		
Family Income in 1972:		
$6000-$8999	.2509	-.3974
	(.2202)	(.2170)
$9000-$11999	-.2944	-.4049
	(.3145)	(.3137)
$12000-$14999	-.3250	-.1975
	(.4372)	(.4059)
>=$15000	-.6078	-.5506
	(.5873)	(.5459)
(Less than $6000 omitted)		
Female	-.6412**	-.3437
	(.3147)	(.2991)
Hispanic	-.0393	-.9375
	(.7752)	(.7173)
Number of siblings	-.0616	-.0964**
	(.0445)	(.0455)
Birth Order	-.0672	.1501**
	(.0600)	(.0607)
Achievement (predicted)	.0341	.0590*
	(.0342)	(.0331)
Opportunity Wage	-.5176**	.0988
	(.1452)	(.0907)
College/High School Earnings Differential	.000002	-.000001
	(.000003)	(.000002)
Unemployment rate for 18-21 yrs.	5.8056	-1.0733
	(2.4602)	(2.0907)
Dad's Duncan index (socioeconomic index)	-.0082	-.0036
	(.0095)	(.0092)
Average state resident tuition for 4 yr. college	.0026**	-.0003
	(.0008)	(.0003)
State & Local HE expenditures (per capita)	.0043**	-.0003
	(.0008)	(.0006)

Table 3b (continued)

Variable	Probability of attending: [a] Two-year school	Four-year school
Miles to closest 2 yr. school (DISTANCEl)	-.0129** (.0048)	.0013 (.0023)
Miles to closest 4 yr. school (DISTANCE2)	.000009 (.0036)	-.0087** (.0031)
Constant	-10.4572** (8.2584)	-15.2794* (8.0263)
Log L -1106.4		
N=1203		
56.5% predicted correctly		

Notes:
a. Coefficients for High School set to zero
** significantly different from zero at a 5% level of significance
* significantly different from zero at a 10% level of significance

Table 4
Predicted Probabilities of Post-secondary school attendance:
Estimates based on parameters from multinomial logit

Case	Probability of attending: Two-year school ( )[a]		Four-year school ( )[a]	
1. Base case A: (uses white sample means) Mom's education=High School Dad's education=High School Family Income =$9-12K Number of siblings=2.7 Birth Order=2.27 Opportunity Wage=$8.39 Earnings Differential=$199,810 Unemployment rate=.097 Dad's Duncan index=45 Avg. state resident tuition=$461.09 State & Local HE expenditures=$564.97 Miles to closest 2 yr. school=19 (DISTANCEl) Miles to closest 4 yr. school=20.9 (DISTANCE2)				
White Male (achieve=312.75)	.142	(.140)	.476	(.484)
White Female (320.59)	.214	(.224)	.331	(.278)
Black Male (268.94)	.059		.488	
Black Female (271.96)	.038		.459	
Hispanic Male (286.47)	.095		.491	
Hispanic Female (289.49)	.063		.470	
2. Base case B (uses nonwhite sample means) Mom's education=Less than High School Dad's education=Less than High School Family Income =$6-9K Number of siblings=4.3 Birth Order=3.0 Opportunity Wage=$7.59 Earnings Differential=$180,180 Unemployment rate=.130 Dad's Duncan index=26.6 Avg. state resident tuition=$356.58 State & Local HE expenditures=$590.92 Miles to closest 2 yr. school=21.6 Miles to closest 4 yr. school=20.1				
White Male (283.59)	.092		.132	
White Female (291.43)	.120		.079	
Black Male (257.21)	.227		.241	
Black Female (260.23)	.153		.235	
Hispanic Male (274.74)	.333		.223	
Hispanic Female (277.76)	.235		.228	

Table 4, continued

Case	Probability of attending: Two-year school		Four-year school	
3. a. Base case A except:				
Mom's education=some college				
Dad's education=some college				
White Male (322.44)	.104	(.125)	.704	(.634)
White Female (330.28)	.179	(.225)	.385	(.412)
b. Base case A except:				
Mom's education=high school				
Dad's education=high school				
Family Income=$12-15K				
White Male (312.48)	.151	(.148)	.479	(.490)
White Female (320.32)	.238	(.237)	.273	(.282)
c. Base case A except:				
Mom's education=some college				
Dad's education=some college				
Family Income=$12-15K				
White Male (322.18)	.llO	(.131)	.704	(.637)
White Female (330.02)	.180	(.236)	.559	(.414)
d. Base case A except:				
Mom's education=college				
Dad's education=college				
Family Income>=$lSK				
White Male (331.49)	.071	(.109)	.835	(.728)
White Female (339.32)	.133	(.215)	.726	(.517)
4. a. Base case B except:				
Mom's education=high school				
Dad's education=high school				
Black Male (263.77)	.198		.339	
Black Female (266.79)	.129		.298	
Hispanic Male (281.31)	.293		.316	
Hispanic Female (284.33)	.206		.323	
b. Base case B except:				
Mom's education=less than high school				
Dad's education=less than high school				
Family Income=$9-12K				
Black Male (256.32)	.144		.256	
Black Female (259.34)	.094		.241	
Hispanic Male (273.85)	.222		.249	
Hispanic Female (276.87)	.149		.243	

Table 4 (continued)

Case	Probability of attending: Two-year school	Four-year school
4. c. Base case B except:		
Mom's education=high school		
Dad's education=high school		
Family Income=$9-12		
Black Male (262.88)	.125	.358
Black Female (265.90)	.082	.340
Hispanic Male (280.42)	.193	.349
Hispanic Female (283.44)	.130	.342
d. Base case B except:		
Mom's education=some college		
Dad's education=some college		
Family Income>=$12-15K		
Black Male (268.96)	.189	.584
Black Female (271.98)	.132	.595
Hispanic Male (286.50)	.275	.536
Hispanic Female (289.52)	.200	.565

Notes: a. probability in parentheses is the direct effect on post-secondary attainment of changing a characteristic. That probability calculation uses the average achievement level and does not reflect the change in predicted achievement of changing the characteristic.

Table 5
"Quality" of post-secondary school for individuals attending a 4 year school: Quality" measured as (log) instructional expenditures/student, 1975-76 academic year. (standard errors)

	Whites		Nonwhites	
Variable	(1)	(2)	(3)	(4)
Mom's education:				
High School	.0436	.1080**	-.0593	-.1612*
	(.0533)	(.0516)	(.1565)	(.0827)
Some College	.2615**	.1930**	.3832**	-.1386
	(.0765)	(.0765)	(.1858)	(.2431)
Four yrs. college or more	.2543**	.1777**	-.2366	-.2768
	(.0921)	(.0811)	(.4184)	(-2273)
(Less than high school omitted)				
Dad's education:				
High School	.1271**	.0614	.1485	.0823
	(.0465)	(.0503)	(.1069)	(.1039)
Some College	.2711**	.0992	.2604*	.1429
	(.0619)	(.0708)	(.1440)	(.1604)
Four yrs. college or more	.3598**	.1458*	.5218**	.1564
	(.0792)	(.0824)	(.2619)	(.2894)
(Less than high school omitted)				
Family Income in 1972:				
$6000-$8999	-.0751	.0258	-.2748	-.0077
	(.0671)	(.0647)	(.1437)	(.1681)
$9000-$11999	-.0461	.0187	-.1997	.0263
	(.0732)	(.0658)	(.1561)	(.1308)
$12000-$14999	-.0321	.0621	-.0615	.0199
	(.0762)	(.0666)	(.1894)	(.0934)
>=$15000	.1351*	.1491**	-.2176	-.1373
	(.0797)	(.0719)	(.2623)	(.1703)
(Less than $6000 omitted)				
Female	-.5694**	-.1457	-.1040	.0658
	(.0899)	(.1043)	(.1328)	(.0856)
Number of siblings	-.0600**	-.0206*	-.0471**	-.0035
	(.0102)	(.0120)	(.0191)	(.0296)
Hispanic			-.5589*	-.1280
			(.3366)	(.3402)
Birth Order	.0346**	.0065	.0787**	-.0048
	(.0148)	(.0128)	(.0252)	(.0484)
Achievement (predicted)	.0218**	.0092**	.0285**	.0113
	(.0037)	(.0041)	(.0148)	(.0165)
Opportunity Wage	-.1191**	-.0096	.1050**	.0789*
	(.0427)	(.0355)	(.0484)	(.0609)
College/High School Earnings Differential	-.000002**	-.0000003	-.000001	.0000005
	(.0000006)	(.0000005)	(.000001)	(.000001)
Unemployment rate for 18-21 year olds	-2.2499	-1.1517	-1.3805	-.1367
	(1.3872)	(1.0671)	(1.1509)	(.9816)
Dad's Duncan index (socioeconomic index)	.0019**	.0019**	-.0005	-.0030
	(.0009)	(.0008)	(.0039)	(.0019)
Average state resident tuition for college	.0005**	.0006**	-.0003*	-.0001
	(.0001)	(.0001)	(.0002)	(.0002)

Table 5 (continued)

Variable	Whites (1)	Whites (2)	Nonwhites (3)	Nonwhites (4)
State & Local HE expenditures	-.0005**	.0002**	-.0007**	.0002
	(.0001)	(.0001)	(.0002)	(.0004)
Miles to closest 2 yr. school	.0022**	.0005	.0019	.0005
	(.0007)	(.0006)	(.0013)	(.0012)
Miles to closest 4 yr. school	-.0012*	-.0005	-.0046**	-.0002
	(.0006)	(.0005)	(.0017)	(.0029)
Constant	-5.4474**	-3.4419**	-7.4672**	-3.7989**
	(1.0897)	(1.2666)	(3.6042)	(4.9402)
Lambda		.6085**		-.0922
		(.2426)		(.8370)
Log L	-4264.3		-697.62	
Adj. R^2		.066		.097
SER		.448		.365
N	7281	3094	1203	382
% predicted correctly	69.83		71.1%	

Notes: Columns 1 and 3 report probit estimates of the probability of attending a 4 year school; columns 2 and 4 report estimates of instructional expenditures/student corrected for sample selection (of attending a 4 year school).

** significantly different from zero at a 5% level of significance

* significantly different from zero at a 10% level of significance

Table 6
Multinomial Logit Estimates of the Probability of Post-secondary school attendance - White sample (asymptotic standard errors) - achievement treated as exogenous

White Sample Variable	Probability of Attending Two-year School	Probability of Attending Four-year School
Mom's Education		
High School	.3089** (.0937)	.3292** (.0827)
Some College	.5447** (.1266)	.9124** (.1074)
Four yrs. college or more	.6582** (.1633)	.9943** (.1356)
(Less than high school omitted)		
Dad's Education		
High School	.2384** (.0931)	.3891** (.0829)
Some College	.3322** (.1178)	.7678** (.1010)
Four yrs. college or more	.3705** (.1477)	1.0228** (.1230)
(Less than high school omitted)		
Family Income in 1972:		
$6000-$8999	.0259 (.1367)	-.0206 (.1239)
$9000-$11999	.0330 (.1363)	.1129 (.1221)
$12000-$14999	.1230 (.1456)	.1809 (.1299)
>=$15000	.1891 (.1478)	.5555** (.1299)
(Less than $6000 omitted)		
Female	.1200 (.2014)	-.7704** (.1679)
Number of Siblings	-.0035 (.0224)	-.1179** (.0202)
Birth Order	-.0731** (.0325)	.2998 (.0285)

White Sample Variable	Probability of Attending Two-year School	Probability of Attending Four-year School
Table 6, cont.		
Achievement (observed)	.0109** (.0010)	.0294** (.0010)
Opportunity Wage	.1219 (.1014)	-.1377 (.0848)
College/High School Earnings Differential	.000004** (.000001)	-.0000005 (.000001)
Unemployment Rate for 18-21 year olds	-1.4986 (3.2156)	-.2196 (2.7520)
Dad's Duncan index (socioeconomic index)	.0012 (.0019)	.0066** (.0016)
Average state resident tuition for 4 yr college	-.0005 (.0003)	.0002 (.0003)
State & Local HE expenditures (per capita)	.0027** (.0003)	-.0003 (.0002)
Miles to closest 2 yr. school (DISTANCE1)	-.0101** (.0021)	.0026* (.0014)
Miles to closest 4 yr. school (DISTANCE2)	-.0040** (.0015)	-.0024* (.0013)
Constant	-7.7057** (.8656)	-8.5877** (.7276)
Log L -5936.7 N=7281 64.6% predicted correctly		

Notes: Coefficients for High School set to zero
** significantly different from zero at a 5% level of significance
* significantly different from zero at a 10% level of significance

Table 7
Multinomial Logit Estimates of the Probability of Post-secondary school attendance - Nonwhite sample (Blacks & Hispanics) (asymptotic standard errors) - achievement treated as exogenous

Nonwhite Sample Variable	Probability of Attending Two-year School	Probability of Attending Four-year School
Mom's Education		
High School	-.1640 (.2174)	.1259 (.1798)
Some College	.2393 (.3297)	1.0259** (.2595)
Four yrs. college or more	.0897 (.5660)	.4018 (.4038)
(Less than high school omitted)		
Dad's Education		
High School	.1356 (.2361)	.2876 (.1918)
Some College	.5212* (.3072)	.5777** (.2597)
Four yrs. college or more	.0777 (.6309)	.7173* (.4287)
(Less than high school omitted)		
Family Income in 1972:		
$6000-$8999	.4111** (.2062)	-.1418 (.1870)
$9000-$11999	-.1253 (.2696)	-.1577 (.2258)
$12000-$14999	-.1152 (.3979)	.1017 (.3017)
>=$15000	-.2184 (.4420)	-.0093 (.3255)
(Less than $6000 omitted)		
Hispanic	.4010 (.4216)	-.2879 (.3639)
Female	-.4677** (.2125)	-.0972 (.1752)
Number of Siblings	-.0688* (.0410)	-.1074** (.0353)

Nonwhite Sample Variable	Probability of Attending Two-year School	Probability of Attending Four-year School
Table 7, cont.		
Birth Order	.0693 (.0555)	.1598** (.0469)
Achievement (observed)	.0081** (.0024)	.0210** (.0021)
Opportunity Wage	-.4988** (.1429)	.1326 (.0865)
College/High School Earnings Differential	.000002 (.000003)	.000003 (.000002)
Unemployment Rate for 18-21 year olds	5.6395** (2.4767)	-1.4514 (2.1734)
Dad's Duncan index (socioeconomic index)	-.0018 (.0052)	.0060 (.0040)
Average state resident tuition for 4	.0027**	-.00009
State & Local HE expenditures (per capita)	.0044** (.0008)	-.0002 (.0005)
Miles to closest 2 yr. school (DISTANCE 1)	-.0132** (.0048)	.0011 (.0025)
Miles to closest 4 yr. school (DISTANCE2)	.0003 (.0036)	-.0084** (.0032)
Constant	-4.1530** (1.2987)	-6.2325** (1.0926)
Log L -1050.3 N= 1203 58.7% predicted correctly		

Notes: Coefficients for High School set to zero
** significantly different from zero at a 5% level of significance
* significantly different from zero at a 10% level of significance