Using Geographic Variation in College Proximity to Estimate the Return to Schooling

Article · November 1993

Source: RePEc

1 author:

David Card
University of California, Berkeley

NBER WORKING PAPER SERIES

USING GEOGRAPHIC VARIATION IN COLLEGE PROXIMITY TO ESTIMATE THE RETURN TO SCHOOLING

David Card

Working Paper No. 4483

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October, 1993

I am grateful to Charles Thomas and Norman Thurston for outstanding research assistance, and to Michael Boozer, Alan Krueger, and Cecilia Rouse for comments. This research was funded by the Industrial Relations Section of Princeton University. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #4483
October 1993

USING GEOGRAPHIC VARIATION IN COLLEGE PROXIMITY TO ESTIMATE THE RETURN TO SCHOOLING

ABSTRACT

A convincing analysis of the causal link between schooling and earnings requires an exogenous source of variation in education outcomes. This paper explores the use of college proximity as an exogenous determinant of schooling. Analysis of the NLS Young Men Cohort reveals that men who grew up in local labor markets with a nearby college have significantly higher education and earnings than other men. The education and earnings gains are concentrated among men with poorly-educated parents -- men who would otherwise stop schooling at relatively low levels. When college proximity is taken as an exogenous determinant of schooling the implied instrumental variables estimates of the return to schooling are 25-60% higher than conventional ordinary least squares estimates.

Since the effect of a nearby college on schooling attainment varies by family background it is possible to test whether college proximity is a legitimately exogenous determinant of schooling. The results affirm that marginal returns to education among children of less-educated parents are as high and perhaps much higher than the rates of return estimated by conventional methods.

David Card
Department of Economics
Princeton University
Princeton, N.J. 08544
and NBER

One of the most important "facts" about the labor market is that better-educated workers earn higher wages. Hundreds of studies in virtually every country show earnings gains of 5-15 percent (or more) per additional year of schooling.[1] Despite this evidence, most analysts are reluctant to interpret the earnings gap between more and less educated workers as a reliable estimate of the economic return to schooling. Education levels are not randomly assigned across the population; rather, individuals make their own schooling choices. Depending on how these choices are made, measured earnings differences between workers with different levels of schooling may over-state or under-state the "true" return to education.[2]

A convincing analysis of the causal link between education and earnings requires an exogenous source of variation in education choices. In this paper I argue that geographic differences in the accessibility of college are a potential source of such exogenous variation.[3] Using data from the Young Men Cohort of the

---

[1]Studies of the United States are reviewed in Rosen (1977), and Willis (1986). A survey of international studies is presented in Psacharopoulos (1985).

[2]See Griliches (1977) for an overview of the issues.

[3]A similar idea is used by Kane and Rouse (1993) to control for the endogeneity of choice between a four-year college and a two-year college.
Mallar (1979) used proximity to a training site to estimate the effect of the Job Corps program.

National Longitudinal Survey I find that men who were raised in local labor markets with a nearby 4-year college have significantly higher levels of education and earnings. This differential persists even after controlling for regional and family background factors (including parental education and family structure). The effects of a nearby college are largest for men with the lowest predicted levels of schooling attainment, suggesting that the presence of a local college lowers the costs and/or raises the perceived benefits of education among children with relatively poor family backgrounds.

When college proximity is taken as an exogenous determinant of schooling the implied instrumental variables estimates of the return to education are 25-60 percent higher than the corresponding ordinary least squares estimates. Contrary to widespread belief (e.g. Ehrenberg and Smith (1991, pp. 320-322)) but consistent with a growing number of studies of endogenous school choice, these findings suggest that the cross-sectional earnings gap between more- and less-educated workers may under-state the economic return to schooling for some groups of workers.[4]

---

[4]See e.g. Angrist and Krueger (1991a), Ashenfelter and Krueger (1992), Kane and Rouse (1993) and Butcher and Case (1993). All four of these studies report instrumental variables estimates of the return to schooling that exceed the conventional ordinary least-squares estimate in the same data set.

Since the effect of a nearby college on schooling attainment varies with family background it is possible to test whether college proximity is a legitimately exogenous determinant of schooling -- i.e., whether growing up near a college has a direct effect on earnings or only an indirect effect through the education decision. Specifically, one can include college proximity in the earnings equation and use the interaction of college proximity with a indicator for low parental education as an instrumental variable for education. This identification strategy relies on the extra boost to education and earnings among children with poor family backgrounds. The resulting estimates are still substantially higher than the ordinary least squares estimates, and provide no evidence against the hypothesis that college proximity is an exogenous determinant of schooling.

Preliminary Analysis of Earnings and Schooling in the NLS Young Men Cohort

The data in this paper are drawn from the National Longitudinal Survey of Young Men (NLSYM). The NLSYM began in 1966 with 5525 men age 14-24 and continued with follow-up surveys through 1981. Some descriptive statistics for

---

Interestingly, Griliches (1977) concluded that ordinary least squares estimates of the return to education were probably not downward-biased, once measurement error in schooling was taken into account.

the original sample and two subsamples are presented in Table 1. Like other longitudinal surveys initiated in the mid-1960s, the NLSYM was not a random sample of the U.S. population: rather, men from neighborhoods with a high concentration of non-white residents were over-sampled.[5] As shown in column (1), the NLSYM sample contains a relatively high fraction of men from the Southern region (41% versus approximately 32% for a nationally representative sample) and a high fraction of blacks (28% versus approximately 10% for a nationally representative sample).

In the baseline interview individuals were asked the composition of their family when they were age 14: 77 percent lived with both their father and mother; 12 percent lived with only their mother; the remainder lived with other relatives or at least one step-parent (row 5). Individuals were also asked their father's and mother's education, although a relatively large fraction of the sample report missing values for these variables (22% are missing father's education, 11% are missing mother's education). For observations with missing data I have assigned the overall mean of father's or mother's education. The statistical models reported below include dummies indicating whether either parent's education level is imputed.

---

[5]See Hall and Turner (1970).

The 1966 interview also included a 28 item test of "Knowledge of the World of Work" (see row 7).[6] The overall score on this test is correlated with completed education and wage rates in later waves of the survey, and the test has been used as a measure of "ability" in several previous studies of education and earnings (e.g. Griliches (1976, 1977)).

Finally, the NLSYM data set contains a number of characteristics of the respondent's local labor market in 1966.[7] Among these is an indicator for the presence of an accredited 4-year college in the local labor market (row 8).[8] About 70 percent of individuals lived in a labor market area with a nearby college. The college proximity rate varies by region (lower in the South and Mountain regions), by urban versus rural location (higher for individuals living in a Standard Metropolitan Statistical Area), and is correlated with race and parental education (see below).

---

[6]The test items were questions on the job activities of 10 specific occupations, the education requirements for these 10 occupations, and the relative earnings of 8 different pairs of occupations.

[7]These are based on the county of residence in 1966.

[8]An indicator for the presence of a 2-year college is also included in the NLSYM, but this variable turns out to be only weakly correlated with education or earnings. See below.

Like other longitudinal surveys the NLSYM is affected by sample attrition. Approximately 20 percent of the sample dropped out in the first 3 years of the survey, and only 65 percent of the original sample were interviewed in the final (1981) wave. In selecting a cross-section from the NLSYM there is evidently a tradeoff between response rates and the age of the respondents. Earlier waves have higher response rates but relatively young sample members whereas later waves have lower response rates but older sample members. I compromise by using labor market information from the 1976 interview. In 1976 the youngest respondents are 24 years of age and the available sample is still relatively large (3694 observations or 71 percent of the original sample). An important advantage of the 1976 data is that all respondents were directly asked their educational attainment as of the 1976 interview.

Column (2) of Table 1 reports the characteristics of individuals who were interviewed in 1976 and who provided valid education responses. These men have the same age and regional distributions as the original NLSYM sample but are slightly less likely to be black. The mean level of reported education in 1976 is 13.2 years. One-third of the sample report exactly 12 years of schooling, 23% report some college, and 27% report 16 or more years of education.

Eighty-three percent of men interviewed in 1976 report a valid wage observation. The characteristics of this working subsample are reported in column (3) of Table 1. Comparisons with the mean characteristics in columns (1)-(3) show few differences between the original sample, the subsample of 1976 interviewees, and the subsample with 1976 wages.

To begin an investigation of the returns to schooling in the NLSYM Table 2 presents a variety of conventional earnings functions estimated by ordinary least squares (OLS). All models include a linear education term, a quadratic function of potential experience (age-education-6), a race indicator, and dummies for residence in the South and in a metropolitan area (SMSA) in 1976. The specification in column (2) adds 8 indicators for region of residence in 1966 and another for residence in an SMSA in 1966. The models in columns (3)-(5) add an increasing set of family background characteristics: measures of father's and mother's education (column 3); interactions of father's and mother's education (column 4); and indicators for family structure at age 14 (column (5)). As shown by the test statistics in row 13, the full set of family background variables are never jointly significant, although the family structure indicators are marginally significant by themselves. The estimated education coefficient (in row 1) is remarkably stable across specifications and implies a

7.3-7.5% earnings advantage for each additional year of education, controlling for experience and other factors.[9]

Despite their stability across specifications the estimated education coefficients in Table 2 may give a biased estimate of the true economic return per year of education. To facilitate discussion of the econometric issues involved, consider a simple two-equation system describing schooling (S_i) and log wages (y_i) for individual i (in 1976):

$$(1)\quad S_i = X_i\gamma + v_i,$$

$$(2)\quad y_i = X_i\alpha + S_i\beta + u_i\ .$$

Here X_i is a vector of observed attributes (with $E(X_iu_i) = E(X_iv_i) = 0$) and β has the interpretation of the "true" return to education.[10] A conventional earnings equation estimated by OLS gives a consistent estimate of β if and only if u_i and v_i are uncorrelated (i.e. if S_i is econometrically exogenous in (2)).

---

[9]Note that the estimated coefficient of a linear education variable is only strictly interpretable as a "rate of return" to schooling under very rigid conditions (see Mincer (1974)). I use the terminology "rate of return to schooling" to refer to the education coefficient in conventional human capital model.

[10]If the return to education varies across individuals then the coefficient β in equation (2) should be interpreted as the average return to education. Specifically, suppose $y_i = X_i\alpha + S_i\beta_i + \epsilon_i$, where β_i is the marginal return to education for i. Then equation (2) holds with $\beta = E(\beta_i)$ and $u_i = \epsilon_i + S_i(\beta_i - \beta)$.

There are a variety of reasons why schooling may be correlated with the unobserved component of earnings. One that has received considerable attention in the literature is "ability bias" (see e.g. Griliches (1977)). Suppose that some individuals have an unobserved characteristic ("ability") that enables them to earn higher wages at any level of education. If these individuals acquire higher-than-average schooling then the OLS estimate of β will be upward-biased. The fact that individuals with higher test scores (on IQ or achievement tests) tend to have higher earnings and more schooling is often interpreted as evidence of ability bias.

Another important source of correlation between u_i and v_i is measurement error in schooling. Measurement error induces a negative correlation between the error components of earnings and observed schooling, leading to a downward bias in OLS estimates of β (see Griliches (1977)).[11] A similar negative bias arises if the true return to schooling varies across the population and if individuals with lower levels of schooling have higher returns to schooling. Such a negative correlation is implied by a model of school choice in which individuals with different

---

[11]Estimates in the literature (cited by Griliches) suggest that 10% of the variance in measured education is due to measurement error. In this case the OLS estimate of the return to education is downward biased by 10-15 percent, depending on what other covariates are included in the model.

discount rates invest in schooling until the marginal return to schooling equals the discount rate (see Card (1993) and Lang (1993)).

A consistent estimate of the true return to education can be obtained if there is a component of the vector X_i that affects schooling but not earnings.[12] If schooling were randomly assigned, for example, then the realization of the randomizing process could be used as to estimate equation (2) by instrumental variables (IV).[13] In the absence of "pure" random assignment, however, one needs to identify a causal determinant of schooling that can be legitimately excluded from the earnings equation. The presence of a nearby college may be such a variable. Students who grow up in an area without a college face a higher cost of college education, since the option of living at home is precluded.[14] One would expect this higher cost to reduce

---

[12]If the true rate of return to education varies across the population then one can obtain a consistent estimate of the average return to education for some subset of the population. See Angrist and Imbens (1993).

[13]Something like this idea is used by Angrist and Krueger (1991b), who use draft-lottery status as an instrument for schooling of men who could have served in the Vietnam war.

[14]Tabulations of the October 1973 Current Population Survey show that in the early 1970s 34% of college students age 18-24 lived with their parents while attending school. The fraction is higher (39%) for black students.

investments in higher education, at least among children from relatively low-income families.[15]

To check this basic insight I fit a linear model to years of completed schooling (in 1976) for the subset of men who grew up in local labor markets without an accredited 4-year college. The determinants of schooling include region and urban/rural indicators (measured as of 1966), age and race dummies, and family background factors (family structure and parental education).[16] I then divided the overall sample into quartiles of predicted education in the absence of a nearby college and calculated the mean levels of education by quartile of predicted education for men who grew up in areas with and without a local college. Figure 1 plots the mean levels of education. In every quartile the mean level of education is higher for those who grew up near a college. For men in the three highest predicted quartiles of education the effect of college proximity is modest (0.2 to 0.4 years). For men in the lowest quartile, however, the difference in mean education is 1.1 years. As expected, the presence of a nearby college has its strongest effect on men with lowest propensities to continue their education (e.g. men from

---

[15]See Anderson, Bowman, and Tinto (1972) for a review of the sociological literature on the effects of college accessibility on attendance probabilities.

[16]The R-squared of the regression is 0.30.

single-headed families with low parental education in rural Southern areas).

Instrumental Variables Estimates of the Return to Education

Table 3 presents a series of reduced form education and earnings equations and the corresponding structural estimates of the return to education, using college proximity as an instrumental variable for completed education. Columns (1) and (2) show the coefficients of an indicator for college proximity in models for years of schooling. Columns (3) and (4) show the coefficients of the college proximity variable in reduced form wage equations (i.e. models that exclude education). Finally, columns (5) and (6) report the IV estimates of the return to education: these are simply the ratios of the corresponding reduced form coefficients in the earnings and schooling equations. The models in columns (1), (3) and (5) exclude parental education and family structure variables while the models in columns (2), (4), and (6) include these variables.

Two alternative specifications are reported in the upper and lower panels of the table. The models in the upper panel (Panel A) include the conventional measures of experience and experience-squared constructed from observed age and education. If schooling is measured with error, however, then experience is also mismeasured -- suggesting possible biases in the reduced form models in Panel A. By the same token, if education is truly endogenous in the earnings equation, then so is experience, since experience is mechanically related to education. Therefore, in the lower panel (Panel B) I have estimated models that instrument experience and experience-squared with age and age-squared.

Regardless of the inclusion or exclusion of family background variables, and irrespective of the treatment of experience, the conclusions from Table 3 are similar. Growing up near a college has a strong positive effect on both education (0.32 to 0.38 years of schooling) and earnings (4.2 to 4.8 percent). The use of college proximity as an exogenous determinant of schooling yields IV estimates of the return to education in the range of 0.12 to 0.14. These estimates are 50-60 percent higher than the corresponding OLS estimates -- about the same relative ratio as reported by Butcher and Case (1993), Kane and Rouse (1993), and Angrist and Krueger (1993). Nevertheless, the standard errors of the IV estimates are relatively large, and one cannot reject the hypothesis that differences between the IV and OLS estimates are due to sampling error.[17]

[17]Under the null hypothesis that the OLS estimates are consistent the variance of the difference between the IV and OLS estimates of the return to education is the difference in their variances, which is approximately equal to the variance of the IV estimate.

Table 4 presents a series of alternative specifications designed to probe the robustness of the estimates in Table 3. The top row of the table contains OLS and IV estimates of the return to education for the "basic specifications" in Tables 2 and 3 (OLS from column (5) of Table 2; IV from the lower panel of column (6) in Table 3). Row 2 presents estimates from the same specifications, using as a dependent variable the logarithm of 1978 wages for the subset of men with wages in that year.[18] The OLS estimate of the return to education is slightly lower in 1978 than 1976: otherwise, the estimated coefficients and overall fit of the wage equation are similar in the two years. As in the 1976 data, the use of college proximity as an instrument raises the estimated return to education by over 50%.

Row 3 presents OLS and IV estimates of the return to education when a direct measure of "ability" -- the "Knowledge of the World of Work" (KWW) score -- is included in the model. In the OLS model the KWW score is a significant determinant of earnings (t-statistic = 6.9): a 1-standard deviation increase in KWW is associated with a 6.6 percent increase in earnings. The addition of the KWW score leads to a 25% attenuation in the return to education relative to the basic OLS estimate. Since education and the KWW test score are highly correlated,

---

[18]Education, experience, and the current location variables are all defined as of the 1978 survey.

however, some of this attenuation is potentially attributable to the presence of measurement errors in education.[19] When college proximity is used as an instrument for education (row 3 column 2) the estimated return to education rises and the estimated coefficient of the KWW test falls to a small and statistically insignificant value.

A potential criticism of this specification is that the KWW score is treated as an error-free measure of "ability". To address this criticism, the IV specification in row 4 treats both education and the KWW score as "endogenous" (or measured with error) and uses a measure of IQ (taken from school records for a subset of NLSYM respondents) to instrument the KWW score.[20] This lowers the IV estimate of the return to education slightly but raises the standard errors of the education and KWW coefficients to the point where neither is statistically different from 0.

The IV estimates presented in rows 5 and 6 of Table 4 use two alternative measures of college proximity as instruments for

---

[19]Assuming that measurement errors account for 10 percent of the cross-sectional variance in observed schooling (Siegel and Hodge (1968)), and that the true effect of KWW on earnings is 0, the expected attenuation of the schooling coefficient when KWW is added to the model is about 5%.

[20]Note that one could include IQ in the earnings equation and use the KWW score as an instrument. This has no effect on the conclusions from Table 4.

schooling. In row 5 college proximity is defined as living in a local labor market with a public 4-year college.[21] Proximity to a public college has a slightly smaller reduced form effect on education (0.31 years versus 0.32 for proximity to any college) and a slightly larger reduced form effect on earnings (6.2% versus 4.2%). Thus the implied IV estimate of the return to college is higher than the IV estimate using proximity to any college, although the standard error is again relatively large.

The IV estimation in row 6 combines 2 college proximity variables: one for any accredited 4-year college, another for any accredited 2-year college. In the reduced form equations the presence of a nearby 2-year college has small positive effects on schooling and earnings (whether or not an indicator is included for proximity to a 4-year college). Using both indicators as instruments leads to an estimated rate of return to education of 0.12, and a very slight improvement in the standard error of the estimate relative to the baseline estimate in row 1.

One difficulty with these college proximity measures is that they pertain to the place of residence in 1966 rather than the place of residence at age 18 or 19, when the college enrollment decision is typically made. By the time of the 1966 interview

---

[21] Among the men who grew up in local labor markets with accredited 4-year colleges, 73% were in labor markets with a public 4-year college.

some of the older NLSYM respondents could have already moved to be closer to a college, giving rise to a reverse causation between college proximity and schooling attainment.[22] A simple check is to exclude the oldest respondents in the sample (e.g. those over age 19 in 1966). An important caveat to this exclusion is that the narrowing of the age range of the sample makes it more difficult to separately identify the effects of education and experience. Row 7 presents OLS and IV estimates based on the subsample of men age 14-19 in 1966. The OLS estimate of the return to education for the subsample is similar to the baseline estimate. The IV estimate is above the corresponding OLS estimate, although at the low end of the range of IV estimates (24% above the OLS estimate).

The results of these specification checks confirm the two main conclusions from Table 3. First, IV estimates of the rate of return to schooling based on college proximity are uniformly higher than OLS estimates. Second, although the IV estimates are imprecise, the range of the point estimates is 25-60 percent above the corresponding OLS estimates.

---

[22]For unmarried respondents enrolled in college and living away from home in 1966 the place of residence was defined as the place of residence of their parents. Thus there should be no reverse-causation for these individuals.

Is College Proximity a Legitimate Instrument?

For college proximity to serve as a legitimate instrument for completed education it must affect individual schooling decisions but have no direct effect on earnings. There are at least three reasons why men who grew up near a college may have higher earnings than other men, controlling for education, geographic information, and parental background. First, families that place a strong emphasis on education may choose to live near a college. Children of these families may have higher "ability" or may be more highly motivated to achieve labor market success. Either factor could induce a positive correlation between college proximity and the unobserved determinants of wages (i.e. $u\Omega$ in equation (2)). Second, the presence of a college may be associated with higher school quality at nearby elementary and secondary schools. Card and Krueger (1992) show that higher school quality is associated with higher earnings. The omission of direct information on the quality of schools attended by men in the NLSYM may then lead to an error component in wages that is correlated with college proximity. Finally, if only imperfect indicators are available for the place of residence in 1976, and if men who grew up in areas with a nearby college tend to live in higher-wage areas, then college proximity may be correlated with unobserved geographic wage premiums.

The interpretation of college proximity as a factor that lowers the cost of higher education suggests that growing up near a college should have a bigger effect on the education outcomes of children from poorer families. The pattern of education differentials in Figure 1 confirms this notion. Letting X_{1i} denote the components of X_i <u>other than</u> college proximity, the implied model for schooling is:

(1b) $S_i = X_{1i}\gamma_1 + C_i\delta_0 + C_i{*}P_i\delta_1 + v_i$,

where C_i is an indicator for growing up near a college, P_i is an indicator for low family income, and the coefficients δ_0 and δ_1 are both positive. In this case, even if C_i is included directly in the earnings equation:

(2b) $y_i = X_1\alpha_1 + C_i\alpha_0 + S_i\beta + u_i$,

the <u>interaction</u> $C_i{*}P_i$ of college proximity and poor family background can be used as an instrumental variable for education. The maintained assumption in this identification strategy is that the <u>direct</u> earnings effects of living near a college (e.g., unobserved geographic wage differentials) do not vary by family background.

Table 5 presents reduced form and structural estimates of the return to education based on equations (1b) and (2b). Low family background is defined by neither parent graduating from high school.[23] The reduced form coefficients in columns (1) and (2) confirm that the effects of living near a college are bigger for men with poorly-educated parents. The corresponding IV estimate of the return to education is presented in column (3), along with the direct earnings effect of living near a college. The estimated return to schooling is slightly smaller than the IV estimates in Table 3, and the estimated standard error is slightly larger. On the other hand, the point estimate of the direct earnings effect of college proximity is small and insignificantly different from 0. Although imprecise, these estimates provide no evidence against the assumption that college proximity is an exogenous determinant of schooling.

One potential criticism of the specification in columns (1)-(3) is the arbitrary classification of family backgrounds into only 2 categories. An alternative is to interact college proximity with a broader set of parental education indicators. The results in column (4) use interactions of college proximity with indicators for 8 parental education classes (the same indicators used in the earnings models in Tables 2-4). The expansion of the instrument

---

[23]This definition of low family background was derived by comparing mean education levels of men in the 8 parental education classes used in the models in Tables 3 and 4. The means show a discrete drop for men from the two lowest parental education categories. I therefore combined the two categories as a "low family background" indicator.

set has the effect of lowering the standard error of the IV estimate, while raising the point estimate slightly. An over-identification test for the mutual consistency of the available instruments is insignificant (p-value =0.28). As in column (3), the estimate of the direct earnings effect of living near a college is small and statistically insignificant.

Another alternative is to interpret predicted education in the absence of a nearby college (i.e. the predicted education level used to generate the quartiles in Figure 1) as a continuous indicator of "family background". Using the interaction of predicted education and college proximity as an instrument, and including college proximity directly in the earnings equation, the IV estimate of the return to education is 0.122, with a standard error of 0.075.

Regardless of the method of classifying family background, IV estimates based on the interaction of family background and college proximity are similar to IV estimates based on college proximity alone. Furthermore, estimates of the direct effect of college proximity on wages are uniformly small and statistically insignificant. Assuming that college proximity can be excluded from the earnings equation, both college proximity and its interaction with family-background indicators can be used as instruments for schooling. For example, using 9 parental education indicators interacted with college proximity as instruments for schooling, the IV estimate of the return to schooling is 0.115, with a standard error of 0.034.[24] Although this estimate is 57% above the corresponding OLS estimate, the Hausman-Wu statistic is 1.24 -- not large enough to reject the hypothesis of no simultaneity bias at conventional significance levels.

Discussion of Results

Although imprecise, the instrumental variables results presented in Tables 3-5 suggest that a conventional OLS estimation strategy yields a downward-biased estimate of the "true" return to education. This finding echoes the conclusion reached in a number of recent studies of endogenous schooling (cited above), and seems directly at odds with the widely accepted notion that individuals with higher education would have above-average earnings at any level of education. One possible explanation for the positive gap between IV and OLS estimates of the return to education is that the latter are downward-biased by measurement error in schooling. In light of the estimated reliability of survey measures of education, however, the potential downward bias in the OLS estimates is on the order of 10-15 percent. The differences between the IV and OLS estimates in

[24]The over-identification test statistic for this estimate (with 8 degrees of freedom) has a probability value of 0.38.

this paper and in other recent studies is substantially above this range.

An alternative possibility, discussed in some detail in Card (1993), is that the "true" rate of return to education varies across the population, and that the increase in education associated with college proximity occurs for individuals with relatively high rates of return to schooling. Algebraically, the IV estimate of the return to schooling is the ratio of the differences in average wages and average education between individuals who grew up in labor markets with and without a nearby college.[25] If the presence of a nearby college affects <u>only</u> the education decisions of men with poor family backgrounds, then the IV estimate depends only on the marginal return to schooling in this subset of the population. Thus one explanation for the relatively high IV estimates of the return to education in Tables 3-5 is that the marginal return to education among men with poor family backgrounds is relatively high.

Why do men with poorly educated parents have high returns to schooling? According to the simplest economic model of

---

[25]Specifically, let y_1 and y_2 represent mean wages of individuals who grew up in labor markets with and without a nearby college (adjusted for other covariates), and let S_1 and S_2 represent mean years of schooling for the same 2 groups (again, adjusted for other covariates). Then the IV estimate of the return to schooling is $(y_1-y_2)/(S_1-S_2)$.

school choice (Becker (1967)), individuals have decreasing marginal returns to schooling and invest in education until the marginal return to the last year of schooling equals their marginal discount rate.[26] If most of the variance in education outcomes is attributable to differences in individual-specific discount rates then on average the less-educated population will be mainly composed of individuals with high discount rates. Since low-income families presumably face higher interest rates than high-income families, this line of reasoning suggests that marginal returns to schooling are highest for the children of poor families. In effect, many less-educated workers stopped their schooling "too soon" because they faced high marginal costs of funds for further education.

This interpretation of the less-educated labor force stands at odds with a more conventional view that the less-educated are less "able" or have low benefits of schooling. At a minimum, the results in Tables 3-5 suggest that marginal returns to schooling among the less-educated are as high as typical OLS estimates of the return to schooling. Taken in combination with the results in other recent studies of endogenous education -- all of which find downward bias in the OLS estimates -- the results here suggest

---

[26]This is a condensed version of the argument developed in Card (1993). See also Lang (1993).

that the economic value of education for many children may be significantly understated.

Conclusion

Any credible analysis of the causal link between education and earnings requires an exogenous source of variation in education choices. In this paper I explore the use of college accessibility as an exogenous determinant of schooling. An analysis of education and earnings outcomes for men in the NLS Young Men Cohort shows that men who grew up in areas with a nearby 4-year college have significantly higher schooling and significantly higher earnings. These effects are concentrated among men with poorly-educated parents -- men who would otherwise stop schooling at relatively low levels. The implied instrumental variables estimates of the earnings gain per year of additional schooling (10-14%) are substantially above the earnings gains estimated by a conventional ordinary least squares procedure (7.3%).

These inferences are robust to minor changes in specification, including the addition of measured test scores to the earnings model and changes in the definition of college proximity. Nevertheless, they rely on the restrictive assumption that living near a college has no effect on earnings apart from the effect through education. To test this assumption I use the fact that college proximity has a larger impact on the schooling choices of men with poorer family backgrounds. Thus, an interaction of college proximity and low family background can be used as an instrumental variable for observed schooling even in earnings models that include a direct college proximity effect. The results of this test give rise to estimates in the same range as the simpler instrumental variables estimates based on college proximity alone.

While none of the instrumental variables estimates of the return to education is very precise, they all point toward relatively high returns to schooling for children of poorly-educated parents. This pattern is consistent with a simple economic model of endogenous schooling in which differential access to funds leads to relative under-investment in schooling among children of lower-income families.

References

Anderson, C. Arnold, Mary Jean Bowman, and Vincent Tinto. Where Colleges Are and Who Attends. New York: McGraw Hill, 1972.

Angrist, Joshua D. and Guido W. Imbens. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." Unpublished Discussion Paper, Harvard University Department of Economics, 1993.

Angrist, Joshua D. and Alan B. Krueger. "Does Compulsory Schooling Affect Schooling and Earnings?" Quarterly Journal of Economics 106 (November 1991): 979-1014. (1991a)

Angrist, Joshua D. and Alan B. Krueger. "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery." Princeton University Industrial Relations Section Working Paper #290, August 1991. (1991b)

Angrist, Joshua D. and Alan B. Krueger. "Split Sample Instrumental Variables". Unpublished Discussion Paper, Princeton University Department of Economics, July 1993.

Ashenfelter, Orley and Alan B. Krueger. "Estimates of the Economic Return to Schooling for a New Sample of Twins." Princeton University Industrial Relations Section Working Paper #304, July 1992.

Becker, Gary S. Human Capital and the Personal Distribution of Income. Ann Arbor: University of Michigan Press, 1967.

Butcher, Kristin F. and Anne Case. "The Effect of Sibling Composition on Women's Education and Earnings."

Unpublished Discussion Paper, Princeton University Department of Economics, June 1993.

Card, David and Alan B. Krueger. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." Journal of Political Economy 100 (February 1992): 1-40.

Card, David. "Earnings Schooling and Ability Revisited". Unpublished Paper, Princeton University Department of Economics, August 1993.

Ehrenberg, Ronald and Robert Smith. Modern Labor Economics (4th edition). New York: Harper Collins, 1991.

Griliches, Zvi. "Wages of Very Young Men." Journal of Political Economy 84 (1976): 569-586.

Griliches, Zvi. "Estimating the Returns to Schooling: Some Econometric Problems." Econometrica 45 (January 1977): 1-22.

Hall, George E. and Anthony Turner. "Sampling, Interviewing, and Estimating Procedures". Appendix B in Career Thresholds: A Longitudinal Study of the Education and Labor Market Experience of Male Youths. United States Department of Labor Manpower Administration. Washington DC: USGPO, 1970.

Kane, Thomas J. and Cecilia E. Rouse. "Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?" Princeton University Industrial Relations Section Working Paper #311, January 1993.

Lang, Kevin. "Ability Bias, Discount Rate Bias, and the Return to Education." Unpublished Discussion Paper, Boston University Department of Economics, May 1993.

Mallar, Charles D. "Alternative Econometric Procedures for Program Evaluations: Illustrations from an Evaluation of Job Corps." 1979 Proceedings of the Business and Economic Statistics Section, American Statistical Association. Washington DC: American Statistical Asociation, 1979, pp. 317-321.

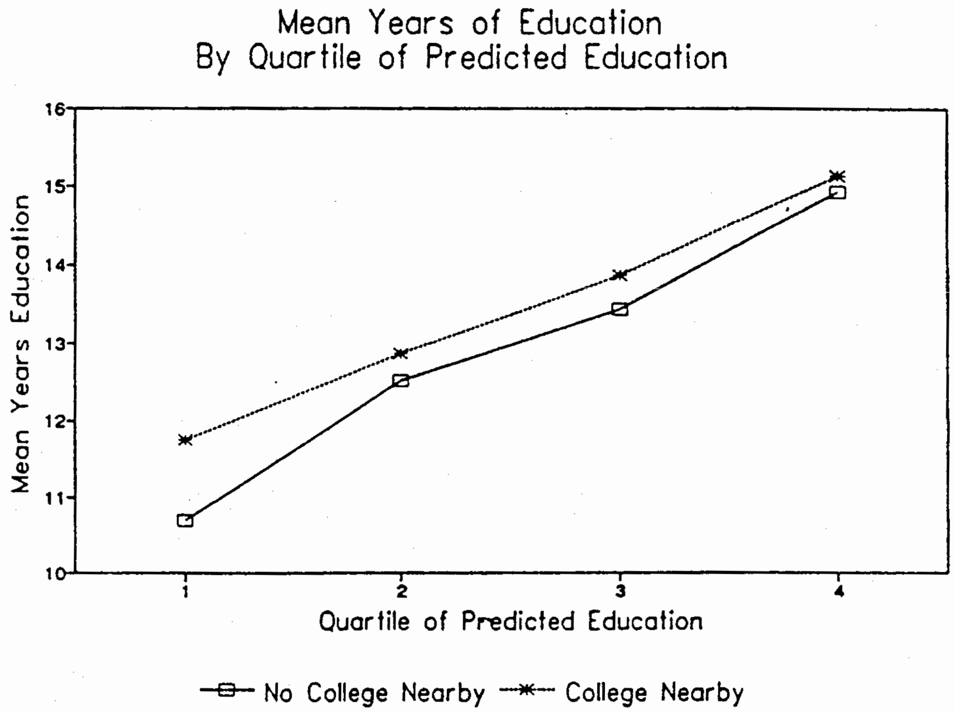
Mincer, Jacob. Schooling, Experience, and Earnings. New York: National Bureau of Economic Research, 1974.

Psacharopoulos, George. "Returns to Education: A Further International Update and Implications." Journal of Human Resources 20 (Fall 1985): 583-604.

Rosen, Sherwin. "Human Capital: A Survey of Empirical Research." In Ronald Ehrenberg, editor, Research in Labor Economics volume 1. Greenwich Connecticut: JAI Press, 1977.

Seigel, Paul and Robert Hodge. "A Causal Approach to the Study of Measurement Error." In Hubert Blalock and Ann Blalock, editors, Methodology in Social Research New York NY: McGraw Hill, 1968.

Willis, Robert J. "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions." In Orley Ashenfelter and Richard Layard, editors, Handbook of Labour Economics. New York NY: North Holland, 1986.



Note: prediction equation is fit to subsample with no college nearby

Figure 1

Table 1: Sample Characteristics for Overall Sample and 1976 Subset of National Longitudinal Survey of Young Men

	Overall NLS-YM Sample	Subset Interview in 1976: Valid Education	Valid Wage & Education
1. Age Distribution in 1966:			
Age 14-15 (%)	25.9	25.3	25.5
Age 16-17	24.9	23.8	24.1
Age 18-20	23.1	24.1	24.6
Age 21-24	26.1	26.7	25.8
2. Regional Distribution in 1966:			
Northeast (%)	20.2	20.0	20.7
Midwest	25.4	26.3	26.0
South	41.1	41.3	41.4
West	13.3	12.5	11.9
3. Lived in SMSA 1966 (%)	66.0	64.3	65.0
4. Lived Near 4-year College in 1966 (%)	69.2	67.8	68.2
5. Family Structure at Age 14:			
Mother & Father (%)	76.8	79.2	78.9
Mother Only (%)	11.8	10.0	10.1
6. Average Parental Education			
Mother's Education (yrs)	10.3	10.4	10.3
Father's Education (yrs)	9.4	10.0	10.0
7. Percent Black	27.5	23.0	23.0
8. Average Score on KWW Test	33.0	33.5	33.5
9. Interviewed in 1976 (%)	70.7	100.0	100.0
10. Mean Education in 1976	13.2	13.2	13.3
11. Live in South in 1976 (%)	39.6	40.0	40.3
12. Sample Size	5225	3613	3010

Notes: Means are based on all available valid observations in any subsample.

Table 2: Estimated Regression Models for Log Hourly Earnings

	(1)	(2)	(3)	(4)	(5)
1. Education	0.074 (0.004)	0.075 (0.003)	0.073 (0.004)	0.074 (0.004)	0.073 (0.004)
2. Experience	0.084 (0.007)	0.085 (0.007)	0.085 (0.007)	0.085 (0.007)	0.085 (0.007)
3. Experience-Squared /100	-0.224 (0.032)	-0.229 (0.032)	-0.230 (0.032)	-0.226 (0.032)	-0.229 (0.032)
4. Black Indicator	-0.190 (0.017)	-0.199 (0.018)	-0.194 (0.019)	-0.194 (0.019)	-0.189 (0.019)
5. Live in South	-0.125 (0.015)	-0.148 (0.026)	-0.146 (0.026)	-0.145 (0.026)	-0.146 (0.026)
6. Live in SMSA	0.161 (0.015)	0.136 (0.020)	0.136 (0.020)	0.137 (0.020)	0.138 (0.020)
7. Region in 1966 (8 indicators)	no	yes	yes	yes	yes
8. Live in SMSA in 1966	no	yes	yes	yes	yes
9. Parental Education[a] (main effects)	no	no	yes	yes	yes
10. Interacted Parental Education Classes[b]	no	no	no	yes	yes
11. Family Structure[c] (2 indicators)	no	no	no	no	yes
12. R-squared	0.291	0.300	0.301	0.303	0.304
13. P-value for family background effects	--	--	0.235	0.462	0.165

Notes: Standard errors in parentheses. Sample size is 3010. The dependent variable in all cases is the log of hourly wages in 1976. The mean and standard deviation of the dependent variable are 6.262 and 0.444.

[a] Variables representing years of education of mother and father, plus indicators for missing mother's or father's education.

[b] Indicators for 8 classes of mother's and father's education.

[c] Indicators for father and mother present at age 14, and single mother at age 14.

Table 3: Reduced Form and Structural Estimates of Education and Earnings Models

	Reduced Form Models: Education		Reduced Form Models: Earnings		Structural Models of Earnings	
	(1)	(2)	(3)	(4)	(5)	(6)
A: Treat Experience and Experience Squared as Exogenous						
1. Live Near College in 1966	0.320 (0.088)	0.322 (0.083)	0.042 (0.018)	0.045 (0.018)	--	--
2. Education	--	--	--	--	0.132 (0.055)	0.140 (0.055)
3. Family Background Variables[a]	no	yes	no	yes	no	yes
B: Treat Experience and Experience Squared as Endogenous[b/]						
4. Live Near College in 1966	0.382 (0.114)	0.365 (0.105)	0.047 (0.019)	0.048 (0.019)	--	--
5. Education	--	--	--	--	0.122 (0.046)	0.132 (0.049)
6. Family Background Variables[a]	no	yes	no	yes	no	yes

Notes: standard errors in parentheses. Sample size is 3010. The dependent variable in columns 1 and 2 is completed education in 1976 (mean and standard deviation: 13.263 and 2.677). The dependent variable in columns 3-6 is the log of hourly wages in 1976 (mean and standard deviation: 6.262 and 0.444). All models include a black indicator, indicators for southern residence and residence in an SMSA in 1976, indicators for region in 1966 and living in an SMSA in 1966, as well as experience and experience squared.

[a] 14 variables representing mother's and father's education, indicators for missing father's or mother's education, interactions of mother's and father's education, and dummies for family structure at age 14.

[b] In these models, experience is treated as endogenous. Instruments for experience and experience squared are age and age squared.

Table 4: OLS and Instrumental Variables Estimates of the Return to Education: Alternative Specifications

	OLS Estimate	IV Estimate[a]
1. Basic Specification (N=3010)	0.073 (0.004)	0.132 (0.049)
2. Use 1978 Wages and Education (N=2639 with 1978 data)	0.066 (0.006)	0.117 (0.061)
3. Include KWW Test Score (N=2963 with valid KWW)	0.055 (0.004)	0.136 (0.078)
4. Include KWW Test Score Instrument KWW with IQ[b] (N=2040 with valid KWW and IQ)	0.061 (0.005)	0.089 (0.085)
5. Use Proximity to Public College as instrument for education	as in row 1	0.194 (0.059)
6. Use Proximities to 2-year and 4-year colleges as instruments for education	as in row 1	0.117 (0.047)
7. Use Subsample Age 14-19 in 1966 (N=2037)	0.076 (0.006)	0.094 (0.064)

Notes: The dependent variable in row 1 and rows 3-7 is the log of hourly wages in 1976. The dependent variable in row 2 is the log of hourly wages in 1978. Reported estimates are coefficients of linear education variable in models that also include a black indicator, indicators for southern residence and residence in an SMSA in 1976, indicators for region in 1966 and living in an SMSA in 1966, experience and experience squared, and 14 variables representing mother's and father's education, indicators for missing father's or mother's education, interactions of mother's and father's education, and dummies for family structure at age 14.

[a] In these models education and experience are treated as endogenous. Instruments for experience and experience squared are age and age squared. Instrument for education is proximity to a 4-year college unless otherwise noted.

[b] In the IV estimation KWW is treated as endogenous and IQ score is added to the instrument list.

Table 5: Instrumental Variables Estimates of the Return to Education Based on Interaction of Parental Education and Proximity to College

	Reduced Form Models: Education	Reduced Form Models: Earnings	Structural Models of Earnings	
	(1)	(2)	(3)	(4)
1. Live Near College in 1966	0.154 (0.135)	0.029 (0.024)	0.015 (0.029)	0.013 (0.024)
2. Live Near College * Low Parental Education[a]	0.462 (0.186)	0.043 (0.032)	--	--
3. Education[b]	--	--	0.093 (0.065)	0.097 (0.048)
4. Family Background Variables[c]	yes	yes	yes	yes

Notes: standard errors in parentheses. Sample size is 3010. The dependent variable in all models is the log of hourly wages in 1976. The mean and standard deviation of the dependent variable are 6.262 and 0.444. All models include a black indicator, indicators for southern residence and residence in an SMSA in 1976, indicators for region in 1966 and living in an SMSA in 1966, as well as experience and experience squared. Experience and experience squared are treated as endogenous, with age and age squared used as instruments.

[a] Interaction of indicator for living near a college in 1966 and indicator for both parents having less than high-school education.

[b] In column 3 the instrument for education is an interaction of an indicator for low parental education with an indicator for living near a college in 1966. In column 4 the instruments are interactions of 8 parental education class indicators with an indicator for living near a college in 1966.

[c] 14 variables representing mother's and father's education, indicators for missing father's or mother's education, interactions of mother's and father's education, and dummies for family structure at age 14.