This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Mental Ability and Higher Educational Attainment in the 20th Century

Volume Author/Editor: Paul Taubman and Terence Wales

Volume Publisher: NBER

Volume ISBN: 0-87014-243-7

Volume URL: http://www.nber.org/books/taub72-1

Publication Date: 1972

Chapter Title: Conceptual and Statistical Problems

Chapter Author: Paul Taubman, Terence Wales

Chapter URL: http://www.nber.org/chapters/c5258

Chapter pages in book: (p. 11 - 16)

3. Conceptual and Statistical Problems

CAUSALITY VS. DESCRIPTION

A question that is of some importance in statistical considerations is the interpretation of the education-ability relationship. That is, does ability "cause" the educational attainment—or vice versa—or does the relationship arise for other reasons.

Let us assume that students and their families have a demand function for educational attainment—for both the consumption and investment aspects. Regardless of whether students want either or both of these aspects, plausible arguments can be made that the demand depends upon the student's ability. Indeed, whether one uses students' educational plans or their actual realization of these plans, a substantial body of evidence exists suggesting that the demand for education is a function of ability.[1] This demand will also depend on such other factors as the family's income level, job and scholarship opportunities, tuition, etc.[2]

On the other hand, educational authorities try to weed out people with low ability levels. What is considered too low may depend upon the physical and budget capacity of the institutions or governments involved. In any event, evidence exists that willingness to promote students to higher grades, to encourage them to stay in school, or to permit them to go to a higher educational institution has varied over time.[3] Thus any observed relationship between educational attainment and ability is the outcome of the factors that affect supply and

[1] The plans may be more relevant because one reason that students do not fulfill their plans is that the educational authorities exclude those with low ability. That is, the realization in part reflects supply conditions.

[2] The family's income level affects the demand relation because of imperfect capital markets, differences in tastes for present versus future consumption, and the luxury nature of the consumption of education.

[3] See, for example, Folger and Nam (1967) on the trends in the number of students who were not in the normal school grade of their age group. Consider, also, state and federal provision of support for college facilities.

demand. Shifts in these factors can alter the observed relationship without implying any causation; therefore, we conclude that the data on education and ability should not be interpreted in a causal sense. We will generally use the term "descriptive" to characterize this relationship.

The fact that we interpret the education-ability relation as descriptive provides no guidance for deciding which variable to use as the dependent one in regressions. However, there are two major reasons for using ability as the dependent variable.[4] First, the education-ability relation enables us to correct the bias (of the education coefficient) arising from the omission of ability in income equations. For this purpose we require the education-ability equation to be formulated with education as the independent variable.[5] Second, errors in measuring ability will not bias the coefficient in the regression if ability is used as the dependent variable. There will be a bias if ability is used as the independent variable. On the other hand, there is a rationale for using education as the dependent variable when dealing with certain nonlinear functional relations. That is, one way to test for nonlinearities is to include the independent variable in squared form. This can be accomplished only if ability is the independent variable.[6]

In general, there appear to be no sound reasons for preferring a particular functional form to relate education to ability.[7] For simplicity, we used the linear form. We have, however, tested for nonlinearities by regressing education on ability and ability squared. Where the nonlinearities are significant we indicated the extent to which our conclusions are affected. We also experimented with the logarithmic form but have not presented the results, since this form does not fit well in the tails of the distrubitions, and the estimated coefficients appear to be very sensitive to the scaling of the ability variable—for example, using the midpoints or endpoints of the decile ranks.

[4] In addition, for samples in which individuals have different amounts of education when tested, it may be possible to correct the bias when ability is the dependent variable.

[5] This is necessary because, as shown in footnote 11, Chapter 1, in equation (2) the estimate of k is obtained from estimating (by least squares) an equation in which education is the independent variable.

[6] The education variable cannot be included in both unsquared and squared forms as the independent variable because it is obtained by aggregating a zero-one variable, which when squared is still a zero-one dummy variable.

[7] However, for purposes of analyzing the relationship between income, education, and ability, it is necessary that the functional form for the side relation correspond to that of the basic relation. If a dummy variable for college entrance is used in the income analysis, then our linear equation is appropriate. If different dummies are used to represent various educational levels, then our linear equation provides the first step in determining the bias.

There is one minor statistical point that can be dispensed with now. We have been talking interchangeably of the education variable as representing a situation in which an individual does or does not enter college and as representing the fraction of high school graduates entering college. These two concepts can be reconciled as follows. We define a variable D_i as 1 if the ith person enters college and as zero otherwise. Our linear equation for the ith individual is therefore $A_i = h + kD_i$, where A_i is again the ability of the individual. Suppose that we now order the data by ability class and average the observations in each ability group. The education variable then becomes the percentage of people in each ability class who enter college (E_{12}) and the ability variable becomes the average ability level in the class (A).[8]

The linear equation that we estimate for the different samples is therefore $A = h + kE_{12}$. It may be useful at this point to interpret the coefficients h and k. The coefficient h indicates the level of ability at which the fraction of high school graduates entering college is zero. Since in nearly all our samples some students enter college at all ability levels, our estimates of h are generally negative. An alternative interpretation of h may be obtained by solving this equation for E_{12} to give: $E_{12} = -\frac{h}{k} + \frac{1}{k} A$. Provided that h is negative, some students will continue to college even at the lowest ability levels. From this equation, $1/k$ can be interpreted as the increase in the fraction of students entering college for each unit increase in A.

EFFECT OF EDUCATION ON MENTAL ABILITY

Our main interest is in determining the relationship between the percentage of high school graduates entering college and their mental ability at the time of college entrance. The ability measures that we use are various IQ and achievement test scores. These are determined in part by the amount of schooling the individual received prior to taking the tests.

The pioneering study of Learned and Wood (1938) clearly demonstrates the extent to which even IQ measures are affected by years of schooling. In this study nearly 28,000 high school seniors were given a twelve-hour examination in 1928. One part of the examination was the Otis IQ test. Those students who went on to college were retested in eight-hour examinations in 1930 and 1932. Moreover, exactly the same Otis test was given on the last two occasions. Comparing test scores for

[8]Formally, this can be accomplished by multiplying by a grouping matrix G whose elements in the ith row (which corresponds to the ith ability group) are zero for all observations not in that ability group and $1/n_i$ for the n_i observations in the group. This gives: $GA_i = hG + kGD_i$. In the ith ability group $GD = n^*_i/n_i$ where n^*_i is the number of people who enter college in that group. Since in the ith group D has n^*_i entries of one and $n_i - n^*_i$ values of zero, its average is n^*_i/n, which is equal to the percentage of people in that ability class who entered college. We denote this percentage as E_{12}.

those in the sample in 1928 and 1930 and those in the sample in 1930 and 1932, it was found that the average score on this test rose 7-1/2 percent from 1928 to 1930 and 5 percent from 1930 to 1932. In other words, the Otis (and presumably all other IQ tests) appear to measure educational attainment as well as mental ability.

Consequently, data from samples in which individuals are subjected to tests after having completed their formal education must be treated differently from those in which all individuals are tested as high school seniors. From a statistical viewpoint, the former problem may be analyzed as an error-in-variables.[9] In nontechnical terms, the problem may be described as follows. People with more education will score higher on tests because of this additional education. Thus it is difficult to distinguish between the effect of education on test scores and the relationship between the mental ability of students at, say, the end of high school and after additional educational attainment. In this case our regression analysis yields biased estimates of the parameters of the

[9]We wish to estimate the (descriptive) relationship between educational attainment, S, and mental ability, A. Let the true relationship be expressed as:

(1) $S = \gamma A + u$

Suppose, however, that instead of observing A, we measure IQ where

$IQ = A + z$ and where $E(u,z) = E(A,z) = 0$ but $E(S,z) > 0$

If we use ordinary least squares to estimate the equation $S = gIQ + v$, then:

(2) $\text{plim}\,(\hat{g}) = \text{plim}\ \dfrac{\Sigma(S,z) + \gamma\Sigma A^2}{\Sigma(A^2 + z^2)}$

Hence $\hat{g}$ from (2) will, in the limit, exceed γ provided that:

(3) $\dfrac{\Sigma(S,z)}{\Sigma z^2} > \gamma$

But the left-hand side of (3) can be interpreted as the least squares estimate of λ in the equation $S = \lambda z + v$. Thus our estimate of $\hat{g}$ exceeds or falls short of γ, as λ exceeds or falls short of γ, and not even the direction of the (asymptotic) bias is determinable without further information. However, studies such as Learned and Wood contain information on the change in z due to a change in education, and hence we can estimate λ using first differences. This permits us to determine the sign but not the extent of the statistical bias, which in general requires knowledge about $\Sigma A^2 / \Sigma(A^2 + z^2)$.

This ambiguity in the sign of the bias is removed if we postulate the relationship as:

(4) $A = \delta S + w$

Once again we measure A as IQ and regress IQ = ds, which yields:

(5) $\text{plim}\,(\hat{d}) = \delta + \text{plim}\ \dfrac{\Sigma(S,z)}{\Sigma S^2} > \delta$

Thus with S as the independent variable, our estimate of δ will be biased upward and $\hat{d}$ can be used as an upper limit of δ. Of course it will only be possible to estimate the extent of bias if $\Sigma(S,z) / \Sigma S^2$ is known. But this term is the least squares estimate of ψ in $z = \psi S + v$, and as such measures the contribution of schooling to knowledge or scores on tests. It may be possible to estimate this relationship from data in Learned and Wood.

equation relating ability and education. However, as shown in footnote 23, when education is the independent variable, it may be possible to correct the estimate on the basis of a regression of IQ on additional education.

On the other hand, the relation between ability and education is not obscured if it is estimated from a sample in which IQ's are measured for individuals with the *same* amounts of education at the time of the test.[10] This condition is satisfied by a follow-up survey. By a follow-up survey we mean one in which individuals with the same amount of schooling are tested at a point in time, then their further educational attainment is determined by a future survey. Since all the students will have had the same amount of schooling when they are tested, there can be no differences in the IQ scores that are due to differences in years of schooling.

[10]That is, in terms of the errors in variables analysis, the bias arises because z varies between individuals. If z is constant for all individuals then $(z - \bar{z})$ will be equal to zero for each person and all sums involving z's will also be zero.