Wages of Very Young Men

Zvi Griliches

The Journal of Political Economy, Vol. 84, No. 4, Part 2: Essays in Labor Economics in Honor of H. Gregg Lewis. (Aug., 1976), pp. S69-S86.

Stable URL:

The Journal of Political Economy is currently published by The University of Chicago Press.

Wages of Very Young Men

Zvi Griliches

Harvard University

I. Introduction

Over the past decade there has been much interest in and a large amount of work done on estimating the economic returns to formal schooling and on trying to untangle such returns from the contributions of native ability, family background, discrimination, and nepotism. The rather large literature that has emerged has been discussed and surveyed by a number of authors (Jencks 1972; Welch 1975; Griliches 1975*a*; Rosen 1975; among others). Different methodologies and different sets of data have produced very little agreement. It is not the purpose of this paper to review and revive all the debates again. Instead, it will attempt to replicate the results of an earlier study of "Education, Income, and Ability" (Griliches and Mason 1972) on a new set of data, the National Longitudinal Survey of Young Men, focusing on the estimation of the economic returns to schooling in the presence of individual differences in ability.

The NLS data base is of interest because it is the most representative data set combining information on earnings, schooling, and measures of ability. It contains data on *two* measures of ability: IQ scores collected from the high schools attended by the respondents and scores on a test of "knowledge of the world of work" (KWW) administered at the time of the initial interview in 1966. Data were also collected on parental background, wage rates (rather than just total income or earnings), and work experience. These data are also of interest because of the availability of repeated observations on the same individuals and the ability to match family members across surveys. I shall not pursue, however, the last two

topics here; instead, I shall concentrate on the replication of the Griliches and Mason (1972) study as a way of entering and reconnoitering this new and rather large body of data.[1]

No data set is perfect, and this one has at least two major shortcomings: many data are missing (IQ scores were collected for only 65 percent of the respondents), and the respondents are very young. In the nonenrolled portion of our sample, the average age is only 22 in 1969, though the range is from 17 through 27. It is well known that the full effects of schooling or ability are not easily observable at this early an age. There are intimations of the individual futures to come but far from a clear reading of them. Also, much of the labor force behavior of youths in the earlier part of this age range is characterized by search, experimentation, and often a lack of "seriousness." Thus, observations on their encounters with the labor market during these early years cannot be taken as reflecting unequivocally their underlying potential. Nevertheless, a glimmer of it is there, and that is what I will be analyzing.

II. The Model

The theoretical and statistical problems of simultaneously estimating the *net* effects of both schooling and ability are discussed in some detail in Griliches and Mason (1972) and Griliches (1975b). A brief recapitulation will suffice here. We are interested in estimating the following equation:

$$y_1 = LW = \alpha_1 + \beta_1 S + \gamma_1 A + X\delta_1 + e_1, \tag{1}$$

where LW is the logarithm of the wage rate, S is schooling, A is a measure of ability, X is a vector of other wage determining variables such as experience, region, and city size, and e is a disturbance summarizing the effects of other, hopefully random, sources of differences in wage rates. The issue in Griliches and Mason (1972) and in much of the rest of the literature is focused on the "importance" of A and the bias introduced into our estimates of β_1 by ignoring it. The analysis proceeds by finding measures of A, introducing them into (1), and seeing how much the estimate of β_1 is changed thereby. A number of issues arise immediately: (1) Since we have two measures of ability, KWW and IQ, which should be used? (2) What if these measures are subject to error? (3) Can one take schooling as predetermined and error free? We shall explore these issues in turn.

The bias in the estimated schooling coefficient due to the omission of the ability dimension is given by $\gamma_1 \cdot b_{AS \cdot X}$, where γ_1 is the true net coefficient of ability, while $b_{AS \cdot X}$ is the auxiliary regression coefficient of the left-out ability measure on schooling, holding the other variables in the model (X) constant. Even if one is willing to assume the constancy of

---

[1] Additional analyses of these data are presented in Griliches (1975b).

γ_1, the second component of the bias, which summarizes the relationship in the sample between ability and schooling and reflects the "selectivity" of the particular schooling system, time, and place, need not be constant across samples, subsamples, and time periods. Thus, there is no good reason to expect that any particular study can answer this question definitively for all other past and future studies. Moreover, β_1 itself is unlikely to be constant in the face of different X's. For example, in this age range it matters greatly whether the effect of schooling is estimated holding age or experience constant. The former procedure leads to a significantly lower estimate of β_1. Hence, the same estimate of absolute "ability bias" may imply different estimates of relative (percentage) bias. Also, given that we do not have a direct measure of the output of schooling but only a measure of the years spent (input) in schools, we cannot really interpret the estimated IQ coefficients as reflecting solely the net effects of early ability. The missing component of school quality and individual input into the schooling process during those years is hopelessly confounded with the measured IQ scores and a fortiori with the KWW scores. Recognizing that years of schooling are not an error-free measure of "education achieved" implies that the reported ability coefficients and associated estimates of bias in the schooling coefficients may themselves be biased upward.

Since the KWW test cannot be taken as independent of schooling, we describe its relationship to schooling by[2]

$$y_2 = \text{KWW} = \alpha_2 + \beta_2 S + \gamma_2 A + X\delta_2 + e_2. \tag{2}$$

To represent the dependence of schooling on ability, we write

$$y_3 = S = \alpha_3 + \gamma_3 A + X\delta_3 + e_3, \tag{3}$$

and to allow for the possibility that IQ may not be a perfect measure of the relevant ability concept, we define

$$y_4 = \text{IQ} = \alpha_4 + A + e_4. \tag{4}$$

Assuming that either IQ or KWW are errorless measures of the relevant concepts amounts to assuming that e_2 and e_4 have zero variance. If one were to use KWW as a measure of ability, one would get (substituting [2] in [1] and ignoring constant terms)

$$LW = (\beta_1 - \beta_2/\gamma_2)S + \gamma_1/\gamma_2 \text{ KWW} + X(\delta_1 - \delta_2/\gamma_2) + e_1 - \gamma_1/\gamma_2' e_2. \tag{5}$$

[2] Actually, KWW depends on schooling in 1966 (SC66) rather than on SC69. But in the not-enrolled portion of our sample, SC66 and SC69 are too closely intercorrelated to distinguish between them. In a separate study using expectational variables for income and schooling, this distinction is made (see Griliches 1975b).

Thus, unless schooling does not affect KWW ($\beta_2 = 0$), introducing KWW into the wage equation will eliminate some of the "ability bias" in the schooling coefficient but in its turn introduce another downward bias ($-\beta_2/\gamma_2$) into the estimated schooling coefficient. If KWW itself is subject to error of measurement ($\sigma_2^2 \neq 0$), the above conclusion will be attenuated but will retain its qualitative implications. If both IQ and KWW are subject to independent measurement error, equation (1) can be estimated using instrumental variable methods. If one considers the possibility that schooling itself may be determined on the basis of expected income and that it too may be measured with error, this would imply a correlation of e_1 with S and suggest the use of instrumental variables for it too.

Note that I have not discussed family background variables explicitly. They are contained in the definition of X and allowed to be correlated with the true A. In what follows I shall assume, for the most part, that family background variables such as mother's education or father's occupation do not affect wages directly (i.e., do not enter eq. [1] separately) but only indirectly via their effects on ability and schooling. This is a testable restriction on the model. For some purposes, however, one is more interested in the "total" effect of family background or ability on wages and then in their "net" effects holding schooling constant. Such "total" effects can be discerned from the reduced-form version of (1), derived by substituting (3) into it:

$$LW = (\gamma_1 + \beta_1\gamma_3)A + X(\delta_1 + \beta_1\delta_3) + e_1 + \beta_1 e_3. \tag{6}$$

In the following sections, I will concentrate first on estimating equation (1), assuming that IQ and KWW are error-free measures. This is most comparable to our earlier study (Griliches and Mason 1972) and to most of the other work in this area. Two-stage least-squares estimates of the wage equation which allow for errors of measurement in IQ (or KWW) *and* for the endogeneity of schooling and experience will be presented briefly toward the end of the paper.[3]

III. Data, Major Variables, and Problems

The data used in this study are based on a national sample of the civilian noninstitutional population of males who were 14–24 years old in 1966.[4] Blacks were oversampled in a 3:1 ratio. The original sample consisted of 5,225 individuals of whom 3,734 were white. By 1969 about 23 percent of the original sample was lost, 13 percent of it only temporarily (to the army). Data are currently available from the 1966–69 surveys of these

---

[3] For a more detailed discussion and modeling of such a system of equations, see Griliches (1975b) and Chamberlain (1976).

[4] See U.S. Department of Labor (1970–74, vols. 1–4) for more details on the sample.

individuals. The surveys for 1970–71 should be available in the near future, and plans are in progress for continuing to resurvey these same individuals at 2-year intervals in the future. In addition to the usual direct sociodemographic interview questions, all youths who had completed ninth grade by 1966 were asked to sign waivers letting their schools supply the census their scores on various tests and other background materials. The resulting School Survey yielded data on different mental ability scores for 3,375 individuals. These were rescored at Ohio State University into IQ equivalents.[5] The availability of such scores and of scores on a test of the Knowledge of the World of Work in the original 1966 survey interview allows a comparison with the Griliches and Mason (1972) study, which used Armed Forces Qualification Test scores to analyze the earnings of 25–38-year-old veterans in 1964.

Table 1 describes the major characteristics of the samples and the main variables used in the various analyses. Since we limit ourselves here to those who were interviewed in 1969, we start with a sample of 4,033. Of these, only 3,765 reported earnings in 1969, and another 300–500 observations were lost due to missing data on schooling, occupation, and parental background variables. Given the interest in the postschooling labor force experience, this analysis deals primarily with two subsamples of *not-enrolled* youth, based on 2,062 and 1,362 individuals for the "all" not-enrolled and "real IQ" data subsets, respectively.[6] The not-enrolled sample is somewhat older and slightly more disadvantaged relative to those who stay on in school longer. The subsample with all "real IQs" excludes mainly those who did not continue school past the ninth grade, thus cutting off much of the bottom of the distribution (by economic background and ability).

The major dependent variable (LW69) is the logarithm of the wage rate (hourly or converted to hourly equivalent) earned in 1969 on the current or last job.[7] The major independent variables used (besides ability and family background) were: schooling in 1969 (S69) in years, a nonlinear function of cumulated work experience ($XBT = exponent - 0.1 \cdot EXP69$, EXP69 in years), time spent in the armed forces (AFEX) in years, and a current location (CL) set, consisting of binary variables for SMSA, current location in the South (RNS), and the interaction of being black and currently in the South (BRNS).

Background (BKG) variables include measures of father's occupation

---

[5] U.S. Department of Labor 1970–74, vol. 4, appendix E.

[6] In the earlier version of this paper (Griliches 1974) I reported similar results for somewhat larger samples. The samples here are limited to those individuals who also reported their expectations of educational and occupational achievement at age 30. The expectational variables are analyzed in Griliches (1975*b*).

[7] We also looked at the logarithm of total earnings for the previous year, with rather similar results. To refrain from getting into issues of labor force participation, I am not reporting them here (see Griliches 1974 for details).

TABLE 1

CHARACTERISTICS OF DIFFERENT SUBSAMPLES OF YOUNG MEN
FROM THE NATIONAL LONGITUDINAL SURVEY:
MEANS AND STANDARD DEVIATIONS (IN PARENTHESES)

	ALL		NOT ENROLLED IN 1969	
VARIABLE	All	Valid IQ Scores	All	With Valid IQ Scores
N	4,601	3,025	2,026	1,362
Age 69	21.2	21.5	22.2	22.3
	(3.2)	(3.0)	(3.2)	(3.2)
S69	...	...	11.6	12.5
			(2.4)	(1.9)
S66	10.7	11.5	10.8	11.6
	(2.4)	(1.9)	(2.4)	(2.0)
LW69	...	...	5.60	5.68
			(0.426)	(0.398)
KWW	33.3	35.5	33.0	35.1
	(8.6)	(7.6)	(9.0)	(7.9)
IQ	...	101.2	...	97.7
		(15.9)		(15.3)
FOMY14	5,120	5,372	4,826	5,095
	(1951)	(1960)	(1,779)	(1,777)
Black	0.27	0.17	0.28	0.19
Culture	2.2	2.4	2.0	2.3
	(1.01)	(0.8)	(1.0)	(0.9)
Siblings	3.3	2.9	3.6	3.1
	(2.6)	(2.3)	(2.7)	(2.4)
EXP69	...	...	4.0	3.7
			(3.1)	(2.8)
XBT	...	...	0.70	0.72
			(0.27)	(0.18)
SMSA	...	...	0.61	0.65
RNS	0.32	0.33	0.41	0.33

NOTE.—LW69 = logarithm of hourly earnings (in cents) on the current or last job in 1969; KWW = score on the "knowledge of the world of work" test administered in 1966; IQ = score on IQ-type tests collected from the high school last attended by the respondent; FOMY14 = occupation of father or head of household when respondent was 14 scaled by the median earnings of all U.S. males in this occupation in 1959; culture = index based on the availability of newspapers, magazines, and library cards in the respondent's home; EXP69 = postschool work experience estimated on the basis of the work record (in weeks) since 1966 and the date of first job after school and the date stopped school (in years; truncated at age 14 if respondent started working earlier); XBT = $e^{-0.1 \cdot \text{EXP69}}$; SMSA = respondent in SMSA in 1969; and RNS = respondent in South in 1969.

(FOMY14), mother's education (MED), number of siblings, dummy variables for observations with missing values for father's occupation (DFO14) and mother's education (DME), and a culture index based on the availability of newspapers, magazines, and library cards in the respondent's home.

Ability variables include IQ (as recoded at Ohio State University), a dummy variable, DIQ, for those respondents for whom the IQ score was missing, and KWW, the score on a Knowledge of the World of Work test administered by the interviewers in 1966. This last test score is available

for almost all of the sample. It is not an "intelligence test" but, rather, an "occupational information test":

> Our measure of "knowledge of world of work" is a very limited one, consisting of three components. The first of these involves occupational identification. Respondents were asked to select one of three statements that best describes the duties of each of ten occupations—hospital orderly, machinist, acetylene welder, stationary engineer, statistical clerk, fork lift operator, economist, medical illustrator, draftsman, and social worker. The second component involves the typical educational attainment of men in each of these same ten occupations: "How much regular schooling do you think hospital orderlies usually have?" Third, respondents were asked, for each of eight pairs of occupations, which one provides the highest average annual earnings: "Who do you think earns more in a year, a man who is an automobile mechanic or an electrician?" Standards for scoring the second and third components were derived from 1960 census data on occupation by highest year of school achieved and median earnings by occupation.[8]

It is a measure that should reflect both the quantity and quality of schooling, intelligence, and motivation (curiosity about the outside world). The resulting score (based on all three components and running from 0 to 56) is closely related to schooling, race, region, and city size. Considering that the test is very short (only two pages in the original questionnaire) and nothing like a standard IQ test, it was rather surprising to find that it seems to perform rather similarly (and parallel) to the IQ variable. We also use occasionally a set of current situation (CS) variables consisting of dummy variables for marital (MRT), union, and health status of the respondent.

More variables are available on the original survey tapes, and many more could be constructed from them. We did experiment with some other definitions and versions of some of the above variables, but this list should provide an adequate description of the type of data available to us.

There are at least two serious problems in using these data: missing data and the youngness of this sample. As noted before, IQ scores are available only for about 65 percent of the sample, and more of them are missing, relatively, for blacks than for whites. In addition, questions on parental education and occupation were not answered by about 20 percent of the respondents. There are also a variety of other missing and unreasonable entries. We have dealt with the missing data problem in

[8] U.S. Department of Labor 1970–74, 1:120.

two ways. For the major variables of interest, such as wages, schooling, and IQ, we restricted the sample only to those for whom complete and reasonable data were available. For variables of subsidiary interest, such as parental background, we imputed a mean value to the missing observations and added dummy variables corresponding to each set of missing observations for each of the independent variables.[9]

The youngness of the sample requires much more careful attention to the experience variable, since there is much on-the-job training and search at this age and, hence, the wages of different youths of the same age cannot be taken as reflecting the same amount of human capital if they differ in their labor market experience. In this study, we construct the experience variable directly from the work history of the individual; hence, it is not tautologically equal to age minus schooling.[10] Since it is likely that on-the-job training declines with time, the functional form chosen to represent this variable (borrowed from Mincer [1974]) reflects this. I have also used a more general functional form (cubic) for the experience variable, but because later on we will be examining the possibility that schooling and hence also experience should be treated as endogenous variables in a more complete achievement model, I will concentrate attention on a single summary measure of it (XBT).

IV. The Major Results

Table 2 summarizes the major results of applying standard least-squares procedures to the data in the two subsamples. There are a number of findings worth noting:

1. The treatment of experience matters. The schooling coefficients are significantly higher and the fit is somewhat better when the experience variables are used instead of age (compare eqq. [A1] and [A3] or [A2] and [A4] in table 2), supporting Mincer's (1974) position on this.[11] The absolute value of estimated "ability bias" (using KWW) is about the same for either comparison, about .012, but the implied relative (percentage) "biases" differ greatly. Holding age constant results in an estimate of relative ability bias of 35 percent. Holding experience constant, the same

---

[9] This corresponds to a procedure that estimates jointly the mean missing value for each of these variables. This procedure, however, is neither consistent nor fully efficient. Because of our special interest in the IQ variable, we also tried a more complete treatment of the missing-value problem, using a regression of observed IQ scores on all the relevant variables (including schooling and other tests) to estimate the missing values individually. The results were not strikingly different and are not reproduced here (see Griliches [1974] for details). See Dagenais (1973) for a discussion of the econometrics of such "extrapolation" procedures and references to the literature on missing observations.

[10] Age and schooling account for only 80 percent of the variance in the XBT variable.

[11] Age is not statistically "significant" when added on top of the experience variables. This differs from the results reported in the earlier version of this paper, which were based on a cruder measure of experience.

TABLE 2

SCHOOLING AND ABILITY COEFFICIENTS IN LOG WAGE EQUATIONS:
DEPENDENT VARIABLE LW69

EQUATION NUMBER	COEFFICIENTS (AND t-RATIOS) OF:			OTHER VARIABLES IN EQUATIONS	R^2 (SEE)
	SC69	KWW	IQ		
A: $N = 2,062$:					
(A1)	0.034 (10.6)	...	...	CL, age	.356 (0.342)
(A2)	0.022 (6.1)	0.0070 (6.1)	...	CL, age	.367 (0.339)
(A3)	0.063 (18.9)	...	...	CL, cubic in EXP69, AFEX	.383 (0.335)
(A4)	0.050 (12.2)	0.0059 (5.3)	...	CL, cubic in EXP69, AFEX	.391 (0.333)
(A5)	0.051 (12.5)	0.0060 (5.3)	...	CL, XBT, AFEX	.385 (0.334)
(A6)	0.059 (15.5)	...	0.0017 (2.6)	CL, XBT, AFEX, DIQ	.380 (0.336)
(A7)	0.062 (17.0)	...	...	CL, XBT, AFEX, CS	.438 (0.320)
(A8)	0.056 (15.3)	...	0.0020 (3.3)	CL, XBT, AFEX, CS	.442 (0.319)
(A9)	0.056 (13.6)	...	0.0016 (2.4)	CL, XBT, AFEX, DIQ, BKG	.382 (0.336)
B: $N = 1,362$:					
(B1)	0.065 (13.2)	...	...	CL, XBT, AFEX	.309 (0.332)
(B2)	0.053 (9.1)	0.0054 (3.8)	...	CL, XBT, AFEX	.316 (0.330)
(B3)	0.059 (10.7)	...	0.0019 (2.8)	CL, XBT, AFEX	.313 (0.331)
(B4)	0.058 (11.0)	...	0.0026 (3.2)	CL, XBT, AFEX, black, CS	.382 (0.315)
(B5)	0.053 (9.5)	...	0.0020 (2.8)	CL, XBT, AFEX, CS, BKG	.386 (0.314)

NOTE.—Variables and variable sets: SC69 = schooling completed in 1969 (in years); KWW = score on test of "knowledge of world of work"; IQ = score on an IQ-type test; DIQ = dummy variables for missing IQ score; XBT = $e^{-0.1 \cdot \text{EXP69}}$ (EXP69 = cumulated work experience in 1969, in years); AFEX = service in the armed forces (in years); DIQ = IQ missing; CL = current location: RNS, BRNS, SMSA (RNS, region now South; BRNS, black in South now; SMSA, in SMSA in 1969); CS = current situation: health 68, union, MRT (health 68, health impaired in 1968; union, member of a union in 1969; MRT, married); BKG = background: black, siblings, culture, MED, DME, FOMY14, DFO14 (siblings, number of siblings; culture, index of home culture [library card, magazines, etc.]; MED, mother's education; DME, MED missing; FOMY14, father's occupation when R14, scaled by the 1959 media earnings of males in the particular occupation; DFO14, FOMY14 missing); SEE = standard error of estimate (estimated standard deviation of the residuals).

absolute bias is seen to be only a 21 percent relative bias. Parameterization matters.

2. The single exponential measure of experience, XBT, appears to provide an adequate approximation to more complex functional forms (compare [A4] and [A5]).

3. As expected, the implied ability bias is higher when KWW is used instead of IQ, .012 versus .006 ([A3] and [A4] or [B1] and [B2] vs. [B1]

and [B3]). IQ-based estimates of the "ability bias" are only on the order of 10 percent (holding experience constant). These estimates are robust to the inclusion of additional variables, such as marital or union status of the respondent.

4. The estimated schooling coefficients move between about .05 and .06, somewhat lower than the .06–.08 reported by Mincer (1974) and others but not out of line with other estimates limited to younger ages and the earlier portion of the age-earnings profile. They are statistically very significant and quite robust to the introduction or deletion of other variables (except to the choice of age vs. experience).

5. The contribution of the ability measures (IQ or KWW) to the fit of the various equations is miniscule. Their introduction leads to some realignment in the estimated coefficients of schooling, but rarely does it lead to a substantive reduction in the unexplained variation of wage rates (though given our large samples, the actual reduction is often "statistically" significant). The standard deviation of the residuals is rarely changed by more than in the third decimal place, and their variance is reduced by less than 2 percent. Whatever their merit in reinterpreting the role of schooling, the available ability measures do not noticeably improve our ability to explain the observed dispersion in wage rates or earnings, at least in this age group.[12]

6. Family background variables are not statistically significant as a group when added on top of the schooling and ability variables. (Compare the estimated standard errors in [A6] and [A7] or [B4] and [B5].) This is consistent with our assumption that the background variables work primarily via schooling and measured ability and have no substantive direct contribution of their own in the wage equation.

7. There is little difference in the results between the larger sample and the "real IQ only" subset. Apparently the selectivity of the missing IQ values (concentrated among low-schooling individuals and blacks) does not bias our results seriously.

Table 3 presents all the coefficients of one of the better fitting equations (B4) and illustrates some of the additional results of this study. Note that the three most "significant" variables are schooling, experience, and union membership. The first years of experience appear to be quite valuable, roughly on the order of the contribution of a year of schooling, starting at .06 per year at zero experience and dropping slowly to .03 after 7 years of experience and to only .015 per additional year after 14 years of experience.

While the number of those with military service is relatively small in this sample (about 18 percent among those not enrolled in school) and the estimated effect therefore not very precise, it appears that a year of mil-

---

[12] See Griliches (1974) for parallel results on earnings.

TABLE 3

LOG WAGE 69 EQUATION ($N = 1,362$)

Variable	Coefficients	t-Ratios
SC69	0.058	11.0
IQ	0.0023	3.2
XBT	−0.617	10.2
AFEX	0.016	3.5
RNS	−0.061	2.7
SMSA	0.106	5.7
BRNS	−0.181	4.0
Black	0.026	0.8
Health 68	−0.057	2.0
Union	0.203	10.2
MRT	0.118	5.7

NOTE.—$R^2 = .382$; SEE = 0.315.

itary service has a significantly smaller effect on wages (1.6 percent) than a year of work experience in the civilian sector. This differs from but is not greatly inconsistent with the results of Griliches and Mason (1972), who found no effect of differences in the length of military service on the income of veterans in a sample where everybody had served in the armed forces.

Black-white wage and income differentials are not a major focus of our analysis. They are dealt with at some length in the current work of Freeman (1974), and I shall not poach extensively on his territory. The following major facts do emerge though from our analysis: with slightly over a quarter (27 percent) of our sample black, we cannot really detect a statistically significant wage differential except for blacks still located in the South. Holding schooling completed constant, we find no evidence of any significant wage discrimination against young black males outside of the South.[13] In the South there persists about a 15 percent wage differential in favor of whites.

In general, there are rather strong regional and city size effects in the sample. Being in a metropolitan area adds about 11 percent to one's wage rate. Being in the South currently subtracts about 6 percent, while not having been healthy in 1968 leads to a 6 percent lower wage rate and up to 20 percent lower earnings in 1969. One of the "strongest" variables in these equations (in terms of t-ratios, β coefficients, or partial correlations) is being married, which is associated with both a higher wage rate and expanded labor force participation. Married men have about a 10 percent higher wage rate per hour and earn about a third more per year

[13] These estimates are based on equations estimated across the whole sample. When estimated separately for blacks and nonblacks, the resulting wage and income equations appear quite similar except that blacks seem to gain more from formal schooling and less from work experience or the sheer passage of time (age) than whites do in this same age range.

than unmarried men of similar background. Another "strong" variable is belonging to a union, which raises wage rates by about 20 percent.

Besides the variables described here, several more were tried in various forms. In particular, we divided schooling into categories and interacted it with the IQ measure with no discernible effect on the fit of the equation. While one might expect that there would be a positive interaction between schooling and IQ, we could not detect it in our data.

V. Reduced-Form Estimates

In the debates about the relative role of schooling and ability in determining the ultimate economic success of an individual, the net contribution of ability (holding schooling constant) is often confused with its total contribution (including that via schooling). In econometric terminology, sometimes one is interested in the parameters of the "structural" equation while at other times one is interested in those of the "reduced form," the equation in which all of the other endogenous (casually subsequent) variables have been solved out to show the total contribution of the remaining exogenous variables.

Table 4 presents a number of estimates of such reduced- (and semi-reduced-) form equations for the major variables of interest and for several subsets of our sample. The first three equations deal with schooling. Since a large fraction of our population have not yet finished their schooling, we introduce a new variable—expected ultimate level of schooling (EXSC) and compare it with actual schooling attained in 1969. By and large, the results are similar: parental background, region of origin, and IQ account for about a third of the variance of schooling. Mother's education appears to be a somewhat stronger variable than father's occupation but not by a great margin. Later ability, as measured by the KWW test, is affected about equally by schooling and by early ability. Number of siblings has a significant negative effect on late ability and on achieved schooling. For given parental status and region of origin, black youngsters do score lower on both early (IQ) and later (KWW) tests, on the order of two-thirds and one-third of a standard deviation for IQ and KWW (holding IQ scores constant), respectively. In spite of this, they have a higher schooling attainment (on the order of half a year) than white young men with equal parental background *and* test scores.

With regard to the estimated log wage equations, several important points emerge: (1) There is a sizable net effect of schooling on wages and income net of family background and measured IQ. (2) This effect is significantly larger, by 50 percent or so, when we adjust for the lower work experience of those with more schooling. (3) There are significant though small negative effects of being black on wage rates (net of family background, IQ, and schooling), particularly for those of southern origin

TABLE 4

REDUCED- AND SEMIREDUCED-FORM EQUATION ESTIMATES

DEPENDENT VARIABLE (AND N)	COEFFICIENTS (AND t-RATIOS) OF:								OTHER VARIABLES IN EQUATION	R^2 (SEE)
	Culture	MED	FOMY14	SIB	Black	IQ	SC66 OR SC69	XBT		
1. EXSC .... $N = 3,025$	0.42 (8.0)	0.115 (8.1)	0.149 (7.1)	−0.08 (4.5)	1.22 (3.9)	0.065 (23.7)	...	...	ROS, DMED, DFO14, age, date	.332 (2.0)
2. EXSC .... $N = 1,362$	0.31 (4.4)	0.138 (6.6)	0.170 (5.0)	−0.06 (2.3)	1.09 (4.6)	0.054 (13.4)	...	...	RB, date, DMED DFO14	.292 (1.9)
3. SC69 .... $N = 1,362$	0.14 (2.5)	0.116 (7.0)	0.142 (5.2)	−0.07 (3.3)	0.43 (2.3)	0.049 (15.2)	...	...	RB, age, DMED, DFO14	.365 (1.5)
4. KWW .... $N = 1,362$	1.06 (5.1)	0.091 (1.4)	0.096 (0.9)	−0.22 (3.0)	−1.7 (2.7)	0.135 (10.3)	0.101* (8.0)	...	RN, age, age squared, DMED, DFO14	.465 (5.8)
5. LW69 .... $N = 2,062$	0.03 (2.9)	0.006 (2.1)	0.015 (2.9)	−0.006 (1.9)	−0.145 (5.0)	0.0017 (2.5)	...	...	RB, age, DMED, DFO14, DIQ	.295 (0.358)
6. LW69 .... $N = 2,062$	0.01 (1.4)	0.002 (0.7)	0.010 (1.8)	−0.003 (1.0)	−0.126 (4.4)	0.0004 (0.6)	0.031 (7.0)	...	RB, age, DMED, DFO14, DIQ	.312 (0.354)
7. LW69 .... $N = 2,062$	0.01 (0.8)	0.002 (0.7)	0.013 (2.6)	−0.002 (0.6)	−0.078 (2.7)	0.0008 (1.2)	0.064 (11.0)	−0.79 (8.4)	RB, age, DMED, DFO14, DIQ AFEX	.337 (0.348)
8. LW69 .... $N = 1,362$	0.01 (0.8)	0.004 (1.7)	0.008 (1.3)	−0.005 (1.2)	−0.010 (0.2)	0.0024 (3.4)	...	...	RB, age, DMED, DFO14	.272 (0.342)
9. LW69 .... $N = 1,362$	0.01 (0.7)	0.003 (0.8)	0.007 (1.1)	−0.004 (1.0)	−0.013 (0.3)	0.0021 (2.5)	...	...	RB, age, DMED, DFO14	.273 (0.341)
10. LW69 .... $N = 1,362$	0.01 (0.5)	0.004 (1.0)	0.009 (1.5)	−0.003 (0.8)	0.010 (0.2)	0.0019 (2.5)	0.045 (5.3)	−0.71 (6.0)	RB, age, DMED, DFO14, AFEX	.293 (0.337)

NOTE.—EXSC = expected schooling to be completed ultimately (as of 1969 or 1966); date = dummy variable if EXSC is as of 1966; RB = region before: ROS, BROS, and POC (ROS, region South when 14; BROS, black and ROS; POC, in large city or suburb of large city when 14); and NE = not enrolled in school in 1969.
* SC66; others are SC69.

who stayed in the South and especially in the larger, economically poorer sample.[14] (4) Parental background variables have little direct effect on wages once schooling and IQ are allowed for.

One way to look at these results is to consider what they imply for two youngsters who are 1 standard deviation apart on each of the listed family background variables and IQ.[15] Equations (1) and (5) would predict that they would find themselves, other things equal, about 0.75 and 0.25 of a standard deviation apart on schooling and wages, respectively, imply-ing a rather strong regression toward the mean. Now, if a youngster with lower family background and IQ managed somehow to acquire 4 extra years of schooling (e.g., went on to and completed college), which would equal an additional 1.5 standard deviation units of schooling, it would more than wipe out his original handicap (using the schooling coefficients of eq. [6]).[16] If, in fact, he had an equal IQ to start with, he would need only 2 more years of schooling to compensate him for his lower social class start.

Thus, while schooling by itself does not appear to be a variable that accounts for a great deal of the observed variance in wages or income, the model does predict that in the United States in the 1960s additional schooling could be used to overcome social class handicaps. Compen-satory education would not eliminate much of the observed income inequality, the residual standard errors are on the order of a third to a half in the wage and earnings equations, respectively, and more than half of the observed variance is left unaccounted for by our variables, but it could be used to compensate for and eliminate some of the systematic sources of the observed differences in wages and earnings.

[14] They are larger for earnings than for wages, but much of that is due to the lower work experience of blacks. This disagrees with Adams and Nestel (1973), who do not find a "Southern rural origin" effect for blacks but do find a negative effect of having grown up black in a non-South city. We tried such southern rural and northern city origin dummies but found them not particularly statistically significant though they did occasionally have sizable coefficients. In the various versions of our model we do get, from time to time, sizable negative estimated effects of having grown up (residence at 14) black in the rural South but very little consistent additional effect of having done so in a northern city. The difference in results may be due to our use of a later year, a longer list of other included variables, and our control for experience. There is some indication in our data that young black men of northern city origin worked 5–7 percent less in 1968–69 than those of southern rural origin. Since Adams and Nestel do not control for migration status, their major effect might be interpreted as a premium on having moved from the South to the metropolitan North and could be related to the selectivity of migration.

[15] Since the various family background and IQ scores are not highly intercorrelated, being 1 standard deviation apart on each of these measures implies that they are quite a bit farther apart on a more inclusive measure of the distribution of social and genetic inheritance.

[16] Since he would be predicted to have about 2 years less schooling than his more fortunate counterpart, he would need 2 years more than was predicted for him to achieve equal schooling and 2 additional years to wipe out the remaining family background and IQ deficits.

VI. Errors in Variables and the Endogeneity of Schooling[17]

One of the objections that can be raised against the results reported above is the use of rather poor test scores as a proxy for true ability. Both the IQ and KWW measures may be subject to significant errors, which could bias their coefficients downward and the schooling coefficient upward. Since we have two test scores, we can use one of them as an instrument for the other (assuming that their errors are independent of each other) together with the parental background variables to estimate the "true" ability coefficient by instrumental variables (TSLS) methods. Such estimates are presented in rows 1, 3, and 5 of table 5. They indicate that allowing for errors in the test measures raises their coefficients significantly while reducing the schooling coefficient only by a bit more (about another 10 percent from a level of approximately .06). The results of equations (3) and (5) compared with table 2, equations (B2) and (B3), imply that about half of the independent variance of IQ (or about 30 percent of the total IQ variance) and almost 60 percent of the independent variance of KWW (or 35 percent of the total) are due to errors of measurement.

Given such high estimates of error variances in the tests, one is hard pressed to maintain the assumption that such "errors" are uncorrelated with schooling or with each other. These errors are computed by finding that part of the test score variance that is correlated jointly with the other test, parental background, schooling, and wages. The rest is then labeled "error." If the rest is large, it is quite likely that it may be related to success in school (e.g., test wiseness) even if it does not contribute directly to earnings. This would imply that schooling cannot itself be treated as independent of the disturbance in the wage equation once an erroneous test measure is introduced instead of the "true" underlying ability dimension. It too must be treated as endogenous.

Moreover, since the schooling decision itself is made, at least in part, in anticipation of economic returns, the piece of the disturbance in the wage equation which is anticipated by the individual (even though unobservable to the researcher) will affect the outcome of this decision. This may result in a correlation between the unobserved variables in the wage equation (the disturbance) and completed schooling, suggesting again the use of simultaneous-equation methods in estimating the schooling coefficient.[18]

---

[17] See Griliches (1975b) and Chamberlain (1976) for more detailed discussion of the issues raised in this section and for additional empirical results.

[18] There are three reasons why schooling might be endogenous: (1) errors of measurement, (2) correlation between the disturbances in the income and schooling equations (since schooling is optimized with respect to expected income), and (3) test wiseness: the presence of another "ability" component in the test scores which is correlated across tests and with schooling, inducing a correlation between the "errors" in these equations. The procedure discussed in the text and the results presented in table 5 deal more or less adequately with reasons (1) and (2). Reason (3) would prevent us from using one of

TABLE 5

Log Wage 69 Equation: tsls Estimates

	SC69	KWW	IQ	Other Variables in Equations and Additional Instruments	SEE
A: $N = 2,062$:					
1. KWW endogenous ...	0.046 (6.4)	0.0086 (2.9)	...	CL, XBT, AFEX; instruments: IQ, DIQ, BKG, DIQ · BKG, age, age squared	0.335
2. KWW, SC69, and XBT endogenous ...	0.076 (5.7)	0.0044 (0.1)	...	Same as above	0.339
B: $N = 1,362$:					
3. IQ endogenous.	0.052 (7.0)	...	0.0038 (2.4)	CL, XBT, AFEX; instruments: KWW, BKG, age, age squared	0.332
4. IQ, SC69, XBT endogenous	0.096 (5.7)	...	−0.0016 (0.6)	CL, XBT, AFEX; instruments: BKG, age, age squared	0.339
5. KWW endogenous ...	0.036 (3.8)	0.0130 (3.6)	...	CL, XBT, AFEX: instruments: BKG, IQ, age, age squared	0.334
6. KWW, SC69, XBT endogenous	0.078 (2.7)	0.0027 (0.4)	...	CL, XBT, AFEX; instruments: IQ, BKG, age, age squared	0.333

Such two-stage least-squares estimates of the wage equation, treating both schooling and experience (XBT) as endogenous, in addition to the test variables, are presented in table 5: rows 2, 4, and 6. Allowing for the endogeneity of schooling raises its coefficient significantly above the original starting point. It appears that the downward bias due to the endogeneity of (and possible errors in) schooling is larger than the originally feared upward bias due to the disregard of the role of ability.[19] In any case, there is no evidence of a "net" ability bias when the estimation method treats schooling and experience symmetrically with test scores.

the tests as an instrument for the other. The model is still identified but not very precisely. Chamberlain and I are analyzing a more general two-factor model, encompassing all three reasons, using data on brothers from the same surveys. Adding information on family structure allows one to relax and test some of the more dubious assumptions (such as no correlation of the test errors) imposed in this paper.

[19] See Griliches (1975b) for a more detailed discussion of these results. A downward bias might arise not only from errors of measurement but also if "ability" is taken to reflect the initial level of human capital. Then it is easy to show that the optimized level of schooling will be related negatively to such a measure or component of ability.

Appendix on the Literature

The studies most comparable to the results of Sections III and IV of this paper are those of Kohen (1973), Bulcock, Fagerlind, and Emanuelsson (1974), and Sewell and Hauser (1974). Kohen estimates a similar model based on the 1966 data for these same young men. His results are not too different, though substantially weaker. In his equations, IQ contributes a bit more and schooling a bit less than in ours, but schooling still outperforms IQ by a significant margin. The main drawback, in our view, of his study is the omission of the very important age-experience variables and the elimination of the lower quarter of the sample due to missing IQ scores in that portion. Also, we are observing the same individuals 3 years later, giving them a bit more time to mature and find their way in the labor marketplace.

Another important study of the labor market success of young men is that of Sewell and Hauser and their associates (1974) in which the 1957 cohort of Wisconsin high school seniors has been followed for 10 years. The results of their studies are too rich to summarize here except to note that they put more emphasis on the effects of parental income on the subsequent success of sons and that they get higher effects for ability and lower effects for schooling than we do. The main drawback of their study is the restricted range of their sample. By focusing on one cohort of high school seniors, they cut out a third or more of the total schooling distribution and reduce the estimated effect of schooling thereby. Also, they do not allow for the differential experience of individuals in the labor market and underestimate thereby the ultimate effect of schooling on income.[20] As a result of this restriction of range and the omission of experience and current location variables, the overall fit of their models is quite poor (the R^2s for the income equations are on the order of .05–.07 vs. .4–.6 for our samples and models).

Bulcock et al. (1974) have been reanalyzing the extended and updated Malmo (Sweden) data set which had been previously analyzed by Husen (1969), Griliches (1970), Hause (1972), and de Wolff and Slijpe (1973), among others. The data set originates with a 1938 sample of 10-year-olds in the city of Malmo. These have been followed up through 1972. Data are available on childhood IQ, parental background, adult IQ (army test scores), schooling completed, and occupation and income in 1971, for approximately 500–700 males. The main difference between their study and ours (and the Sewell-Hauser one) is that the Malmo respondents were significantly older in 1971 (about 43 years old). This leads to a much higher estimated effect of "occupation" on income than was found either by us or by Sewell and Hauser. They find a larger effect of "late ability" measures on current income than we do, but they too do not allow for differences in the length of work experience, which tends to result in an underestimate of the net effect of schooling. In spite of this, their estimated total effect of schooling, both via occupation and the adult ability measure, is quite high.

Our study is the first, however, to deal seriously with the problem of errors, of both measurement and concept, in the available measures of ability and schooling.

[20] Given that they are following a single age cohort, they cannot really allow separately for the effects of experience, since in their sample it is almost perfectly negatively colinear with schooling.