223-74

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

BIASES IN MEASUREMENT OF THE PRODUCTIVITY BENEFITS OF HUMAN CAPITAL INVESTMENTS

John H. Bishop

UNIVERSITY OF WISCONSIN ~ MADISON

BIASES IN MEASUREMENT OF THE PRODUCTIVITY BENEFITS OF HUMAN CAPITAL INVESTMENTS

JOHN H. BISHOP

September 1974

Some of the ideas incorporated in this paper are a consequence of sitting in on two econometrics courses taught by Arthur Goldberger. The author would like to express gratitude to Larry Suiter of the Census Bureau for making available early tabulations of the 1970 CPS-Census match and to Arthur Goldberger and Barbara Zoloth for commenting on early versions of this paper. The author is Research Associate at the Institute for Research on Poverty, University of Wisconsin. The research was supported by funds granted to the Institute by the Office of Economic Opportunity, pursuant to the provisions of the Economic Opportunity Act of 1964 and by funds from Grant #51-55-73-04, Manpower Administration, Department of Labor under provisions of Social Security Act 81 Stat. 888 and by funds from Ford Foundation Grant #700-0309. The opinions expressed and the errors that remain are solely the responsibility of the author.

ABSTRACT

Two important sources of downward bias in measures of the returns to education are examined: (1) errors in measuring years of educational attainment and (2) discrepancies between reported earnings and the concept appropriate for analysis of social policy, marginal value product. The recently completed CPS-Census match study has found that schooling reports were correlated only .887. This plus methodological errors implies that the corrections for unreliability and therefore the estimated true effect of education used in Jencks, et. al.'s Inequality for their models of status attainment are low by about 9 percent.

Using an approach suggested by Theil, an estimator of the errors in variables bias in income education relationships is derived. Using the 1970 Census-CPS match data, the errors in variables regression bias in Census and CPS data (estimated coefficient over true) is estimated to lie between .85 and .94 depending on degree to which errors in predicting income are positively correlated with errors in measuring schooling. Biases appearing in Census income education tabulations are also examined. The nonrandom character of reporting error causes census tabulations to exaggerate the benefits of the first 2 years of high school and to underestimate the benefits of high school graduation and the first 2 years of college.

When adjustments for the incompleteness of reported money earnings are made 1970 Census social productivity benefits of schooling turn out to be underestimated by 6 percent. Since the value of student time is underestimated even more, coverage bias does not significantly effect private or conventionally calculated social rates of return to years of schooling. Coverage

bias is an important bias, however, for evaluations of health programs and school quality changes and for returns to years of schooling when student's time costs are foregone leisure.

BIASES IN MEASUREMENT OF THE PRODUCTIVITY BENEFITS OF HUMAN CAPITAL INVESTMENTS

Much of the recent work estimating returns to education has been a search for the appropriate downward adjustment of the gross effect of education for upward bias introduced by lack of controls for ability and family background.[1] In the process two important sources of downward bias have been neglected: (1) errors in measuring years of educational attainment and (2) discrepancies between reported money earnings and the theoretical concept we would like to measure, the marginal value product. The nature of these biases will be discussed, their size estimated for the 1960 and 1970 Census and for the yearly Current Population Survey, and rules-of-thumb suggested for correcting rate-of-return and benefit-cost ratios calculated from reported earnings.

The first source of downward bias is simply the familiar errors in variables problem appearing in the income by education contingency tables. Some people who report themselves at the top of the education scale actually have less. Some who report themselves at the bottom actually have more. Earnings differences between people grouped by reported education typically understate the true differential. This produces a bias in both internal rates of return and benefit-cost ratios. Data generated by the reinterview program of the 1950, 1960, and 1970 Census will be used to estimate the size of the bias that results. In this discussion it is shown that the conventional assumption that reporting errors are uncorrelated with the true level is not valid for education, and a way of obtaining estimates of the bias is suggested.

The other source of downward bias also results from measurement difficulties. The reported earnings in decennial census publications do not include fringe benefits, employer social security tax payments and excise taxation. The size of these discrepancies are estimated by comparing the aggregates implied by the census and CPS with adjusted National Income Account aggregates and by applying the appropriate tax rates.

Not all sources of potential bias will be discussed here. We will neglect education's impact on the value of or loss of leisure time[2] as well as possible discrepancies between social productivity and observed income differentials due to screening, queuing for jobs, and labor market restrictions.

I. Errors-in-Variables Bias: Errors in Measuring Education

People report education more accurately than they report other socio-economic status measures such as occupation and income. Nevertheless, when reinterviewed, Census respondents reported a different number of years almost 40 percent of the time.[3] When expressed as a scale, two separate reports of years of schooling for the same person had a correlation of .88, .915 and .86 in the 1970, 1960 and 1950 Censuses respectively.[4]

The Census obtains data on the response variance by returning to a sample of those enumerated in the Census and asking the questions over again. Different techniques have been used in each study. The 1950 and 1960 Post Enumeration Surveys (PES) attempted to obtain more accurate answers (1) by using a more detailed questionnaire with extensive filtering, (2) by obtaining wherever possible answers about an adult from the adult himself, and (3) by asking the respondent to clear up discrepancies between the Census and PES answer when they appeared. While lowering the total error variance of the follow-up survey, this last characteristic of the PES tends to increase the positive correlation between errors in the two surveys.[5]

Two other studies conducted in 1960 used techniques that closely paralleled the Census. The same questionnaire was used and any responsible member of the household could answer for the others. The time lag was only 2 or 3 weeks. In contrast both the PES follow-ups were 5 months later. In the Reinterview Study information on 5000 households was obtained by personal interview. In the Requestionnaire Study the information for 1000 households was obtained by mailing out the questionnaire, asking that it be mailed back, and following up nonresponses and internal inconsistencies--a procedure identical to the one used in the Census. The 1970 study matched approximately 16,000 final edited records from the 20 percent questionnaire of the Census with corresponding records of the March Current Population Survey (CPS).[6]

A. Bias in Standardized Regression Coefficients

Much of the work on correcting education coefficients for errors in measurement has been designed to improve status attainment modeling where standardized regression or path coefficients have been the main interest.[7] As a consequence, errors in variables modeling has customarily used a correlation approach.

The observed correlation between two imperfect measures of education, r'_{cp}, has the following relationship with correlations between the imperfect reports and the true level of education:[8]

$$(1) \qquad r'_{cp} = r_{tc} r_{tp} + r_{ucp} \sqrt{\left(1 - r_{tc}^2\right) \left(1 - r_{tp}^2\right)}$$

where r_{tc} is the correlation of the census report with true level;

r_{tp} is the correlation of the CPS report with true level;

r_{ucp} is the correlation of the errors in census and CPS report;

r'_{cp} is the observed correlation between the two reports. It has been estimated to be .887 and .875 for males females respectively in the 1970 CPS-Census Match and .9149 from the 1960 reinterview-census match.

The Occupational Change in a Generation study conducted by the CPS is the primary data base used by Jencks, et al. and Bowles in their modeling of the status attainment process with estimated "true" correlation matrices.[9] They were, therefore, interested in the accuracy of the CPS education report. Both Jencks and Bowles have assumed the CPS was analogous to the PES and that it and the PES education report were substantially more accurate than the ensus report. As a consequence they obtained rather high correlations, .98 and .958 respectively between the OCG education report and the true level.

In fact, however, the CPS and PES education responses were obtained in entirely different manners. The CPS uses the same question as the Census. The PES uses a longer and much better designed question. In two-thirds of the 1960 PES households reconciliation of discrepancies was attempted. No such reconciliation has been attempted between CPS and other education reports. The CPS obtained its answers from "any responsible adult" which in most cases is the wife, since families are contacted during the day. The PES obtained answers about an adult from the adult himself, wherever possible. As a result, the CPS is substantially less accurate than the PES.

For male educational attainment the CPS may even be less accurate than the Census. All the population in 1970 and 82 percent in 1960 received their census questionnaires in the mail and were asked to fill them out on their own time. As a consequence, there was an opportunity for each adult to fill out his own part of the questionnaire. Furthermore, "a respondent in a personal interview situation (as in CPS) may be more likely to erroneously report education at a terminal category than is true when the person is actually completing a questionnaire."[10] For education, these advantages seem to outweigh the disadvantages of massive scale and inexperienced interviewers and supervisors that are reported to lower the reliability of other census data.

Comparison of the March 1970, CPS with the 1970 decennial census tabulations is consistent with this view. The CPS shows more people (especially men) than does the Census having completed terminal grades (8, 12, and 16) of a given level of schooling.[11] This heaping suggests that guesses are being made by one respondent about some other family member's attainment. Furthermore, income differences (especially those of men) between educational levels were greater in the Census.[12] This is exactly what one would expect if reporting errors for education were greater in the CPS.

The lower accuracy of the CPS is further supported by the fact that it correlates with the 1970 Census only .887 while the 1960 PES and Census had a correlation of .934 and the 1960 Reinterview and Census correlated .915.

If the best is used--the newly available CPS-Census--and the CPS and census are assumed equally accurate, and Jencks' assumption that $r_{ucp} = 0$ is adopted, the correlation between true and CPS report is .942, not the .98 Jencks used. With Bowles' more reasonable assumption that errors in reporting education are correlated .5, the correlation between the CPS report and true level is .880. Our preferred assumption for r_{ucp} is .4. This implies the true level and CPS reports are correlated .90. Jencks' et al. use of .98 to correct estimates of paths to and from education results in their understating the partial effects of education by about 9 percent and the size of indirect paths through education by about 19 percent.

B. <u>Bias in Unstandardized Regression Coefficients</u>

We shall attempt to characterize the bias that occurs in regression estimates of the income-education relationship by first positing a true model, then getting expressions for the coefficients obtained when education is measured with error and then applying the information available from follow-up studies to produce estimates of the bias.

Define the following terms,

Y = reported income

E = true level of education

C = census report of education

P = the CPS or follow-up report of education

Let the true model be:

(2) $Y = \beta_0 + \beta E + \varepsilon$

(3) $P = \alpha_0 + \alpha E + u$

(4) $C = \lambda_0 + \lambda E + v$

E, ε, u and v are independent. The mean of all variables and errors is zero. Except for the fact that α and λ are not necessarily 1, this is the traditional errors in variables model. When the CPS report is regressed on true schooling, the coefficient on true schooling, E, is α. Both α and λ are expected to be less than one because education is a scale with upper and lower bounds. An error by those truely in the bottom category of the scale can only be positive and errors by those at the top are necessarily negative. Another way of writing 3 and 4 is

(5) $(\alpha-1)E + u = P-E-\alpha_o$ = CPS report discrepancy

(6) $(\lambda-1)E + v = C-E-\lambda_o$ = Census report discrepancy

The difference between the reported and the true level of schooling is negatively related to the level of education.

Since E is unobservable, our data consists of sample estimates, $\hat{b}$ and $\hat{\gamma}$, from the regressions:

(7) $Y=b_2 + \hat{b}_p P$ and $Y=b_1 + \hat{b}_c C$

(8) $C=a_2 + \hat{\gamma}_{cp} P$ and $P=a_1 + \hat{\gamma}_{pc} C$

Since equation (7) represents the manner in which the income-education relationship has been estimated in the literature, we are interested in how $\hat{b}_p$ and $\hat{b}_c$ relate to β. This will give us an indication of the nature and degree of bias in traditional estimates of the return to education.

From the assumptions of our model equations (2),(3), and (7) yield:

$$(9)\quad \text{plim}\ \hat{b}_p = b_p = \frac{C(Y,P)}{V(P)} \doteq \frac{\beta\alpha\ V(E)}{V(P)} = \frac{\beta\alpha\ V(E)}{\alpha^2\ V(E) + V(u)}$$

where C(Y,P), V(P), V(E), etc are population variances and covariances. The sample variances and covariances from the follow-up studies are consistent estimators of the population variances and covariances of the observable variables.

A conventional way of estimating $\frac{b}{\beta}$ is to use (1) to estimate r_{tp}^2, the squared the correlation between the CPS report and the true level, and then assume $\alpha = 1$ so that $r_{tp}^2 = \frac{V(E)}{V(P)}$ To obtain r_{tp} from the single parameter r'_{cp} in the manner Jencks and Bowles have, it was necessary to make an assumption about the relative accuracy of the two measures of education C and P, and about r_{ucp}. According to (9), using r_{tp}^2 as an estimate of b/β involves the further assumption that $\alpha = 1$ (ie that there are no boundary effects). This last assumption is unneeded for correcting path coefficients but is required if corrections of unstandardized regression coefficients are to be calculated in this manner.

An alternative approach exists that can produce unique estimates of the bias in unstandardized regression coefficients and that makes fuller use of the follow-up sample data while imposing less a priori structure on the model. According to Theil if ϕ_p is defined as the coefficient on P in the auxiliary regression $E = \phi_o + \phi_p P$, the biased estimates probabilility limit, b, is equal to $\beta \cdot \text{plim}\ \phi_p$.[13] If $C = \lambda_o + E + v$ (in other words $\lambda=1$), $\hat{\gamma}_{cp}$ would be a consistent estimator of plim ϕ_p. Since λ is believed generally to be less than one, regressing C on P provides an estimator of

$$(10)\quad \gamma_{cp} = \text{plim}\ \hat{\gamma}_{cp} = \frac{\text{Cov}(P,C)}{V(P)} = \frac{\alpha\ \lambda\ V(E)}{\alpha^2\ V(E) + V(u)} \doteq \lambda \frac{b_p}{\beta} < \frac{b_p}{\beta}$$

Thus if λ is known, a consistent estimator of b_p/β is provided by $\frac{\hat{\gamma}_{cp}}{\lambda}$. We no longer need to assume $\alpha = 1$ or that the ratio of census and CPS error variance is known. The least squares estimates of $\hat{\gamma}_{cp}$ (or alternatively $\lambda\phi_p$) and $\hat{\gamma}_{pc}$ that have been calculated from the follow-up studies of the last three censuses are presented in Table 1. Using the 1970 CPS-Census match our estimate of γ_{cp} for males on the CPS is .89.

What can be said about λ and α? They are less than one because of the boundaries on the scaling of education. If the sample in which the income education relationship were estimated included only those truly in the middle of the distribution, errors could appear in both directions, boundary effects would disappear and α and λ would be equal to 1. A number of data bases that have been used to estimate income education relationships (army veterans) are of this type. Estimating γ_{cp} on samples limited to those who reported 3 to 15 years of schooling in the CPS would tend to replicate this situation. Estimates of bias from samples restricted to those who reported 3 to 15 years schooling (in the survey whose bias is being measured) are indicated by an asterisk: γ_{cp}^*, γ_{pc}^*, b_p^*, b_c^*. They are also reported in Table 1.

If we assume that the boundary effects on the criterion schooling report are negligible (ie. $\lambda = 1$ when $\gamma_{cp}^* + b_p^*$ are being estimated), $\hat{\gamma}_{cp}$ provides a consistent estimate of $\frac{b_p^*}{\beta}$. An examination of (9) reveals that since V(E) has restricted variance and α is closer to 1, $\hat{b}_p^*$ will generally be smaller than $\hat{b}_p$. As a result $\frac{b_p^*}{\beta}$ is not an unbiased estimator of b_p / β.

C. <u>Bias when Measurement Errors are Correlated</u>

Now let us investigate the implications of relaxing the assumption that errors in reporting education in two separate surveys are uncorrelated. The model specified earlier is retained with only two changes. Assumptions 3 and 4 are modified to provide for some common error variance:

$$3') \quad P = \alpha_0 + \alpha E + w + u$$

$$4') \quad C = \lambda_0 + \lambda E + w + v$$

where w is independent of E, u, v, and ε.

The coefficient in a regression predicting Y with P is now:

$$(11) \quad b_p = \frac{\beta\alpha V(E)}{\alpha^2 V(E) + V(w) + V(u)} .$$

When C is regressed on P, $\hat{\gamma}_{cp}$ is obtained:

$$(12) \quad \gamma_{cp} = \text{plim}\ \hat{\gamma}_{cp} = \frac{\alpha\lambda V(E) + V(w)}{\alpha^2 V(E) + V(w) + V(u)} = \frac{\lambda b_p}{\beta} + \frac{V(w)}{V(P)}$$

or

$$(13) \quad \frac{b_p}{\beta} = \frac{\gamma_{cp}}{\lambda} - \frac{V(w)}{\lambda V(P)} .$$

If P is related to C only through their joint dependence on the true level of schooling, $V(w) = 0$ and the second term of (12) and (13) vanishes, leaving $\frac{b_p}{\beta} = \frac{\gamma_{cp}}{\lambda}$ as before. If the two reports have equal boundary effects, $\lambda = \alpha$, and $V(w) = k \cdot V(u)$, then $\frac{V(w)}{V(P)} = k\ (1 - \gamma_{cp})$ and

$$(14) \quad \frac{b_p}{\beta} = \frac{\gamma_{cp}}{\lambda} - \frac{k(1-\gamma_{cp})}{\lambda} .$$

Note that the second term of (13), $\frac{V(w)}{\lambda V(P)}$, tends to cause γ_{cp} to overestimate $\frac{b_p}{\beta}$ while λ being less than 1 tends to cause γ_{cp} to underestimate $\frac{b_p}{\beta}$. If the common error variance is four-tenths of P's total error variance estimates of $\frac{b^*}{\beta}$ for restricted samples are $\frac{b^*_p}{\beta} = .8952 - \frac{2}{3}\ (1-.8952)$ $= .825$ for the CPS and $\frac{b^*_c}{\beta} = .9079 - \frac{2}{3}\ (1-.9079) = .846$ for the census. If we make the further assumption that λ and α $= .95$, the full sample correction factors may be estimated. For the CPS, $b_p/\beta = [.8903 - \frac{2}{3}(1-.8903)]/.95$

$= .817/.95 = .86$. For the full Census, ${}^{b}c/\beta = [.8833 - \frac{2}{3}(1-.8833)]/.95 =$

$\frac{.8055}{.95} = .848$.

While these estimates of the $\frac{b}{\beta}$ might be considered lower bounds [because the criterion boundary parameter is close to 1 and $V(w)/(V(w) + V(u)$ is not likely to be larger than .4], their size is nevertheless startlingly large. They imply that when CPS or 1970 Census mean earnings or income are regressed on reported years of schooling, we expect the true effect to be larger than the estimated coefficients by approximately 18 percent.

D. Bias When Equation and Measurement Error are Correlated

Finally let us examine the assumption of zero correlation between the error in reporting education (w+u) and the error in predicting income (ε). If Cov(ε, w+u) is not assumed to be zero, (11) and (13) become

$$(11') \quad b_p = \frac{\beta\alpha\, V(E) + \mathrm{Cov}\,(\varepsilon,\, w+u)}{\alpha^2 V(E) + V(w) + V(u)}$$

$$(13') \quad \frac{b_p}{\beta} = \frac{cp}{\lambda} - \frac{V(w)}{\lambda V} + \frac{\mathrm{Cov}\,(\varepsilon,\, w+u)}{\beta\lambda V(P)} \, .$$

The bias is smaller (larger) and ${}^{b}p/\beta$ becomes larger (smaller) if random errors in measuring education are positively (negatively) correlated with the error in predicting income. Thus the bias is smaller if those who exaggerate their education also tend to report a higher income than most others with the same level of education. The resulting higher-than-expected income could be due to chance, a reporting error, or an unmeasured attribute that raises income.

If Cov(ε, w+u) is not zero, most psychological theories would predict it to be positive. Note that this positive correlation tends to raise ${}^{b}p/\beta$ relative to $\frac{\gamma_{cp}}{\lambda}$, while the positive covariance between errors in reporting education, V(w) tends to lower ${}^{b}p/\beta$ relative to $\frac{\gamma_{cp}}{\lambda}$. These two effects will cancel out if:

$$\beta = \frac{Cov(\varepsilon, w+u)}{V(w)} = \frac{Cov(\varepsilon, w)}{V(w)} + \frac{Cov(\varepsilon, u)}{V(w)} = \frac{Cov(\varepsilon, w+u)}{V(w+u)} \Big/ \frac{V(w)}{V(w+u)}$$

The interpretation of this condition is that the true income education coefficient, β, is equal to the coefficient of a regression of ε on the error in reporting education, divided by the coefficient predicting one error in reporting education with another. If β is greater than this ratio, $b_p/\beta < \frac{\gamma_{cp}}{\lambda}$. Actually going to school for an extra year certainly raises income by more than systematically saying one went to school for an extra year when one has not. Since w is the education reporting error that occurs in both questionnaires, $\beta > \frac{Cov(\varepsilon, w)}{V(w)}$. Further, we can most likely safely assume that education reporting error which does not recur in the second survey has a negligible correlation with ε. Thus, if $\beta > \frac{Cov(\varepsilon,w) + Cov(\varepsilon,u)}{V(w)}$, γ_{cp}/λ places an upper bound on b_p/β. Retaining our earlier assumption that $\lambda = \alpha = .95$, the upper bound estimate of the CPS's $b_p/\beta = \frac{.8903}{.95} = .937$. The upper bound for the Census $b_p/\beta = \frac{.8833}{.95} = .930$.

A common way of estimating human capital models is to use individual observations and the log of earnings as the dependent variable. Mincer has observed that the coefficient on education in such regressions (which can be interpreted as a rate of return) is consistently lower than the rates of return estimated directly.[14] Predicting the log of earnings with individual data means one is assigning to each educational class its geometric mean earnings. A 10 percent increase from \$1000 to \$1100 affects the estimate to the same degree as a 10 percent increase from \$20,000 to \$22,000. In positively skewed distributions, like income, the regression line passes below each group's arithmetic mean and differences between groups are understated. Since skewness increases as education increases, estimates of the mean dollar difference between groups are understated even more. Rates of return derived from this specification are not comparable to rates of return on other assets and, without adjustment, should not be used for setting policy.

E. Combining Omitted Variable with Errors in Variables Bias

What does the $\hat{b}_P$ obtained from a regression of Y on P tell us when the true model includes other variables like ability that are correlated with years of schooling? The true model for this situation is:

(2') $Y = \beta_0 + \beta_1 E + \beta_2 A + \varepsilon$

(3') $P = \alpha_0 + \alpha E + w + u$

(4') $C = \lambda_0 + \lambda E + w + v$

(16) $T = A + t$

(17) $A = \mu_0 + \mu E + r$

where ε, u, v, w, t, r, are uncorrelated with each other and with E and A

A is true ability prior to schooling [(17) is associational not causal)]

T is test score prior to schooling

(18) $$b_p = \text{plim } \hat{b}_p = \frac{\sigma_{PY}}{\sigma_{PP}} = \frac{\beta_1 \alpha \sigma_{EE}}{\alpha^2 \sigma_{EE} + \sigma_{ww} + \sigma_{uu}} + \frac{\beta_2 \alpha \mu \sigma_{EE}}{\sigma_{PP}}$$

for a regression of T on P, $$\text{plim}(\tilde{\mu}) = \frac{\sigma_{PT}}{\sigma_{PP}} = \frac{\alpha \mu \sigma_{EE}}{\alpha^2 \sigma_{EE} + \sigma_{ww} + \sigma_{uu}}$$

(19) $$b_p = \left(\frac{\gamma}{\lambda} cp - \frac{\sigma_{ww}}{\lambda \sigma_{pp}} \right) \beta_1 + \beta_2 \tilde{\mu}$$

(20) $$\beta_1 = (b_p - \beta_2 \tilde{\mu}) \div \left(\frac{\gamma}{\lambda} cp - \frac{\sigma_{ww}}{\lambda \sigma_{pp}} \right)$$

or if the true μ is known,

$$(21)\quad \beta_1 = b_p \div \frac{\gamma_{cp}}{\lambda} - \frac{\sigma_{ww}}{\lambda\sigma_{pp}} - \beta_2\mu.$$

Combining the two corrections is straightforward. The first requirement is an estimate of the true effect of ability on earnings, β_2. Next using the population from which $\hat{b}$ was estimated, the amount by which those with higher reported schooling have higher prior test scores is calculated. The product of these parameters is subtracted from $\hat{b}$ and the adjustment factor derived earlier in (13) is applied.[15]

F. Bias in Rates of Return Calculated from Census Tabulations

Most human capital studies use published census tabulations that group people into a few classes by years of education attained. Grouping on the independent variable generally reduces the size of the errors-in-variables problem.[16]

If one is interested in an estimate of the true earnings gain from a particular level of schooling, estimates of an errors-in-variables correction derived from samples encompassing the full range of education may not be appropriate. The process that generates reporting errors may not be regular (i.e., characterizable simply by $C - E = \lambda_o + (\lambda-1)E + u$ where the distribution of u is independent of E). The availability of detailed cross tabulations of Census and follow-up education reports allows us to calculate estimates of the bias $\frac{\hat{\beta}}{b}$ for adjacent education classes without assuming u is independent of E.

Using tabulations of Census and follow-up education reports for 1960 and 1970, an estimate was made of the true earning increments between adjacent education classes. A true earnings level was assigned to each year of educational attainment. This level was assumed to be the mean for the corresponding criterion education report. A weighted average of these means for each census reported education class corresponds to the earnings averages that census tabulations produce.[17] This procedure drops the assumption that the process that generates errors in the census report is regular and thereby allows for heaping effects. The results presented in Table 2 assume (1) the error in the criterion report is uncorrelated with the true level of schooling ($\alpha = 1$) and with the census reporting error ($\sigma_{uv} = 0$) and (2) the error in the census report is uncorrelated with earnings ($\sigma_{\varepsilon v} = 0$).

The assumption of no boundary effects in the criterion report is almost certainly violated for intervals adjacent to the boundary (0-7 versus 8 and 16 versus 17+). The bias estimates for these intevals are, therefore, too large.[18]

As expected the results in Table 2 imply that different schooling increments have different biases. The large estimated biases for elementary and postgraduate schooling exaggerate the true bias because of boundary effects on the criterion schooling report. Census tabulations seem to consistently overestimate the earnings gain from the first two years of high school. For women in 1970 and everyone in 1960 the earnings gain of the first two years of college are underestimated in the Census by almost 20 percent. The CPS seems to underestimate the male earnings increments for 8 through 16 years of schooling by more than the Census does. For females, in contrast, census and the CPS bias estimates are quite similar.

Combining errors in variables with omitted variables bias adjustments of earnings increments estimated from census tabulations is straightforward. The effect of adjusting the return to college graduation for ability differences will be examined. The ratios of high school graduate to college graduate earnings were .747 for ages 25 to 34, .624 for ages 35 to 54, and .599 for ages 55 to 64. Prior to college, college graduates typically had IQ's one standard deviation above those who did not enter college. If we adopt 7 percent as the true reduction in earnings that occurs if a college graduate has the same ability as the typical high school graduate, the earnings increment for ages 25 to 34 is reduced by 28 percent and the earnings increment for 35 plus is reduced by approximately 18 percent.[19] The errors in variables adjustment is then multiplied by the proportion of the gross effect remaining [(i.e., for 25 to 34, $\beta \div \hat{b} = 1.056\ (1 - .747 - .07) \div (1 - .747) = .76$)].

II. Coverage Bias: Errors in Measuring a Person's Contribution to Output

The second major source of bias in estimating social returns derives not from random errors, but from systematic undercounting. Reported earnings do not provide complete coverage of a worker's total compensation and do not include taxes paid by employers on output or on labor input. The bias from this source will be called coverage bias. The contribution to total output of a marginal increment in a given factor of production is its marginal physical product times the price consumers pay for the product. A profit maximizing firm will arrange its use of factors so that the total compensation paid including taxes for a marginal increment of a factor equals that factor's marginal revenue product. Discrepancies arise between reported earnings and total cost of labor input due to under or over reporting incomplete coverage, and employer paid taxes on employee wages. Discrepancies may arise between marginal revenue and consumer price due to excise taxes or monopoly power.

It is possible to determine the average degree of under or overreporting to Census interviewers for each type of income by comparing national income aggregates derived from establishment sources with the aggregates implied by the Census household data. Doing this for the March 1970 and March 1971 CPS we find that the unreported component of money earnings was 5.4 percent of the money earnings reported to Current Population Survey interviewers. While 96 percent of wage and salary income was reported, only 52 percent of farm income was reported. This will cause significant understatement of mean incomes and income differentials of agricultural states. The 1970 Census comes substantially closer to its control aggregates than the CPS for that year. In Table 3 we reconcile CPS and Census aggregates to National Income Account aggregates. The percent of aggregate earnings missed was only 1.5 percent in 1970 Census and seven-tenths of a percent in the 1960 Census.

Census and CPS aggregates may be low either because people are missed or because on average each person understates his earnings. The Bureau of the Census has developed estimates of the amount by which the nation's population was understated in the Census and the CPS.[21] Applying age, sex, and race specific undercount rates to 1970 Census estimates of their earnings aggregates and assuming that those missed earn two-thirds the average, our adjustment for the undercount increases Census aggregates by 2.1 percent in 1960, and 2.8 percent in 1970. This implies in turn that per capita earnings on the Census overstated national accounts per capita earnings by 1.47 percent in 1960 and 1.34 percent in 1970. The CPS understated national account per capita earnings by 2.8 percent.

Another source of discrepancy between reported earnings and total compensation received are food and housing received as pay, farm products consumed at home, and employer contributions to private pension plans. Estimates of the amount of each of these received by each income class are available in Roger A. Herriot and Herman P. Miller's, "Who Paid the Taxes in 1968."[22] The major element of these imputations, employer contributions to pension plans, was assumed to be proportional to wages and salaries. Together these imputations average 5 percent of reported income. Employer contributions to private pension and welfare funds has been a rising proportion of total employee compensation in the last few decades. As a result the imputation adjustment for 1959 is 1.3 percentage points less than the one described in Table 3 for 1968 and 1969.

The final discrepancy between reported earnings and the marginal revenue product are the Social Security and Unemployment Insurance taxes paid by employers. These taxes declined as a proportion of earnings because in 1969 the social security tax was paid only on the first $7800 of wages and the unemployment insurance taxes were paid only on the first $3400 of wages. The statutory rate (4.8 percent for social security and 1.4 percent for unemployment insurance) was used for earnings brackets below the maximum taxable wage.[23] Above that the average social security tax rate was the maximum tax, $374, divided by the midpoint of the earnings interval and the average unemployment insurance tax rate was $47.60 divided by the midpoint. Since 1969 the Social Security tax rate has risen from 4.8 to 5.85 percent and the maximum taxable wage has risen from $7800 to $10,800. The 1969 rates are used because the adjustments are intended for use with decennial census data. If the earnings data being used are for a period of higher tax rates, and employers have had time to adjust, a larger adjustment for labor input taxation would be in order. In 1959 Social Security tax rates were 2.5 percent on the first $4800. Therefore, calculations of coverage bias for 1959 use the lower tax rate applicable then.[24]

The final adjustment must take us from marginal revenue product to the consumer's valuation of that output. Both monopoly power and taxes cause consumer price to exceed marginal revenue. State and federal sales and excise taxes, not including alcohol and tobacco taxes, totaled $27.6 billion in 1968 or approximately 3.2 percent of GNP. Alcohol and tobacco taxes are excluded because they are assumed to reflect the negative externalities a person's use of these products imposes on others. The dollars of excise revenue generated by a person's work were calculated by multiplying the sum of labor input taxes and reported, unreported, and imputed earnings by .033, $(\frac{1}{.968} - 1)$.[25]

Whether monopoly power makes a further correction desirable depends upon the source of the monopoly and which factor of production is receiving the monopoly rents. If monopoly rents add equal percentage increments to workers' wages and to capital's return, no problem is created, for our compensation data has already captured them. If a firm faces close to infinitely elastic long-run demand curve at its limit price, but nevertheless receives monopoly rents because of the ownership of some unique factor of production (e.g., patents, control of best raw material sources, government licenses), no adjustment is required. An add on is required only where P > LRMR = LRMC and where the monopoly rents do not get paid to labor.

How large might such monopoly profits be? Harberger's upper bound estimate of the welfare impact of monopoly implied that one-third of manufacturing profits were excess profits.[26] Assuming that the share of monopoly rents [(P - LRMC)q] in corporate profits is one-third for manufacturing and one-tenth for nonmanufacturing, we obtain an upper bound estimate of $17.7 billion for 1969 or 1.9 percent of GNP.[27] The results presented in Table 3-5 do not include an adjustment for monopoly distortions or for systematic economies of scale.[28] The reader may make his own adjustment for monopoly with his own assumption about monopoly by simply multiplying the average and marginal ratios of social benefit to reported income in Table 4-6 by a number between 1 and 1.019.

Putting all these adjustments together we find that the average social productivity benefit of a person's work--the sum of after tax earnings and taxes generated--averages about 120 percent of reported earnings. As earnings rise, the ratio of social benefit to reported earnings tends to fall from 1.20 to 1.13. The fall in the ratio is a consequence of imputations not rising as fast as income and the zero marginal social security tax on wages above $7800.

What portion of this total or social return can the individual be expected to take into account when he makes his own decisions? Splitting the social return into private and public components is necessarily more arbitrary than calculating the total return. Table 4 presents lower bound estimates of the private share of the total return. It is based upon the assumption that extra earnings do not, on the margin, place any additional burden on the government's provision of services. This is a valid assumption for pure public goods--defense, foreign affairs, space, and police and fire protection. Providing an individual with more of a pure public good inevitably means everyone else gets more.

However, for many government services provided at zero or nominal cost, providing the service to one person means it must be denied to someone else. If usage of such services rises with income, extra after tax income places an additional burden on other taxpayers. Directly provided services of this kind are education, libraries, airports, congested highways, recreation, sewers, water supply, and garbage collection.[29] Usage of certain other services--Food Stamps, directly subsidized housing, Medicaid, unemployment insurance and AFDC--go down as earnings rise. If one takes a life cycle perspective, however, the largest of transfer programs, social security, provides larger dollar benefits to people with higher earnings.

The net impact of earnings on usage of government programs is likely to be positive, though it may be small (i.e., 3 or 4 percent).[30] In exceptional cases, an education or health intervention that increases after tax earnings may actually decrease total government expenditures. For Black females, the present value of expected welfare payments is about $7000 less for high school graduates than for dropouts. This reduction is likely to be larger than the increase in services that are positively associated with education and earnings.

By neglecting these impacts of government programs, the private benefit of labor market productivity can be estimated simply by subtracting personal income taxes and the employee's share of Social Security taxes from the sum of reported earnings, unreported earnings, and imputations. In almost all cases, this places a lower bound on the estimate of the private return. The average incidence of federal and state income tax payments were taken from Herriot and Miller's "Who Paid Taxes in 1968." The incidence of Social Security taxes on earnings has already been described. The sum of these two taxes rises from 8.9 percent of reported earnings in the $2000-$4000 bracket to 17.4 percent in the $15,000-$25,000 bracket. The ratio of disposable earnings (private returns) to reported earnings, therefore, falls from 1.01 to .905 as one moves from low to high brackets. The ratio of all taxes generated to reported earnings rises from .19 to .23 as earnings rise.

Marginal ratios of after tax earnings to reported earnings and of taxes generated to reported earnings were estimated by the following procedure. Average ratios were calculated separately for each bracket. After tax income and total tax generated were calculated for the representative family in the interval using average ratios and the midpoints of the intervals as family rates of taxes generated as reported income increases were above their averages and were generally declining with income. They decline because the drop in the marginal rate of Social Security tax from 9.6 percent to zero outweighs the progressivity of the personal income tax. The marginal rates of private return after tax earnings, were below the average and tended to fall with income from a high of .935 to .89. Largely because of the declining ratio of Social Security taxes to earnings as earnings increase, the marginal ratio of social or total return to reported earnings is also below the average. It falls from 1.19 in low brackets to 1.12 in high brackets.

III. Implications of Coverage Bias

Most studies that have attempted to measure social and private returns to education have neglected the effects of underreporting, fringe benefits, and employer paid taxes on wages and value added. To what extent does this bias their results? Table 4 presents recommended adjustments for coverage bias and progressive income taxation for each census and for the CPS.

Since most of the calculations of returns to education have been presented as rates of return, no problem is created if the measures of both cost and benefits are biased to the same degree. Measures of foregone earnings are subject to the same type of coverage bias as benefit measures. In fact, because foregone earning changes occur in the lower range, their coverage bias is proportionally larger. Also the split between taxes generated and private costs is different. The young people who are investing time in their own education face lower or zero rates of income taxation. As a consequence private rates of return (especially those of the lowest schooling levels) are lowered by adjustments for coverage bias and taxation.

Social costs include, however, a large government expense component which does not suffer coverage bias. If for college graduation instructional costs are equal to foregone earnings, a CPS coverage bias adjustment would make costs 8 percent higher and benefits 13 percent higher. Thus, the true rate of return will be 1.05 times the conventionally calculated return. This is not a very large bias especially when one considers that estimates of returns on physical capital are affected by one of the biases enumerated above (value added taxation).

Coverage bias is a serious problem when costs of a student's time are actually leisure foregone or when the student's time commitment does not change. This latter situation occurs when a government educational intervention is being evaluated. If we are calculating the costs and benefits of improving the quality of education, providing a specialized service for problem children, or subsidizing some college instructional programs more than others, current practice counts all the costs but only some of the benefits. The ratios presented in the social productivity benefit section of Table 4 are the correction factor that should be applied to the benefit cost ratios of increases in the quality of education. Therefore, coverage bias produces a systematic downward bias of 8 to 18 percent in the calculated social benefit-cost ratio depending on which level of education the intervention is at.

If the time spent by the student on school work actually comes at the expense of leisure rather than employment the social cost of that time is less. Most young people have control over the number of hours they work for wages. They, therefore, adjust their work time until on the margin an hour of leisure is worth to them approximately what they get from one more hour of employment, the after tax wage rate. When a student's time comes at the expense of involuntary unemployment rather than voluntary leisure, the social cost is even less.

The extra leisure time a nonstudent has does not produce taxes, however, so the social cost of his lost leisure is equal to the private cost. Thus when time spent in school is at the expense of leisure yet that time has been valued at the money wage, the understatement of the social rate of return is the greatest.

Errors in reporting educational attainment produces an identical bias in private and social rates of return to extra years of schooling. They affect different educational increments differently, however. The understatement of rates-of-return and of benefit-cost ratios is largest for grade school and the first few years of college. The returns to starting high school are not understated and may in fact be overstated.

Table 5 presents the estimated ratio of corrected to uncorrected benefit measures for different increments in education and for different data sources. In most cases, whether or not an educational project is undertaken will not depend on a bias of this size. However, given the simplicity of the adjustment required, there is no reason not to use the correct productivity benefit concept. It is interesting to note that the understatement bias is greater for investments in low wage workers. Thus, benefit-cost ratios for educational investments in discriminated minorities, women, secondary education and low-skill workers are understated more than investments in college or graduate education. The benefits of health investments are understated more as well because here average ratios of true productivity to reported earnings are relevant.

TABLE 1

Measure of Bias in Income Education Coefficients

	Reliability Coef[1] r_{PC}	Bias in 1970 Census: Coef Predicting CPS[2] $\hat{\gamma}_{PC}$	Bias in 1970 Census: Coef Predict CPS W/o Boundary Effects[3] $\hat{\gamma}^*_{PC}$	Bias in CPS: Coef Predicting Census[4] $\hat{\gamma}_{CP}$	Bias in CPS: Coef Predicting Census W/o Boundary Effects[5] $\hat{\gamma}^*_{CP}$
All Males (n~10,000)	.887	.8833	.9079	.8903	.8952
All Females (n~10,000)	.875	.8603	.9019	.8905	.9281
White Males (n~9000)	.888	.8854	.9084	.8911	.8927
White Females (n~9000)	.879	.8615	.8983	.8972	.9320
Black Males (n~1000)	.834	.8200	.8442	.8476	.8887
Black Females (n~1000)	.843	.8363	.9259	.8494	.8708
		Bias in 1960 Census		Bias in 1960 Follow-ups	
Full PES (n~5000)	.9343	.9422	.9598	.9264	
White PES (n~4500)	.9333	.9368	.9548	.9298	
Black PES (n~500)	.9225	.9626	.9862	.8841	
Reinterview (n~7500)	.9149	.9130	.9305	.9168	
Requestionnaire (n~1500)	.9146	.9132	.9118	.9160	
		Bias in 1950 Census		Bias in 1950 PES	
All (n~4000)	.8603	.8546	.8985	.8660	.9140

[1] The correlation of the Census and follow-up reports.

[2] Obtained from the auxiliary regression $P = a_1 + \hat{\gamma}_{PC}\,C$ run on the full range of Census reported educational attainment. C is Census report and P is follow-up report.

[3] Obtained from the auxiliary regression $P = a_3 + \hat{\gamma}^*_{PC}\,C$ run on people who reported 3 to 15 years schooling in the Census.

[4] Obtained from the auxiliary regression $C = a_2 + \hat{\gamma}_{CP}\,P$ run on the full range of follow-up reported schooling.

[5] Obtained from the auxiliary regression $C = a_4 + \hat{\gamma}^*_{CP}\,P$ run on people who reported 3 to 15 years of schooling in follow-up.

TABLE 2

Effect of Errors in Reporting Education on Measures of Return to Education from Census and CPS Tabulations

	Years of Education						
	0-7	8	9-11	12	13-15	16	17+
Assume True Earnings[1,2] (1969)	$6,485	$7,841	$8,589	$9,634	$11,169	$14,776	$17,009
Weighted Average for Reported Educ in Census[1] (1969)	$6,990	$7,945	$8,697	$9,698	$11,199	$14,576	$16,273

	Finish Elem.	Elem. to H.S.D.O.	H.S. GRAD	H.S. GRAD to COLL.D.O.	COLL. GRAD	POST GRAD
Increase for Census Reported Educ.[1] Categ. (1969)	955	752	1,002	1,500	3,377	1,697
True Increase[1] (1969)	1,356	748	1,048	1,535	3,607	2,233
Ratio of True to Tabulated Incr. for Rept. Educ. (β:b)						
Males for Census Report with CPS as Criterion (1969)	1.42	.99	1.046	1.02	1.068	1.316
Females for Census Report with CPS as Criterion (1969)	1.49	.806	1.13	1.20	1.016	1.456
Males for CPS Report with Census as Criterion (1969)	1.36	1.018	1.081	1.106	1.095	1.07
Females for CPS Report with Census as Criterion (1969)	1.58	.801	1.069	1.178	1.049	1.188
$\hat{\beta}$:b for Census if PES Criterion (1959)	1.19	.93	1.04	1.19	.99	1.19
$\hat{\beta}$:b for Census if Reinterview is Criterion[3] (1959)	1.17	.94	1.15	1.02	1.01	1.44
$\hat{\beta}$:b for Census if Requestion is Criterion[3] (1959)	1.33	.99	1.03	1.16	.96	1.15
Preferred Ratio of True to Reported Increase[4] (1959)	1.19	.95	1.07	1.12	.99	1.19

[1]Estimate is derived by: (1) using CPS responses as the criterion for measuring bias in the census. (2) For those who reported a given level of education in census, a weighted average of census mean incomes was calculated using the frequency distribution of true educational levels from the CPS. This produces a series of mean incomes by education class that would have been observed in the census if education and income reporting errors are statistically independent and the 1970 cell means are taken as the true means.

[2]Row is based on Table 1 of Earnings by Occupation and Education, 1970 Census of Population, U.S. Bureau of the Census, Subject Reports, PC(2)-8B. For education intervals numbers do not exactly correspond because an earnings figure was assigned to each individual year and the number presented is a weighted average of those figures using the CPS marginals as weights. Earnings are for males 25-64 who worked 50-52 weeks.

[3]Ratio of true increase in earnings to the increase for census reported education categories using the reinterview and the requestionnaire responses as standard.

[4]Except for the education increments nearest the bounds of the scale, the preferred estimate of bias in 1960 is a simple average of the results from the three sources. For the increments at the boundary, the P.E.S. estimated bias is used because it is believed to have the smallest tendency to overstate the bias.

Table 3

Reconciliation of Census and CPS Money Income Aggregates
With National Income Account Aggregates

	1959 Census	1969 Census	1969 CPS	1970 CPS
National Income Account Wages and Salaries	258.2	509.7	509.7	541.9
less payments of military overseas[1]	2.4	4.9	4.9	4.9
less payments to decedents	1.5*	2.9*	2.9*	3.0
less imputed food and lodging and civilians overseas	1.8*	3.5*	3.5*	3.5
plus payments of military reservists and directors fees	.9*	1.9*	1.9*	1.9
NIA Wages and Salaries Adjusted to Census Concepts	253.4	500.3		
less payments of military living on post			5.0*	5.5
NIA Wages and Salaries Adjusted to CPS Concepts			495.5	526.8
Total Wages and Salaries rept. to Census or CPS[2]	248.1	500.1	478.9	508.2
Ratio of National Account to Census or CPS	1.021	1.000	1.035	1.037
Business and Professional Earnings in NIA	35.1	50.5	50.5	49.9
NIA Business and Professional Earnings Adjusted	34.4*	49.5*	49.5*	48.9
Business and Professional Earnings in Census or CPS[2,3]	41.8	47.7	43.6	45.0
Ratio of National Account to Census or CPS	.822	1.038	1.135	1.087
Farm Self Employment Income in NIA	11.4	16.7	16.7	16.9
NIA Farm Income Adjusted to Census and CPS Concept[2,3]	10.6*	15.4*	15.4*	15.6
Farm Income in Census or CPS	6.3	8.8	8.45	7.9
Ratio of NIA Farm Income to Census or CPS	1.66	1.75	1.82	1.97
Total NIA Earnings Adjusted to Census or CPS Concept	298.4	565.2	560.2	591.3
Total Earnings Reported in Census or CPS	296.3	556.6	531.0	561.1
Ratio of NIA Earnings to Census or CPS	1.007	1.015	1.055	1.054
Ratio of Other Labor Income to Census or CPS Earnings[4]	.038	.051	.053	.057
Personal Income in National Income Accounts[5]	383.5	750.3	750.3	803.6
Personal Money Income Adjusted to Census or CPS Concept[5]	351.4	686.0*	681.0*	726.4
Personal Money Income in Census or CPS	331.7	631.5	613.9	646.9
Ratio of Adjusted Personal Money Income to Census or CPS	1.059	1.086	1.109	1.123
Ratio of NIA Personal Income to Census or CPS	1.156	1.188	1.222	1.242

*Adjustments for 1959 and 1969 follow those used by Dorothy Projector and Judity Bretz in "Measurement of Transfer Income in the Current Population Survey" presented at conference on Research in Income and Wealth, October 3-4, 1972. Prior year adjustments were assumed to be the same proportion of national account totals.

[1]Estimated by the smaller of Direct Military Expenditure overseas from the balance of payments or number of men stationed overseas times basic pay plus allowances.

[2]Earnings are for the 14 and above age group. Taken from Table 99 of Volume 1 of 1960 Census and the comparable Table 243 for 1970 Census and Table 38 of P60#75 for the 1969 CPS aggregate.

[3]Self employment income was divided into its farm and nonfarm components by using the PC(2)-8C, Income of the Farm Related Population for 1970 and PC(2)-4C Sources and Structure of Family Income for 1960. Table 38 provides disaggregation for the CPS. Nonfarm self employment income is sometimes exaggerated by reporting a gross rather than a net figure.

[4]Other labor income is primarily fringe benefits--employer contributions to private pension plans and compensation for injuries.

[5]The difference between personal income in the national accounts and adjusted personal money income is primarily imputations for owner occupied housing, services provided free by financial intermediaries, and payments to fiduciaries and decedents. 1960 estimates of adjusted OBE personal income are from Herman Miller, Income Distribution in the United States, p. 173.

Table 4: Coverage Bias in Estimates of Benefits and Costs of Education[1] Ratios of True Productivity Benefit or Cost to Reported Money Earnings

	Productivity Benefits						Student's Time Costs					
	Social			Private			Social[2]			Private[3]		
	Census 1959	Census 1969	CPS	Census 1959	Census 1969	CPS	Census 1959	Census 1969	CPS	Census 1959	Census 1969	CPS
Finish Elementary[4]	1.08	1.09	1.14	.87	.86	.89	1.12	1.15	1.20	1.02	1.01	1.05
Elem. to HSDO[4]	1.07	1.08	1.13	.85	.86	.90	1.12	1.15	1.20	.98	.97	1.01
HSG	1.06	1.07	1.12	.84	.85	.89	1.08	1.11	1.16	.88	.87	.91
Col. Dropout	1.05	1.06	1.11	.84	.85	.89	1.08	1.11	1.16	.88	.87	.91
Col. Grad.	1.04	1.06	1.10	.83	.84	.88	1.08	1.11	1.16	.88	.87	.91
Grad. School	1.04	1.05	.09	.82	.83	.87	1.08	1.11	1.16	.87	.86	.90

[1]By a simple manipulation of these factors, rates of return and benefit-cost ratios may be adjusted for taxation and coverage bias. Calculate the cost adjustment by taking a weighted average of the student time cost factor and one. The weights are the conventionally calculated foregone earnings and either instructional cost or out of pocket tuition and books costs. This average is divided into the productivity benefit adjustment factor.

[2]If time spent in schooling would have been spent working.

[3]The social cost if time at school comes at the expense of leisure or the private cost no matter how the school time would have been spent.

[4]Foregone time costs adjustment for students through 10th grade use the average ratio rather than the marginal ratio that is assumed for all other levels of schooling. In other words, until the 10th grade, it is assumed that those in school hardly earn anything at all. For all others, it is assumed those in school already work some and that the effect of dropping out is to increase the amount of work and leisure from an already existing base and that therefore, marginal rates of taxation and coverage bias apply. It is further assumed that elementary students pay only social security taxes on their earnings.

Table 5: Approximate Adjustments to Productivity Benefit Measures to Correct for Coverage and Errors in Variables Bias in Census Data

	Private Rate of Return[1]			Conventional Social Productivity Rate of Return[2]			Social Rate of Return[3] if Schooling Time is at the Expense of Leisure		
Extra Years of Schooling	1960 Census	1970 Census M	1970 Census F	1960 Census	1970 Census M	1970 Census F	1960 Census	1970 Census M	1970 Census F
Finish Elementary	1.01	1.21	1.27	1.28	1.54	1.62	1.15	1.35	1.41
Elem. to HS Dropout	.82	.88	.72	.96	1.00	.81	.91	.93	.76
HSDO to HS Grad.	1.02	1.02	1.10	1.08	1.05	1.13	1.05	1.01	1.09
HSG to Col. Dropout	1.04	.97	1.14	1.12	1.01	1.19	1.09	.98	1.15
Col. Dropout to Grad.	.91	1.00	.95	.99	1.05	1.00	.96	1.01	.96
Col. Grad. to Post Grad.	1.12	1.27	1.40	1.18	1.29	1.43	1.15	1.24	1.37
Improvements in Quality of College[4]	.84	.85	.85	1.05	1.06	1.06	1.19	1.22	1.22
Death Reduction Programs[5]									
$5,000 yearly earnings	.93	.93	.93	1.11	1.14	1.14	1.29	1.32	1.32
$20,000 yearly earnings	.87	.87	.87	1.06	1.09	1.09	1.27	1.31	1.31

It is assumed the uncorrected benefit measures standardize for hours worked, unemployment, mortality, expected productivity growth, and the confounding influence of ability and family background. The covariance of measurement error is assumed to be zero.

[1]The Uncorrected is not adjusted for progressive taxation and we assume that only social security taxes need be subtracted from the foregone earnings of elementary students. Students above grade 10 would have had their foregone earnings taxed at the marginal rates for the appropriate brackets. Tuition and Books costs are assumed to be 20 percent of the private costs of an undergraduate education and zero percent for high school and graduate education (fellowships not reported as earnings equal tuition paid).

[2]It is assumed that the only cost of the time spent in school is foregone earnings which are 5 percent of social costs of elementary school, 50 percent of the social costs of the first 2 years of high school and 63 percent of the rest of high school and college. These percentages are derived from Hines, et. al., op. cit.

[3]It is assumed that the time spent in school comes at the expense of leisure which has been valued at the money wage rate, and that valued in this way, lost leisure is equal to instructional cost.

[4]No change in time committed by student. In first 6 columns private return assumes the improved quality results dollar for dollar increases in tuition. Last 3 colums are for an improvement in earnings accomplished by studying more hours in college.

[5]The costs of the health program are private expenditures in the first 3 columns, public expenditures in the middle 3, and leisure in the last 3. Benefits are money earnings only.

[6]Because they are typically married, women are assumed to have the same coverage bias as men.

FOOTNOTES

[1]Griliches and Mason's, "Education, Income and Ability," and John C. Hause's, "Earnings Profile: Ability and Schooling," in Investment in Education Supplement to May/June 1972 Journal of Political Economy, Robert Hauser, Kenneth Lutterman and William Sewell, "Socioeconomic Background and the Earnings of High School Graduates," in Education, Occupation, and Earnings: Achievement in the Early Career ed. William Sewell and Robert M. Hauser (New York, The Academic Press, 1974); Taubman and Wales, "Higher Education, Mental Ability and Screening," Journal of Political Economy, Jan. 1973, 81:1, p. 28-55.

[2]C. M. Lindsey, "Measuring Human Capital Returns," Journal of Political Economy 79:6, November 1971, p. 1195-1215; R. S. Eckaus, "Returns to Education with Hourly Standardized Incomes," Quarterly Journal of Economics, February 1973, p. 121-131. Donald Parsons, "Costs of School Time, Foregone Earnings and Human Capital Formation," Journal of Political Economy, March 1974, 82:2 p. 251-266. Michael has looked at the effect of education on the efficiency of consumption. If one accepts his model and results the improvement in total real consumption is three or four times the dollar amount of the earnings impact alone. No significant adjustments to costs are required by his approach so the rate of return to education is effectively tripled. Robert T. Michael, "Education in Non-Market Production," Journal of Political Economy, March/April 1973, p. 306-327.

[3]Paula Schneider and Joseph Knott "Accuracy of Census Data as Measured by the 1970 CPS - Census - IRS matching Study" mimeo, U.S. Bureau of the Census.

[4]A 1960 correlation of .915 derived from the reinterview program is preferred over a .93 correlation derived from the P.E.S. because techniques were almost identical to the census, no attempt was made to reconcile discrepancies and interviewers had no knowledge of the respondent's earlier answer.

Evaluation and Research Program of the U.S. Censuses of Population and Housing, "Accuracy of Data on Population Characteristics as Measured by Reinterviews," Series ER60 #4; "Effects of Different Reinterview Techniques on Estimates of Simple Response Variance," Series ER60, #11, 1960 (Washington, D.C., Government Printing Office 1964). The Post Enumeration Survey: 1950 Technical Paper #4, Bureau of the Census (Washington, D.C., Government Printing Office, 1960).

[5]The reconciliation of answers was done in two-thirds of the 1960 PES sample and all the 1950 sample. For one-third of the PES respondents the interviewer was aware of the census response when the first PES interview was conducted. Barbara Bailor has shown that this knowledge increases the consistency of unreconciled responses with the original census report. Barbara Bailor, "Recent Research in Reinterview Procedures," Journal of American Statistical Association. March 1968, p. 41-63.

[6]The use of final edit records means that in contrast to the earlier reports, errors are the sum of coding, key punching, and response error and are "therefore" measurements of errors in the public use tapes and in publication level statistics. Schneider & Knott, Ibid., p. 2.

[7]The classic work on the effect of measurement error on the education income relationship is by Paul Siegel and Robert W. Hodge, "A Causal Approach to the Study of Measurement Error" in Hubert M. Blalock and Ann B. Blalock, Methodology in Social Research (New York McGraw-Hill, 1968).

For a large and varied set of alternative assumptions about the nature of the error they calculated the bias in estimators of standardized regression coefficients, $\frac{b\sigma_E}{\sigma_Y}$, or partial correlations. There was a great deal of variance in the resulting estimates. Results for biases in standardized regression coefficients depend upon the bias in the sample estimates of σ_E and σ_Y as well as in the coefficient b.

[8]Samuel Bowles shows that the equivalence of the reliability approach to the errors in variables model in "Schooling and Inequality from Generation to Generation," Journal of Political Economy, 80:3, May-June, 1972, p. S240.

[9]Jencks reports adopting Siegel & Hodge's measure of education's reliability. Our reading of Siegel & Hodge is that no choice amongst alternative estimates was made. Jencks et al., Inequality, p. 333. The .958 reported for Bowles is from his most recent article, Bowles and Nelson "The 'Inheritance of IQ' and the Intergenerational Reproduction of Economic Inequality" Review of Economics and Statistics 56:1, Feb/74, p. 39-51.

[10]Schneider & Knott, Ibid., p. 7.

[11]"Educational Attainment, 1972," Current Population Reports, Series P20, #243.

[12]Income differentials for males over 25 with between 8 and 12 and between 12 and 16 years of schooling were $3,012, and $5,187 in the Census and $3,018 and $4,431 in the March 1970 C. P. S.

[13]Henri Theil, Principles of Econometrics (John Wiley, New York 1971), p. 607.

[14]Jacob Mincer, "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," The Journal of Economic Literature, March 1970, p. 8.

See Finis Welch's "Black White Differences in Returns to Schooling" for a method of calculating the arithmetic means from a log linear specification., AER, Dec. 1973, p. 901.

[15] If the relationship between reported education and test scores was not estimated from the same set of data, we must assume that the patterns of errors in measurement in both sets are the same. If the education and test score relation is estimated in data without errors in measurement of education, the correction factor in (13) is applied first and the product of $\beta_2\mu$ is subtracted last

[16] M. S. Bartlett, "The Fitting of Straight Lines if Both Variables are Subject to Error," Biometrics 1949, Volume 5, p. 207-242; J. W. Hooper and H. Theil, "The Extension of Wald's Method of Fitting Straight Lines to Multiple Regression," Review of International Statistical Institute, 1958.

[17] The model is: $f(Y) = \beta E + \varepsilon$, $C = E + u$ and $P = E + v$. We drop all assumptions about u, but v is independent of u and E, and ε is independent of u. Then $E(E|C) = E(P|C)$ so $E(f(Y)|C) = \beta\, E(P|C)$.

[18] The assumption of uncoorelated reporting errors can be relaxed as well. Assume $v|\,E_i \sim N(o,V(v_i)$ and $w\,|E_i \sim N(o,kV(v_i)$, $C-E = v+w$, and $P-E = u+w$. Then corrected bias estimates can be obtained by multiplying our tabulated estimates by $(1+k)$.

[19] When the relationship between early test scores and college graduation in a sample used for estimating an earnings function and the national population are not the same, the adjustment calculated above will not be equal to the proportionate reduction in education's coefficient that occurs when ability is entered.

A number of studies have measured the effect of an early IQ test scores on later earnings of college graduates. We record for a few of these studies the effect of 15 IQ points on earnings. 6.17 for 1958 Wisconsin high school seniors and 24-28 who had graduated from college; Janet Fisher, Kenneth Lutterman and Dorothy Ellegaard "Post High School Earnings; When and for Whom does Ability Seem to Matter" SSRI workshop paper 7312, University of Wisconsin. 7.8 percent for all 16-26 year old whites in the 1966 Parnes data; Charles Link and Edward Rattedge, "Social Returns to Quantity and Quality of Education," 4.8 percent for both age 33 and 47 in NBER-Thorndike sample, and 13.8, 10.5, 8.0 and 11.1 for 44, 39, 34 and 29 for Rogers sample; John C. Hause, "Earnings Profile: Ability and Schooling," Journal of Political Economy, 80:3 part 2 pp. S108-138.

[20] Dorothy Projector and Judity S. Bretz, "Measurement of Transfer Income in the Current Population Survey," paper given at Conference on Research in Income and Wealth, National Bureau of Economic Research, October 1972.

[21] Jacob Siegel, "Estimates of Coverage of the Population by Sex, Race and Age in the 1970 Census", Demography, 11:1, Feb/74, p. 1-23. and Jacob Siegel "Completeness of Coverage of the Nonwhite Population in the 1960 Census and Current Estimates and some Implications" Social Statistics and the City, ed. David Heer, Report of Conference June 1967, Joint Center for Urban Studies of MIT & Harvard.

[22] Roger Herriot and Herman P. Miller, "Who Paid the Taxes in 1968," Bureau of the Census, mimeo.

[23] The maximum taxable wage and tax rate for unemployment insurance are the weighted averages of differing state rates.

[24] No adjustments was required for changes in maximum taxable wage for $4,800 and $7,800 appear at the same point in the income distribution in their respective years.

[25] Returns to investments in physical capital (i.e., highways) must be corrected in the same manner.

[26] Arnold C. Harberger, "Monopoly and Resource Allocation," American Economic Review, May 1954, p. 77-87. Regressions run by Michael Klass on 1958 profit after tax plus interest over assets had 2.22 percent as an intercept when concentration ratio is the sole dependent variable. When the independent variable is the concentration ratio only when it exceeds 30 the intercept was 3.04. The mean profit rate was 3.19 percent. None of the industries had concentration ratios below 12.7. Evaluated at a concentration r ratio of 15 the first model predicts 2.9 percent. If this were interpreted as the measure of the normal profit rate monopoly profits are only 10 percent of total profits. See Michael Klass, Inter-Industry Relations and the Impact of Monopoly, Ph.D. dissertation, University of Wisconsin, Madison, 1970.

[27] While the adjustment implied by the one-third assumption is small, this is not an inconsequential issue. If a large part of what is accounted as a return to capital is really return to market power, and the incremental investments, the rate of return on physical capital to which investments in human capital are compared is substantially reduced.

[28] In industries facing long run economies of scale (a production function of degree greater than one), the sum of the marginal products is greater than the average product. The firm, however, cannot pay to its factors more than its total revenue. With average cost pricing the firm must pay the worker less than the price of the product times his marginal physical product. Only if the firm uses a multi part tariff so that the marginal price is lower than the average price, can the consumer price be lowered to a point where the marginal value product equals the wage. The industries with the strongest economies of scale do price in this manner (and may in fact have marginal prices below long run marginal cost according to the Averich-Johnson literature), so no adjustments will be made for this.

[29] Whether extra use of these programs places a burden on other taxpayer depends upon whether taxation is a way of forcing us to pay for something we are provided free or whether it is a way of buying the externalities others produce by using the government programs. In the former case it is approximate to add the extra use of these programs to the after tax earnings changes when calculating private returns. If in the latter case the marginal subsidy of a given service exactly equals the marginal externality benefit produced by the extra use of the service that is induced by higher earnings, no addition is required. If on the other hand, we subsidize education or health because we feel no one should be denied these things simply because they lack funds, there might be zero externality benefit from extra education or health services received as a consequence of an individual being richer. Thus services that are effectively in-kind transfers, should be added to earnings to calculate private return. Which type education is a debatable issue.

[30] State and local benefits per family have a slope of .032 to .036 when regressed on family income. (All programs are treated as if they create zero marginal externalities when extra use is induced by higher income.)