Volume Title: Income Inequality: Regional Analyses within a Human Capital Framework

Volume Author/Editor: Barry R. Chiswick

Volume Publisher: NBER

Volume ISBN: 0-870-14264-X

Volume URL: http://www.nber.org/books/chis74-1

Publication Date: 1974

Chapter Title: The Schooling Model

Chapter Author: Barry R. Chiswick

Chapter URL: http://www.nber.org/chapters/c3671

Chapter pages in book: (p. 31 - 47)

PART B

Income as a Function of Schooling

3

The Schooling Model

The theoretical model on which Part B rests, relating labor market income (earnings) to years of schooling, is presented in the following pages. We note its statistical characteristics and see how it can be used to analyze interregional differences in income inequality.

An individual's earnings are assumed to depend upon the earnings he would receive without any training, on his dollar investment in training, and on the rate of return received from his investment. Training is defined in terms of both years of formal schooling completed and years of labor market experience.

Further, the chapter is devoted to methods of computing the explanatory power of schooling for the two levels of aggregation in the income-schooling relationships that form the basis of our discussion. The first is intraregional, and relates the natural logarithm of earnings to an individual's level of schooling and to his average rate of return from investments in schooling. The second is interregional, and relates a measure of the relative variance of earnings in a region to the variance of schooling and the level of the rate of return from schooling in that region. Thus, the schooling model is used to explain (a) *individual* differences in earnings within regions via years of schooling; and (b) *regional* differences in the relative inequality in earnings via the rate of return from, and the inequality in the years of, schooling.

The schooling model shows, under simplifying assumptions, how the natural log of labor market income can be expressed as a

TABLE 3-1

Annual Earnings during and after Training

Years of Training	Year							
	1	2	3	...	$N-1$	N	$N+1$	∞
0	E_0	E_0	E_0	...	E_0	E_0	E_0	E_0
1	0	$E_0(1+r)$	$E_0(1+r)$	...	$E_0(1-r)$	$E_0(1+r)$	$E_0(1+r)$	$E_0(1+r)$
2	0	0	$E_0(1+r)^2$	...	$E_0(1+r)^2$	$E_0(1+r)^2$	$E_0(1+r)^2$	$E_0(1+r)^2$
...				...				
N	0	0	0	...	0	0	$E_0(1+r)^N$	$E_0(1+r)^N$

Note: E_0 = earnings in the absence of investment in training;
0 = zero earnings in years in which investment is undertaken;
r = rate of return on investment in training,
N = number of years of training.

linear function of the product of two terms: the individual's level of schooling, measured in years, and the (adjusted) average rate of return on the investment in his schooling. If the rate of return is assumed constant across individuals, the variance of the natural log of income across individuals in a region (for example, a state) becomes a function of the square of the rate of return from schooling and the variance in years of schooling in that region. Since average rates of return from schooling are not readily available on a regional basis, they are computed from a linear regression of the natural log of income on schooling with individual or microdata within each region.

THE THEORETICAL MODEL

Individual Earnings Function

Let us designate the perpetual annual earnings after N years of training as E_N, and the perpetual earnings if there were no training as E_0.[1] It is assumed initially that all persons are of equal ability, that the only private costs of training are forgone earnings, and that during the training period there are no earnings. With these assumptions, Table 3-1 will help clarify the derivation of the relation between training and earnings.

A person without training would earn E_0 every year, as shown in row 1 of Table 3-1. A person who invests in training for one year is assumed to have forgone the amount E_0; that is, no earnings were received during that year. This is shown by the zero in the second row of the first column. If a rate of return of r were received on his investment, he would earn $E_1 = E_0 + rE_0 = E_0(1 + r)$ in year two and all subsequent years, where rE_0 is the perpetual return on the investment E_0. This is shown in the second row of Table 3-1. If the rate of return were the same for all years of training, a person with two years of training would have received no earnings during years one and two, and after that an amount equal to

$$E_2 = E_0 + r(E_0) + r(E_0 + rE_0) = E_0(1 + r)(1 + r) = E_0(1 + r)^2,$$

where $r(E_0 + rE_0) = rE_1$ is the perpetual annual return to the investment in the second year of $E_0 + rE_0 = E_1$. A person with N

1. The assumption that earnings do not rise with age greatly simplifies the model. The effects of a rise in earnings with age on the analysis of schooling data are discussed below.

years of training would receive nothing during the first N years and

$$E_N = E_0 + r(E_0) + rE_0(1 + r) + \cdots + rE_0(1 + r)^{N-1},$$

or

$$E_N = E_0(1 + r)^N \tag{3-1}$$

after the investment period.

If the rate of return were not the same for all years of training, the product terms in the equations above could not be combined, and the postinvestment income stream would be represented by

$$E_N = E_0 \prod_{j=1}^{N} (1 + r_j), \tag{3-2}$$

where Π is the mathematical symbol for multiplication.

The assumptions that there are no direct costs of training and no earnings during the period of investment are not realistic. A year of schooling ordinarily leaves the summer free for working, and, for some levels of schooling, direct costs (tuition, school supplies, and other expenses necessitated by schooling) are far from negligible. Those engaged in on-the-job training usually receive positive incomes in excess of direct costs, in contrast to the past, when payments for the privilege of being in an apprenticeship program were quite common.

Let C_j equal the direct plus forgone-earnings costs of the investment in the j^{th} year of training. E_{j-1} is the income which would be received after $j - 1$ years of training if no further investments were undertaken. Furthermore, let us designate by K_j the ratio C_j/E_{j-1}. That is, K_j equals the proportion of potential income during year j that is invested. We previously assumed that the only cost of education was a full year of forgone earnings, so that $C_j = E_{j-1}$ and $K_j = 1$. Now K_j may differ from unity. If the total costs of the investment were greater than potential earnings during the year of training, K_j would be greater than one. If the potential earnings exceeded the total costs, K_j would be less than one.

The introduction of the investment-income ratio, K, modifies the earnings equation. If there were no investment, E_0 would still be earned. If, in year one, the amount $C_1 = K_1 E_0$ were invested at a rate of return of r_1, the postinvestment income would be

$$E_1 = E_0 + r_1(K_1 E_0) = E_0(1 + r_1 K_1). \tag{3-3}$$

If N years of investments were undertaken,

$$E_N = E_0 \prod_{j=1}^{N} (1 + r_j^*), \tag{3-4}$$

where $r_j^* = r_j K_j$ is the "adjusted" rate of return to the j^{th} year of education.

Individual differences may be introduced into equation (3-4) by the inclusion of a residual U_i^* and an allowance for differences in rates of return to a given level of training. The earnings equation then becomes

$$E_{N,i} = E_0 \prod_{j=1}^{N} (1 + r_{ij}^*)\, U_i^*, \tag{3-5}$$

where r_{ij}^* is the adjusted rate of return to the i^{th} individual for the j^{th} year of training. Taking logarithms of both sides of equation (3-5) and using the relation $\ln (1 + a) \approx a$ when a is small results in

$$\ln E_{N,i} = \ln E_0 + \sum_{j=1}^{N} r_{ij}^* + U_i, \tag{3-6}$$

where $U_i = \ln U_i^*$ and the "approximately equal to" sign has been replaced by the symbol for "equal to."[2]

Earnings Inequality

Since the purpose of this study is to analyze the effects of regional differences in schooling on regional differences in inequality of income, we must first select a measure of income inequality. Let it be stated at the outset that no one measure is ideal.[3] In referring to several measures of inequality, Lydall writes: "As has frequently been pointed out, they are not unambiguous

2. Rates of return (r) tend to range from 5 per cent to 20 per cent, and K generally does not greatly exceed unity. Hence, rK is sufficiently small to keep the approximation appropriate. Individual differences in the zero investment level of earnings may be considered to be in the residual.

3. This problem is not unique to the dispersion of a distribution—there are several measures of a central tendency, i.e., the mode, the median, the arithmetic mean, and the geometric mean. The ranking of a series of distributions by a measure of central tendency will be sensitive to the measure selected if the distributions have different shapes. There are also several different ways of measuring skewness, the asymmetry of a distribution.

indicators of the degree of inequality, since distributions of various shapes may have the same concentration coefficient. If the distribution is exactly log normal, of course, this problem does not exist; and we can measure the degree of relative dispersion either by the coefficient of concentration, or by the standard deviation of the logarithm of income, or by several other coefficients. . . .The use of a single index of inequality is, therefore, not an ideal arrangement, except where one can be fairly confident that the essential shape of the distribution, i.e., its functional form, is constant."[4] Although the distribution of labor market incomes within regions is not precisely log normal, it does have a universal positive skewness.[5]

The measure of dispersion I use in this study is the variance of the logarithm of income—the square of the standard deviation of the log of income. It has several advantages. First, it is a measure of *relative* inequality and is therefore devoid of units. This permits a comparison of income inequality across regions even if the measuring units (U.S. dollars, Canadian dollars, et cetera) differ. Second, there is probably more social concern about relative income inequality than about absolute income inequality. For example, if all incomes and all prices doubled, there would be no real change in relative wealth or in the equity of the income distribution. The variance in logs would, in fact, remain unchanged. Yet the absolute variance of income would quadruple.[6] These two reasons argue for a measure of relative inequality, but not necessarily for the variance in logs.

There is, however, a third advantage, which dictates the use of the variance in the log of income as the measure of dispersion. The human capital model I use here relates income to investments. When investments are measured in dollars, the appropriate measure of income is also in dollars.[7] However, when investments are measured in *time equivalents* (such as years of schooling, years of

4. See Harold Lydall, *The Structure of Earnings*, Oxford, Clarendon Press, 1968, pp. 137-138. For a discussion of various measures of income inequality, see also Mary Jean Bowman, "A Graphical Analysis of Personal Income Distribution in the United States," *American Economic Review*, September 1945, pp. 607-628.

5. See, for example, H. P. Miller, *Income of the American People*, New York, John Wiley, 1955, p. 3; and Barry Chiswick, "An Interregional Analysis of Schooling and the Skewness of Income," in W. L. Hansen, ed., *Education, Income, and Human Capital*, New York, NBER, 1970.

6. If the variance of X is Var(X), the variance of $2X$ is Var($2X$) = 4 Var (X). The variance of the log of $2X$ is Var(ln$2X$) = Var(ln$X$).

7. See Gary Becker, *Human Capital*, New York, 1964, Chapter III.

labor market experience), the appropriate measure of income is the natural log of income. This latter measure is a linear function of years of investment. Taking the variance of both sides of the relation, it is the variance of the natural log of income that is related to the variance in years of schooling. While data on dollar investments in human capital are very scarce, data on time-equivalent investments in one form of human capital, namely schooling, abound. Since the availability of data, as mentioned before, determines much of the form of this analysis, the measure of dispersion—the variance of the natural logarithm of income—is related to investments in schooling measured as years of schooling completed.

Effect of Training on Earnings Inequality

To find the effect of training on the inequality of earnings, the variance of both sides of equation (3-6) is computed. This results in

$$\text{Var}(\ln E) = \text{Var}\left(\sum_j r_j^*\right) + \text{Var}(U) + 2\,\text{Cov}\left(U, \sum_j r_j^*\right), \quad (3\text{-}7)$$

where Var means variance and Cov means covariance.

The sum of the adjusted rates of return $\Sigma_j r_{ij}^*$ can be rewritten as

$$\sum_j r_{ij}^* = \bar{r}_i^* N_i = \bar{r}^* N_i + (\bar{r}_i^* N_i - \bar{r}^* N_i),$$

where $\bar{r}_i^*$ is the i^{th} person's average adjusted rate of return and $\bar{r}^*$ is the average $\bar{r}_i^*$ for the population.[8] If it were assumed that

8. In mathematical terms,

$$\bar{r}_i^* = \sum_{j=1}^{N_i} \frac{r_{ij}^*}{N_i} = \sum_{j=1}^{N_i} \frac{r_{ij} K_{ij}}{N_i},$$

where N_i is the number of years of training and

$$\bar{r}^* = \sum_{i=1}^{p} \frac{\bar{r}_i^*}{p},$$

where p is the size of the population.

deviations from the population's average adjusted rate of return appear in the residual U', $[U_i' = U_i + (\bar{r}_i^* - \bar{r}^*)N_i]$, equation (3-6) could be rewritten as

$$\ln E_{N,i} = \ln E_0 + \bar{r}^* N_i + U_i' \tag{3-8}$$

and

$$\text{Var}(\ln E) = (\bar{r}^*)^2 \text{ Var}(N) + \text{Var}(U') + 2\bar{r}^* \text{ Cov}(U', N). \tag{3-9}$$

The model generates as a parameter a commonly used measure of relative inequality, the variance of the log of earnings. This measure of income inequality is related to the rate of return from, and the inequality of investments in, years of training (see equation 3-9). The statistical analysis developed below rests on this equation.

STATISTICAL IMPLEMENTATION OF THE MODEL

Data on money investments in schooling are scarce. There is, however, considerable information on the number of years of schooling. For this reason it is the measure chosen for the empirical analysis in Part B, despite the fact that it masks the effect of differences in the money cost and quality of a given level of schooling. There is also considerable public interest in the role of years in school in determining the distribution of income. A model which explicitly includes postschool training is presented and analyzed in Part C of this volume.

Years of Schooling

Years of training can be separated into two components, schooling and postschool (on-the-job) training. Thus, the earnings function (equation 3-6) becomes

$$\ln E_{N,i} = \ln E_0 + \sum_{j=1}^{S_i} r_{S_{ij}}^* + \sum_{j=1}^{J_i} r_{J_{ij}}^* + U_i, \tag{3-10}$$

where S_i and J_i are the number of years of schooling and postschool training, respectively, and N_i equals $S_i + J_i$.

If schooling were the only explanatory variable, the relevant earnings equation would be

$$\ln E_{S,i} = \ln E_0 + \sum_{j=1}^{S_i} r_{S_{ij}}^* + U_{S,i}. \tag{3-11}$$

Since $\bar{r}_i^*$ was previously defined as the i^{th} person's average adjusted rate of return, equation (3-11) could be rewritten as

$$\ln E_{S,i} = \ln E_0 + (\bar{r}_i^*) S_i + U'_{S,i}. \quad (3\text{-}12)$$

For simplicity's sake, let us temporarily neglect individual differences in the residual $U'_{S,i}$. Then, if we calculate the variances of both sides of equation (3-12) we obtain

$$\text{Var}(\ln E) = \text{Var}(\bar{r}_i^* S_i). \quad (3\text{-}13)$$

Thus, the relative variance of income depends on the absolute variance of the product of the adjusted rate of return and the number of years of schooling.

Schooling and the Rate of Return

The variance of the product of two independent random variables, $\bar{r}_i^*$ and S_i, can be expressed as[9]

$$\text{Var}(\bar{r}_i^* S_i) = \bar{r}^{*2} \text{ Var}(S_i) + \bar{S}^2 \text{ Var}(\bar{r}_i^*) + \text{Var}(S_i) \text{ Var}(\bar{r}_i^*). \quad (3\text{-}14)$$

Thus, if $\bar{r}_i^*$ and S_i were independent, the relative variance of income would be positively related to both the average level and the variance of each of the two variables.[10] There are theoretical reasons which make this assumption plausible.

With wealth held constant, those with higher marginal rates of return for a given level of schooling have a greater incentive to invest. This implies a positive correlation between schooling level and rate of return. For a given level of ability, those with greater wealth have a lower discount rate and therefore a greater incentive to invest. As a consequence, they receive a lower rate of return. This implies a negative correlation between schooling level and rate of return. Thus, using a priori analysis, the sign of the correlation between the average rate of return to an individual and his level of schooling is ambiguous.[11] Empirically, Mincer has shown that the rate of return from schooling is uncorrelated with the person's level of schooling (holding experience and weeks worked in the year constant).[12]

---

9. Leo Goodman, "On the Exact Variance of Products," *Journal of the American Statistical Association*, December 1960, pp. 708-713.

10. This differs from Lydall's view that, to explain income dispersion, "what matters is the *inequality* (*sic*) of environment and education, not its average level." (See Lydall, *The Structure of Earnings*, 1968, p. 10).

11. See Gary Becker, *Human Capital and the Personal Distribution of Income*, Ann Arbor, 1967.

12. Mincer, *Schooling, Experience, and Earnings*, Part 2.

The variance of a product of two variables that are not independent can be evaluated,[13] but this is not necessary for the present purpose. The foregoing implies that the intraregional relative variance of income is positively related to the average levels and variances in both years of schooling and rates of return from schooling, even if, for individuals, the level of schooling and the rate of return are not perfectly independent. Data on regional differences in the average rate of return are scarce, but a procedure used in this study permits the computation of estimates for many regions. Data on regional differences in the variance in rates of return are nonexistent. Thus, equation (3-14) is of restricted applicability in an empirical analysis.

Returning to equation (3-12), substituting the population's average adjusted rate of return from schooling into the equation and placing deviations from this population average into the residual, $U'' = U' + (r_i^* - \bar{r}^*)S$,

$$\ln E_{S,i} = \ln E_0 + \bar{r}^* S_i + U''_{S,i}. \tag{3-15}$$

Then,

$$\text{Var}(\ln E_{s,i}) = (\bar{r}^*)^2 \text{ Var}(S) + \text{Var}(U'') + 2\bar{r}^* \text{ Cov}(S, U''). \tag{3-16}$$

Although some average rates of return from schooling have been calculated in recent years, their number is still very small, and relying on the income inequality formulation presented in equations (3-14) and (3-16) would severely limit an empirical analysis. Therefore, rates of return from schooling are computed specifically for this study for many regions. Two methods of estimation, named for their method of computation, are employed: the "regression estimate" of the rate of return, computed for the United States, Canada, Mexico, and Puerto Rico; and the "overtaking age" estimate of the rate of return, computed indirectly only for the states of the United States.[14]

Biases in Computing Regression Rates of Return

The regression estimate of the rate of return is obtained via equation (3-15) by regressing the natural logarithm of earnings on

13. Goodman, "On the Exact Variance of Products," 1960.

14. The procedure used to compute the "overtaking age" estimate is discussed in Appendix A-2. Because it is computed indirectly it is subject to considerable measurement error. Mincer developed a shortcut for estimating the overtaking age rate of return; see his *Schooling, Experience, and Earnings.*

years of schooling completed:

$$\ln E_{S,i} = (\widehat{\ln E}_0) + \hat{r} S_i + \hat{U}_i, \tag{3-17}$$

where $\hat{r}$ and $(\widehat{\ln E}_0)$ are the least-squares linear regression estimates of the average adjusted rate of return from schooling and the zero schooling level of earnings, respectively, and $\hat{U}$ is the residual. Unbiased estimates of the adjusted rate of return are obtained when S_i and U_i' from equation (3-15) are uncorrelated, that is, when schooling and the omitted variables are uncorrelated and when there are no errors of measurement. The residual contains the effects of differences in luck, tastes, ability, investments in human capital other than schooling, and wealth. By definition, luck is uncorrelated with schooling. There is reason to believe, however, that the other variables may be correlated with schooling. For example, we would expect a positive, although not perfect, correlation between schooling and family wealth.

Due to the secular increase in schooling, those with low levels of schooling tend to be older and are receiving their return on earlier investments in postschool training. Thus, a regression of the log of earnings on years of schooling in which all age groups are pooled results in a downward-biased estimate of the slope coefficient of schooling, and hence of the regression estimate of the rate of return. The downward bias would not be fully eliminated by restricting the regressions to specified age groups. For a given age, an additional year of schooling implies one year less of experience (investment in postschool training). Since years of schooling and of experience are negatively correlated in the cross section, the omission of experience from the regression equation results in a downward bias in the slope coefficient of schooling.

An individual's dollar investments in health and migration are positively correlated with his level of schooling.[15] This generates a positive correlation between years of schooling and the component of the residual reflecting the money return from those forms of capital, since it is unlikely that their rates of return are sufficiently (if at all) negatively correlated with the level of schooling.

Another component in the residual is differential ability, as

15. See Michael Grossman, *The Demand for Health: A Theoretical and Empirical Investigation*, NBER, 1972; Selma Mushkin, "Health as an Investment," *Journal of Political Economy*, October 1962, pp. 129-157; Rashi Fein, "Educational Patterns in Southern Migration," *Southern Economic Journal*, July 1965, pp. 106-124; and June O'Neill, "The Effect of Income and Education on Inter-Regional Migration," Ph.D. dissertation, Columbia University, 1970.

reflected in differences in the rate of return from investments in schooling. For simplicity of presentation, let us assume that $K_i = 1$ for all i in equation (3-12). Then, for each individual,

$$\ln E_{S,i} = \ln E_0 + \bar{r} S_i + (d_s S_i + U'_{S,i}), \qquad (3\text{-}18)$$

where $(d_s S_i + U'_{S,i})$ is the residual and $d_s = (\bar{r}_i - \bar{r})$ is the difference between the rate of return received by the i^{th} person and the average rate of return. Since the expected value of d_s is zero, S and $d_s S$ would be uncorrelated and there would be no bias from this source if S and d_s were independent of each other.[16] It is not clear a priori or empirically whether d_s and S are positively or negatively correlated.[17]

Errors in measurement of the variables may also bias the regression coefficient. If there were random errors in both S and $\ln Y$, and these errors were uncorrelated, the effect would be a downward bias. If these errors were positively correlated (for example, if in sample data there were a tendency for those who overreport their level of schooling to overreport their earnings), the effect would be unclear a priori.[18]

To summarize, it appears that wealth, migration, and health are positively correlated with earnings—with schooling held constant—and positively correlated with schooling. Thus, omitting these variables biases the slope coefficient of schooling upward. The effect of ability and errors of measurement is unclear. Omit-

16. It is not sufficient for S and d_s to be uncorrelated. If $\bar{r}_i$ is uncorrelated with S_i, Cov $(d_s,S) = 0$. We know that $E(d_s) = 0$. Then Cov $(d_s,S) = E[d_s (S - \bar{S})] = E(d_s S) = 0$. Thus, Cov $(S,d_s S) = E[(S - \bar{S})\,(d_s S)] = E(d_s S^2) - \bar{S}\,E(d_s S) = E(d_s S^2)$.

Hence, Cov $(S,d_s S) = 0$, if Cov $(d_s,S^2) = 0$, which necessarily holds when d_s and S are independent.

17. If ability were measured by the average rate of return from schooling for a given level of schooling, and if average and marginal rates of return were positively correlated, then (since those with higher marginal rates of return have a greater incentive to invest in schooling) there would be a positive relation between ability (d_s) and schooling (S) if all individuals had the same supply curve of funds. If, however, differences were greater in supply conditions than in demand conditions, and if those with greater ability had lower investment costs, d_s and S would be negatively related.

18. Let $y = a + b\,x + U$ be the true relation and let $X = x + w$ and $Y = y + v$ be the observed values, where w and v are normally distributed with zero expectation and constant variance, and are independent of x, y, and U. Then $Y = a + bX + (U - bw + v) = a + bX + z$, where $z = U - bw + v$ and $E(z) = 0$. Since $E(X) = x$, $w = X - E(X)$, and Cov $(X,z) = E[(X - E(X))(z)] = E[(w)\,(U - bw + v)] = -b$ Var (w) + Cov (w,v). If x and y were positively correlated, a sufficiently strong positive correlation between w and v could produce an upward bias of the slope coefficient.

ting years of experience biases the slope coefficient downward because the variable is negatively correlated with schooling in cross-sectional data.[19] Indeed, in empirical work, regression-estimated rates of return (when schooling is the only explanatory variable) are lower than directly estimated internal rates of return and lower than regression-estimated rates of return, with investments in experience held constant.[20]

The downward bias notwithstanding, these estimates are employed in the empirical analysis that follows because experience cannot be held constant for the data sets studied.

Schooling as an Explanatory Vehicle

Taking the variance of both sides of regression equation (3-17) results in

$$\text{Var}\,(\ln E_{s,i}) = (\hat{r})^2\ \text{Var}\,(S_i) + \text{Var}\,(\hat{U}_i) \qquad (3\text{-}19)$$

for each region.[21] The ratio of the explained variance ($\hat{r}^2$ Var S) to the total variance (Var(ln E)), called the "coefficient of determination," indicates the proportion of the variability in the dependent variable in that region "explained" by the variation in the explanatory variables. If the regression slope coefficient were an unbiased estimate of the average adjusted rate of return and if the two variances were unbiased, the coefficient of determination would be an unbiased estimate of the explanatory power of schooling. A bias in the slope coefficient, however, produces a bias in the same direction in the coefficient of determination.

If an unbiased estimate of the adjusted rate of return ($\bar{r}^*$) were known, then equation (3-16) could be used to calculate

$$1 = \frac{(\bar{r}^*)^2\ \text{Var}\,(S)}{\text{Var}\,(\ln E)} + \frac{\text{Var}\,(U'')}{\text{Var}\,(\ln E)} + \frac{2\bar{r}^*\ \text{Cov}\,(S,U'')}{\text{Var}\,(\ln E)}\ . \qquad (3\text{-}20)$$

The first ratio indicates the direct explanatory power of schooling, the second measures the direct explanatory power of other variables, while the third shows the explanatory power of the covariation of schooling and these other variables. The latter would be

19. The negative correlation is due to both secular trends in schooling and the negative correlation between schooling and experience for a given age, as discussed above.

20. For a comparison of internal and regression rates of return, see Table 4-2. For an analysis of the effect on the regression estimate of including experience, see Table 5-3, and Mincer, *Schooling, Experience, and Earnings.*

21. The regression forces Cov(S,U) = 0 for each region.

positive (or negative) if, on balance, S and U'' were positively (or negatively) correlated. If the covariance term were zero [Cov $(S,U'') = 0$], the direct contribution of schooling would be the same as the total contribution and the regression estimate of the contribution.

For each region the coefficient of determination shows the proportion of differences in earnings within that region that can be explained by differences in schooling, given the regression estimate of the adjusted rate of return. We are also concerned, however, with the proportion of the variation across regions in the inequality of earnings [Var($\ln E$)] that can be explained by differences in the educational component of income inequality [$\hat{r}^2$ Var(S)] and the variation in "all other variables" [Var($\hat{U}$)]. There is no unique estimate of these values because of the covariation of the education component and the residual variance. We can, of course, estimate the separate explanatory powers of schooling, the residual, and the covariation of schooling with the residual. If for the i^{th} region

$$v_i = \text{Var}\ (\ln E)_i,$$

$$s_i = \hat{r}_i^2\ \text{Var}\ (S)_i,$$

and

$$t_i = \text{Var}\ (\hat{U})_i,$$

then from equation (3-19), $v_i = s_i + t_i$. Taking variances across regions,

$$\text{Var}(v) = \text{Var}(s) + \text{Var}(t) + 2\ \text{Cov}(s,t), \tag{3-21}$$

and the interregional explanatory power of

$$\left.\begin{array}{l} \text{(a) the education component is } \dfrac{\text{Var}\ (s)}{\text{Var}\ (v)}, \\ \text{(b) the residual variance is } \dfrac{\text{Var}\ (t)}{\text{Var}\ (v)}, \\ \text{and} \\ \text{(c) the covariation is } \dfrac{2\ \text{Cov}\ (s,t)}{\text{Var}\ (v)}. \end{array}\right\} \tag{3-22}$$

An alternative procedure for estimating the interregional explanatory power of schooling is to regress the variance of the log of income on the education component. Then the regression's adjusted coefficient of determination is schooling's interregional

explanatory power of income inequality. This procedure forces the covariance term in equation (3-22) to equal zero, and biases upward the education component's direct explanatory power of schooling.

Schooling is measured in years and is used in the same units for all regions.[22] The variances of the log of earnings [Var(ln$E$)] and the residual [Var($\hat{U}$)] in equation (3-19) are both pure numbers and can be compared across countries without converting to a common currency. In subsequent chapters these parameters are analyzed within and among regions in order to ascertain what patterns of relationships exist.

22. Note, however, that, while the unit is "years," the length and quality of a school year vary within and across regions. No attempt is made here to adjust for these differences.