Part II

EMPIRICAL ANALYSIS

3

Schooling and Earnings

3.1 QUANTITATIVE ANALYSIS

Human capital models have been employed in empirical analyses of income distributions in attempts to explain differences in level, inequality, and skewness of earnings of workers who differ by schooling and age, to interpret shapes of age-earnings profiles of individuals, and to explain differences in earnings distributions among regions and countries.[1] Though sketchy in many respects, these studies tend to provide at least qualitatively consistent interpretations of some of the apparently bewildering variety of features of income distributions.

There is as yet no evidence of quantitative explanatory power of the human capital model to match the promise indicated by the qualitative or comparative analyses. As yet, no serious attempts have been made at a full quantitative accounting of the effects of the distribution of investment in human capital on observed earnings inequality. The only available empirical estimates of the extent of in-

1. Mincer (1957, 1958, 1962), Becker (1964, 1967), Ben-Porath (1967), Chiswick (1967), Lydall (1968).

come inequality[2] that can be attributed to investments in human capital are limited to investments in schooling, measured by years of schooling.

Applying the "schooling model" (equation 1.3) in a simple regression of 1959 log earnings of men aged 25–64 in the experienced labor force on their years of schooling, Chiswick found coefficients of determination varying between 10–20 per cent within U.S. regions and states. The coefficients are 10 per cent and 18 per cent for earnings of white men in the non-South and South, respectively. Within states, Chiswick applied regressions to incomes of men aged 25 and over, instead of earnings, which were not available in the published 1960 Census data.

Low as they are, the coefficients of determination are overstated, because they are based on data grouped by income and schooling intervals. Application of the same regression to individual observations of 1959 earnings of all U.S. white, nonfarm, nonstudent males,[3] aged 15–64 yields a coefficient of determination of barely 7 per cent.

The inadequacy of the schooling model as an explanation of inequality, which is measured here by the variance of log earnings, is apparent not only in the low coefficients of determination but also in the small slope coefficients of the regression. According to equation (1.3) these coefficients are supposed to represent estimates of average rates of return on investments in schooling. But as Chiswick's data and my Table 4.4, below (first row) show, the regression slopes are substantially lower, almost half the size of internal rates calculated directly from age profiles by Becker, Hansen, and Hanoch.

The disappointing performance of the schooling model need not cast doubt on the relevance or importance of human capital analysis. As the discussion in Part I indicates, the schooling model represents an incomplete specification of human capital theory of the distribution of earnings. The model cannot adequately explain in-

2. Though the human capital model applies strictly to labor incomes, the empirical literature often describes total incomes rather than earnings.

3. These were males with some earnings in 1959. Earnings were defined as wages and salaries plus self-employment income, provided wages and salaries were the major source of earnings. The 1/1,000 sample of the 1960 U.S. population Census, used in this study, contained 31,093 men in this category. The earnings of over 95 per cent of them consisted of wages and salaries alone. This is the basic body of data used in our empirical analyses.

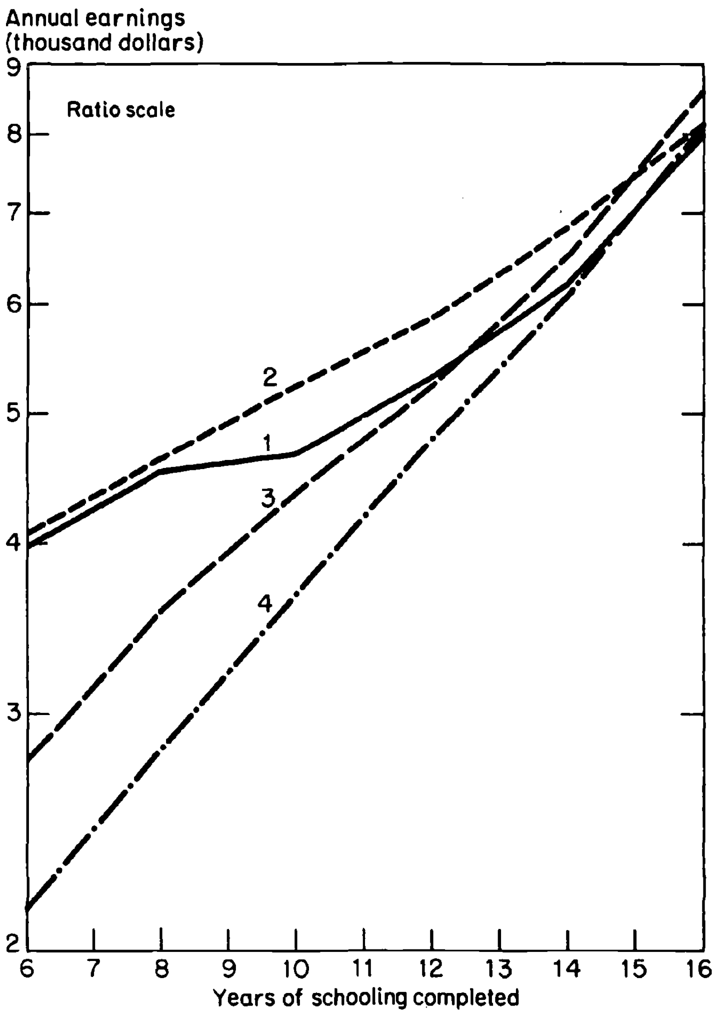
equality of earnings among individuals who differ not only in schooling but also in other behavioral characteristics including, in particular, other forms of investments in human capital. In the empirical analyses that follow, it will be seen that when the human capital model is expanded to include post-school investments its explanatory power is greatly increased. In the expanded model biases in the regression estimates of the schooling model are removed. Although the inclusion of an undifferentiated and indirect concept of post-school investments constitutes only an initial step toward a more complete analysis, it provides a unified interpretation for a variety of qualitative and quantitative aspects of the structure of earnings.

3.1.1 GROUPED DATA

Before proceeding to incorporate post-school investment behavior into the empirical analysis, it is useful to consider the applicability of the schooling model somewhat more closely. As we have seen, the schooling model is too blunt an instrument for analyzing the ungrouped distribution of earnings. Evidently, variation in earnings within schooling groups is a major part of total inequality. With grouped data, the positive relation between schooling and earnings does, of course, emerge clearly. Still, the model does not fit properly in one respect: The slope of line *1*, Chart 3.1, that is, the regression slope of average earnings (in logs) on years of schooling, is again too flat, as it was in the ungrouped regression. Apparently, grouping does does not eliminate the problem of within-group variation of earnings. These earnings have been averaged in each schooling group, but the average depends on the age distribution in the groups, given the existence of pronounced age-earnings profiles. As is well known, earnings at later stages of work experience are substantially higher than at early stages. Because of strong secular trends in schooling, average age is older in the lower schooling groups, younger in the higher schooling groups (Table 3.1, column 2, below). Consequently, earnings differentials among schooling groups, shown as the slope of line *1*, Chart 3.1, are understated. But, even if earnings of a fixed age group (e.g., age 32–33, line *2*, Chart 3.1) are compared, the downward bias in the slope is still not removed.

The basic reason for the persistent bias becomes intuitively apparent if it is assumed that the individual growth curve of earnings is

CHART 3.1
SCHOOLING AND AVERAGE EARNINGS, 1959
(schooling groups of white, nonfarm men)



NOTE:
Curve *1:* average earnings of all workers, age 15–64.
Curve *2:* average earnings at age 32–33.
Curve *3:* average earnings with 10 years of experience.
Curve *4:* average earnings with 7–9 years of experience.
SOURCE: 1/1,000 sample of U.S. Census, 1960. Estimates are shown in Table 3.1.

largely a function of post-school investment, such as on-the-job and other forms of training and experience. The earnings profile is a function of work experience rather than of age: Since less schooled persons enter the labor force earlier, they spend more time acquiring work experience; at a given age, they will reach higher relative levels of their earnings profiles than persons of the same age, but with more schooling. This is why earnings differentials are still understated in line *2*. On this post-school investment hypothesis, the more appropriate standardization is for years of experience rather than age. Empirical support for the argument is found in line *3* of Chart 3.1, where earnings are shown at ages corresponding to a decade after completion of schooling. The slope of line *3* is almost double that of lines *1* and *2*, and is indeed well within the usual range of directly calculated internal rates of return (about 12 per cent).

In the absence of direct information in the 1960 Census, years of work experience were measured by subtracting the age of completion of schooling from reported age. Average ages of school leaving were estimated by Hanoch (1968) from the same data (cf. Table 3.1, column 2). Conceptually, age is not irrelevant, since it is a factor in the depreciation of human capital stock. Separate estimates of age and experience effects on earnings require individual data on job experience. Such estimates as are available indicate that experience, rather than age, is the dominant factor in earnings.[4]

The intuitive argument in support of a standardization by years of experience does not indicate the particular stage of experience at which earnings of different schooling groups should be compared. But the decade of experience chosen for line *3* is not entirely arbitrary. The argument and evidence can be more rigorously stated, paying closer attention to the concepts implicit in the schooling model (1.3):

$$\ln Y_s = \ln Y_0 + rs.$$

Implementation of this model is a problem not only because the variation in earnings within schooling groups is omitted, but also because data for the (dependent) earnings variable are not available. According to the derivation of the schooling models, Y_s represents a hypothetical concept of earnings a person would receive after completion of schooling, if he did not incur any further growth-producing

4. See discussion in Chapter 4, below.

TABLE 3.1
SCHOOLING, AVERAGE ANNUAL EARNINGS, AND RATES OF RETURN, 1959
(U.S. white, nonfarm men)

		Age at First Year of Experience	Average Annual Earnings			At Overtaking		Rate of Return	
Years of Schooling (1)	Median Age (2)	(3)	All Ages (4)	At Age 32–34 (5)	In 10th Year of Experience (6)	Earnings (Y_s) (7)	Year (8)	Used (r) (9)	Implicit (r_s) (10)
0–4	52	14	$3,350	$3,390	$ 2,520	$1,910	7		
5–7	48	14	4,000	4,070	2,740	2,130	7	(17)	
8	45	16	4,520	4,600	3,580	2,830	7	16	19.6
9–11	36	18	4,660	5,250	4,360	3,660	7–8	(14)	
12	34	20	5,330	5,870	5,280	4,800	8	13	13.2
13–15	33	23	6,240	6,850	6,520	6,100	8–9	(11)	
16	35	25	8,020	8,160	8,600	7,950	8–9	10	10.1
17+	37	28	9,200	8,710	10,200	9,900	9	(9)	7.3

NOTE:

Col. 3: Estimates of Hanoch (1967): Mean age at the terminal school year plus 1. These estimates were modified in the lowest two groups by the assumption that boys did not enter the labor force before the age of fourteen. Also, an average of five rather than six years was estimated as the average duration of college studies.

Cols. 4–7: 1/1,000 sample of U.S. Census, 1960.

Cols. 7–8: Uses estimate of r in column 9 to equate the present values of Y_s in column 7 with the present values of the observed profiles.

Col. 9: Values in parentheses are extrapolated.

Col. 10: $r_s = (\ln Y_{s_2} - \ln Y_{s_1})/\Delta s$.

self-investments. Values of Y_s are not observable, but as was shown in Chapter 1, they can be approximated if certain assumptions are accepted.

The two basic assumptions are that rates of return to schooling are not very different from rates of return to post-school investment, and that earnings profiles Y_s with no further (net) investment remain largely flat for most of the working life. Both assumptions are empirical, and some evidence in their support is considered in later discussion.

Recall the expanded earnings function (1.4):

$$Y_{sij} = Y_{si} + \left(r \sum_{t=0}^{j-1} C_{ti} - C_{ji}\right),$$

Here Y_{sij} denotes net earnings of person i with s years of schooling who is in his jth year of work experience; C_j, dollar costs of post-school investments; and r, rates of return to post-school investments. Since the first expression on the right is Y_{si}, gross earnings after completion of schooling, it equals the observed earnings $Y_{si\hat{j}}$ at the stage of experience $j = \hat{j}$, when the second right-hand term is equal to zero.

As was demonstrated in equation (1.10), $\hat{j} < 1/r$. If r is not very different from the rates of return as usually calculated, the "overtaking" year of experience at which observed earnings $Y_{si\hat{j}} = E_s$ should be a decade or less. As a rough guess, earnings at ten years after completion of schooling were used in line *3* of Chart 3.1.

A more direct approach is to estimate Y_s as that amount of annual earnings in a constant income stream whose present value equals the present value of the actual earnings profile. The present values must be taken at the start of working life, and the rate of return is used as the rate of discount. Such estimates of earnings Y_s are utilized in line *4* of Chart 3.1. Its slope is somewhat steeper than that of line *3*, because higher rates of return, hence earlier "overtaking" years of experience and lower earnings than in line *3*, were assigned to the lower schooling groups.[5] The overtaking years of experience run from 7 in the lower to 9 in the higher schooling groups.

The earnings figures (Y_s), the estimated years at overtaking ($\hat{j}$), internal rates (r) used for estimating them,[6] and the slopes of the lines (r_s), are shown in Table 3.1 (columns 7, 8, 9, and 10, respectively). Note that the slope r_s in the schooling model (1.3) is an estimate of the rate of return *to schooling only,* while the rate as usually calculated (r) from age profiles, although often called a rate of return to education or schooling, is a rate on a mix of schooling and post-

5. The causes of the differences in slopes of the four lines in Chart 3.1 are perhaps best visualized by inspection of Chart 4.3 in Chapter 4, which shows the age profiles of log earnings in the several schooling groups. The slope of line *1* corresponds to the vertical distance (per school year) between points at mean ages; the slope of line *2* corresponds to the distance *ABC* at age 33; while the slopes of lines *3* and *4* correspond to the distances between the estimated overtaking points ($A^1B^1C^1$). The last is the best estimate. It is necessarily the steepest.

6. These rates were calculated from the earnings profiles shown in Charts 4.1 and 4.2. Direct costs and student earnings were conveniently ignored in the calculation, on the assumption of their rough cancellation at higher levels and unimportance at lower levels. In this I follow Hanoch (1968). The assumption is not tenable in general, but rough estimates suffice for the present analysis.

TABLE 3.2
SHORT-CUT AND STANDARD ESTIMATES OF RATES OF RETURN [a]
TO SCHOOLING, 1939, 1949, 1958
(U.S. white, nonfarm men)

	High School		College [b]	
Year	Short Cut (1)	Standard (2)	Short Cut (1)	Standard (2)
1939	15.1%	12.5%	12.8%	11.0%
1949	13.5	11.8	10.6	10.6
1958	14.4	15.1	10.2	11.5

SOURCE: Mincer (1962).

a. For 1939, data are based on earnings; for 1949 and 1958, based on income.

b. For 1939 and 1949, refers to persons having more than sixteen years of schooling; for 1958, sixteen years.

school investment. It is a weighted average of the rates on schooling (r_s) and on post-school investments (r_p).[7] In constructing lines *3* and *4* of Chart 3.1, it was assumed that r_s and r, hence r_s and r_p, do not differ. A rough check of consistency appears in the results in Table 3.1. A comparison of r_s estimated by the slopes of line *4* and of the r utilized for its construction shows them to be very close at the college and high school levels (Table 3.1, columns 9 and 10). The small discrepancy at these levels suggests that it is not misleading to label internal rates of return calculated from earnings profiles as "rates of return to education." Lines *3* and *4* are not only steeper, but also straighter than line *1*. Evidently, the closer the correspondence of the data to the concepts of the model, the better the empirical fit.[8] Actually, linearity is not required by the model, since r may differ by level of schooling. Nonetheless, the broken shape of line *1* is more likely to reflect a bad fit than erratically different values of r.

The experiments reported above indicate that although the schooling model is incomplete, it is relevant to the analysis of

7. Cf. Becker (1964, p. 42).

8. Lydall (1959) attempted to test the "goodness of fit" of the semilog form of the schooling model, using line *1*. This, as we have seen, is not the most appropriate test. Nevertheless, he would not have rejected the model had he not mistakenly used a double-log form in his test (p. 95).

earnings differentials. Moreover, its proper empirical implementation gives rise to a useful by-product: a quick and easy, though rough, method of assessing rates of return to *schooling*. Data for fewer than the first ten years of earnings are needed for the purpose, a major advantage in up-to-date analysis, compared to procedures which require information on a whole working life of earnings.

Table 3.2, column 1, shows estimates of rates of return to *schooling* calculated by assigning ages 23, 28, and 33–34 as the periods of overtaking to elementary school, high school, and college graduates. The ages are taken from Table 3.1 (column 3 + column 8). This calculation, the same as in column 10 of Table 3.1, utilizes only one earnings figure in each schooling group. In contrast, the rates of return shown in column 2 of Table 3.2 were calculated from complete age profiles of earnings. The similarity is rather close, a strong suggestion of the feasibility of "short-cut" estimation.

3.1.2 UNGROUPED DATA

The schooling model will now be explored in ungrouped, individual data. Overtaking values of earnings Y_s which were estimated for schooling groups can also be estimated, under somewhat stronger assumptions, for individuals whose schooling is known. Since, at $\hat{j}$,

$$r_p = C_{\hat{j}} \Big/ \sum_{t=0}^{j-1} C_t,$$

if all individuals in a schooling group are assumed to have the same rate of return to, and proportionate time distribution of, post-school investments, the overtaking year of experience ($\hat{j}$) would be the same for all. On this assumption we may select a set of individuals in our sample whose years of work experience correspond to the overtaking years which were used in the grouped data. The distribution of earnings of these individuals can be viewed as an estimate of the latent distribution of earnings that would be received if no further human capital were invested after completion of schooling.

As indicated in Table 3.3, below, I selected several subsets of the sample to approximate the distribution of earnings at overtaking. The findings in Table 3.3 do not vary much among the samples. As expected, earnings inequality in the overtaking sets is smaller than aggregate inequality. Indeed, the earnings at this stage of the life cycle are an estimate of lifetime earnings, since the present value of Y_s approximates the present value of the observed earnings profile. The variance of log earnings in this group is about 0.50 (Table 3.3, column 6) compared to 0.68 in the aggregate. Thus, at the start of working life, expected lifetime inequality measured in relative terms (in logs), is about 25 per cent less than aggregate inequality. The difference in dollar dispersions is greater. The dollar variance of the aggregate cross-sectional distribution of annual earnings is about twice the size of the dollar dispersion in the overtaking set.

The earnings distribution at overtaking serves two purposes: As suggested above, it provides a base for assessing the contribution of post-school investments to aggregate inequality. More directly, it serves as a testing ground for the schooling model, since the latter can be directly applied only to earnings of this particular population group. However, for several reasons, the inequality estimated in the overtaking set cannot be fully explained by differences in *years* of schooling alone:

a. The distribution of schooling investments is only partly measured by the distribution of *years* of schooling. The dispersion in years of schooling fails to reflect variation in initial earning capacity and in expenditures of time and money of students attending schools of the same quality, as well as schools of differing quality.[9]

b. The empirical definition of the "overtaking" set is quite rough. In the absence of specific information each individual was assigned the average age of school-leaving in his schooling group. Actual dispersion in those ages is not negligible.[10]

c. Overtaking years differ among people with the same amount of schooling and experience, if their rates of return differ, and if their dollar investment profiles are not proportional. The observed residual variances in the regressions of (log) earnings on years of schooling in the empirical overtaking sets, as presented in Table 3.3, column 5, must, therefore, overstate the true residual variation.

9. Information on direct costs and earnings of students can be incorporated into the calculation of investment ratios k during school years, instead of assuming that each $k = 1$.

10. National Science Foundation data for 1966 from the National Register of Scientific and Technical Personnel indicate standard deviations of 2 to 3 years for ages at which B.A. and higher degrees were obtained (Weiss, 1971).

TABLE 3.3
REGRESSIONS IN OVERTAKING SETS

Years of Experience (1)	Number of Observations (2)	Regression Equation (3)	R^2 (4)	$\sigma^2(u)$ (5)	$\sigma^2(\ln Y)$ (6)	$\sigma(r)$ (7)
8	790	(1) $\ln Y = 6.36 + .162s$ (16.4)	.306	.333	.48	.046
	$\bar{s} = 12.1$ $\sigma^2(s) = 7.4$	(2) $\ln Y = 2.14 + .115s + 1.27 \ln W$ (15.1) (21.0)	.575	.204		.036
6–10	3,689	(1) $\ln Y = 6.75 + .133s$ (36.1)	.261	.422	.56	.052
	$\bar{s} = 12.2$ $\sigma^2(s) = 7.9$	(2) $\ln Y = 2.07 + .104s + 1.31 \ln W$ (34.0) (43.4)	.511	.279		.042
7–9	2,124	(1) $\ln Y = 6.30 + .165s$ (26.5)	.328	.353	.52	.048
	$\bar{s} = 12.2$ $\sigma^2(s) = 7.7$	(2) $\ln Y = 1.89 + .121s + 1.29 \ln W$ (24.6) (30.6)	.596	.218		.037
		(3) $\ln Y = 4.78 + .424s - .010s^2$ (10.0) (−6.1)	.347			
		(4) $\ln Y = 1.60 + .183s - .002s^2$ (5.3) (−1.7) $+ 1.270 \ln W$ (29.7)	.602	.215		

NOTE: Figures in parentheses are t ratios; Y = earnings in 1959 of white nonfarm men; s = years of schooling; $\sigma^2(s)$ = variance of years of schooling; W = weeks worked in 1959; R^2 = coefficient of determination; $\sigma^2(u)$ = residual variance; $\sigma^2(\ln Y)$ = aggregate variance; $\sigma(r)$ = standard deviation of rates of return.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

Regressions were run in several alternative subsets of the sample, representing approximations to the overtaking stage of experience. Experiments were carried out with subgroups of different sizes, running from 790 in a single experience year ($j = 8$) to 3,689 individuals in an aggregated (6–10) year-group. The coefficients of determination (R^2) and the regression slopes differ somewhat depending on which interval of experience is chosen. The R^2 in these regressions run from 0.26 to 0.33, while the slopes of the schooling variable, which are estimates of the (average) rate of return to schooling, vary between 0.13 and 0.16.

Table 3.3 contains results for three subgroups varying in level of aggregation, but centering around $j = 8$. The regression slopes in Table 3.3 are estimates of rates of return to schooling. The size of the slope is affected by the number of weeks worked during the year. When the regression is expanded to include the number of weeks (in logs) worked during 1959 (W) as a second variable, the partial coefficients of schooling (at s) are several percentage points lower than were the simple coefficients. This is because (logs of) W are positively correlated with schooling: on the average, longer-schooled individuals work more weeks during the year. The coefficients for W are above unity, implying a positive correlation between weekly earnings and weeks worked during the year even for workers with the same schooling attainment.

If the positive correlation between weeks worked and schooling and between weeks worked and weekly earnings reflected primarily a positively sloped labor supply curve, then the coefficient at s based on weekly earnings would be the more appropriate estimate of rates of return to schooling. These correlations may be, however, a consequence of a greater incidence of turnover, unemployment, seasonality, and illness at lower levels of schooling and earnings. In that case coefficients at s based on annual earnings would be the more appropriate estimates, if the reduction of such incidence is an effect of schooling.[11]

Estimates of rates of return directly calculated from age profiles of earnings (such as those of Becker, Hansen, and Hanoch) are usually higher at lower levels of schooling. A statistical test of this inverse relation between r and s is performed in regressions (3) and (4) in Table 3.3. A quadratic term in s is added to the regression to allow for a systematic change in r with changing levels of s. A significant negative coefficient at s^2 means that rates of return are lower at higher levels of schooling. This is, indeed, the case in regression (3). However, the same test performed in regression (4), where weeks worked are included, yields a negative sign but a small and statistically insignificant coefficient at the quadratic term. It appears, therefore, that differences in the amount of time worked during the

11. For further discussion of the working-time variable, see Chapter 7, below.

year almost fully account for the higher rates of return at the lower levels of schooling.

A comparison of regressions (2) and (3) suggests that about half of the rate of return to elementary school graduates can be attributed to their greater amount of employment during the year compared to people with less schooling. The employment factor accounts for about a third of the rate of return at the high school level, and is of little importance at the college level: From quadratic regression (3) estimates of marginal r_s are:

$$\frac{d(\ln Y)}{ds} = .424 - .021s.$$

$$\begin{aligned} \text{For } s &= \ 8,\ r_s = .256 \\ s &= 12,\ r_s = .172 \\ s &= 16,\ r_s = .088 \end{aligned}$$

The explanatory power of schooling investments in the distribution of earnings at overtaking is underestimated by the regressions of Table 3.3. Variations in quality of schooling and in ages of school-leaving are left in the residual. The latter may account[12] for 0.01 to 0.04 in $\sigma^2(u)$, but the former is likely to be more important. According to figures quoted by Becker (1964, p. 108) the coefficient of variation in expenditures on a college education in New York State alone was no less than the coefficient of variation in the national distribution of years of schooling. Solmon and Wachtel (1972) adjusted years of schooling for "quality" by expressing expenditures per student as a ratio to estimated student opportunity costs and adding these time-equivalents to each student's reported years of schooling. For students with at least a college education in the NBER-Thorndike sample,[13] the variance in the "quality-adjusted" years of schooling was three times the size of the variance of unadjusted years of schooling. According to the same data the dispersion in high school quality was smaller, but still quite considerable. At any rate, a conservative guess would be that the "quality-adjusted" variance of schooling at all levels exceeds the unadjusted variance by a third. If so, R^2 corrected for schooling quality could increase from the ob-

12. If the standard deviation of ages at school-leaving is 1 to 2 years within schooling groups (judging by data of Weiss, 1971).

13. For a description of NBER-TH, see Juster (1972).

served 33 per cent to over 40 per cent in the regressions which do not include the weeks-worked variable, and from the observed 60 per cent to perhaps 70 per cent in those that do include it.

Variation in rates of return, which cannot be observed, is probably the main component of residual variation in these regressions. The correlation between these rates and the quantity of schooling investment across individuals is evidently weak, as experiments with the inclusion of s^2 in the regressions (Table 3.3) suggested. If so, the assumption of independence between r_i and s_i across individuals can be used and provides a way of estimating upper limits for the dispersion of individual rates of return $\sigma^2(r)$. In this case the schooling model, equation (1.3), in variance form is:

$$\sigma^2(\ln Y_s) = \bar{r}^2\sigma^2(s) + \bar{s}^2\sigma^2(r) + \sigma^2(s)\sigma^2(r) + \sigma^2(v), \tag{3.1}$$

where v is a residual due to other unmeasured factors. The residual variance in the regressions of Table 3.3 is:

$$\sigma^2(u) = \bar{s}^2\sigma^2(r) + \sigma^2(s)\sigma^2(r) + \sigma^2(v), \tag{3.2}$$

with $\sigma^2(v)$ larger in regressions (1) than (2), since the effects of weeks worked are in the residuals of (1). Therefore,

$$\sigma^2(r) < \frac{\sigma^2(u)}{\bar{s}^2 + \sigma^2(s)}. \tag{3.3}$$

The values of the upper limit for $\sigma(r)$ are shown in column 7 of Table 3.3. They range from 4 per cent in the regressions which are standardized for weeks worked to 5 per cent in those that are not. The coefficient of variation in individual average rates of return is therefore at most a third in each of the regressions.

It is difficult to judge whether the estimated (upper limit) coefficient of variation is "small" or "large." It is apparently much smaller than the coefficient of variation in corporate rates of return, observed by Stigler (1963) in annual data (1947–54).[14] It should be noted that the dispersion of rates of return to schooling is not a good measure of risk to the extent that abilities and opportunities underlying this dispersion are known to the individual.

14. Note also that year-to-year instability is far greater in business incomes than in earnings of male adults: The interyear correlations in corporate earnings decay rapidly over time (Stigler, p. 71) in contrast to the slow decline in panel correlations of individual earnings shown in Table 7.1, below.

TABLE 3.4
CORRELATION OF LOG EARNINGS WITH SCHOOLING WITHIN EXPERIENCE OR AGE GROUPS

	Coeff. of Det. (r^2)			
Years of Experience	All [a] (1)	Year-round (2)	Years of Age	Coeff. of Det. (r^2) (3)
1–3	.31	.25		
4–6	.30	.27	20–24	.02
7–9	.33	.30	25–29	.04
10–12	.26	.30	30–34	.11
13–15	.20	.25	35–39	.14
16–18	.17	.20	40–44	.16
19–21	.16	.18	45–49	.12
22–24	.13	.17	50–54	.12
25–27	.13	.15	55–60	.09
28–30	.12	.14	60–64	.08
31–33	.07	.14		
34–36	.05	.07		
37–39	.07	.09		
Aggregate	.07	.08	Aggregate	.07

SOURCE: 1/1,000 sample of the U.S. Census, 1960.
a. All workers, including both year-round and those whose work was part time, seasonal, or otherwise intermittent.

Without standardization for weeks worked and without adjustment for quality, the schooling model explains a third of the inequality of earnings in the overtaking subset of the earnings distribution. This is a great deal more than the 7 per cent found in the simple regression of log earnings on schooling in the aggregate distribution. The greater applicability of the schooling model to the overtaking period than to subsequent stages of experience is shown clearly in Table 3.4.

As measured by simple coefficients of determination, the effects of schooling on earnings decay continuously in successive three-year experience groups after the first decade of experience. This is shown in columns 1 and 2 of Table 3.4.

The decay of the coefficient of determination (R^2) reflects the growing importance of accumulated experience in the determination of earnings. R^2 is the ratio of "explained" to total variance of log earnings. In the overtaking set

$$R_{\hat{j}}^2 = \frac{r^2\sigma^2(s)}{r^2\sigma^2(s) + \sigma^2(u)}. \tag{3.4}$$

The content of the residual variance $\sigma^2(u)$ was already discussed. At later stages when $j > \hat{j}$, assuming little or no correlation between time-equivalents of schooling and post-school investments:

$$R_j^2 = \frac{r^2\sigma^2(s)}{r^2\sigma^2(s) + \sigma^2(u) + \sigma^2(r \sum_t^{j-1} k_t - k_j)}. \tag{3.5}$$

R_j^2 declines because the denominator grows with experience, since the right-hand term in it must grow. The decline in R_j^2 may be strengthened for additional reasons: The coefficient of schooling (r) may decline over time—a possibility suggested by a "vintage" hypothesis of schooling effectiveness.[15] A random shock structure in the residual u would give rise to a growing $\sigma^2(u)$, thereby increasing the rate of decay in R^2.

The systematic effects of accumulated experience are obscured when the schooling model is applied to age groups (Table 3.4, column 3): The coefficient of determination at its highest is half the size of that found in the overtaking group.[16] Its peak is reached in the 40–44 age group, and it is quite small before age 30. The weaker fit of the schooling model in age groups compared to experience groups is due to the negative correlation between schooling and post-school investments at given ages. This is most pronounced at the early post-school ages, when investment in experience is heaviest. The later decay is due to the accumulation of post-school investments, as already discussed.

During the first decade of experience, the coefficients of determination are relatively high but somewhat less than at overtaking. It is plausible though not necessary that the denominator in the expres-

15. Welch (1972) observes declines in regression coefficients of schooling over experience in both 1960 Census data and 1967 data of the U.S. Department of Labor Survey of Economic Opportunity, and interprets them as "vintage" effects.

16. The contrast is somewhat overstated, as the age intervals are wider.

sion for R^2 decline during the first decade, as suggested in the discussion in Chapter 2.

In a longitudinal study of over 1,500 men who were 30–39 years old in 1968, Blum (1971) also found that the correlation between schooling and earnings was higher after ten years of work experience ($R^2 = .24$) than in the initial year ($R^2 = .16$). Differential post-school investments of individuals can account for the difference.[17]

3.2 SOME QUALITATIVE IMPLICATIONS

When applied to the proper data, the schooling model can be a useful tool for quantitative analysis. More generally, though less rigorously, the model also yields several important qualitative implications about distributions of earnings.

1. A tendency toward positive skewness of earnings is produced by the transformation of absolute differences in years of schooling into relative differentials in earnings. Clearly, by equation (1.3) a symmetric distribution of years of schooling implies a positively skewed distribution of earnings. Unless the skew in the distribution of schooling is highly negative, a positive skew will be imparted to the distribution of earnings. Because of the finite lower limit (zero, or a legal minimum) empirical schooling distributions are more likely to be positively skewed when the average level of schooling is low. Skewness may change from positive to negative as the average level of schooling reaches high levels. Thus the U.S. distribution of schooling has become negatively skewed in the cohorts below age 40, as shown in Table 3.5. Even so, negative skewness in schooling is not sufficient to create negative skewness in earnings. It will be recalled (Chapter 2, note 1) that, according to the schooling model, positive skewness in earnings obtains so long as $1 - (d_2/d_1) < rd_1$, where d_1 is the schooling interval (in years) between the median and a lower (say tenth) percentile and d_2 is the interval between the median and a corresponding upper (ninetieth) percentile. Given rates of return r in excess of 10 per cent, the above condition is empirically satisfied in Table 3.5 in all age groups. A fortiori (cf. section 2.4), the aggregate

17. The notion that schooling has a positive effect on earnings merely as a "credential" is difficult to reconcile with the pattern of correlations observed in Table 3.4 and in the longitudinal study.

TABLE 3.5
DISTRIBUTION OF YEARS OF SCHOOLING (s), BY AGE GROUPS, 1959
(U.S. white, nonfarm males)

Age	P_{10} (1)	Md (2)	P_{90} (3)	$\bar{s}$ (4)	σ_s (5)	$d_u - d_1$ (6)
14–19	6.0	9.8	11.8	10.5	2.2	−1.8
20–24	8.1	11.9	15.2	12.1	2.8	−0.5
25–29	6.8	12.0	16.0	12.2	3.2	−1.2
30–34	6.4	11.8	16.0	11.7	3.4	−1.2
35–39	6.3	11.7	15.9	11.7	3.4	−1.2
40–44	5.7	11.3	15.6	11.2	3.4	−1.3
45–49	5.3	10.5	15.4	10.5	3.6	−0.3
50–54	5.1	9.5	15.0	10.1	3.6	+1.1
55–59	4.7	8.5	14.2	9.4	3.7	+1.9
60–64	4.4	7.8	13.9	8.8	3.7	+2.7
65 or older	3.5	7.4	13.2	8.5	4.0	+1.9
All	6.2	10.5	15.7	10.9	3.5	+0.9

P_{10} = 10th percentile. $\bar{s}$ = arithmetic mean schooling.
Md = median. σ_s = standard deviation.
P_{90} = 90th percentile. $d_u = P_{90} - Md$; $d_1 = Md - P_{10}$.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

distribution of earnings is likely to be positively skewed. As the U.S. level of schooling is the highest in the world, its distribution is more negatively (less positively) skewed than that of any other country. Hence positive skewness in earnings is likely to be universal.

2. The schooling model implies that relative dispersion of earnings is larger the larger the absolute dispersion in the distribution of schooling and the higher the rate of return. In terms of the schooling regression, where the variance in r is suppressed:

$$\sigma^2(\ln Y_s) = r^2\sigma^2(s) + \sigma^2(u). \tag{3.6}$$

Chiswick's (1967) regional comparisons of income inequality do indeed show that inequality and skewness of income are larger the larger the variance in the distribution of schooling and the higher the rate of return as measured by the size of the regression slope in (1.3). According to Chiswick, these factors jointly explain over a third of the differences in inequality among regions,[18] with the rate of return *apparently* the *more important factor.*

Rapid upward trends in years of schooling attainment in the United States are reflected in Table 3.5 in the systematically different distributions of years of schooling in the separate age groups of employed men in 1959. The typical (median) 25-year-old was a high school graduate in 1959, while the typical 60-year-old was an elementary school graduate. Dispersion in the distribution of schooling, as measured by a percentile range or a standard deviation, narrowed somewhat from the older to the younger cohorts, while skewness changed from positive to negative as the level rose.[19]

The systematically larger dispersion and skewness of the schooling distribution with increasing age is paralleled by increases in relative dispersion and skewness in earnings in the age groups, as shown in Tables 3.6 and 6.3. However, the consistency of this phenomenon with predictions of the schooling model is only qualitative: The actual rate of increase of earnings inequality with age is far too strong to be attributable in the main to the mild increase in the dispersion of schooling. The schooling model in variance form (equation 3.6) predicts a smaller percentage increase in $\sigma^2(\ln Y)$ than in $\sigma^2(s)$, if r^2 and $\sigma^2(u)$ do not increase. The variance of schooling is only about 20 per cent larger in the 55–59 age group than in the 30–34 age group (Table 3.5, column 5), yet the variance of relative (log) earnings is 70 per cent greater in the older compared to the younger group[20] (Table 6.3, column 1). The variance of schooling is about 25 per cent larger in the 55–64 age group than in the 25–34 age group in Table 3.5, but the variance of income doubles in this range in every annual

18. In his current work, Chiswick greatly increases the explanatory power of the earnings function by expanding it to include post-school investments. Lydall (1968), who did not employ the rate of return as an explicit variable, found the dispersion in the distribution of schooling to be a significant factor in explaining differences in the inequality of earnings among a set of countries.

19. In their survey of trends in educational attainment of the U.S. *population,* Folger and Nam (1967) found that "educational attainment is more evenly distributed in the population than it used to be." The data in their Chapter 5 show mild trends in dispersion, as well as a pronounced change from positive to negative skewness in the distribution of schooling.

20. As shown in Table 6.3, the relative variance of earnings has a U-shaped age pattern with low values in the 30–34 age group. The age and experience patterns of dispersion are more fully analyzed in Chapter 6.

TABLE 3.6
COHORT AND CROSS-SECTIONAL CHANGES IN INCOME INEQUALITY, ALL U.S. MEN, 1947–70
(variance of logs of income)

Year	Age 25–34 (1)	Age 35–44 (2)	Age 45–54 (3)	Cross Section (col. 3 less col. 1)	Cohort [a]
1947	.352	.494	.541	.189	
1957	.420	.518	.697	.277	
1967	.387	.459	.546	.159	.194
1948	.355	.445	.585	.230	
1958	.445	.489	.727	.282	
1968	.389	.454	.567	.178	.212
1949	.379	.538	.680	.301	
1959	.442	.478	.692	.250	
1969	.418	.469	.572	.154	.193
1950	.378	.471	.642	.254	
1960	.428	.554	.719	.291	
1970	.458	.486	.585	.097	.207
Average				.225	.204

SOURCE: Schultz (1971, Table 2).
a. In each twenty-year span, column 3 of the last year of the span minus column 1 of the first year.

cross section (1947 to 1970) in Table 3.6. The observed age gradient in earnings inequality cannot be ascribed to cohort differences in the distribution of schooling. Rather, it is a phenomenon connected with aging of the same cohort whose distribution of schooling is, of course, fixed.

The within-cohort changes can be observed directly in the repeated cross sections of Table 3.6: Individuals who were 35–44 years old in 1959 were in the 25–34 age group in 1949, and in the 45–54 age group in 1969. Income variances can be compared in the three survey years to detect changes within fixed cohorts. This procedure was applied to variances of logs of income of men in decade age intervals, which were calculated by T. P. Schultz (1971) from Current Population Surveys for the years 1947–70. The results shown in Table 3.6 indicate that the cross-sectional age differences in income inequality were mainly a consequence of changes within the same cohorts.[21] The cross-sectional changes are shown in the rows, the within-cohort changes along the diagonals. As the last two columns show, the cross-sectional increase in inequality is only slightly greater than the within-cohort increase. At the same time there are no clear trends in inequality within fixed age groups.[22] Apparently, the cross-sectional increases in inequality with age are produced, in the main, by factors other than the secular change in the distribution of schooling. The theoretical analysis suggested that an explanation for much of the age difference in parameters of earnings distribution would be found in the distribution of post-school investments. We now proceed to an empirical exploration of the age and experience differences in earnings.

---

21. Very similar results are produced by comparing Gini coefficients, calculated from the same data by H. P. Miller (1963, Table 12). Though the data underlying Table 3.6 are incomes of all men, rather than earnings of nonfarm white men, it is not likely that the conclusions are affected by this inaccuracy. Age patterns of log variances are not very different under the two definitions, though levels of income exceed levels of earnings by 20–30 per cent in each age group.

22. For an analysis of these trends see Chiswick and Mincer (1972).