Volume Title: Schooling, Experience, and Earnings

Volume Author/Editor: Jacob A. Mincer

Volume Publisher: NBER

Volume ISBN: 0-870-14265-8

Volume URL: http://www.nber.org/books/minc74-1

Publication Date: 1974

Chapter Title: Analysis of Residuals: Distributions of Earnings Within Schooling and Age Groups

Chapter Author: Jacob A. Mincer

Chapter URL: http://www.nber.org/chapters/c1768

Chapter pages in book: (p. 97 - 114)

6

Analysis of Residuals: Distributions of Earnings Within Schooling and Age Groups

6.1 VARIANCES

It was possible to make indirect estimates of the contribution of human capital investments to total earnings inequality by assuming that the latent residuals (u_i) in the earnings function (equation 5.6) were homoscedastic. Is this assumption consistent with the empirical data? This question is an invitation to explore the structure of earnings distributions within groups defined by years of schooling and years of age (experience). Since the within-group variation in earnings is quite large, such an exploration is of interest in its own right, and not merely as a test of particular assumptions.

The estimated values of the regression equations, which were shown in Table 5.1, are estimates of means in the schooling-experience groups. The within-group distributions are, therefore, distributions of residuals[1] (v_i) (equation 5.7). Since $v_i = u_i + (v_i - u_i)$,

$$\sigma^2(v_{it}) = \sigma^2(u_i) + \sigma^2(v_i - u_i) + 2\rho\sigma(u_i)\sigma(v_i - u_i).$$

1. Apart from the sampling errors of the means in each cell.

The observed residual variance $\sigma^2(v_i)$ in a schooling group changes over the life cycle (t) only if $\sigma^2(v_i - u_i)$ changes, assuming $\sigma^2(u_i)$ is homoscedastic and fixed.

By definition of equations (5.6) and (5.7) of Chapter 5, $v_i - u_i$ contains unobserved individual differences in returns to post-school investment. The human capital model (Chapter 2) predicts that the residual variance in a schooling group, $\sigma^2(v_i)$, will change systematically with age and experience.

6.1.1 EXPERIENCE PROFILES OF DOLLAR AND LOG VARIANCES OF EARNINGS

To analyze the experience profiles of both dollar variances $\sigma^2(Y_i)$ and log variances $\sigma^2(v_i) = \sigma^2(\ln Y_i)$, consider three points in the working life: the initial stage $t = 0$, the overtaking stage $t = \hat{j}$, and the peak earnings stage $t = t_p$. The arbitrariness and difficulty of deriving functional forms of profiles of variances is avoided by using this procedure, while making it possible to determine whether the profiles are monotonic.

First the expressions for dollar variances at the three points are derived:

$$Y_{si} = E_{si} - C_{0i};\ \therefore\ \sigma^2(Y_s) = \sigma^2(E_{si}) + \sigma^2(C_0) - 2\rho_{C_0,E_s}\sigma(E_s)\sigma(C_0). \quad (6.1)$$

$$Y_{\hat{j}i} = E_{si};\ \sigma^2(Y_{\hat{j}i}) = \sigma^2(E_s). \quad (6.2)$$

$$Y_{pi} = E_{s_i} + rC_T;\ \sigma^2(Y_p) = \sigma^2(E_s) + r^2\sigma^2(C_T) + 2r\rho_{C_T,E_s}\sigma(E_s)\sigma(C_T). \quad (6.3)$$

In each equation, C_0 is initial-period post-school investment; C_T, the sum of positive post-school net investments; E_s, initial post-school earning capacity; Y_p, peak earnings; ρ, correlation coefficient; and r, the rate of return to post-school investments.

In general, $\sigma^2(Y)$ must vary over the life cycle. The pattern of variation depends on the dispersion in post-school investments and on the correlation between the dollar volumes of post-school investment and earning capacity E_s. If, as appears from intergroup analysis (Chapter 2), the correlation between (dollar) schooling and post-school investment is positive, ρ is positive and dollar variances must rise from overtaking to peak earnings. In addition, dollar variances will rise throughout if $\sigma^2(Y_0) < \sigma^2(Y_{\hat{j}})$, which must be true if

$$\rho(C_0, E_s) > \frac{1}{2}\frac{\sigma(C_0)}{\sigma(E_s)}.$$

Chart 6.1 and Table 6.1 show that dollar variances indeed increase monotonically and sharply throughout working life. The standard deviations more than double between the eighth and thirtieth year of experience in the middle and upper schooling groups, but increase at a slower rate in the lower ones. The same data also show that the profiles of variances of the more educated are systematically higher at all stages of experience. A sufficient condition for this phenomenon can be found by comparing the variances at overtaking: Here $\sigma^2(Y_{\hat{j}}) = r_s^2\sigma^2(C_s) + \sigma^2(u)$, when the variation in E_0 and r_s is impounded in the residual. $\sigma^2(C_s)$ evidently increases with level of schooling. This is quite plausible: over time, total costs of schooling cumulate with level of schooling, and so do individual differences in total costs.

The rate of growth of variances from $\hat{j}$ to t_p is obtained by dividing (6.2) into (6.3):

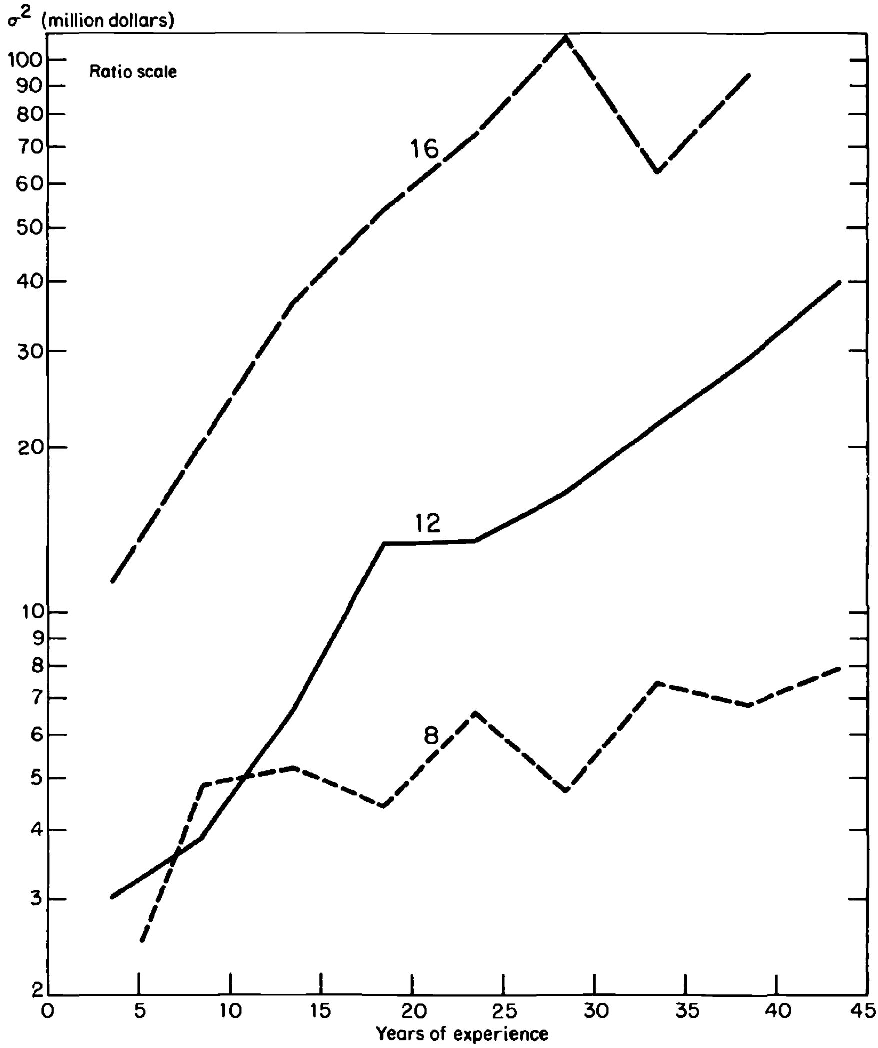
$$\frac{\sigma^2(Y_p)}{\sigma^2(Y_{\hat{j}})} - 1 = r^2\frac{\sigma^2(C_T)}{\sigma^2(E_s)} + 2r\rho\frac{\sigma(C_T)}{\sigma(E_s)}. \tag{6.4}$$

The weaker growth of variances at lower levels of schooling suggests either a weaker correlation (ρ) between earnings capacity (E_s) and post-school investments (C_T), or a smaller ratio $\sigma(C_T)/\sigma(E_s)$. Define the regression slope of C_{Ti} on E_{si}, which is equal to $\rho[\sigma(C_T)/\sigma(E_s)]$ as the "marginal propensity to invest" (*MPI*). Evidently, *MPI* tends to be smaller at lower levels of schooling.[2]

The important conclusion resulting from the analysis of dollar variances is that the usually observed increases of variances with experience and age are strongly influenced by the staggering of post-school investments over individual working lives. A large enough dispersion of post-school investments and a positive correlation between dollar schooling and post-school investments can explain the sharp age gradients. The increases of dollar variances with education are likely to reflect the almost necessarily larger residual dollar dispersion of total schooling costs at higher levels of schooling.

2. Cf. the findings of Solmon (1972) that the marginal propensity to save is also smaller at lower levels of schooling.

CHART 6.1
EXPERIENCE PROFILES OF VARIANCES OF ANNUAL EARNINGS OF WHITE, NONFARM MEN, 1959



NOTE: Figures on curves indicate years of schooling completed.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

TABLE 6.1
AGE PROFILES OF DISPERSION IN EARNINGS, 1959
(white, nonfarm men)

	Years of Schooling					
	5–8	12	16	5–8	12	16
Age	Standard Deviation			Variance of Logs		
			Annual Earnings			
20–24	$2,200	$1,730	($ 2,360)	.562	.454	–
25–29	2,280	1,970	3,270	.457	.258	.360
30–34	2,110	2,590	4,530	.336	.231	.251
35–39	2,570	3,650	6,040	.358	.240	.317
40–44	2,200	3,680	7,340	.371	.272	.421
45–49	2,730	4,070	8,590	.360	.339	.555
50–54	2,620	4,710	10,550	.392	.403	.626
55–59	2,820	5,390	8,920	.424	.451	.612
60–64	3,360	6,340	9,700	.525	.460	.933
			Weekly Earnings			
20–24	$53.3	$ 46.3	$ 46.1	.489	.363	–
25–29	47.1	39.0	70.6	.320	.205	.235
30–34	44.7	46.0	75.1	.263	.183	.212
35–39	43.3	65.0	102.2	.266	.203	.277
40–44	46.0	69.0	121.7	.275	.226	.336
45–49	56.2	77.1	144.4	.310	.270	.424
50–54	53.2	82.3	176.6	.292	.312	.436
55–59	55.1	93.0	153.3	.328	.317	.552
60–64	63.3	107.8	162.8	.409	.369	.748

SOURCE: 1/1,000 sample of U.S. Census, 1960.

A positive correlation between means and variances of economic variables is a frequently encountered empirical phenomenon. It might be taken for granted as an arithmetical necessity, which it is not. The structure of means and variances of earnings in these schooling-age cells is an example of it. In this case, however, the human capital model provides an explanation: higher levels of earnings represent returns cumulated by additional investment. Thus if H_1 is a lower stock of human capital and $H_2 = H_1 + \Delta H$ is a higher one, earnings $E_1 = rH_1$ and $E_2 = r(H_1 + \Delta H)$. Then $E_2 > E_1$ and $\sigma^2(E_2) >$

$\sigma^2(E_1)$ so long as the correlation between H_1 and ΔH is not excessively negative. Indeed, a positive correlation is expected, since the determinants of investment in current and past periods are likely to persist for a given individual.

6.1.2 ANALYSIS OF MARGINAL VARIANCES OF EARNINGS

Dollar standard deviations of earnings in marginal distributions, that is, in distributions of all earnings in a given row or column of the two-way classification of the population by schooling and experience (or age), are shown in Table 6.3, column 2, below. The total variance in such a group (say an age group) is, by (2.12):

$$\sigma_T^2 = \frac{1}{n} \sum_{i=1}^{s} n_i(\sigma_i^2 + d_i^2), \tag{6.5}$$

where n is the number of observations in the age group; n_i, the number of observations in the ith schooling cell; σ_i^2, the within-cell variance; and $d_i = (\bar{X}_i - \bar{X}_a)$, the differential between the mean in the cell and the overall mean of the age group.

Clearly, marginal variances σ_T^2 (therefore, standard deviations) must increase with experience and with age, because within-cell variances σ_i^2 increase, and because d_i, differentials among profiles of means, also increase, as we learned in Chapter 4. The increase of σ_T^2 is sharper in age groups than in experience groups, because the intergroup differential d_i grows more rapidly in the former: age profiles of mean earnings diverge more strongly than experience profiles.

Similarly, variances in the marginal distributions by schooling must increase with schooling, again because cell variances σ_i^2 and mean age differentials d_i increase with schooling.

These statements are based on the assumption that the relative frequencies n_i/N are the same in each marginal row or column, that is, the age distributions are the same in all schooling groups, and the schooling distribution is the same in all age groups. This would be the case in a cohort which is followed over its working life, or in the cross section if there were no secular trends in schooling. The effect of such secular trends is, of course, that the weights n_i/N differ systematically in the cross section: they are bigger in older age cells at lower levels of schooling, and in higher schooling cells at younger ages. For this reason the age increases in marginal variances are likely to be less in the cross sections than in cohorts. In the cross-sectional age comparisons, there is, however, an offsetting effect due to secular trends in the dispersion of schooling. Whatever the difference from cohorts, the cross-sectional gradients are quite strong, as shown in Table 6.3, below.

6.1.3 ANALYSIS OF LOG VARIANCES OF EARNINGS

We now turn to a similar analysis of relative variances, $\sigma^2(\ln Y_i)$, in groupings of the earnings distribution. The observations are shown in Table 6.1 and in Charts 6.2–6.4. Again, consider three points in the working life:

$$\ln Y_{si} = \ln E_{si} + \ln (1 - k_{0i}); \tag{6.6}$$

$$\therefore \sigma^2(\ln Y_s) = \sigma^2(\ln E_s) + \sigma^2 \ln (1 - k_0) + 2\rho_1\sigma(\ln E_s)\sigma \ln (1 - k_0).$$

$$\ln Y_{\hat{j}i} = \ln E_{si};\ \sigma^2(\ln Y_{\hat{j}}) = \sigma^2(\ln E_s). \tag{6.7}$$

$$\ln Y_{pi} = \ln E_{si} + rK_{Ti};\ \sigma^2(\ln Y_p)$$

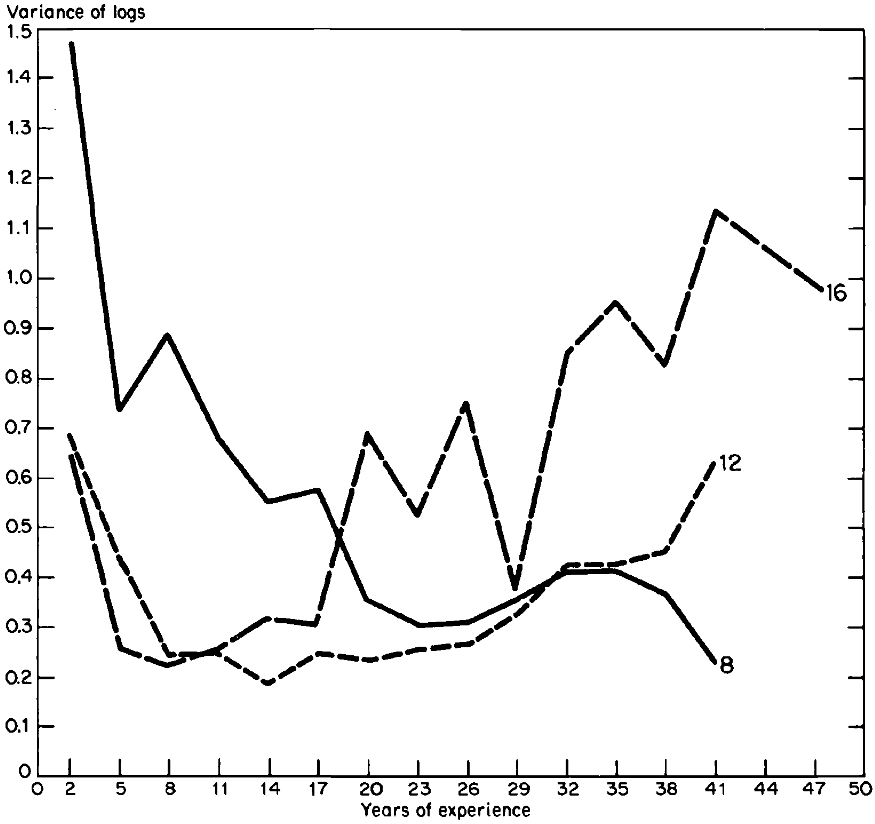
$$= \sigma^2(\ln E_s) + r^2\sigma^2(K_T) + 2\rho_2 r\sigma(\ln E_s)\sigma(K_T). \tag{6.8}$$

Now, the change in log variances over the working life depends on the size of the dispersion in cumulated post-school investments ratios K_T and on the correlation between $\ln E_s (= \ln E_0 + rs)$ and K_T. A positive correlation between time-equivalent post-school investment K_T and initial post-school earning capacity $\ln E_s$ implies a negative correlation between $\ln E_s$ and $\ln (1 - k_0)$. If the correlations are weak, $\rho_2 = \rho_1 = 0$ and the profile of log variances is U-shaped, with the bottom at overtaking. The U-shape is preserved if the correlations are within a specified interval bracketing zero.[3] A more pronounced negative value of ρ_2 implies a monotonic decline in log variances over the working life, while a stronger positive ρ_2 implies a monotonic growth in log variances.

A zero correlation between the investment ratio and initial post-

3. The intervals are $|\rho_1| < \frac{1}{2}\frac{\sigma[\ln (1 - k_0)]}{\sigma(\ln E_s)}$, and $|\rho_2| < \frac{1}{2}\frac{\sigma(K)}{\sigma(\ln E_s)}$.

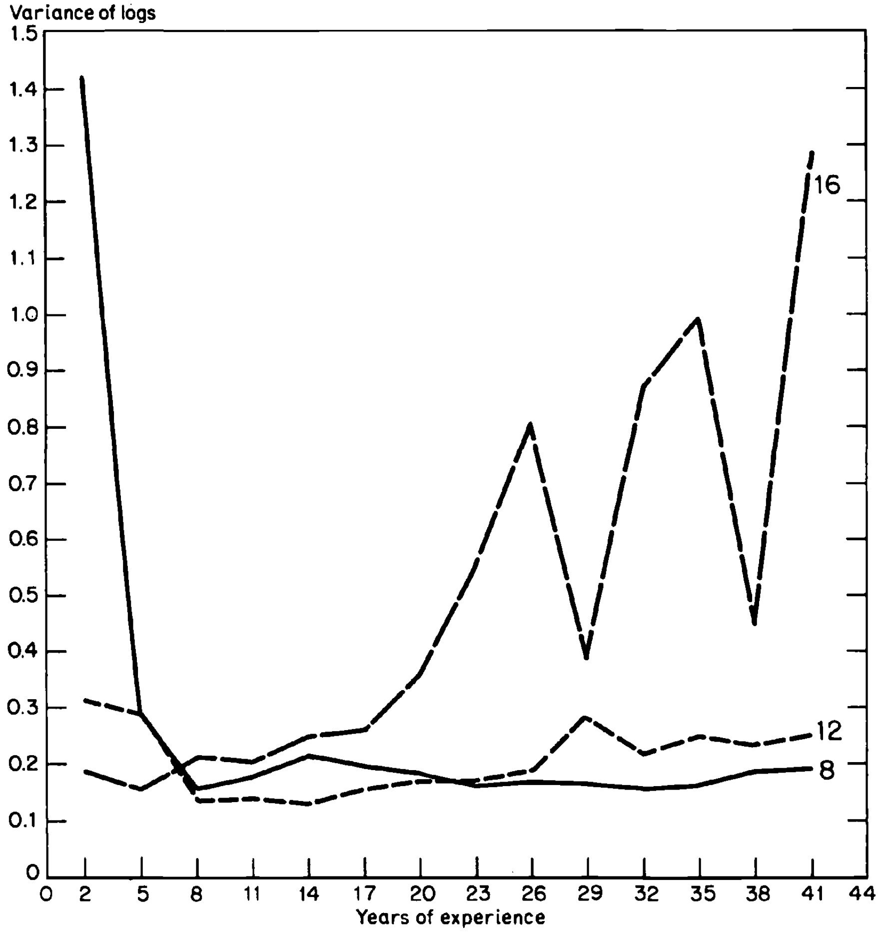
CHART 6.2
EXPERIENCE PROFILES OF LOG VARIANCES OF ANNUAL EARNINGS OF WHITE, NONFARM MEN, 1959



NOTE: Figures on curves indicate years of schooling completed.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

school earning capacity may be due to a unitary elasticity of dollar post-school investment with respect to initial earning capacity. Positive correlations may be caused by elasticities above 1; negative correlations, by elasticities below 1. Charts 6.2 and 6.3 indicate that experience profiles of log variances are largely U-shaped in the central (12 years) schooling groups, suggesting a weak correlation between post-school investment ratios and earning capacity within this schooling level; tend to be positively inclined (show pronounced growth) in the upper schooling groups, suggesting a positive correlation; and are negatively inclined (decline, by and large) at lower

CHART 6.3
EXPERIENCE PROFILES OF LOG VARIANCES OF ANNUAL EARNINGS OF WHITE, NONFARM MEN WORKING YEAR-ROUND, 1959



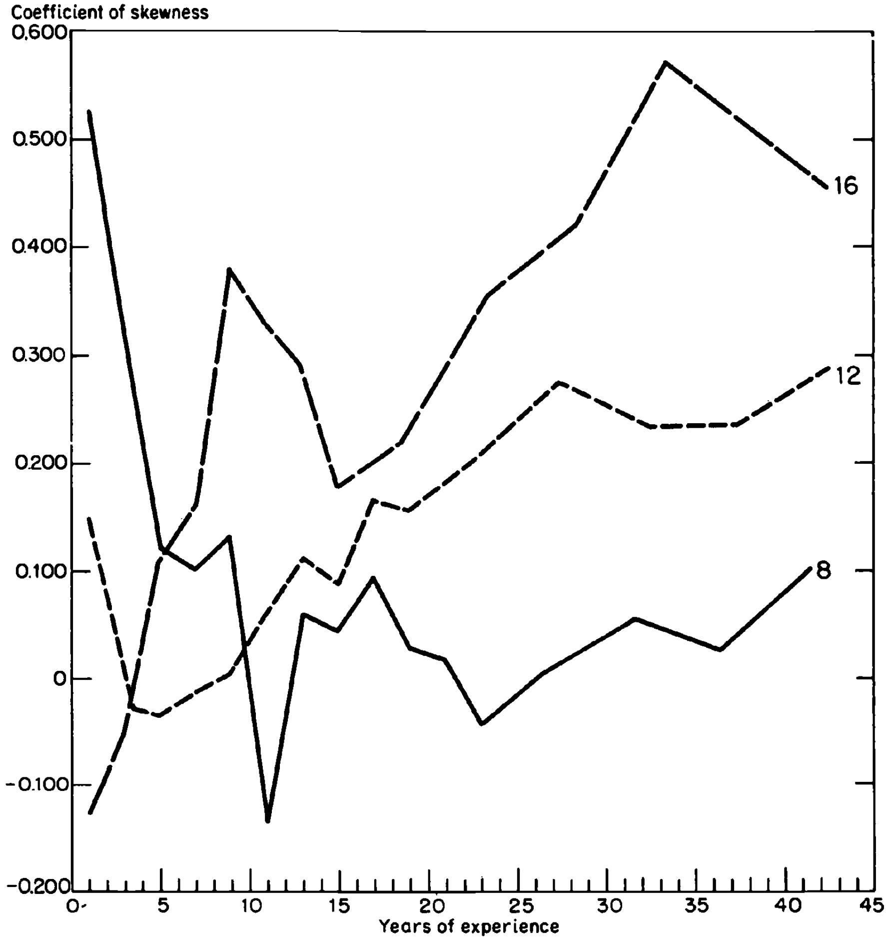
NOTE: Figures on curves indicate years of schooling completed.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

levels, suggesting a negative correlation.[4] Apparently, the within-group elasticity of post-school investment is a positive function of schooling. This finding is formally consistent with the other findings:

4. The stronger growth of both dollar and relative variances at higher than at lower levels of skill (occupation or education) was noted in different data by several analysts. Cf. Adams (1958), Hill (1959), Lydall (1968), Mincer (1957), and Morgan et al. (1962).

CHART 6.4

PROFILES OF RELATIVE SKEWNESS OF ANNUAL EARNINGS OF WHITE, NONFARM MEN, 1959



NOTE: Figures on curves indicate years of schooling completed.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

The ratio k_t is an "average propensity to invest." According to the analysis of log-experience profiles of mean earnings in Table 4.1, it tends to decline from lower to higher schooling groups. At the same time, the profiles of dollar variances suggest that the "marginal propensity to invest" increases with the level of schooling. Elasticities, therefore, increase correspondingly and more strongly.

Another explanation of the difference in log-variance profiles among schooling groups may be found in the serial correlations of investments (the correlation between C_T and E_8 is an example of it). Less stability in investment (and in employment) behavior over the life cycle results in weaker growth of variances at lower levels of schooling. It is also possible, as we noted before, that the mean earnings profile overstates the size of post-school investments in the lower schooling groups. If so, the dispersion of post-school investments is likely to be a less important component in explaining patterns of earnings inequality at lower than at middle and higher levels of schooling.[5]

While dollar variances were larger at higher schooling levels at overtaking, with the plausible interpretation that $\sigma^2(C_s)$ increases with schooling, no comparable statement can be made a priori about $\sigma^2(s)$, unless the ranking of the variation in schooling quality were known. The observed log variances (Chart 6.3) indeed differ very little among schooling groups below college at that stage in the distribution of full-time earnings. Relative (time equivalent) variation in college quality evidently exceeds that at lower levels of schooling. In all earnings (Chart 6.2), variances of the schooling groups below high school are inflated, an effect of large variation of weeks worked during the year (cf. Chapter 7). Because the differently inclined profiles of variances intersect in the second decade of experience, there is a reversal of ranking in inequality by level of schooling: inverse at first and direct in the later parts of the working life. This pattern is not changed much by shifting from the experience comparisons to comparisons based on age.

It is easy to see that these configurations, together with the structure of mean log-earnings profiles, produce U-shaped patterns of *marginal* relative variances by age. These are shown in Table 6.3, column 1. The strong growth of mean differentials d_t with age contributes to the stronger age gradient of inequality and to the earlier reversal of it by age than by experience. In published empirical research, this reversal was noted as a persistent feature of relative earnings structures.[6]

The distinction between cross-sectional and cohort patterns of

5. Some evidence on the particular importance of the variation in weeks worked at lower levels of schooling is seen in the comparison of Charts 6.2 and 6.3 and in Table 7.2, below.

6. Cf. Morgan (1962).

marginal relative variances can be analyzed by means of equation 6.5, above, as were the dollar variances. The implications are similar, except that because of differences in the numbers included in the old and young age groupings within schooling groups, the cross-sectional relation of inequality to schooling has a more negative tilt than that in the cohort, though still somewhat U-shaped, as Table 6.3, column 1, indicates.

Since the purpose of this study was to relate the structure and inequality in the distribution of earnings to the distribution of *amounts* invested in human capital, individual variation in rates of return was ignored. However, in studying the patterns of residuals v_i, variation in rates of return cannot be entirely ignored. But a hypothesis that the observed residual variances contain mainly variation in rates of return rather than variation in post-school investment can be rejected. It was shown, in Part I, that the assumptions $\sigma^2(r) > 0$, while $\sigma^2(C_T) = 0$, leads to a monotonic increase in the residual dollar variances over working life, provided $C_T > 0$. However, $\sigma^2(C_T) = 0$ means that the same dollar amount is invested by each individual, so K_T, the investment ratio, is perfectly negatively correlated with earnings E_s. But that would produce sharp monotonic *decreases* in log variances over the life cycle. The empirical evidence contradicts such a hypothesis.

The hypothesis that $\sigma^2(r) > 0$ and $\sigma^2(k_t) = 0$, while $k_t > 0$, means that there is a perfect positive correlation between C_t and E_s. However, the strong decay of the correlation between schooling and earnings shown in Table 3.4 (Chapter 3) contradicts this hypothesis.

I conclude that post-school investment varies among persons with the same schooling both in dollars and in time-equivalents. The variation in rates of return has no effect on the profiles of residual variances, unless there is post-school investment and it has a nonzero variance. Indeed, the latter is a sufficient explanation of the profiles of residual variance shown in Charts 6.1–6.3. This, of course, does not deny the existence of dispersion in rates of return.[7]

The negative ranking of inequality with respect to schooling seen in the profiles of relative variance in the earlier stages of working life and the reverse ranking later are in no obvious way related to

7. Indeed, as I argued in Chapter 1, post-school investment has no effect on the distribution of earnings in the overtaking set. Hence the residual variance in that distribution, after correction for schooling quality and weeks worked, can be interpreted as resulting from individual variation in rates of return.

secular trends in human capital, such as the upward trend in schooling. The presence of these trends does affect our understanding of the cross section, as we noticed, only when we aggregate the two-way groupings (age *and* schooling) into marginal, that is, one-way groups (age *or* schooling). Because of these trends, higher schooling groups are more prevalent at younger ages, and conversely. In aggregating cross sections, therefore, parameters of the more educated groups receive greater weight in the younger age groups, and those of the less educated receive greater weight in the older ones. Given the reversal of profiles of inequality, therefore, the stronger the upward trend in schooling, the greater the attenuation of aggregate inequality, as larger weights are attached to the smaller relative variances. Or, to put it differently, if there were no trends in schooling, and the distribution of schooling in each age group were the same as the currently observed distribution among all earners, regardless of age, aggregate inequality would be larger than currently observed. The hypothetical distribution would, in effect, be a distribution over the working life of a fixed cohort. A simulation experiment utilizing 13×9 experience parameters and frequencies of schooling groups shows that, under these assumptions, the aggregate log variance in the cohort, that is, in the trendless cross section, would be 0.805 compared to 0.668. Thus the growth of schooling reduces the aggregate inequality observed in the cross section by about 17 per cent, and this effect is obtained not by narrowing the distribution of schooling but by diminishing the importance of groups whose post-school behavior generates a great deal of dispersion in earnings.

The same experiment which keeps the distribution of schooling in each experience group the same as at overtaking ($j = 7$–9 years) yields an aggregate variance of logs of 0.721. This is an estimate of the inequality in the cohort which was at overtaking in 1959. The fraction of aggregate inequality attributable to human capital investment based on this figure is an estimate which abstracts from secular trends in schooling. It is a few percentage points higher than the estimates based on the observed cross section.

6.2 SHAPES OF RESIDUAL DISTRIBUTIONS

Shapes of the within-group earnings distributions are portrayed in Chart 6.4, which shows experience profiles of asymmetry (relative

TABLE 6.2
AGE PROFILES OF SKEWNESS IN EARNINGS, 1959
(white, nonfarm men)

	Years of Schooling					
	5–8	12	16	5–8	12	16
Age	Bowley's Coefficient			Ratio: Mean to Median		
20–24	.107	.039	.232	1.100	1.015	1.178
25–29	−.009	−.025	.119	1.046	1.002	1.101
30–34	.077	.154	.281	1.043	1.045	1.167
35–39	−.028	.211	.256	1.028	1.095	1.121
40–44	−.078	.180	.362	1.022	1.100	1.200
45–49	.057	.222	.463	1.035	1.131	1.182
50–54	.017	.200	.527	1.023	1.132	1.271
55–59	.074	.252	.328	1.049	1.173	1.220
60–64	−.005	.295	.584	1.051	1.238	1.410

SOURCE: 1/1,000 sample of U.S. Census, 1960.

skewness) of earnings in each of the schooling groups. The measure of skewness is Bowley's coefficient:

$$RSk = \frac{(P_{90} - Md) - (Md - P_{10})}{P_{90} - P_{10}},$$

where P denotes percentile and Md median of the distribution. Quite similar results are obtained when the ratio of mean to median is used as a measure of skewness (Table 6.2).

Skewness grows montonically in the upper schooling groups, its profile is U-shaped in the high-school group, and it first rapidly declines and then levels off in the lowest group. Its ranking is directly related to schooling level, except during the first decade of experience, when the ranking is inverse. The pattern resembles the profiles of log variances and can be interpreted in much the same fashion: A strong positive correlation between investment ratios k_{ti} and earning capacities E_{si} within higher levels of schooling, a weak correlation in the middle, and a negative correlation at the lower levels of schooling.

The similarity of the behavior of skewness measured by the mean-to-median ratio and of the log variances is theoretically assured when the distributions are log-normal.[8] More generally, consider the experience profile of gross earnings: $E_{t+1} = E_t(1 + rk_t)$.

By a theorem of C. C. Craig (1936), the distribution E_{t+1} is more positively (less negatively) skewed than the distribution of E_t, if the correlation between E_t and k_t is zero or positive. During the first period of observed earnings, $Y_{s0} = E_s(1 - k_0)$. If the correlation of k_t with E_t is strongly positive, so is that of k_0 with E_s. In that case, skewness of initial earnings (Y_{s0}) is likely to be smaller than at overtaking (E_s). The U-shaped result of a near-zero correlation, and a declining profile of skewness due to a negative correlation are deduced in the same way.

The association between schooling and skewness, at given stages of experience, appears positive more often than the association between schooling and inequality (log variance), since the reversal of ranks takes place earlier in the working life (Chart 6.4). This is partly because the greater incidence of employment instability reduces positive skewness at lower levels of schooling.[9] In consequence, in the marginal distributions skewness generally increases with age and with schooling (see Table 6.3). This is not only because within-group skewness is larger at higher schooling levels and older ages. As noted before, in the aggregation process, the positive correlation between group variances and group means augments skewness and sharpens the gradient.

Positive skewness is a persistent feature of aggregate income distributions. Its presence has drawn a great deal of attention, starting with Pareto. Many of the theories of income distribution,[10] particularly the stochastic models, which will be discussed in the next section, were concentrated almost exclusively on this feature of the distribution. As we have seen, in human capital models, skewness is analyzed and explained at several levels. In the schooling model, skewness is made conditional on the shape of the distribution of schooling, and is not predicted as an inherent and persistent feature: the shape of the schooling distribution is exogenous to the model and does change secularly. Already the schooling distributions of the younger cohorts in the United States are negatively skewed. Why then does positive skewness persist?

8. See Aitchison and Brown (1957, pp. 22–23).

9. See discussion section 7.2.

10. Cf. Mincer (1970).

TABLE 6.3
INEQUALITY AND SKEWNESS IN MARGINAL DISTRIBUTIONS OF EARNINGS, 1959
(white, nonfarm men)

	Log Variance (1)	Dollar Standard Deviation (2)	Skewness: Logs [a] (3)	Skewness: Dollars [a] (4)	Skewness: RM [b] (5)
Age					
25–29	.433	2,610	−.206	−.020	1.015
30–34	.343	3,060		.140	1.068
35–39	.388	4,050	−.131	.157	1.107
40–44	.426	4,380		.167	1.119
45–49	.498	4,880	−.151	.172	1.163
50–54	.506	5,180		.190	1.174
55–59	.590	4,620	−.193	.176	1.162
60–64	.671	5,490		.185	1.188
Schooling (years)					
Under 8	.740	2,120		.018	1.070
8	.682	3,020	−.350	.066	1.057
9–11	.542	3,280		.128	1.072
12	.397	3,740	−.242	.178	1.111
13–15	.503	5,160		.349	1.211
16 or more	.534	6,810	+.046	.440	1.314

a. Bowley's measure of skewness.
b. Ratio of mean to median.

The answer lies in the positive correlation between dollar means and variances in the age-schooling groups of the earnings structure. As was pointed out before, this correlation reflects persistence in human capital accumulation: individuals who accumulate more capital over a lifetime invest larger amounts in most of the successive time periods.

Some of the stochastic or mathematical theories of income distribution generate Pareto or log-normal "equilibrium" distributions. Both forms are positively skewed. The Pareto distribution is also positively skewed in logs, while the log-normal one is symmetric in logs. Observed distributions, however, are typically positively skewed in dollars and *negatively* skewed in logs (Table 6.3, column 3).[11]

11. See also the findings of T. P. Hill (1959).

In the 1959 earnings data, the dollar distribution of earnings at overtaking was practically symmetric. It was, therefore, negatively skewed in logs, as was the distribution of schooling in years. Since dollar skewness grows with experience and age, logarithmic negative skewness diminishes correspondingly, as is shown in Table 6.3, column 3.

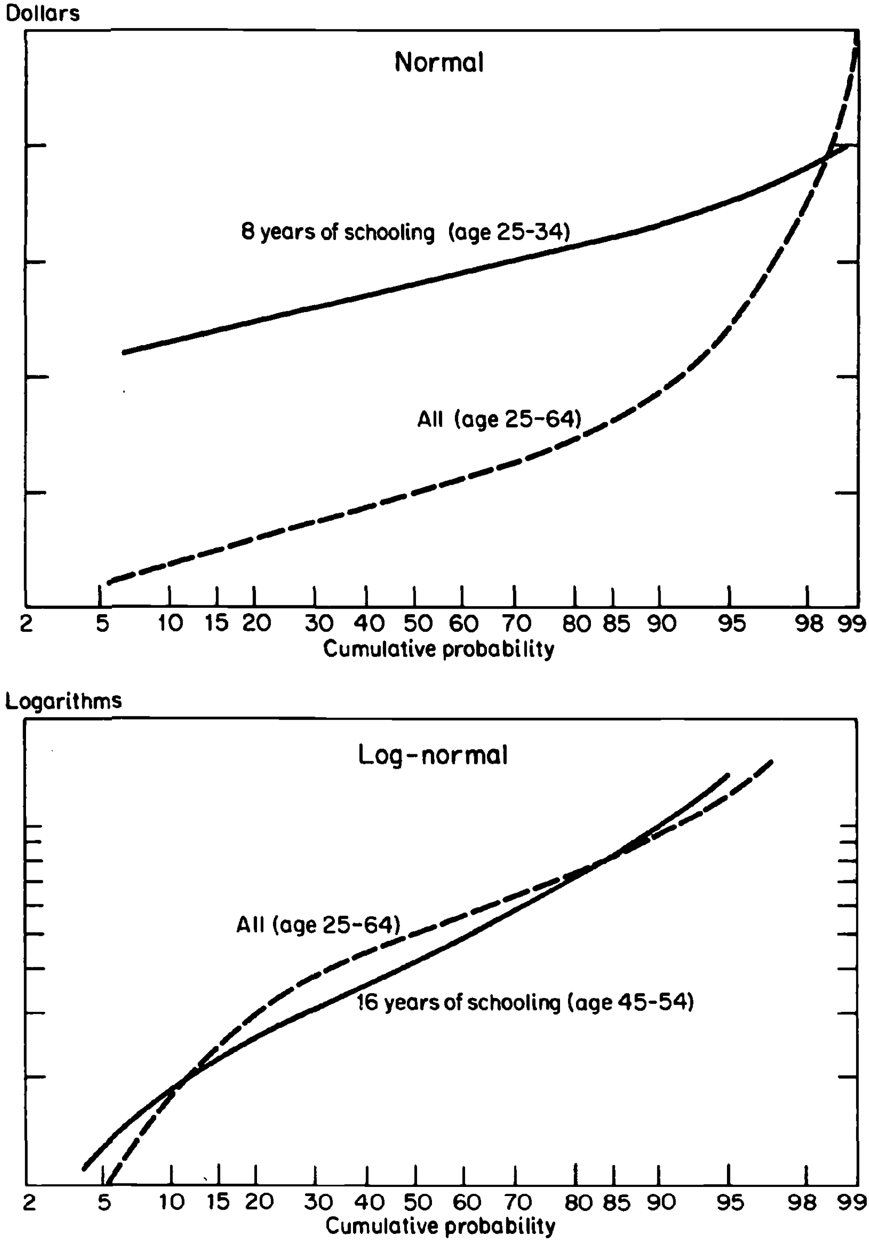
It also appears from the analysis of profiles of relative variances (Charts 6.2 and 6.3) that the sign of the overall correlation between group means and variances in logs is neither clearly positive nor negative: Group variances are positively related to means at upper levels of schooling and experience, negatively at lower levels. There is, therefore, little reason for the aggregation process to produce positive skewness or symmetry in logs in the overall distribution, when this is not true of the components. There is an implication, however, that aggregation within upper schooling-experience groups, hence upper earnings groups, tends to impart positive logarithmic skewness, while aggregation within lower earnings groups does the opposite. This may well explain the leptokurtic shape of the overall distribution, which has been observed in a number of studies (Rutherford, 1955; Bjerke, 1969; Lydall, 1968). When drawn on log-normal probability paper, the graph of the cumulative distribution is S-shaped, strongly concave at lower levels of earnings, and convex at upper levels. This reflects strong negative skewness (in logs) at lower levels and some positive skewness at upper levels of earnings. When drawn on normal probability paper, the lower level of the graph is linear (zero dollar skewness), the upper sharply convex (strong positive skewness in dollars). Chart 6.5 shows that the graph of a component distribution at a lower age-schooling group shows a relatively good fit to the normal distribution, while the graph of a higher age-schooling group shows a closer fit to the log-normal.

Summing up: If earnings distributions are to be classified on a scale of skewness somewhere between normal distributions (zero dollar skewness), log-normal (zero log skewness), and Pareto (positive log skewness), they fit between the normal and log-normal. As indicated by the curves in Chart 6.5, shapes of component distributions when ranked by average level of human capital systematically range from symmetry in dollars to symmetry in logarithms. This is entirely consistent with the theoretical conjecture in Chapter 2.

CHART 6.5

FIT OF ANNUAL EARNINGS DISTRIBUTIONS TO NORMAL AND LOG-NORMAL CURVES, 1959

(white, nonfarm men)



NOTE: Figures on each curve show age (in parentheses) and years of schooling completed.

SOURCE: 1/1,000 sample of U.S. Census, 1960.