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: Interregional Applications

Chapter Author: Barry R. Chiswick

Chapter URL: http://www.nber.org/chapters/c3672

Chapter pages in book: (p. 48 - 82)

4

Interregional Applications

The relationships between income inequality and the rate of return from schooling, the variance of schooling, and the residual variance—the focal points of the model developed in the preceding chapter—are analyzed here empirically on an interregional basis for the United States, Canada, and the Netherlands. A summary of the findings and a discussion of the relationships among the relevant parameters conclude Chapter 4.

The empirical analyses—here as well as in Chapter 5—rest on the theoretical model of Chapter 3. A linear regression (equation 3-17) of the log of income on years of schooling is run for each region of the United States and Canada and for Puerto Rico and Mexico. This regression provides data on the regression estimate of the rate of return ($\hat{r}$), the residual variance (Var ($\hat{U}$)), and the intraregional explanatory power ($\bar{R}^2$) for each region. These data, together with the inequalities of income and schooling, are then used as the inputs in the second level of the analysis. Here interregional differences in income inequality are related to interregional differences in the rate of return from schooling, the inequality of schooling, the education component ($\hat{r}^2$ Var (S)), and the residual variance (Var ($\hat{U}$)).

Two estimates of the explanatory power of schooling are calculated: (1) the proportion of individual differences in the log of income ($\ln E$) within a region that can be explained by years of schooling (S) and (2) the proportion of the variations in income inequality (Var ($\ln E$)) across regions that can be explained by the education component ($\hat{r}^2$ Var (S)). As will be shown below, there is a systematic downward bias in the fraction of the differences in the log of income explained by schooling within each region. If these biases are similar in each region, the fraction of interregional differences in income inequality explained by schooling may be unaffected by the intraregion bias.

As to coverage, it is a truism that, as a prerequisite for an empirical analysis, the data must first be delimited, the relevant population and income concepts defined. All inhabitants of a region cannot be included in the data for the purposes of this study. Students, for example, must be removed because the model is designed to cover those who have already completed their investment in schooling. Wives, whose labor force behavior is strongly influenced by their husband's income and the number and age distribution of their children, should also be excluded from the data base. Finally, the aged should be excluded, since many of them have low labor force participation rates due to ill health, discrimination, and pension income. Thus, we approximate the desired group by restricting the data to males between the ages of twenty-five and sixty-four.[1] Where possible and practicable, separate calculations are made to remove the effects of racial differences.

On the income side, the dependent variable used is market earnings. This is dictated by the fact that the model is developed for labor income and data on nonmarket and psychic earnings are unavailable. The total annual money earnings of those with some earnings reflect unemployment, but persons unemployed for the entire year are omitted from the study.[2]

THE UNITED STATES

In examining the effects of schooling on the intraregional and interregional differences in labor market income inequality, we first look at the two major regions of the United States—the South and the non-South. Next comes an extensive interstate analysis,

1. The model was developed under the assumption of an infinite working life. However, for positive rates of return and long labor force participation, the difference in the rate of return between assuming an infinite and a finite working life is trivial.

2. Unemployment compensation and home production during the period of unemployment are omitted from the earnings data. The proportion of adult males unemployed for an entire year, however, is very small.

TABLE 4-1

Regression of Natural Log of Earnings on Schooling: Males, Twenty-five to Sixty-four, South versus Non-South, 1959

	Summary Statistics				Regression: $\ln E$ on S_2			
	SD($\ln E$)	SD(S)	AV($\ln E$)	AV(S)	$\ln E_{0,1}$	$\hat{r}_1$	$\bar{R}_1^2$ R_1^2	Var(U)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Non-South, white	.65	3.36	1.66	10.78	.96 (.26)	.06 (.02)	.10 .11	.38
2. South, white	.74	3.90	1.43	9.96	.60 (.23)	.08 (.02)	.18 .20	.44
3. Non-South, total	.65	3.41	1.63	10.67	.94 (.23)	.06 (.02)	.10 .11	.38
4. South, total	.76	4.03	1.32	9.42	.47 (.20)	.09 (.02)	.22 .23	.45

Note: The following definitions hold for Table 4-1 and all subsequent tables:

E = annual earnings in thousands of dollars.
Y = annual income in thousands of dollars.
S = years of schooling attended.
U = residual income or earnings.
$\hat{r}$ = regression estimate of adjusted rate of return, slope computed from regression equation (3-17).
$\ln E$ = natural log of E, similar for $\ln Y$.
$\ln E_0$ = zero education level of earning, intercept computed from regression equation (4-17), similar for $\ln Y_0$.
$\bar{R}^2$ = adjusted coefficient of determination; R^2 = unadjusted, computed from regression equation (3-17).
SD = standard deviation.

based, for the most part, on money income rather than earnings data (since for the individual states the latter are available only for males of fourteen years and over). Attention is focused on the rate of return, and a procedure is developed to improve the model's explanatory power by correcting for the downward bias inherent in the regression estimate of the rate of return.

To test the general validity of the relationships found, the interstate analysis is repeated for the income inequality of white males alone, for the earnings inequality of all males, and for another measure of the rate of return from schooling called the "overtaking age rate of return."[3] Finally, the relationship between the level of schooling and income inequality is also examined.

3. See Chapter 3, footnote 14.

TABLE 4-1 (Concluded)

Regression: ln*E* on ₵, $, and *H*					
$\ln E_{0,3}$	$\hat{r}_{\text{₵}}$	$\hat{r}_{\$}$	$\hat{r}_H$	$\bar{R}_3^2$ / R_3^2	$\mathrm{Var}(U)_3$
(9)	(10)	(11)	(12)	(13)	(14)
1.09	.05	.06	.08	.07	.39
(.67)	(.09)	(.06)	(.06)	.11	
.66	.07	.09	.09	.16	.46
(.50)	(.08)	(.07)	(.06)	.20	
1.06	.05	.06	.08	.08	.39
(.58)	(.08)	(.05)	(.05)	.12	
.50	.08	.09	.09	.20	.46
(.40)	(.07)	(.07)	(.06)	.23	

AV = average.
Var = variance.

Subscripts 1, 3, ₵, $, *H*: 1 is used when schooling is treated as a single variable; 3 and ₵, $ and *H* are used when it is treated as three variables; ₵ or "low" education is defined as 0–8 years of school, $ or "medium" as 8–12 years, and *H* or "high" as more than 12 years.

In calculating the adjusted coefficients of determination for the regression of the natural log of earning on schooling, the number of degrees of freedom was assumed equal to the number of cells minus the number of parameters estimated. For the regions there were 7 schooling intervals and 11 income intervals for a total of 77 cells.

Standard errors are in parentheses.

Source: U.S. Census of Population: 1960, Subject Reports, Occupation by Earnings and Education, Tables 2 and 3.

SOUTH VERSUS NON-SOUTH

The relationships between schooling and income inequality in the South and the non-South are compared by regressing the natural log of earnings on years of schooling for males between the ages of twenty-five and sixty-five in the Census Bureau's "South" and "non-South."[4] The results appear in Table 4-1, with rows 1

4. The "South" consists of sixteen Southern states plus the District of Columbia. The remainder of the country, the "non-South," is also referred to as the "North."

The data are from a 5 per cent sample of the population grouped into seventy-seven cells. The midpoint of each closed earnings interval represents the earnings of all persons in the interval. The shape of the lower end of the earnings distribution is not definitely known, and with census definitions and 2 reporting the results for white males only, and rows 3 and 4, for all males—white and nonwhite.

Reading down columns (1) through (14), we see that the inequalities of earnings, "residual" earnings and schooling, and adjusted rates of return from education are higher in the South than in the non-South.[5] The education component ($\hat{r}^2$ Var (S)) and the coefficient of determination are also larger in the South, while levels of earnings and education are lower there. A comparison of rows 1 and 2 with rows 3 and 4 indicates that the same patterns exist when both whites and nonwhites are covered in the data, but that the inclusion of nonwhites widens the regional differences.[6]

Regression estimates of the rate of return for college and secondary education are presented in Table 4-2. These can be isolated by assuming that the appropriate investment-income ratios are $\overline{K}_H = 1.0$ and $\overline{K}_{\$} = 0.75$.[7] Clearly, the regression technique

---

there can never be negative earnings. Therefore, the class mean of the lower open-end interval is considered to be the midpoint between zero earnings and the upper limit of the interval. The effect of any error introduced in this estimate would be small since the lowest earnings group contains only a small part of the samples analyzed. The Pareto equation, which provides a fairly good fit for the income distribution in the higher income ranges, is used to estimate the mean income in the upper open-end interval. See M. J. Bowman, "A Graphical Analysis of Personal Income Distribution in the United States," *American Economic Review*, September 1945; and N. O. Johnson, "The Pareto Law," *Review of Economics and Statistics*, February 1937, pp. 20-26.

5. Since direct estimates of rates of return from schooling by state are not available for the states, equation (3-17) is used here to generate estimates from census data. The data for the South and non-South are from the 1960 *Census of Population* and cover 1959 earnings of all males in the experienced labor force between the ages of twenty-five and sixty-four.

The pair-wise product moment correlations among the explanatory variables when schooling is divided into three components are not large, but due to the definitions of the variables, the multicollinearity is substantial. Note that the standard error of the slope coefficient increases from 0.02 to a low of 0.05 and a high of 0.09.

Hanoch's estimates also show higher internal rates of return from schooling in the white South. See G. Hanoch, "Personal Earnings and Investment in Schooling," Ph.D. dissertation, University of Chicago, 1965, pp. 71 and 84.

6. The effects of the inclusion of nonwhites are discussed in Appendix B-1 with special reference to the interstate analysis.

7. For college education in the United States, Becker estimates that direct costs are approximately equal to potential summer earnings, or $\overline{K}_H = 1.0$. (See his *Human Capital*, Chapter 4.) A low estimate for $\overline{K}_{\$}$, the high-school investment income ratio, is 0.75. This is based on the assumption of a nine-month academic year and no direct costs of schooling. Positive direct costs would raise the ratio. No estimate has been made of $\overline{K}_{\not{p}}$, and therefore $\hat{r}_{\not{p}}$ cannot be separated into its components $\overline{r}$ and $\overline{K}_{\not{p}}$.

TABLE 4-2

Calculated Rates of Return for Adult Males in the United States and Canada (per cent)

	Elementary School (0–8 Years)	High School (9–12 Years)	College (13+ Years)
United States			
Becker (white)	—	20.0	13.0
Hansen (all)			
I	15.0	11.4	10.2
II	∞	14.5	10.1
Hanoch (white, North)			
I	>100.0	16.1	9.6
II	33.6	18.0	10.0
Regression estimate			
(all)	—	9.2	8.0
(white)	—	8.9	7.8
(North)	—	8.2	7.6
(white, North)	—	8.4	7.6
Canada			
Podoluk	—	16.3	19.7
Regression estimate	—	10.0	7.6

Sources:

United States

G. S. Becker, *Human Capital*, New York, NBER, 1974, Chapters 4 and 6. Marginal internal private rates of return based on earnings for graduates. No rate of return was calculated for elementary education. College is for 13 to 16 years of education. Data from *U.S. Census of Population: 1950.*

W. L. Hansen, "Total and Private Rates of Return to Investment in Schooling," *Journal of Political Economy*, April 1963, pp. 128–140, for income of adult males as reported in the *U.S. Census of Population: 1950.* College refers to 13 to 16 years of schooling. I: p. 134, Table 3. Internal rate of return on total resources. II: p. 137, Table 5. Internal rate of return to private resources after taxes. Infinite rate of return to elementary education, given assumption of costless education to the individual through the eighth grade.

G. Hanoch, "Personal Earnings and Investment in Schooling," Ph.D. dissertation, University of Chicago, 1965. Private internal rates of return from two methods of calculation, based on earnings from *U.S. Census of Population: 1960.* Hanoch assumes that potential earnings equal the cost of the investment. I: p. 71, Table 6. II: p. 84, Table 7.

Regression estimate: Adjusted rates of return divided by $\overline{K}_H = 1.0$ for college and $E_g = 0.75$ for high school. Estimated rates of return are not presented for elementary school because the appropriate value for K_g is unclear. Calculated from regressing log earnings on three education variables for males aged 25 to 64. *(U.S. Census of Population: 1960, Subject Reports, Occupation by Earnings and Education*, Washington, Bureau of the Census, Tables 1, 2, and 3.) See Table 6-1.

Canada

J. R. Podoluk, *Education and Earnings*, Ottawa, Dominion Bureau of Statistics, Central Research and Development Staff, 1965, pp. 60–65. Rates of return to private investments in schooling for completing secondary school

(Continued)

TABLE 4-2 (*Concluded*)

and receiving a university degree. Based on earnings before taxes for nonfarm males as reported in the *Census of Canada: 1961*.

Regression estimate: Adjusted rates of return from regressing the natural log of earnings on three education variables for nonfarm males, age 25 to 64, divided by $K_H = 1.0$ for college and $K_g = 0.75$ for high school. (*Census of Canada: 1961, Population Sample. Income of Individuals*, Bulletin 4.1-2, Ottawa, Dominion Bureau of Statistics, Table B6). See Table 6-1.

results in lower estimates of rates of return to secondary and higher education than the internal rate of return method.[8] The differences are quite large in the case of secondary schooling in the United States and Canada and of university education in Canada.

The regression slope coefficient and therefore the intraregional explanatory power of schooling appear to be downward-biased. A negative covariation term (i.e., a negative correlation between schooling and the residual) would result if the unbiased rates of return were used. The downward bias notwithstanding, column 7 of Table 4-1 indicates that the explanatory power of schooling is substantial, especially in the South.

The percentage of the variance of the log of earnings which is attributable to the education component and the percentage attributable to the residual variance can be calculated on the basis of the following equation. According to equation (3-21), where $v = \text{Var}(\ln E)$, $s = \hat{r}^2 \text{ Var}(S)$, and $t = \text{Var}(\hat{U})$, with two regions and with s and t positively related, then necessarily $R_{s,t} = 1.0$ and

$$\text{Var}(v) = \text{Var}(s) + \text{Var}(t) + 2R_{s,t}\text{SD}(s)\text{SD}(t)$$

$$\text{Var}(v) = [\text{SD}(s) + \text{SD}(t)]^2$$

$$\text{SD}(v) = \text{SD}(s) + \text{SD}(t).$$

8. This is not due to the pooling of age groups. The adjusted rate of return and explanatory power are low for all age groups, as seen in the table below, but rise when experience rather than age is held constant (see Mincer, *Schooling, Experience, and Earnings*).

Regression of Natural Log of Earnings in 1959 on Schooling

Age Group	Non-South Males		Non-South White Males	
	$\hat{r}$	R^2	$\hat{r}$	R^2
25-34	.051	.063	.049	.061
35-44	.077	.135	.077	.134
45-54	.078	.126	.078	.121
55-64	.070	.096	.070	.094

Note: Based on 77 cells.

Source: United States Census of Population: 1960, Subject Reports, Occupation by Earnings and Education, Tables 2 and 3.

The proportion of the variation in income inequality explained by the education component is SD (s)/SD (v), and the proportion explained by the residual is SD (t)/SD (v).

Applying this method to the data on Table 4-1, we find that the interregional explanatory power of the education component is approximately 48 per cent for white males and 54 per cent for all males. (See Table 4-5.) Thus, the interregional explanatory power of schooling is very large, both in absolute terms and relative to the intraregional explanatory power. This may result from a number of important factors. For example, differences in ability and luck vary considerably within regions but presumably have fairly similar distributions across regions. It seems plausible that the smaller the variation of these traits, the larger the explanatory power of schooling. Note that the inclusion of nonwhites in the data tends to increase this explanatory power.

INTERSTATE ANALYSIS

In contrast to the preceding two-region, South–non-South comparison, we turn to variations in the effects of schooling on income distribution among fifty states and the District of Columbia (fifty-one states).

Income Inequality

Regression equation (3-17) is computed for each state, with the natural log of personal income ($\ln Y$) for 1959 as the dependent variable.[9] The data cover males of twenty-five years and over with some income.[10]

Table 4-3 shows the correlation matrix for the parameters under study. Here each of the fifty-one states represents an observation. Column 1 of Table 4-3 indicates a significant positive correlation of the log variance of income with the rate of return

9. While it would be preferable to restrict the state data to earnings of adult males under sixty-five, such statistics are unobtainable. However, including males over sixty-five and income other than earnings does not disturb the qualitative regional results, although the values of the parameters are altered (see Appendix B-2).

10. The income data come from the *U.S. Census of Population: 1960, Vol. 1, Characteristics of the Population*, Parts 2-52, Table 138; they are based on a 25 per cent sample cross-classified by schooling and income in seventy-two cells—eight intervals for schooling and nine intervals for income.

TABLE 4-3

Matrix of Correlation Coefficients for Fifty-one States

	$\mathrm{Var}(\ln Y)$ (1)	$\hat{r}_1$ (2)	$\ln Y_{0.1}$ (3)	$\overline{R}^2$ (4)	$\mathrm{Var}(S)$ (5)	$\mathrm{Var}(U)$ (6)	$\hat{r}_1^2$ (7)	$\hat{r}_1\,\mathrm{Var}(S)$ (8)	$\mathrm{AV}(\ln Y)$ (9)
$\hat{r}_1$	.77								
$\ln Y_{0.1}$	-.86	-.94							
$\overline{R}^2$	.79	.77	-.68						
$\mathrm{Var}(S)$	.48	.24	-.19	.76					
$\mathrm{Var}(U)_1$	.90	.78	-.77	.46	.18				
$\hat{r}_1^2$	.77	1.00	-.94	.79	.27	.76			
$\hat{r}_1^2\ \mathrm{Var}(S)$	.88	.86	-.79	.98	.69	.62	.87		
$\mathrm{AV}(\ln Y)$	-.79	-.79	.94	-.67	-.32	-.66	-.81	-.76	
$\mathrm{AV}(S)$	-.73	-.66	.70	-.78	-.62	-.51	-.68	-.81	.84

Note: The null hypothesis is that the correlation coefficient in the population is zero. The probabilities represent the chance that sample estimates of R will be further from zero than the values given (one-tailed test). The probabilities are based on the number of degrees of freedom equal to, or nearest to, the number of observations minus two. See R. A. Fisher and F. Yates, *Statistical Tables*, London, Oliver and Boyd, 1938, Table VI, pp. 36–37. The critical values for the correlation coefficient (R), under alternative type I errors (α), are $R(\alpha = .05) = .23$, $R(.025) = .27$, $R(.01) = .32$. The critical values are based on 50 degrees of freedom.

Source: See Table A-1.

($\hat{r}$), the inequality of schooling (Var (S)), and the education component ($\hat{r}^2$ Var (S)), but a negative correlation with the level of income and schooling.

The relationship between the rate of return ($\hat{r}$) and the variance of schooling (Var (S)) (Table 4-3, row 4, column 2) is quite weak (0.24), and the hypothesis that no statistically significant correlation exists cannot be rejected. There is a strong positive correlation between the variance of residual income and the variance of log income. However, the intrastate explanatory power of schooling [$R^2 = 1 - \text{Var}(U)/\text{Var}(\ln Y)$] is significantly positively correlated with income inequality. This implies that interstate differences in the residual variance (Var (U)) are smaller than interstate differences in the variance of income. In other words, education increases interstate differences in the variance of income.

The explanatory powers of the education component, the residual, and their covariation are shown in Table 4-5, row 3. It appears that each of the components explains approximately one-third of the interstate variances in income inequality. Again, while education explains an important part of *interstate* differences in the variance of income, it has less *intrastate* explanatory power. The adjusted coefficient of determination ranges from 32 per cent for Mississippi to 11 per cent for Nevada, with a mean value of 18.4 per cent.[11]

Regression Estimate of Rate of Return Corrected

The results of the regression of Var ($\ln Y$) on $\hat{r}^2$ Var (S) across fifty-one states appear in row 1 of Table 4-4. The slope $b = 1.58$ suggests that the estimate of the average rate of return across the states is $\overline{r}_c = \bar{\hat{r}}\sqrt{b} = (0.102)\sqrt{1.58} = 0.126$.[12]

11. The relatively low coefficients of determination reflect the large variation of individual income and rates of return within a given level of schooling, possibly due to individual differences in rates of return, experience, and employment. The variance and coefficient of variation in rates of return from schooling have received little attention in the literature. Two exceptions are Mincer's *Schooling, Experience, and Earnings* (Part 2) and my "Racial Differences in the Variation in Rates of Return from Schooling" in G. von Furstenberg et al., eds., *Patterns of Racial Discrimination*, Vol. 2, D. C. Heath, 1974.

12. Recall equation (3-16): $\text{Var}(\ln E_{S,i}) = (\bar{r}^*)^2\ \text{Var}(S) + \text{Var}(U'') + 2(\bar{r}^*)$ Cov (S, U''). The regression across states may be run:

$$\text{Var}(\ln E_{S,i}) = a + b[(\hat{r})^2\ \text{Var}(S)] + V_i,$$

where a is the average Var (U'') + 2($\hat{r}$) Cov (S, U'') and V_i is a random residual. The value $b\ (\hat{r})^2$ is our estimate of $(\bar{r}^*)^2$.

TABLE 4-4

Regression Analysis of U.S. Data for Fifty-one States

Regression	Slope Coefficient[a]	Coefficient of Determination
Var(ln Y) on $\hat{r}^2$ Var(S)	1.58 (0.11)	0.77
Var(ln Y) on r_M^2 Var(S)	0.52 (0.05)	0.68
Var(ln$E$) on $\hat{r}^2$ Var(S)	1.30 (0.14)	0.63
Var(ln$E$) on r_M^2 Var(S)	0.45 (0.05)	0.61

Source: Data from Tables A-1 and A-4.
[a]Standard errors are in parentheses.

An adjustment for the corrected rate of return $\hat{r}_{ci} = \hat{r}_i\ (\sqrt{1.58})$ alters the interregional and intraregional explanatory power of schooling. This correction raises the education component's direct explanatory power of interstate differences in income inequality to (1.58) × (30.5) = 48.2, or approximately 50 per cent.[13] The proportion of differences in income within states explained by schooling is also increased if each state's rate of return were corrected by the factor $\sqrt{1.58}$. Rankings would not change, but the adjusted coefficient of determination would range from (1.58) × (32) = 51 per cent for Mississippi to (1.58) × (11) = 17 per cent for Nevada, with a mean value of (1.58) × (18.4) = 29 per cent.

Interstate Analysis of White Males

In an attempt to determine whether the interstate results are the consequence of different proportions of nonwhites in the data, the regression analysis was performed for the fifty-one states, but with nonwhites deleted from the data of seventeen states. These states include all those in which nonwhites constitute eight per cent or more of the relevant population.[14]

The regression analysis for white males substantiates the findings for all males, but shows generally weaker results. The average

13. Under the assumption that the covariance between $(r_c)^2$ Var (S) and the residual variance is nonnegative.

14. The results are discussed in Appendix B-1.

TABLE 4-5

The Explanatory Power of Schooling in the United States and Canada (per cent)

Region	Education Component (1)	Residual (2)	Covariation of Education Component and Residual (3)	Average Intra-Area Explanatory Power (4)
North-South (whites) (2 areas)	47.6	52.4	–	14.1
North-South (2 areas)	54.0	46.0	–	16.1
States (51 states)	30.5	32.0	37.5	18.4
Corrected $\hat{r}$	48.2	51.8		29.1
States (17 states adjusted for nonwhites)	22.3	43.2	34.5	17.4
Non-South (34 states)	13.3	57.9	28.8	15.4
South (17 states)	48.2	40.0	11.8	24.4
Non-South (34 states adjusted for nonwhites)	16.5	57.0	26.5	15.0
South (17 states adjusted for nonwhites)	18.8	66.5	14.7	21.9

Note: If $s = \hat{r}^2 \operatorname{Var}(S)$, $t = \operatorname{Var}(\hat{U})$, and $v = \operatorname{Var}(\ln Y)$, then $v = s + t$ and

$$1 = \frac{\operatorname{Var}(s)}{\operatorname{Var}(v)} + \frac{\operatorname{Var}(t)}{\operatorname{Var}(v)} + \frac{2\operatorname{Cov}(s,t)}{\operatorname{Var}(v)}.$$

The three ratios are the interregional explanatory powers of schooling, the residual, and their covariation, respectively. If there were only two regions and the education component and the residual variance were positively correlated, $1 = \frac{SD(s)}{SD(v)} + \frac{SD(t)}{SD(v)}$ where the two ratios would be the explanatory powers of schooling and the residual.

Source: Tables 4-1, A-1, A-2, A-3.

explanatory power of schooling within the states decreases slightly from 18.4 per cent to 17.4 per cent, while a comparison of rows 3 and 5 of Table 4-5 indicates that the interstate explanatory power of schooling decreases by a larger proportion. Schooling is still a very important variable, and more so interstate than intrastate, even though both explanatory powers are reduced.

Intra-South and Intra-non-South

In the South-non-South analysis (p. 51) we saw a clear regional difference. To ascertain whether the interstate conclusions are due solely to North-South differences, correlation matrices are cal-

TABLE 4-6

Matrix of Correlation Coefficients, All Males, Seventeen Southern States
(17 observations)

	Var(ln Y) (1)	$\hat{r}_1$ (2)	ln $Y_{0,1}$ (3)	$\overline{R}^2$ (4)	Var(S) (5)	Var(U) (6)	$\hat{r}_1^2$ (7)	$\hat{r}_1^2$ Var(S) (8)	AV(ln Y) (9)
r_1	.79								
ln $Y_{0,1}$	-.85	-.91							
$\overline{R}^2$	.58	.82	-.70						
Var(S)	.15	-.08	.04	.43					
Var(U)	.73	.28	-.44	-.14	-.19				
$\hat{r}_1^2$	.79	1.00	-.93	.81	-.08	.28			
$\hat{r}_1^2$ Var(S)	.79	.89	-.83	.96	.39	.15	.89		
AV(ln Y)	-.84	-.78	.95	-.61	-.08	-.50	-.79	-.77	
AV(S)	-.80	-.74	.81	-.68	-.23	-.40	-.73	-.79	.89

Note: The critical values for the correlation coefficient (R), under alternative type I errors (α), are $R(\alpha = .10) = .33$, $R(.05) = .41$, $R(.025) = .48$, $R(.01) = .56$. The critical values are based on 15 degrees of freedom. See notes to Table 4-3.

Source: Table A-1.

culated for the seventeen Southern and thirty-four non-Southern states; they appear in Tables 5-6 and 5-7.

Within the South and non-South, the variables in equation (3-19), the adjusted rate of return ($\hat{r}$), and the education component ($\hat{r}^2$ Var (S)) show a significant positive correlation with the variance of log of income (Var $(\ln Y)$). The variance of the residual (Var $(\hat{U})$) and the adjusted coefficient of determination also exhibit a significant positive correlation with the inequality of income. The variance of schooling does not fit the previously observed patterns, is not significantly related to income inequality, and has a significant negative correlation with the rate of return in the non-South. The levels of schooling and income appear to be negatively related to income inequality, especially in the South.

The average intrastate explanatory power of schooling in the South is 24.4 per cent, compared with 15.4 per cent in the non-South. An adjustment for the corrected rate of return, $\hat{r}_{ci} = (1.26)\, \hat{r}_i$, raises the average explanatory power to 38.6 per cent in the South and to 24.3 per cent in the non-South. In the non-South (row 6 of Table 4-5) schooling explains 13.3 per cent and the residual, 57.9 per cent of the variation in income inequality, compared with 48.2 per cent and 40.0 per cent, respectively, in the South (row 7). The correlation between the education component and the residual variance is insignificant in the South but significant in the remainder of the country (row 7, column 6 in Tables 4-6 and 4-7). This results in a much smaller interstate explanatory power of their covariation in the South than in the non-South.

In the white South (Table 4-8), income inequality is significantly positively correlated with the adjusted rate of return, the education component, and the residual variance.[15] It is not significantly positively correlated with the inequality of schooling and the adjusted coefficient of determination. Schooling inequality is almost significantly negatively related to the adjusted rate of return.

The average coefficient of determination adjusted for degrees of freedom for the white South is 21.9 per cent. It is clear from column 4 of Table 4-5 that this is below the figure for the total

15. In only three of the seventeen Southern states was the proportion of nonwhites among adult males very small. These were Kentucky (7.0 per cent), Oklahoma (7.7 per cent), and West Virginia (4.5 per cent). The data for all males in these three states were combined with the data computed for whites in the other fourteen states in order to ascertain the parameters for the "white" South.

TABLE 4-7

Matrix of Correlation Coefficients, All Males, Thirty-four Non-Southern States (34 observations)

	Var(ln Y) (1)	$\hat{r}_1$ (2)	ln $Y_{0,1}$ (3)	$\overline{R}^2$ (4)	Var(S) (5)	Var(U) (6)	$\hat{r}_1^2$ (7)	$\hat{r}_1^2$ Var(S) (8)	AV(ln Y) (9)
$\hat{r}_1$	.89								
ln $Y_{0,1}$	-.75	-.89							
$\overline{R}^2$	.45	.38	-.14						
Var(S)	-.12	-.40	.51	.61					
Var(U)	.95	.86	-.78	.16	-.34				
$\hat{r}_1^2$	.90	1.00	-.89	.40	-.36	.86			
$\hat{r}_1^2$ Var(S)	.78	.68	-.45	.90	.36	.56	.70		
AV(ln Y)	-.48	-.57	.86	.02	.36	-.54	-.58	-.20	
AV(S)	-.18	-.06	.14	-.30	-.38	-.11	-.09	-.26	.46

Note: The critical values for the correlation coefficient (R), under alternative type I errors (α), are $R(\alpha = .05) = .30$, $R(.025) = .35$, $R(.01) = .41$. The critical values are based on 30 degrees of freedom. See notes to Table 4-3.

Source: Table A-1.

TABLE 4-8

Matrix of Correlation Coefficients, Whites, Seventeen Southern States (17 observations)

	$\mathrm{Var}(\ln Y)$ (1)	$\hat{r}_1$ (2)	$\ln Y_{0,1}$ (3)	$\overline{R}^2$ (4)	$\mathrm{Var}(S)$ (5)	$\mathrm{Var}(U_1)$ (6)	$\hat{r}_1^2$ (7)	$\hat{r}_1^2\,\mathrm{Var}(S)$ (8)	$\mathrm{AV}(\ln Y)$ (9)
$\hat{r}_1$	.60								
$\ln Y_{0,1}$	-.62	-.92							
$\overline{R}^2$	.17	.77	-.61						
$\mathrm{Var}(S)$	.00	-.31	.35	.15					
$\mathrm{Var}(U_1)$	.91	.24	-.33	-.26	-.04				
$\hat{r}_1^2$	.60	1.00	-.92	.76	-.32	.25			
$\hat{r}_1^2\,\mathrm{Var}(S)$	.62	.92	-.81	.87	.08	.23	.91		
$\mathrm{AV}(\ln Y)$	-.11	-.36	.51	-.42	-.04	.09	-.37	-.41	
$\mathrm{AV}(S)$	.18	-.20	.29	-.48	-.22	.39	-.19	-.32	.88

Note: The critical values for the correlation coefficient (R), under alternative type I errors (α), are $R(\alpha = .10) = .33$, $R(.05) = .41$, $R(.025) = .48$, $R(.01) = .56$. The critical values are based on 15 degrees of freedom. See notes to Table 4-3.

Source: See Tables A-1 and A-2.

TABLE 4-9

Matrix of Correlation Coefficients, Whites, Thirty-four Non-Southern States
(34 observations)

	Var(ln Y) (1)	$\hat{r}_1$ (2)	ln $Y_{0,1}$ (3)	$\overline{R}^2$ (4)	Var(S) (5)	Var(U_1) (6)	$\hat{r}_1^2$ (7)	$\hat{r}^2$ Var(S) (8)	AV(ln Y) (9)
$\hat{r}_1$	.87								
ln $Y_{0,1}$	-.76	-.92							
$\overline{R}^2$	.45	.58	-.38						
Var(S)	.06	-.01	.19	.77					
Var(U_1)	.93	.75	-.71	.10	-.26				
$\hat{r}_1^2$	.88	1.00	-.91	.56	-.03	.77			
$\hat{r}_1^2$ Var(S)	.75	.77	-.58	.92	.60	.47	.77		
AV(ln Y)	-.59	-.73	.90	-.19	.32	-.59	-.72	-.37	
AV(S)	-.40	-.54	.54	-.45	-.13	-.29	-.52	-.48	.69

Note: The critical values for the correlation coefficient (R), under alternative type I errors (α), are $R(\alpha = .05) = .30$, $R(.025) = .35$, $R(.01) = .41$. The critical values are based on 30 degrees of freedom. See notes to Table 4-3.

Source: Tables A-1 and A-2.

South but above that for the North. In the white South, schooling explains 18.8 per cent, the residual, 66.5 per cent, and their co-variation, 14.7 per cent of the differences in income inequality. Although the removal of nonwhites from the data substantially reduces their interstate explanatory power, schooling is still important in explaining differences in income inequality. However, contrary to the findings for all males, for all white males, and for all males in the South, schooling for white males in the South is a less important factor in variations between states than in variations within states.

Calculations were also made for the thirty-four non-Southern states, with nonwhites excluded from the data for Alaska, Hawaii, and New York.[16] The data of Tables 4-7 and 4-9 indicate that the algebraic value of the correlation coefficients between the variance of schooling and the inequality of income, the adjusted rate of return, the residual variance, and the adjusted coefficient of determination are increased by the exclusion of nonwhites. On the other hand, the average intrastate explanatory power is decreased from 15.4 to 15.0 per cent. Rows 5 and 7 of Table 4-5 indicate that the "adjustment" for nonwhites slightly increases the interstate explanatory power of schooling in the non-South.

Additional Analyses

An alternative definition of income inequality and an alternative measure of the rate of return from schooling are used in this section to test the robustness of the model for explaining income inequality. Virtually identical results emerge from calculations based on the income inequality of adult males and the earnings inequality for all males. Similarly, when the overtaking age rate of return is used rather than the regression estimate, there is no fundamental change in the patterns that emerge. A conservative estimate of the interstate explanatory power of schooling for alternative definitions of income and the rate of return appears to be 60 per cent. Thus, the model is found to be robust.

Analysis of Earnings Inequality

Thus far the empirical analysis has been restricted to income rather than earnings. The 1960 census does contain data by state

16. In 1960 the proportions of nonwhites were: Alaska, 17.6 per cent; Hawaii, 68.9 per cent; and New York, 8.0 per cent. All other non-Southern states had very small proportions of nonwhites.

on the distribution of earnings in 1959 of males fourteen years of age and older who were in the experienced labor force in 1960.[17] This made it possible to calculate the variance of the natural log of earnings of males fourteen years of age and older for each state. Unfortunately, neither are the data cross-classified by schooling, nor can youths of fourteen to twenty-four be eliminated. The inclusion of young males tends to raise, and the exclusion of property income tends to lower, the variance of the log of earnings compared to the variance of the log of income of males aged twenty-five and over, the data used in the other interstate analyses. However, the variance of the log of income and the variance of the log of earnings for the states are very similar. Their product-moment correlation coefficient is $R = +.91$, and neither their means nor their variances differ significantly from each other.

As shown in Table 4-10, the regression estimate of the rate of return ($\hat{r}$), schooling inequality, and the education component show a significant positive correlation with the inequality of earnings. The correlations are lower for earnings inequality than for income inequality, but the differences are not statistically significant at a 10 per cent level. The lower correlations may be due to the large investment in schooling and postschool training on the part of nonadult males. It may also be due to the omission, because of data limitations, of young labor force males in the calculation of the variance in schooling and the rate of return.

TABLE 4-10

Correlation Matrix for Males in Fifty-one States, Income and Earnings Inequality, Regression and Overtaking Age Rates of Return

	Var(ln$E$)	Var(ln$Y$)	$\hat{r}$	r_M	Var(S)	$\hat{r}^2$ Var(S)
Var(ln$Y$)	.91					
$\hat{r}$	.71	.77				
r_m	.81	.80	.73			
Var(S)	.36	.48	.24	.19		
$\hat{r}^2$ Var(S)	.79	.88	.86	.73	.69	
r_m^2 Var(S)	.78	.82	.70	.91	.54	.88

Note: The critical values for the correlation coefficient (R), under alternative type I errors (α), are $R(\alpha = .05) = .23$, $R(.025) = .27$, $R(.01) = .32$. The critical values are based on 50 degrees of freedom. See notes to Tables 4-3 and A-4.

Source: Tables A-1 and A-4.

17. *U.S. Census of Population: 1960, Vol. 1, Characteristics of the Population*, Parts 2–52, Washington, D.C., Table 124.

The direct interstate explanatory power of the education component can be calculated by the procedure shown on p. 55. The ratio of the variance of the education component ($\hat{r}^2$ Var (S)) to the variance of the natural log of earnings is .320.[18] Thus, the direct explanatory power on the basis of earnings is approximately equal to that on the basis of income (.305, see Table 4-5).

"Overtaking Age" Estimate of Rate of Return

Up to this point the "regression estimate" has been the only measure of the rate of return discussed in this study. Yet, it has been shown that these rates of return are systematically biased downward. It would be useful to see whether our findings are specific to this measure; that is, would similar conclusions emerge if an alternative method for computing the rate of return were employed? Such an alternative is available. In his study *Schooling, Experience, and Earnings,* Jacob Mincer presents an alternative shortcut technique for calculating an unbiased rate of return from a given level of schooling, which he calls the "overtaking age" rate of return. The overtaking age rate of return from high school employed had to be computed indirectly due to the lack of appropriate data and is therefore subject to measurement error.[19] The findings shown in Table 4-10 are based on this technique.

The overtaking age estimate (r_M) is highly and significantly correlated with the regression estimate of the rate of return ($\hat{r}$) from schooling (R $(\hat{r},\ r_M) = 0.73$). The estimates of the overtaking age rate of return, however, are consistently larger; and average r_M equals 0.151, while the average $\hat{r}$ is equal to 0.102.

The overtaking age rate of return can be used to analyze the effect of schooling on the variance of the log of earnings of all males and on the variance of the log of income of adult males. The statistic r_M is significantly and highly positively correlated with both the earnings and income inequalities, and is not significantly correlated with the inequality of schooling. (See Table 4-10.)

We can calculate a measure of schooling's intrastate explanatory power of earnings and income inequality with r_M as the measure of the rate of return. Table A-4 contains the ratio of the ex-

18.

Variable	Variance
$\hat{r}^2$ Var(S)	.0047
Var$(\ln E)$	.0116

Source: Tables A-1 and A-4.

19. Estimating procedure and data sources are presented in Appendix A-2.

plained (r^2 Var (S)) to total variation for income and earnings in the case of r_M, and for earnings in the case of $\hat{r}$. The ratio for income and $\hat{r}$ (with the trivial adjustment for degrees of freedom) is in column 6 of Table A-1. For a given measure of the rate of return but different income concepts there are no significant differences. For a given income concept but different rates of return, the explanatory powers differ, being consistently high when r_M is employed. Note, however, that there is a negative correlation between schooling and the residual when r_M is the regression slope coefficient.

The statistic developed to measure the education component's direct interstate explanatory power of schooling breaks down when r_M is used as the rate of return. The education component r_M^2 Var (S) has a larger variance than Var (ln$E$) or Var (ln$Y$), thereby implying that the education component and the residual are negatively correlated.[20]

An alternative procedure for estimating schooling's interstate explanatory power with respect to income inequality is to look at the coefficient of determination when income inequality is regressed on the education component. The results of regressing Var (ln$Y$) and Var (ln$E$) on $\hat{r}^2$ Var (S) and also on r_m^2 Var (S) are shown in Table 4-4.

If the education component were uncorrelated with the residual in the regression, with the model properly specified, the slope coefficients in Table 4-4 would equal unity. The slope coefficient in row 1 of Table 4-4 is significantly greater than 1. This means that the coefficient of determination of 0.77 is an upward-biased estimate of the true explanatory power of the education component. When earnings inequality is used (row 3), the slope coefficient is also significantly greater than unity. This suggests that the explanatory power of 0.63 is upward-biased.

The slope coefficients obtained using the estimated overtaking age rate of return are significantly less than 1. This suggests the existence of a downward bias in the slope coefficient. The procedure for estimating the overtaking age rate of return for the

20.

Variable	Variance
r^2 Var(S)	.0047
r_m^2 Var(S)	.0353
Var(ln$Y$)	.0140
Var(ln$E$)	.0116

Source: Calculated from data in Tables A-1 and A-4.

states assumed that the shape of the experience log of income profile is the same for each state as it is for the nation as a whole. This introduces errors of measurement in r_M. If these errors are random, there is a downward bias in the estimate of the slope coefficient and the coefficient of determination. The coefficient of determination for the variance of the log of income is 0.68 and for the variance of the log of earnings, 0.61. Thus, the schooling model's interstate explanatory power re inequality may be conservatively estimated at 60 per cent for earnings and somewhat higher for income.

State Differences in the Rate of Return from Schooling

Our analysis of income inequality highlights the importance of the rate of return from schooling as an explanatory variable. The correlation matrices indicate that the rate of return is higher in the poorer states. This negative correlation is not due to chance, but is a consequence of differences in regional mobility across schooling levels.

If workers with high levels of schooling were perfect substitutes for those with low levels of schooling, relative wages would depend solely on the substitution coefficient. There is evidence, however, that high-level manpower (professional) is qualitatively different from nonprofessional manpower and that the two factors are not perfect substitutes.[21] Hence, a downward-sloping demand curve results for high-level relative to lower-level manpower. This negatively sloped relative demand curve plays an important role in the analysis of income distribution.

Let us view each state of the United States as a labor market with a negatively sloped demand curve for college graduates versus high school graduates. The relative wage in each state for the two schooling levels depends on relative factor supplies. Relative factor supplies, in turn, are a function of wage rates and mobility.

Wages of college graduates vary very little across the states because of their high mobility. For college graduates there is, in effect, a national labor market, in contrast to those with less schooling, where the tendency to migrate is weaker and there are

21. Using the states as units of observation and a three-factor constant elasticity of substitution production function for professionals, nonprofessionals, and physical capital, the elasticity of substitution is computed to be 2.5. See Carmel J. Ullman, "The Rise of Professional Occupations in the American Labor Force," Ph.D. dissertation, Columbia University, 1972.

significant state differences in wage rates.[22] Those with more schooling have a higher propensity to migrate for several reasons. First, schooling may increase a person's awareness of other areas and thereby reduce the psychic cost of moving to a new environment. Second, college schooling itself often entails moving to a new area, thus loosening ties to the place of origin. Third, those who acquire more schooling tend to be wealthier: since migration is an investment in human capital and discount rates vary inversely with wealth, those with more schooling also tend to be those who invest more in migration. Fourth, due to the presence of direct costs of migration which are unrelated to skill level, the rate of return from migration tends to be higher for those with more skill,[23] encouraging greater migration on their part.

A higher rate of migration out of the poorer region by skilled workers relative to unskilled workers tends to increase the ratio of skilled to unskilled workers in the wealthier region and to decrease

22. Rates of migration across the states of the United States, across the provinces of Canada, and from Canada to the United States are higher for those with higher levels of schooling. See Rashi Fein, "Educational Patterns in Southern Migration," *Southern Economic Journal*, July 1965, pp. 106-124; June O'Neill, "The Effects of Income and Education on Inter-Regional Migration," Ph.D. dissertation, Columbia University, 1970; Bruce Wilkinson, "Some Economic Aspects of Education in Canada," Ph.D. dissertation, Massachusetts Institute of Technology, 1964, pp. 84-85, 87, and 106; Thomas J. Courchene, "Interprovincial Migration and Economic Adjustment," *Canadian Journal of Economics*, November 1970, pp. 550-577.

23. Suppose the wages of a worker of skill i in region j are written as W_{ij}. The worker can either be skilled (s) or unskilled (u), and reside in the North (n) or South (s), where the South is the poorer region. Suppose there is no migration between the North and the South and wages are uniformly higher by 100 (h) per cent in the North:

$$W_{s,n} = (1 + h)\, W_{s,s};\; W_{u,n} = (1 + h)\, W_{u,s}.$$

Let us designate the fraction of the year devoted to the migration investment by k' and the direct costs of migration by C_d. Then, using the simplified formula for the rate of return, r = annual differential/cost, the rate of return to northward migration for the skilled worker is

$$r_{m,s} = \frac{W_{s,n} - W_{s,s}}{k' W_{s,s} + C_d} = \frac{W_{s,s}\,(h)}{k' W_{s,s} + C_d} = \frac{h}{k' + \dfrac{C_d}{W_{s,s}}}.$$

The rate of return to northward migration for the unskilled worker is

$$r_{m,u} = \frac{W_{u,n} - W_{u,s}}{k' W_{u,s} + C_d} = \frac{W_{s,s}\,(h)}{k' W_{u,s} + C_d} = \frac{h}{k' + \dfrac{C_d}{W_{u,s}}}.$$

it in the poorer region. Given the same negatively sloped demand curve for labor,[24] the ratio of the wages of skilled workers to those of unskilled workers is depressed in the wealthier region and raised in the poorer region. The result is a decline in the rate of return from schooling in the wealthier region and a rise in the poorer region.

Higher rates of return from schooling in the poorer states have already been noted. In the analysis of income distribution in Canada (pp. 73-77), higher rates of return are similarly found in the poorer provinces. This model suggests that among regions with little mobility there is no clear prediction as to the relation between the level of income and the rate of return from schooling. This is supported by studies of international differences in rates of return from schooling, which find no consistent pattern.[25]

The Average Level of Schooling

Thus far, the empirical analysis has been concerned primarily with the rate of return, schooling inequality, and what has been called the education component of income inequality, r^2 Var (S). In some of our analyses the level of schooling has a significant negative simple correlation with income inequality. Equations (3-13) and (3-14), however, suggest a positive partial relative between schooling level and income inequality.

Across the states, the level of schooling tends to be negatively correlated with the rate of return. In addition, schooling inequality is larger in the South. What would be the relation between the

---

If the fraction of the year devoted to migration (k') and direct costs (C_d) do not vary with the skill level, since the wages of unskilled workers in the South are lower than those of skilled workers in the South ($W_{u,s} < W_{s,s}$), $r_{m,s} > r_{m,u}$, or the rate of return from migration is higher for skilled workers. This higher average rate of return to migration induces greater migration out of the poorer region by skilled than unskilled workers.

24. The assumption of the same negatively sloped demand curve is not unrealistic. Information about productive techniques spreads rapidly in developed countries such as the United States. This implies similar coefficients for the aggregate production function in each state. Using the states as units of observation, Ullman (op. cit.) found the data to be consistent with a three-factor constant elasticity of substitution production function. Then, interstate variations in the amount of physical capital play no role in determining the relative wage of highly skilled to lower-skilled manpower.

25. See Martin Carnoy, "Rates of Return From Schooling in Latin America," *Journal of Human Resources*, Vol. 2, Summer 1967, pp. 354-374; and T. Paul Schultz, "Returns to Education in Bogota, Colombia," RAND Memorandum, Santa Monica, Rand Corporation, 1968, Table 9.

TABLE 4-11

Income Inequality and Average Level of Schooling for the United States and Canada

Region (1)	Partial Regression Coefficients AV(S) (2)	$\hat{r}$ (3)	VAR(S) (4)	AV(S) Significance (5)	Simple Correlations Coefficient (6)	Significance (7)	Change in Significance (8)
U.S., total	-0.005	6.864	0.010	NS	-0.733	0.1	+
(51 observations)	(-0.360)	(12.491)	(4.158)				
U.S., white	+0.033	6.945	0.012	1.0	-0.161	NS	+
(51 observations)	(+2.734)	(10.961)	(3.822)				
North, total	-0.002	7.476	0.009	NS	-0.177	NS	0
(34 observations)	(-0.156)	(12.947)	(3.297)				
North, white	+0.017	6.698	0.003	NS	-0.396	5.0	+
(34 observations)	(+1.053)	(8.743)	(0.896)				
South, total	-0.041	2.935	0.004	10.0	-0.799	0.1	+
(17 observations)	(-1.766)	(2.078)	(0.566)				
South, white	+0.042	5.910	0.020	10.0	+0.175	NS	+
(17 observations)	(+2.008)	(3.757)	(1.605)				
Canada	-0.005	7.205	0.041	NS	-0.543	10.0	+
(11 observations)	(-0.148)	(2.647)	(1.557)				

Note: NS = not significant at a 10.0 per cent level two-tailed test. t ratios are in parentheses. Degrees of freedom equal the number of regions minus two (for the simple correlations) or four (for the partial regression coefficients). Change in significance from simple (column 7) to partial (column 5) relation: less negative (+), more negative (-), or no clear change (0). For additional notation, see notes to Table 4-1. Significance levels from H. M. Walker and J. Lev, *Statistical Inference*, New York, Holt, Rinehart and Winston, 1953, pp. 465 and 470.

Source: See Tables A-1, A-2, and A-3.

average level of schooling and income inequality if the rate of return and schooling inequality were held constant?[26] To answer this question, a multiple linear regression of the variance of the log of income was run on the average level of schooling, the adjusted rate of return from schooling, and schooling inequality. The results appear on Table 4-11.

The sign of the slope coefficient of average schooling shows whether income inequality and average schooling are positively or negatively correlated when the rate of return and schooling inequality are held constant. For five of the six divisions of the United States (as well as for the provinces of Canada) the partial relation is either less negative or more positive than the simple relation. The only exception is the total North, where there is no perceptible change.

Thus, the negative simple correlation between income inequality and the level of schooling tends to become a nonsignificant relation when schooling inequality and the rate of return are held constant.

CANADA

Unpublished tables from the *1961 Census of Canada* permit an analysis of the effect of schooling on income for the ten provinces and the Yukon territory (referred to as the eleven provinces).[27] Within each province income is cross-classified by years of schooling for nonfarm males between the ages of twenty-five and sixty-four. A comparison of the means and standard deviations of the variables under study for the states vis-à-vis the provinces (see Table A-4) reveals only two substantial differences: both the standard deviation of years of schooling within the provinces and

26. Samarrie and Miller, as well as Aigner and Heins, analyzed interstate differences in family income inequality through multiple regression analysis. They found a negative partial relation between level of schooling and income inequality. They did not hold constant (either explicitly or through a proxy) the rate of return or schooling inequality. See A. Al Samarrie and H. P. Miller, "State Differentials in Income Concentration," *American Economic Review*, March 1967, pp. 59-72; and D. J. Aigner and A. J. Heins, "On the Determinants of Income Equality," *American Economic Review*, March 1967, pp. 175-184.

27. The census data are from a 20 per cent sample of private nonfarm households. They have been processed in the same manner as for the United States. The intraprovince regression results appear in Table A-3.

the range of the standard deviation of schooling are smaller in Canada than in the United States.[28]

Part of the apparently low education inequality within the provinces may be traced to the grouping of the data. The Canadian data contain fewer intervals than the U.S. data for schooling, particularly primary education.[29] In addition, a higher proportion of Canadian males are in the lower-level schooling category.[30] This

28. The standard deviation of schooling in Canada ranges from 3.0 years for Prince Edward Island to 3.5 years for Quebec. For the United States the range is from 3.2 years in Iowa to 4.8 years in Hawaii. (See Tables A-1 and A-3.)

29. Note the comparison below:

Number of Groups in the Education Data for the United States and Canada

	United States		Canada	
	Total and Regions (1)	States (2)	Total (3)	Provinces (4)
Zero schooling	2	1	1	1
Primary schooling		3		1
Secondary schooling	2	2	2	2
College or university	3	2	2	2
Total	7	8	5	6

Sources:

(1) *U.S. Census of Population: 1960, Subject Reports, Occupation by Earnings and Education*, Washington, D.C., Tables 1, 2, and 3.

(2) *U.S. Census of Population: 1960, Vol. 1, Characteristics of the Population*, Parts 2-52, Washington, D.C., Table 138.

(3) *Census of Canada: 1961, Population Sample. Incomes of Individuals*, Bulletin 4.1-2, Ottawa, Dominion Bureau of Statistics, Table B6.

(4) *Census of Canada: 1961*, Ottawa, Dominion Bureau of Statistics, Table A.11, unpublished.

30. Proportion of Adult Males with No More than Elementary School Education (per cent)

	Proportion of Males
Canada	
(nonfarm, excludes those with zero schooling)	
Total	44.1
Newfoundland	49.4
Prince Edward Island	50.0
Nova Scotia	51.2
New Brunswick	55.3
Quebec	51.5
Ontario	42.2
Manitoba	35.4

is particularly true of the Atlantic provinces, which are the poorer provinces. Thus, there is a greater loss of variability for schooling in Canada than in the United States, and a greater loss in the Atlantic provinces than in Canada's other provinces. Note also that the population under examination in the Canadian case—nonfarm males—is more homogeneous with respect to occupation than the population studied for the United States.

Table 4-12 presents the correlation matrix for the Canadian provinces. Income inequality, the residual variance, the rate of return from schooling, and the education component (r^2 Var (S)) are positively correlated with each other. They are negatively correlated with the levels of income and schooling. However, as in the case of the United States, the significant negative correlation of schooling level with income inequality disappears when the rate of return and schooling inequality are held constant (see Table 4-11). Schooling inequality is not correlated with income inequality, whether the rate of return is held constant or not.

The observed higher rate of return in the poorer provinces (lower levels of income and schooling) may be a consequence of the greater propensity to migrate on the part of those with higher levels of schooling.[31] Schooling inequality, however, is smaller in the poorer provinces. The rate of return appears to be more important than the inequality of schooling for explaining income inequality.

A regression of income inequality on the education component (r^2 Var (S)), with the provinces as the unit of observation, has a coefficient of determination (adjusted for degrees of freedom) equal to 68 per cent.

Saskatchewan	42.3
Alberta	34.5
British Columbia	33.1
Yukon	34.2
United States	
(males with earnings, includes those with zero education)	
Total	34.4
U.S., white	32.0
Non-South	30.8
Non-South, white	29.6
South	43.6
South, white	36.9

Sources: Census of Canada: 1961, Population Sample. Schooling by Age Groups, Bulletin 1.3-6, Ottawa, Dominion Bureau of Statistics, Tables 102 and 103; and *U.S. Census of Population: 1960, Subject Reports. Occupation by Earnings and Education*, Washington, D.C., Tables 1, 2, and 3.

31. See the citations for Canada on p. 70.

TABLE 4-12

Matrix of Correlation Coefficients, Eleven Provinces of Canada
(11 observations)

	Var(ln$Y$) (1)	$\hat{r}$ (2)	ln$Y_{0,t}$ (3)	$\overline{R}^2$ (4)	Var(S) (5)	Var(U) (6)	$\hat{r}^2$ (7)	$\hat{r}^2$ Var(S) (8)	AV(ln$Y$) (9)
$\hat{r}$	.71								
ln$Y_{0,t}$	-.67	-.95							
$\overline{R}^2$	.29	.79	-.67						
Var(S)	-.15	-.63	.73	-.35					
Var(U)	.99	-.62	-.59	.18	-.10				
$\hat{r}^2$	.70	1.00	-.96	.76	-.67	.61			
$\hat{r}^2$ Var(S)	.84	.94	-.85	.76	-.33	.76	.92		
AV(ln$Y$)	-.62	-.83	.94	-.56	.62	-.56	-.84	-.75	
AV(S)	-.54	-.58	.64	-.49	.15	-.49	-.57	-.65	.82

Note: The critical values for the correlation coefficient (R), under alternative type I errors (α), are $R(\alpha = .05) = .52$, $R(.025) = .60$, $R(.01) = .69$. The critical values are based on 9 degrees of freedom. See notes to Table 4-3.

Source: Table A-3.

The intraprovince explanatory power of schooling is low, but this may be due to the small number of schooling intervals.[32]

THE NETHERLANDS

This section of the interregional analysis focuses on the level and inequality of income among the seventy-five "geographic-economic" regions of the Netherlands. It is based on data from Schultz's study of income distribution in the Netherlands.[33]

Across regions, the level and inequality of schooling appear to be positively related to the proportion of males with higher education.[34] For the years 1950, 1955, and 1958, Schultz regressed his measure of income inequality, the Gini concentration ratio,[35] on the proportion of males with higher education between the ages of forty and sixty-four, and on several other variables. The other variables included the number and average wealth of taxpayers paying wealth taxes, measures of relative unemployment, the average number of persons on public relief, and the number of in-

---

32. The explanatory power ranges from 8.7 per cent for British Columbia to 15.9 per cent for New Brunswick (averaging 13.4 per cent). However, this is not corrected for the downward bias in the regression estimate of the rate of return.

33. T. Paul Schultz, "The Distribution of Income: Case Study of The Netherlands," Ph.D. dissertation, Massachusetts Institute of Technology, 1965, Chapter 8.

34. Ibid., pp. 339-340. The nationwide minimum school leaving age truncates the distribution of schooling, particularly in the poorer areas. This may be responsible for the positive interregional correlation between the level and inequality of schooling. See Barry R. Chiswick, "Minimum Schooling Legislation and the Cross-Sectional Distribution of Income," *Economic Journal*, September 1969, pp. 494-507.

35. Ibid., pp. 175-176. The Gini concentration ratio, a measure of relative inequality, is

$$C = \frac{1}{2mN^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |Y_i - Y_j| \, f(Y_i) \, f(Y_j)$$

where m is the arithmetic mean, Y_i $(i = 1, \ldots, n)$ is the average income of the i^{th} income class, $f(Y_i)$ is the class weight, and

$$\sum_{i=1}^{n} f(Y_i) = N.$$

A higher concentration ratio means a larger inequality of income.

TABLE 4-13

Netherlands: Cross-Sectional Regressions on Concentration Ratio (C), by Region (75 observations)

Dependent Variable (1)	Constant (2)	ΔA (3)	Ed_{47} (4)	W_t (5)	W (6)	Unp (7)	Rel (8)	$\bar{R}^2$ (9)	N (10)
C_{1958}	.3828	.096 (.073)	.013 (.003)	.062 (.015)	[a]	[a]	-.660 (.325)	.446	75
C_{1955}	.3691	.541 (.237)	.012 (.003)	.073 (.014)	-.0011 (.0007)	[a]	-.783 (.320)	.544	75
C_{1950}	.3866	[a]	.012 (.003)	.068 (.015)	-.0009[a] (.0008)	[b]	[b]	.419	75

Note: Standard errors are in parentheses.

Source: T. P. Schultz, "The Distribution of Income: Case Study of The Netherlands," Ph.D. dissertation, Massachusetts Institute of Technology, 1965, Table 8-12, p. 352, and Table 8-10, pp. 346–348.

[a]Regression coefficient not statistically significant at .05 level of significance. Where no numbers are reported they were absent in the source.

[b]Data not available.

come taxpayers in the region. Table 4-13, column 4, shows that the partial regression coefficients between income inequality and the proportion with higher education are all positive and significantly different from zero. A measure of wealth, W_t, is the most significant variable. Its significance, however, is only slightly greater than that of education. The variables used are defined below.

Definitions of Variables

- $\overline{Y}_t$ Mean income of taxpayers by region in year t (in thousands of current guilders).
- C_t Concentration ratio of incomes of income tax units by region, in year t.
- $\overline{W}$ Mean wealth of wealth-taxpayer in 1951 (in tens of thousands of current guilders).
- w The number of wealth-taxpayers in 1951 as a proportion of income-taxpayers in region in current year.
- W_t Total wealth taxed in 1951, divided by the number of income-taxpayers in current year (in tens of thousands of guilders).

Ed_{47} The percentage of males with a higher education between the ages of forty and sixty-four in May 1947, by region. Higher education is defined to include doctorates and candidates for the doctorate. The doctorate usually requires six years of higher education.

Unp The quarterly average number of totally unemployed persons as a proportion of the income-taxpayers by region, in current year.

Rel The quarterly average number of persons on public relief works as a proportion of the income-taxpayers by region in current year.

Urb The proportion of the population in a region living in a municipal center as of 1950.

A The annual average rate of change in the number of income-taxpayers in region since previous regional income distribution sample (1946, 1950, 1955, and 1958), rounded to thousands of persons.

Note that the concentration coefficient is significantly positively correlated with W_t, a measure of average wealth, as well as the level of schooling, and significantly negatively correlated with *Rel*, the average number on public relief. Thus, it seems that inequality of income is positively correlated with average level of income, since the independent variables are presumably positively correlated with level of income, except for the inverse relation for the relative number on relief.

Schultz regressed average income on the proportion with higher education and several other variables. As can be seen from Table 4-14, column (4), the coefficient of schooling is positive and very significant. A positive relation between the level of income and the inequality of schooling was found among the Canadian provinces, but not across the states of the United States.

In conclusion, it appears that in the Netherlands there are very significant positive interregional correlations among the levels and inequalities of income and schooling. No correlations can be established between these variables and the rate of return since there are not enough data available to estimate relative rates of return by region.

SUMMARY

Comparisons between the two major regions and among the various states of the United States (for all males and for white

TABLE 4-14

Netherlands: Cross-Sectional Regressions on Regional Mean Income per Taxpayer ($\overline{T}$)
(75 observations)

Dependent Variable (1)	Constant (2)	$\overline{W}$ (3)	Ed_{47} (4)	*Urb* (5)	*Rel* (6)	*Unp* (7)	*w* (8)	$\overline{R}^2$ (9)	*n* (10)
$\overline{Y}_{1958}$	3.989	0.234	0.358	-0.0043	0.577[a]	-18.70	0.035	0.712	75
		(0.057)	(0.057)	(0.0013)	(6.88)	(6.29)	(0.014)		
$\overline{Y}_{1955}$	3.027	0.229	0.256	-0.0040	-15.28	-0.403[a]	0.024	0.609	75
		(0.056)	(0.056)	(0.0013)	(5.09)	(1.733)	(0.013)		
$\overline{Y}_{1950}$	2.190	0.151	0.163	-0.0028	b	b	0.019	0.465	75
		(0.043)	(0.043)	(0.0010)			(0.009)		

Note: Standard errors are in parentheses.
Source: Schultz, "The Distribution of Income," 1965, Table 8-11, p. 349.
[a]Regression coefficient not statistically significant at .05 level of significance.
[b]Data not available.

males) reveal that the inequalities of income, of residual income, and of schooling, the rate of return, the education component (r^2 Var (S)), and the explanatory power of schooling tend to be (1) positively correlated with each other and (2) negatively correlated with the levels of income and schooling. When the rate of return and schooling inequality are held constant, the level of schooling ceases to be correlated with income inequality. The higher rate of return in the poorer states can be attributed to higher rates of migration on the part of those with more schooling. A virtually national labor market appears to exist for highly skilled workers, side by side with a tendency toward more local labor markets for those with less skill.

The relationships found among the states are only partly attributable to North-South differences. Similar, although somewhat weaker, relationships are generally found within the non-South and within the South.

When corrected for the downward bias in the regression estimate of the rate of return, schooling explains from 17 to 51 per cent of the variation in income, with a mean value of 29 per cent. The explanatory power is higher in the South than in the non-South, and higher for all males than for white males alone. It appears that the education component ($\hat{r}^2$ Var (S)) can explain half of the North-South differences in income inequality. Among all the states (on the basis of the regression estimate of the rate of return) the education component itself can explain approximately one-third of interstate differences, the intrastate residual variance (Var (U)) can explain another one-third, and their correlation explains the remaining third. When income or earnings inequality is correlated with the education component (r^2 Var (S) or $\hat{r}_M^2$ Var (S)), schooling's interstate explanatory power is at least 60 per cent. The rate of return appears to be more important than the inequality of schooling in explaining interstate differences in income inequality.

In the case of Canada, the analysis for the provinces indicates positive correlations among the inequality of income, the inequality of residual income, the rate of return, and the education component. Provincial variations in the education component explain ($\overline{R}^2$) 68 per cent of provincial variations in income inequality. The rate of return is more important than the inequality of schooling for explaining income inequality. And, the significant negative effect of the level of schooling on income inequality disappears when the rate of return and schooling inequality are held constant. These results are similar to those obtained for the United

States. The brief analysis for the regions of the Netherlands indicates that income inequality is greater the larger the inequality of schooling.

Thus, the schooling model appears to be a powerful tool for explaining differences in the inequality of personal income among the regions of a country.