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

Volume Author/Editor: Barry R. Chiswick

Volume Publisher: NBER

Volume ISBN: 0-870-14264-X

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

Publication Date: 1974

Chapter Title: The Level of Income

Chapter Author: Barry R. Chiswick

Chapter URL: http://www.nber.org/chapters/c3675

Chapter pages in book: (p. 119 - 142)

7

The Level of Income

How does the level of average income relate to the distributions of schooling, age (experience), and employment in a region? This question is explored in the following pages, in an analysis that applies the human capital earnings function (equation 6-10) to data for the United States and Canada.[1]

Regional differences in the level of income in the two countries have received considerable attention from economists, even aside from the economic development literature.[2] This analysis differs from the others by taking an explicit human capital approach to the examination of state differences in the income (or earnings) of adult males. Most U.S. regional studies in this subject area are concerned with explaining differences among all (or white) males, or the white-nonwhite income ratio. Here, in addi-

1. See Appendix A-2 for a description of the data.

2. For example, see Frank Hanna, *State Income Differentials, 1919-1954*, Duke, 1959; Victor Fuchs, *Differentials in Hourly Earnings by Region and City Size, 1959*, Occasional Paper 101, NBER, 1967; Gerald W. Scully, "The North-South Manufacturing Wage Differential, 1869-1919," *Journal of Regional Science*, Vol. 11, No. 2, 1971, pp. 235-252, and "Inter-State Wage Differentials: A Cross-Section Analysis," *American Economic Review*, 1969, pp. 759-773; A. Hurwitz and C. P. Stallings, "Interregional Differentials in Per Capita Real Income Changes," *Regional Income*, Studies in Income and Wealth, Vol. 21, NBER, 1957; and S. E. Chernick, *Interregional Disparities in Income*, Staff Study No. 14, Economic Council of Canada, August 1966. For an international study, see Anne O. Krueger, "Factor Endowments and Per Capita Income Differences Among Countries," *Economic Journal*, September 1968, pp. 641-659.

tion to these variables, interstate differences in income levels among nonwhites are also explicitly examined.

Using the human capital earnings function to study income levels has various advantages. It suggests which variables are relevant, indicates a form for entering these variables, and provides an economic interpretation of the slope coefficients and, consequently, testable hypotheses about the observed distributions.

The intraregional microequation presented in Chapter 6 is converted into an equation to explain interregional differences in the level of income—the mean log of income or the log of the geometric mean—by computing the mean value of both sides of the microrelation. The characteristics of this model are then explored, followed by the empirical analyses of interregional differences in the income level of white and nonwhite males in the United States and of nonfarm males in Canada. Finally, a summary concludes the chapter.

STATISTICAL IMPLEMENTATION OF THE MODEL

The Level of Income Equation

Let us recall that in Chapter 6 the human capital earnings function was converted to the form

$$\ln Y_i = X + r_i S_i + r_i' (A_i - S_i - 5) + \gamma (\ln WW_i). \qquad (7\text{-}1)$$

This equation relates the years of schooling (S), years of experience ($A_i - S_i - 5$), and log of weeks worked ($\ln WW_i$) of an individual to the natural log of his income, or earnings. Rearranging the experience term,

$$\ln Y_i = (X - 5r_i') + (r_i - r_i') S_i + r_i' A_i + \gamma (\ln WW_i). \qquad (7\text{-}2)$$

If it is assumed that the coefficients of S_i and A_i are random parameters but the coefficient of $\ln WW_i$ is constant across individuals, computing the mean value of equation (7-2),[3]

$$\begin{aligned}(\overline{\ln Y}) = X - 5\bar{r}' &+ (\overline{r - r'})\, \overline{S} + [R_{r-r',S} SD\,(r - r')]\; SD\,(S) \\ &+ \bar{r}'\, \overline{A} + [R_{r',A}\; SD\,(r')]\; SD\,(A) \qquad (7\text{-}3) \\ &+ \gamma\, (\overline{\ln WW}),\end{aligned}$$

3. If A_i and B_i are variables with a nonzero correlation ($R_{A,B}$),

$$(\overline{AB}) = (\overline{A})\,(\overline{B}) + \text{Cov}\,(A,B) = (\overline{A})\,(\overline{B}) + R_{A,B}\, SD\,(A)\, SD\,(B).$$

where SD designates standard deviation and R is the correlation coefficient.[4]

Equation 7-3 relates the mean value of the natural log of income in a region to the distributions of schooling, age, and employment in that region. The mean of the log of a variable is the same as the log of its geometric mean.[5] Thus, its use represents an analysis of regional differences in the geometric mean of income. Converting equation 7-3 to the form of a multiple regression and adding a random residual, U_i,

$$(\overline{\ln Y}) = b_0 + b_1 \overline{S} + b_2 SD\,(S) + b_3 \overline{A} + b_4 SD\,(A) + b_5 (\overline{\ln WW}) + U_i. \tag{7-4}$$

The economic interpretations of the coefficients are:

$$\begin{aligned} b_0 &= (X - 5\bar{r}') & b_3 &= \bar{r}' \\ b_1 &= (\overline{r_i - r_i'}) & b_4 &= R_{r',A}\, SD\,(r') \\ b_2 &= R_{r-r_i',S}\, SD\,(r - r') & b_5 &= \gamma, \end{aligned} \tag{7-5}$$

where r_i and r_i' are the ith person's rate of return from schooling (assuming $k = 1$ for the years of schooling) and the slope of his experience log income profile, respectively, and γ is the elasticity of earnings with respect to the fraction of weeks worked. If the regression slope coefficients are constant across the states, or if they are random variables independent of the explanatory variables, the computed slope coefficients are not biased.[6]

Equation (7-4) serves as the basic regression equation in the

4. An alternative approach which does not explicitly delete the squared experience term of equation (6-9) was tried but discarded because of multicollinearity. If

$$\ln Y = a_0 + a_1 T + a_2 T^2 + \cdots,$$

and the coefficients are assumed constant,

$$(\overline{\ln Y}) = a_0 + a_1 \overline{T} + a_2 (\overline{T^2}) + \cdots,$$

where $(\overline{T^2}) = (\overline{T})^2 + \mathrm{Var}\,(T)$. The simple correlation between $\overline{T}$ and $(\overline{T^2})$ is 0.9969 for data on males in the continental states. (See Appendix A-2 for source.)

5. $\ln[GM\,(Y)] = \ln \left[\left(\prod_{i=1}^{N} y_i \right)^{1/N} \right] = \frac{1}{N} \sum \ln Y = (\overline{\ln Y}).$

6. In the case of one explanatory variable, there is no bias if the slope coefficient is not correlated with either the independent variable or the square of the independent variable. If $Y_i = b_0 + b_i X_i + U_i$ where U_i is a random residual, $Y_i = b_0 + \underline{b} X_i + [(b_i - \bar{b}) X_i + U_i]$. The mean value of $b_i - \bar{b}$ is zero. Then, $\mathrm{Cov}\,[X_i, (b_i - \bar{b}) X_i + U_i]$ equals zero if b_i is not correlated with X_i or X_i^2. For a proof, see Chapter 3, footnote 16.

analysis of state and provincial differences in the level of income and earnings of males. It does not, however, adjust for the effects of race differences on incomes.

Let Y^* be the income of a white worker of a given level of schooling, age, and employment. If nonwhites receive proportionately lower (100*d* per cent lower) incomes, we could describe the income of any male of a given schooling-age-employment class by

$$Y_i = Y_i^* (1 - d)^{(NW_i)}, \quad (7\text{-}6)$$

where NW_i is a dichotomous variable taking the value of 1 for a nonwhite and zero for a white individual. Then,[7] for any class or cell,

$$\ln Y_i = \ln Y^* - (d)(NW_i), \quad (7\text{-}7)$$

and the cell mean would be

$$(\overline{\ln Y}) = (\overline{\ln Y^*}) - dp, \quad (7\text{-}8)$$

where p is the per cent nonwhite (the mean of the variable NW_i).[8] Thus, in the all-male analysis, the variable nonwhite (p) per cent of the male labor force is added to measure the partial effect of the relative presence of nonwhites on the overall level of income.

The Variables

Equation (7-4) is a theoretical equation that relates the level of income to a set of human capital and employment variables. The variables used in the actual statistical implementation of the equation and the predictions of the effects of each of the explanatory variables are discussed in the following pages.

The Dependent Variables

The dependent variables for the United States studied here are the average natural logs of (a) the 1959 income of males, age twenty-five and over, with income, and (b) the 1959 earnings of males, age fourteen and over, with earnings. Neither measure is ideal, since the model pertains to the earnings of males who have completed their schooling and the explanatory variables are computed for males between twenty-five and sixty-four years of age.[9]

7. The relation $\ln(1 - d) \approx -d$ if d is small.

8. If d is not constant across states but a variable independent of p, equation (7-8) still holds, but d is now interpreted as the mean value of the per cent difference in income.

9. See Appendix A-2 for definitions and computation of independent and dependent variables.

Average Schooling

The slope coefficient of the level of schooling, b_1, is expected to be positive and significant. The coefficient is hypothesized to be smaller for nonwhites than for whites. This hypothesis is based on the finding of a lower rate of return from schooling and a lower slope of the cross-sectional experience-earnings profile for nonwhites.[10]

Standard Deviation of Schooling

If individual differences in the rate of return from schooling (r_i) and in the slope of the cross-sectional age-log income profile (r_i') exist, the sign of the slope coefficient of the standard deviation of schooling, b_2, depends on the simple correlation of S_i with ($r_i - r_i'$). A priori arguments do not predict a sign for the correlation coefficient, but there is empirical evidence that suggests it is positive.[11]

Average Age

The model predicts a positive slope coefficient, b_3, for the level of age. Holding schooling and weeks worked constant, a higher level of age implies a greater number of years of experience or postschool training. The greater the postschool training financed by the worker, the lower are observed earnings during the early years of experience and the higher observed earnings in subsequent years. Thus, the level of income is expected to rise with age.[12]

Suppose, however, that there are no investments in postschool training ($k_0 = 0$ in equation 6-9)—age would have no effect on in-

10. For empirical evidence on the lower nonwhite rate of return from schooling around 1960, see Gary S. Becker, *Human Capital*, 2nd. ed., New York, 1974, and Finis Welch, "Black-White Differences in Returns to Schooling," *American Economic Review*, December 1973, pp. 893-907. For empirical evidence on the lower slope in the cross section of the experience-earnings profile for nonwhites, see Chapter 6 of this volume, p. 116-118.

11. Using microdata for the country as a whole, the slope coefficient of schooling from a regression of log of earnings on schooling is approximately the same when the regression is computed for all age groups or within narrow age groups. When log of earnings is regressed on years of low, median, and high levels of schooling for all age groups—but experience is not held constant—the slope coefficient increases with the level of schooling. These findings suggest a positive correlation between S_i and ($r_i - r_i'$). See Chapter 4, Table 4-1.

12. This is reinforced by the accumulation of nonhuman assets with age when income, rather than earnings, is the dependent variable.

come. Now let us introduce the depreciation or obsolescence of the human capital acquired by schooling and the pure labor component of income: here one expects a negative relation between age and income. Suppose, in addition, the quality of a given year of schooling rising over time or the extent of labor market discrimination against new entrants falling over time: in this case, too, one would expect a negative association between age and income.

A positive slope coefficient for age level is predicted for all males and white males, but a lower coefficient is predicted for nonwhites than for whites. This last point is based on assuming a lower level of investment in postschool training by nonwhites, as well as a secular rise in the quality of schooling and in job opportunities for young nonwhites relative to young whites due to a secular decrease in discrimination.[13]

The Standard Deviation of Age

Holding the level of age constant, the model predicts that a larger variance in age generates a lower level of income, since those who are older have lower slopes to their age-log income profile (i.e., A_i and r_i' are negatively correlated).[14] The effect on the level of income of the negative correlation $R_{r',A}$ depends on the dispersion of r'. The larger the dispersion of r', the lower (i.e., more negative) is the slope coefficient of the inequality of age.[15] Since it appears that nonwhites have a flatter cross-sectional age-income

13. See the discussion in Chapter 6 of racial differences in the slope of the age-income profile.

14. Theoretically as well as empirically, age-income profiles are concave; income rises with age, but at a decreasing rate.

15. The effect of a negative correlation of A_i and r_i' can be clarified by an example. Suppose we have two situations, each with two persons who, at age forty, have $\ln Y = 10$.

In situation A, one person is forty-one, and has $\ln Y = 11$ (slope = .10), while the other is thirty-nine, and has $\ln Y = 8$ (slope = .20). Mean $\ln Y$ is 9.5, and there is a negative correlation of age with the slope of the age-log income profile.

In situation B, one person is forty-one, and has $\ln Y = 14$ (slope = .40), while the other is thirty-nine, and has $\ln Y = 8$ (slope = .20). Mean $\ln Y$ is 11.0, and there is a positive correlation between age and the slope of the age-log income profile.

For the same level and dispersion in age, the level of income is lower in the situation in which the correlation of age (A_i) and the slope of the age-log income profile (r_i') is negative.

For a given negative correlation between A_i and r_i', the income level is lower the larger the dispersion in age. Suppose the slope of the age-log in-profile, the absolute dispersion of the slopes of their age-log of income profiles is likely to be smaller and their slope coefficient of the standard deviation of age higher (less negative) than for whites.

Thus, a negative slope coefficient of the standard deviation of age, b_4, is hypothesized for all males and white males, and a less negative one (i.e., with lower absolute value) is hypothesized for nonwhites than for whites.

Furthermore, because of the accumulation of nonlabor income with age, the slope of the age-log income profile does not decline with age as rapidly as the slope of the age-log earnings profile. The slope coefficient of the standard deviation of age will, therefore, be less negative (i.e., smaller absolute value) in the income than in the earnings analysis.

Average Log of Weeks Worked

The model yields the mean log of weeks worked as the employment variable to be used for explaining interregional differences in the level (average log) of income. The regression slope coefficient is $b_5 = \gamma$, where γ is the elasticity of earnings with respect to the fraction of weeks worked. Computed values of γ shall be tested against the hypotheses that the population values are 1.0 and 1.17. A coefficient of unity implies that weekly wages are not correlated with the number of weeks worked,[16] while 1.17 was the value Mincer obtained in a microdata analysis of the 1960 1/1,000 sample.[17]

A third hypothesis is that γ is lower for nonwhites than for whites. This is based on two interrelated points. First, nonwhites may obtain less general training, and presumably also less specific training, than do whites. Therefore, a major factor tending to pro-

---

come profile for ages above forty is .05, but .10 for ages below forty. At age forty, $\ln Y = 10$.

Situation A		Situation B	
Age	$\ln Y$ (approx.)	Age	$\ln Y$ (approx.)
39	9.0	38	8.0
41	10.5	42	11.0
Mean	9.75	Mean	9.50

16. See the discussion in Chapter 6.

17. See Chapter 6, footnote 10. Although Mincer's estimate of $\gamma = 1.17$ is a sample value and therefore has a standard error, the standard error is small, and will be assumed to equal zero. His data cover nonfarm white males with earnings, not enrolled in school, and fourteen years of age or older.

duce a γ greater than unity may be weaker for nonwhites. Second, nonwhite males experience greater seasonality in employment than white males.[18] Due to the forces of competition, seasonal jobs offer higher weekly wages for fewer weeks of employment per year, and hence tend to generate a γ less than unity.[19]

The coefficient γ is expected to be lower in the income regressions than in the earnings regressions. Income data contain nonlabor income, and, with the weekly wage held constant, higher nonlabor incomes tend to reduce the number of weeks worked. This factor would reduce the magnitude of γ, although the causation is from income to work.[20]

This hypothesis also implies that, given the smaller nonlabor income of nonwhite males, the difference between γ estimated from earnings and that estimated from income should be smaller for nonwhite males than for white males.

Race and Region

In the all-male analysis, a race composition variable, the nonwhite percentage of the male labor force, p, is introduced to capture the (average) effect of racial differences in income within schooling, age, and employment cells. This variable has a negative partial slope coefficient if nonwhites have lower weekly incomes, with schooling and age held constant.

A region dummy variable, *NSD*, where $NSD = 1$ in the seventeen Southern states, is also introduced to test for the persistence of regional differences in the level of income. A race-region linear

18. Indices of seasonality exist for employment of white and nonwhite males of twenty and over, but not cross-classified by occupation, schooling, or age. The factors for 1959 are:

	White	Nonwhite		White	Nonwhite
January	98.10	96.61	July	101.10	100.99
February	98.10	96.41	August	101.20	100.91
March	98.60	98.10	September	101.00	102.31
April	99.50	99.29	October	101.00	103.10
May	100.40	101.09	November	100.50	101.71
June	101.20	100.91	December	99.40	98.59

The sum of the absolute deviations from 100 is 12.7 for white males and 22.1 for nonwhite males. The monthly factors were computed from the ratio of "Original Data" to "Seasonally Adjusted Data" for employed white and nonwhite males of twenty and over in 1959. See unpublished Bureau of Labor Statistics employment data (1972), and letter to author from Hyman B. Kaitz, Bureau of Labor Statistics, January 5, 1973.

19. See Chapter 2.

20. The distributions of schooling and age can be thought of as controlling for the weekly wage.

interaction variable ($p \cdot NSD$) is included to test for regional differences in the effect of race composition on the level of income.[21]

Summary

The preceding discussion has shown how the human capital earnings function of Chapter 6 (equation 6-10) can be converted into a relation between the level of income and the distribution of schooling, age, and employment. The dependent variable, the average log of income, is a linear function of the levels of schooling, age, and the log of weeks worked, and of the standard deviations of schooling and age. The slope coefficients have economic meaning.

The dependent variables are the average log of (a) the income of males of twenty-five years and over and (b) the earnings of males of fourteen years and over. The independent variables are for males between the ages of twenty-five and sixty-four. The slope coefficient of average schooling is expected to be positive, but lower for nonwhites than for whites because of the former's lower rate of return from schooling and flatter cross-sectional age-log income profile. The slope of the standard deviation of schooling is expected to be positive on the basis of microdata analysis.

With schooling held constant, higher levels of age imply higher levels of experience, and hence higher incomes. The slope coefficient of average age is expected to be lower for nonwhites than for whites, since nonwhite males have a flatter cross-sectional age-income profile. This may have occurred because nonwhites invest less in postschool training than whites and have experienced a greater secular rise in the quality of schooling and job opportunities than have young whites. The partial slope coefficient of the standard deviation of age is hypothesized to be negative because the slope of the age-log income profile declines with higher levels of experience.

The average log of weeks worked captures the effects on the dependent variable of differences in employment level. The computed regression coefficient (γ) will be tested against alternative hypotheses about the population value. In addition, γ is hypothesized to be lower for nonwhites than whites, and lower in the income analyses than in the earnings analyses.

---

21. If $\overline{(\ln Y)} = a + b_6 p + b_7 NSD + b_8 p \cdot NSD$, then $\dfrac{\partial \overline{(\ln Y)}}{\partial p} = b_6 + b_8 (NSD)$. If b_6 and b_8 are negative, an increase in the fraction nonwhite has a more depressing effect on the level of income in the South than in the North.

Three additional variables are included in the all-male analysis: a race (nonwhite per cent of the male labor force), region (North-South dummy variable), and race-region interaction term. These variables test for the effects of the relative number of nonwhites and the applicable region on interstate differences in the mean log of income.

To sum up, the hypothesized signs for the human capital and employment variables are the following:

Variable	Symbol	Hypothesized Sign
Average schooling	$\overline{S}$	+
Standard deviation of schooling	$SD(S)$	+
Average age	$\overline{A}$	+
Standard deviation of age	$SD(A)$	−
Average log of weeks worked	$(\overline{\ln WW})$	+

EMPIRICAL APPLICATION

The model developed above to explain interregional differences in the level of income (mean log of income)—equation 7-4—is applied here to interstate data for all males, white males, and nonwhite males in the United States, and to an interprovincial analysis for nonfarm males in Canada. The data sources are discussed in Appendix A-2.

U.S. Males

The regression results for the mean log of income or earnings for all males in the fifty states and the District of Columbia—to be referred to as the fifty-one states—appear in columns (1) and (2) of Table 7-1. The model has a high explanatory power for earnings ($\overline{R}^2$ = .71), but does less well for income ($\overline{R}^2$ = .41). All of the variables have the hypothesized sign, except for the average log of weeks worked in the income analysis.

For earnings, the level of schooling and the standard deviation of age are highly significant, while the level of age and the dispersion of schooling have significant positive effects at a 5 per cent level. In the income analysis, schooling level is the only significant variable. As was hypothesized earlier in this chapter, the effect of the dispersion in age is more negative and that of the level of weeks worked more positive in the earnings than in the income analysis, although the differences are not significant.

TABLE 7-1

Regression Results, Log of Geometric Mean of Income or Earnings of All Males (fifty-one states)

Independent Variables	Dependent Variables: Level of Income (1)	Level of Earnings (2)	Level of Income[a] (3)	Level of Earnings[a] (4)
$\overline{S}$	0.2504	0.2589	0.2764	0.2562
	(4.33)	(7.19)	(5.40)	(7.38)
$SD(S)$	0.1546	0.1457	0.1503	0.2206
	(1.18)	(1.79)	(1.15)	(2.51)
$\overline{A}$	0.0429	0.0658	0.0135	0.0569
	(0.69)	(1.70)	(0.25)	(1.56)
$SD(A)$	-0.3142	-0.5349	-0.0572	-0.3804
	(-1.35)	(-3.69)	(-0.26)	(-2.61)
$\overline{(\ln WW)}$	-0.6583	0.1824	-0.7291	-0.1126
	(-0.77)	(0.34)	(-0.95)	(-0.22)
p	-0.2651	-0.1747	0.3338	0.0498
	(-0.87)	(-0.92)	(1.11)	(0.25)
NSD	—	—	0.3332	0.0357
			(3.02)	(0.48)
$p \cdot NSD$	—	—	-1.8692	-0.7758
			(-4.10)	(-2.52)
Constant	—	—	0.7466	6.9272
			(0.27)	(3.65)
N	51	51	51	51
R^2	0.4778	0.7441	0.6293	0.7896
$\overline{R}^2$	0.4066	0.7091	0.5587	0.7495

Note: Student t-ratios are in parentheses.

Source: See Appendix A-2.

[a] Regional dummy and race-region interaction terms are added. The covariances of the slope coefficients are:

	col. (3)	col. (4)
(p, NSD)	0.0116	0.0053
$(p, pNSD)$	-0.0653	-0.0298
$(NSD, pNSD)$	-0.0314	-0.0143

Regional dummy and race-region interaction terms are added in columns (3) and (4) of Table 7-1. The adjusted coefficients of determination are increased to .75 and .56 for earnings and income, respectively. In the earnings analysis, it is worth noting that average age becomes nearly significant at the 5 per cent level and that the dispersion in schooling becomes highly significant. There are no noteworthy changes in the slope coefficients for income.

Of the race-region variables in the earnings equation, only the interaction term is significant. It implies that in the North, differences in the pe.centage of nonwhites have no effect on the level of earnings. In the South, however, earnings are lower in states where a larger proportion of the male labor force is nonwhite. Phrased differently, for a given percentage nonwhite, with schooling, age, and employment held constant, earnings are lower in the South than in the North. Thus, a significant regional difference in earnings persists even after controlling for our human capital and employment variables. In the income analysis, the North-South dummy and the interaction term are significant. Assuming $p = .107$ (the average value for the fifty-one states), the model explains one-third of the regional difference in state means for income, and three-fourths for earnings.[22]

For purposes of comparison, the level of income equations, including the race and region terms, are also computed with Alaska and Hawaii deleted from the data (see Table 7-2). These two states have a large proportion of nonwhites, most of whom are nonblack. In spite of the loss of two degrees of freedom, the adjusted coefficient of determination increases five percentage points (to 0.80) for earnings and one percentage point (to 0.57) for income. In the earnings analysis, the levels of schooling, age, and employment and the dispersions in schooling and age have the expected signs and are significant (or nearly significant) at the 5 per cent level. In the income analysis, only the level of schooling is significant.

In the forty-nine-state analysis, the race, region, and race-region interaction variables are significant at the 10 per cent level. In the Northern states the level of earnings and income rises with the proportion of nonwhites, but this is not true in the South. Going from a Northern to a Southern state, assuming $p = .093$ (the mean value), decreases average earnings by 0.11 points. The human capital, employment, and race variables account for approximately 60 per cent of the North-South difference in the mean log of earnings when the analysis is restricted to the forty-nine coterminous states. Average income, however, is increased by a statistically insignificant .04 points. Thus, schooling level and

22. The observed mean values of the dependent variables (in thousand dollar units) are:

	$Av(\ln Y)$	$Av(\ln E)$
North	1.37	1.35
South	1.18	1.08
Difference	0.19	0.27

Source: Appendix A-2.

TABLE 7-2

Regression Results, Log of Geometric Mean of Income or Earnings of All Males (forty-nine states)

Independent Variables	Dependent Variables: Level of Income[a] (1)	Level of Earnings[a] (2)
$\overline{S}$	0.2529	0.2282
	(4.84)	(7.30)
$SD(S)$	0.0515	0.1442
	(0.35)	(1.65)
$\overline{A}$	-0.0056	0.0582
	(-0.09)	(1.62)
$SD(A)$	-0.0164	-0.3077
	(-0.08)	(-2.39)
$(\overline{\ln WW})$	-0.1866	1.0208
	(-0.19)	(1.75)
p	2.5466	2.7277
	(2.03)	(3.63)
NSD	0.4667	0.2116
	(3.51)	(2.66)
$p \cdot NSD$	-4.0262	-3.4702
	(-3.15)	(-4.53)
Constant	-0.4492	2.2039
	(-0.10)	(0.81)
N	49	49
R^2	0.6452	0.8343
$\overline{R}^2$	0.5742	0.8012

Note: Student t-ratios are in parentheses.
Source: See Appendix A-2.
[a]The covariants of the slope coefficients are:

	(1)	(2)
p, NSD	0.1020	0.0366
$p, pNSD$	-0.0152	-0.5440
$NSD, pNSD$	-0.1210	-0.0433

race differences account for regional differences in the measure of income level.

The slope coefficient of average schooling, to be interpreted as the rate of return from schooling minus the average slope of the age-log income profile, appears to be biased upward.

The slope of the average log of weeks is lower than expected. It is less than unity in all but one of the regressions in Tables 7-1 and 7-2. The slope coefficient is significantly below unity in only one regression (Table 7-1, col. 3). The large standard error relative to the slope coefficient of the mean log of weeks worked appears to be due to the very small interstate variation in this variable for all males.[23]

To summarize: the model can statistically explain over 70 per cent of state differences in the level (log of the geometric mean) of earnings and a somewhat smaller proportion of differences in income. Higher levels of schooling and age and a greater inequality in schooling are associated with higher levels of earnings. A greater inequality of age is associated with a lower level of earnings, presumably because of the decline in the slope of the age-log income profile as age increases. The employment variable has a generally positive slope, but is generally not significant. The low slope coefficient and large standard error appear to be caused by the very small interstate variation in this variable. The variables all have the sign hypothesized by our model.

In the income analysis the relationships are weaker, especially as to level and inequality of age. The slope of the age-log income profile declines less with age than the slope of the age-log earnings profile, which may explain the differential results for the standard deviation of age. This lack of significance of the level of age is surprising.

In the continental North, when the other variables are held constant, a larger proportion of nonwhites than of whites is associated with either no change or a small increase in the level of earnings. This may be due to the concentration of Northern nonwhites in urban industrial areas. In the South, a greater proportion of nonwhites tends to be associated with lower average earnings. This could be due to the greater "ruralness" of Southern states with a larger fraction of nonwhites, or it could reflect a greater racial difference in weekly wages in the South. About 60 per cent of the South–non-South difference in earnings for the coterminous states is due to differences in the schooling, age, and employment variables. There is no significant regional difference for income (when the other variables specified by the model are held constant and the variable per cent nonwhite is given its mean value).

---

23. The interstate means, standard deviations, and coefficients of variation of ln*WW* for the forty-nine continental states and for the thirty-nine states for which race-specific data are available are as follows:

		Mean	SD	CV
49 States:	All Males	3.8022	0.0321	0.0084
	White Males	3.8114	0.0301	0.0079
39 States:	All Males	3.7997	0.0346	0.0091
	White Males	3.8113	0.0327	0.0086
	Nonwhite Males	3.6780	0.0807	0.0219

Source: Appendix A-2.

TABLE 7-3

Regression Results, Log of Geometric Mean of Income or Earnings of White Males

Independent Variables	Dependent Variables			
	Level of Income[a] (1)	Level of Earnings[a] (2)	Level of Income[b] (3)	Level of Earnings[b] (4)
$\overline{S}$	0.2135	0.2506	0.2043	0.2656
	(4.29)	(8.26)	(3.19)	(7.38)
$SD(S)$	0.1694	0.1664	0.1432	0.1167
	(1.73)	(2.78)	(1.07)	(1.56)
$\overline{A}$	0.0762	0.0996	0.0774	0.0863
	(1.22)	(2.62)	(0.94)	(1.86)
$SD(A)$	-0.2840	-0.4657	-0.2726	-0.4928
	(-1.39)	(-3.73)	(-1.06)	(-3.41)
$(\overline{\ln WW})$	-1.1288	0.3561	-1.1913	0.0088
	(-1.17)	(0.60)	(-0.96)	(0.01)
Constant	2.6594	4.4588	2.9180	6.6790
N	49	49	39	39
R^2	0.3023	0.6750	0.2542	0.6916
$\overline{R}^2$	0.2212	0.6372	0.1413	0.6449

Note: Student t-ratios are in parentheses.
Source: See Appendix A-2.
[a] 49 observations.
[b] 39 observations.

U.S. White Males

This section presents the empirical analysis for white males, defined as white males in the thirty-nine continental states for which separate race-specific data exist in the 1960 Census of Population, and all males in the remaining ten continental states.[24] Table 7-3 contains the regression results for the level of earnings and income for the forty-nine and thirty-nine–state samples. The model's explanatory power is higher for earnings than income.[25] All of the variables have the expected sign, except for the insignificant "weeks worked" variable in the income analysis.

24. The ten states, and the per cent of males between twenty-five and sixty-four in 1960 who were nonwhite, are Idaho (1.5), Maine (0.7), Montana (3.6), Nevada (7.6), New Hampshire (0.5), North Dakota (2.0), Rhode Island (2.5), Utah (2.0), Vermont (0.2), and Wyoming (2.2).

25. Adjusted R^2 for white males:

	Income	Earnings
49 states:	.22	.63
39 states:	.14	.64

Source: Table 7-3.

For the earnings data, the four schooling and age variables are all significant at a 5 per cent level, except for the standard deviation of schooling in the smaller sample. On the income side, the only significant variables are the level of schooling and, for the forty-nine-state sample, the standard deviation of schooling.

The level of weeks of employment again does not differ from zero. Moreover, it is significantly lower than 1.0 and 1.17 for income, but not for earnings. Indeed, the coefficient is negative in the income data. The lower partial slope coefficient for the income analysis may be demonstrating the adverse effect of nonlabor incomes on the labor supply to the market when the weekly wage is held constant. The small interstate variation in the weeks worked may also be responsible for the large standard error in the slope coefficient.

As in the all-male regression analysis, the slope coefficient of the average level of schooling appears to be biased upward. The slope of the standard deviation of age is lower in absolute value in the income than in the earnings data; this is consistent with the different changes in the slope of the age-log income and age-log earnings profile as age increases.

A comparison of the income and earnings equations reveals a lower absolute value for the slope of the standard deviation of age, and a less positive slope for the level of weeks worked. (Both phenomena are explained above.) The slope coefficient of schooling is also lower in the income equation, but this is not surprising since the slope is interpreted as $\bar{r} - \bar{r}'$. The slope of the age-income profile ($\bar{r}'$ for income) is expected to be steeper than the slope of the age-earnings profile ($\bar{r}'$ for earnings). A lower slope appears in the income analysis than in the earnings analysis, too, for the average age variable. Note, however, that the income data are for males of twenty-five years of age and over while the earnings data are for males of fourteen years of age and older; a unit increase in average age has a smaller (average) effect on income (or earnings) in a population of twenty-five and over than in one of fourteen and over because the income (or earnings) profile is much steeper at young ages.

Comparing the slope coefficients and t-ratios of the income and earnings regressions, we find generally larger standard errors of slope coefficients for income—a major reason for the poorer performance of the income equation. This suggests that the inclusion of nonlabor income may be increasing "purely random" errors in the dependent variable. Such random errors do not bias regression slope coefficients, but they do enlarge the standard error of the slope coefficients and thereby decrease t-ratios. If this explanation of the poorer performance of the income equation for white males is correct, it should perform better in a population with less nonlabor income in this group. A comparison of the income equation for white males with that for nonwhite males below will provide a test for this hypothesis.

U.S. Nonwhite Males

Here the human capital earnings function is used to analyze interstate differences in the log of the geometric mean of income and earnings of nonwhite males from the thirty-nine continental states for which separate race data were made available in the 1960 Census of Population. In the first part of this chapter certain implications regarding racial differences in the regression slope coefficients were developed on the basis of the following set of assumptions: Nonwhites, compared to whites, (a) receive a lower rate of return from schooling, (b) have a flatter cross-sectional experience-earnings profile,[26] and (c) face greater seasonality of employment. Thus, the regression coefficients are expected to be positive but algebraically lower for nonwhites than for whites for the mean levels of schooling, age, and the log

TABLE 7-4

Regression Results, Log of Geometric Mean of Income or Earnings of Nonwhite Males

Independent Variables	Dependent Variables: Level of Income (1)	Dependent Variables: Level of Earnings (2)
$\overline{S}$	0.1822	0.1968
	(8.41)	(8.60)
$SD(S)$	0.1238	0.1745
	(1.30)	(1.73)
$\overline{A}$	-0.0295	-0.0200
	(-1.13)	(-0.73)
$SD(A)$	-0.2358	-0.1993
	(-1.84)	(-1.47)
$(\overline{\ln WW})$	0.8404	0.6373
	(2.59)	(1.86)
Constant	0.4670	6.1210
N	39	39
R^2	0.8236	0.8062
$\overline{R}^2$	0.7967	0.7769

Note: Student t-ratios in parentheses.
Source: See Appendix A-2.

26. For a discussion of this point, see Chapter 6, pp. 116–118.

of weeks worked, and the slope of the standard deviation of age is expected to be less negative.

Table 7-4 presents the actual regression results for nonwhite males. The model has a high explanatory power both for income ($\overline{R}^2 = 0.80$) and earnings ($\overline{R}^2 = 0.78$). All of the variables, except the level of age, have the sign predicted by the human capital model (see p. 128). The effect of schooling level is highly significant in both equations, and the rate of return implied by the slope coefficient of schooling is high and appears to be an upward-biased estimate, as in the preceding regressions. The nonwhite coefficient is lower, but not significantly so, than the value for whites. The slope coefficient of the standard deviation of schooling differs from zero (at a 5 per cent level) only for the earnings data.

When it comes to the age factor, the slope coefficient of the level of age is not significantly different from zero in either the earnings or the income equation. The magnitude of the coefficient is lower for nonwhites than for whites. The slope of the standard deviation of age is lower in absolute value for nonwhites than whites too, but, whereas *SD* (*A*) proved significant for all males and white males only in the earnings data, it is significant for nonwhites only in the income data. This difference may be a consequence of the slower rise with age of nonlabor incomes for nonwhites.

As to weeks worked, the slope coefficient of the mean log of weeks worked is positive and significant for nonwhites in both regressions. Although the values of the coefficients are less than unity, they are not significantly lower than hypothesized population values of 1.0 or 1.17.[27] The statistical significance of this variable for nonwhite males—but not for all males or all white males—may be due to its much larger interstate variability.[28]

To what extent is the lower income and earnings level of nonwhites compared to whites due to different values in the independent variables? The answer is provided in Table 7-5, on the basis of the thirty-nine states with separate race data. If nonwhites retained the value of their slope coefficients but had the same values whites had for the explanatory variables, the level of income

27. If the null hypothesis is $H_0: \gamma = 1.17$, for the income data $t = \frac{.8404 - 1.17}{.3245} = 1.02$, and for the earnings data $t = \frac{.6373 - 1.17}{.3426} = 1.55$. For a two-tailed test, 10 per cent level of significance and 33 degrees of freedom, the critical value is $t = 1.70$.

28. See footnote 23 on p. 132.

TABLE 7-5

Mean Values of the Log of Geometric Means for Income and Earnings, Observed versus Predicted, Thirty-nine States (in thousands of dollars)

	Income	Earnings
Observed values[a]		
Whites	1.368	1.322
Nonwhites	0.754	0.795
Predicted values[b]		
With nonwhites' slope coefficients, whites' mean values of independent variables	1.241	1.286
With whites' slope coefficients, nonwhites' mean values of independent variables	1.014	1.655

Note: Mean values of independent variables are:

	Whites	Nonwhites
Av(S)	10.37	8.00
SD(S)	3.60	3.82
Av(A)	42.91	42.05
SD(A)	10.87	10.86
Av(ln *WW*)	3.81	3.68

[a]See Appendix A-2.
[b]For slope coefficients: see Table 7-3, cols. 3 and 4, and Table 7-4, cols. 1 and 2. For independent variables: see Appendix A-2.

and earnings of nonwhites would increase substantially. The change would narrow the white-nonwhite gap in the average log of income (Av ($\ln Y$)) by 80 per cent, and in the average log of earnings (Av ($\ln E$)), by 92 per cent. This is mainly due to the higher levels of schooling and weeks worked among whites.

Suppose, however, that nonwhites retained the value of their independent variables but had the white values for the regression slope coefficients. The racial gap would be reduced by almost 50 per cent in observed income and would change its sign in earnings. The finding for earnings stems from the effects of the lower slopes of the standard deviation of age and the mean log of weeks worked for whites. However, the slope of the mean log of weeks worked is highly unstable for whites. Indeed, the regression equation appears to be generally less reliable for whites than for nonwhites.

Thus, it appears that the lower income and earnings level (log of the geometric mean) of nonwhites compared to whites is largely due to racial differences in the explanatory variables (particularly the lower levels of schooling and weeks worked) rather than to the effect of these variables on income or earnings.

TABLE 7-6

Regression Analysis, Log of Geometric Mean of Income in Canada

Independent Variables	Dependent Variable: $Av(\ln Y)$	
	Ten Provinces	Ten Provinces and Yukon Territory
$Av(S)$	0.2345	0.2189
	(10.40)	(11.46)
$SD(S)$	0.7276	0.7109
	(4.95)	(4.65)
$Av(A)$	0.0653	0.0620
	(1.38)	(1.26)
$SD(A)$	-0.5927	-0.3493
	(-2.47)	(-2.42)
Constant	0.1413	-2.0022
df	5	6
R	0.9882	0.9874
$\bar{R}^2$	0.9576	0.9582
R^2	0.9765	0.9750

Note: t-ratios are in parentheses.
Source: Appendix A-2.

Canada

Information provided by the 1961 Census of Canada permits us to study the effects of schooling and age on the distribution of income in Canada. The analysis covers nonfarm males of twenty-five to sixty-four years of age in the ten provinces and the Yukon Territory—eleven units of observation referred to as eleven provinces.[29] The very small sample size reduces the meaningfulness of tests of significance, unless regression residuals are assumed to be normally distributed.

Table 7-6 illustrates the regression analysis of provincial differences in the log of the geometric mean of income for a ten- and eleven-province sample. The hypothesized slope coefficients for the four explanatory variables that appear in the data have high t-ratios, except for the level of age, and the slope coefficient of schooling appears to be biased upward. Both the insignificant effect of average age and the significant effect of the variance in age may be explained by the fact that age (experience) exerts a strong influence on earnings at young and old ages but a weak effect in between.

The model's explanatory power is very high, 96 per cent after

29. Although data on weeks worked do exist, they are not useful for this study because the intervals are too broad. See Appendix A-2.

adjusting for degrees of freedom. Altogether, the Canadian results are very similar to those obtained in the U.S. analysis of earnings.

SUMMARY

Chapter 7 is an analysis of interregional differences in the level of income (and earnings) in the United States and Canada within the theoretical framework of a human capital model of income generation. The mean value of both sides of the human capital earnings equation of Chapter 6, relating the log of an individual's annual income to his level of schooling, age, and log of weeks worked in the year, is computed, and the model is analyzed theoretically.

The dependent variable is the average value of the natural log of income—the same as the log of the geometric mean of income. The five explanatory variables are the means and standard deviations of years of schooling and of age, and the mean log of weeks worked. The slope coefficients of the explanatory variables have economic interpretations.

The slope coefficient of the level of schooling ($b_1 = \bar{r}_i - \bar{r}'_i$) is the mean value across individuals of the difference between the average rate of return from schooling ($\bar{r}_i$) and the slope of the experience log of income profile ($\bar{r}'_i$). The coefficient of schooling level is hypothesized to have a positive sign, and to be lower for nonwhites than whites in the United States. The racial difference is based on the assumption that nonwhites have proportionately lower rates of return from schooling and flatter cross-sectional experience-earnings profiles. On the basis of a microdata analysis, the standard deviation of schooling is expected to have a positive slope coefficient.

The slope coefficient of the level of age, when schooling level is held constant, is the average across individuals of the slope of the age-log of income profile ($\bar{r}'_i$). A positive slope is hypothesized, since an increase in age implies more labor market experience and less current investment in training, and thus higher observed earnings. A lower coefficient is hypothesized for nonwhites than whites because the former have flatter experience-earnings profiles.

The predicted sign of the effect of the model's fourth explanatory variable, the standard deviation of age, is the same as the sign of the correlation of age with the slope of the age-log of income profile. Since the slope of the profile declines with age, the model hypothesizes a negative regression coefficient for the standard deviation of age. The coefficient is less negative the flatter the age log of income profile. Hence, a less negative slope of the standard deviation of age is predicted for nonwhites than whites.

The model's fifth variable for explaining interstate differences in the level of income is the average log of weeks worked. The slope coefficient of this variable is the elasticity of income with respect to the fraction of weeks worked (γ), and is hypothesized to be positive (those who work more weeks per year have higher annual income), but lower for nonwhites than whites. The hypothesized racial differences are based on two related points—a lower level of investment in job-specific training and a greater relative concentration in seasonally sensitive jobs on the part of nonwhites.

In the empirical analysis, two dependent variables are employed: the average log of (a) the income of males twenty-five years of age and over, and of (b) the earnings of males fourteen years of age and over. The former measure contains nonlabor market income, while the latter includes young males. Since the model was developed to study the earnings of males who have completed their schooling, it is clear that neither measure of the dependent variable is a perfect fit to the theoretical concept of income or the definition of the independent variables. The data are from the 1960 U.S. Census of Population and the 1961 Census of Canada, and the states and provinces are the respective units of observation.

In the U.S. analysis for all males, the human capital and employment model can explain ($\bar{R}^2$) over 70 per cent of interstate differences in the level (log of the geometric mean) of earnings and a somewhat smaller proportion (40 per cent) of differences in income. Higher levels of schooling and age, a greater inequality of schooling, and a smaller inequality of age are associated with higher levels of earnings. These variables have the predicted signs (p. 128). The employment variable, the log of the geometric mean of weeks worked, has a generally positive slope in the earnings analysis, and is significant in the coterminous forty-nine states, but not in the fifty-one states. The employment slope does not differ from unity or 1.17, a value for the elasticity of earnings with respect to weeks worked that was found in microdata for white males. The relationships are weaker for income—the level of schooling is the only significant variable.

About 60 per cent of the South-non-South difference in earnings is explained by differences in the schooling, age, employment, and race composition (per cent nonwhite) variables.

Separate regressions are computed for data restricted to white males. The model has a high explanatory power for earnings ($\overline{R}^2 \approx .64$), but makes a poor showing for income ($\overline{R}^2 = .14$ to .22). All of the variables have the signs hypothesized by the human capital model, except the mean log of weeks worked in the income analysis. The schooling and age variables are significant in the earnings equation, but only schooling level is significant in the income equation. The conclusions regarding white males are similar to those for all males.

In the analysis of data for nonwhite males, the model has a high explanatory power for both earnings ($\overline{R}^2 = .78$) and income ($\overline{R}^2 = .80$). On the assumption that nonwhite males have little nonlabor income, the income data are a better fit to the model's theoretical income concept and the independent variables than are the earnings data. Unlike the results for all males and white males, the adjusted coefficient of determination for the nonwhite income equation is very high, and larger than the explanatory power for earnings. This suggests that the model's relatively poor performance for the income of all males and white males may be due to the inclusion of nonlabor market income.

Levels of schooling and the log of weeks worked have significant positive slopes for nonwhite income and earnings. The racial difference in the significance of the weeks-worked coefficient appears to be due to interstate variations in this variable. The coefficient of variation of the average log of weeks worked is more than twice as large for nonwhite males as for white males or all males. The standard deviation of schooling has a positive partial effect.

As predicted, the absolute value of the slope coefficients of the levels of schooling and age and the standard deviation of age are lower (and less significant) for nonwhites than for whites. Indeed, the two age variables are not significant for nonwhites, and this is consistent with a fairly flat nonwhite age-log of income profile.

Tests are performed to discover whether the lower level of income and earnings of nonwhite males is due to different values of the independent variables (schooling, age, and weeks worked), or to the effects of these variables on income and earnings (i.e., to different values of the regression coefficients). The racial difference in the mean log of income and earnings is found to be largely (80 to 90 per cent) due to racial differences in the explanatory variables, in particular, the lower levels of schooling and weeks worked per year for nonwhites.

Thus, our human capital model of interstate differences in the level of income and earnings of males appears to do quite well ($\bar{R}^2$ = .60 to .80) for the earnings of all males, white males and nonwhite males, as well as for the income of nonwhite males. The model's poor performance in respect to the income of all males and white males may be due to the inclusion of nonlabor income. Empirically, in the case of all males and white males, earnings are an increasing function of the levels of schooling and age and the standard deviation of schooling, and a decreasing function of the standard deviation of age. In the case of nonwhites, the levels of income and earnings are rising functions of the levels of schooling and the log of weeks worked, and, for earnings, also of the standard deviation of schooling. 80 to 90 per cent of the racial difference in the mean logs of income and earnings can be explained by the racial difference in the levels of schooling and weeks worked.

The interprovincial analysis for Canada is limited by the small sample size, but the results are interesting. The four explanatory variables, the levels and standard deviations of schooling and age, have the hypothesized sign, and except for the level of age, they have very high *t*-ratios. The adjusted coefficient of determination is 96 per cent. The Canadian results are similar to those obtained in the analysis of earnings in the United States.