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: Appendix A: Data Sets

Chapter Author: Barry R. Chiswick

Chapter URL: http://www.nber.org/chapters/c3677

Chapter pages in book: (p. 179 - 193)

# Appendix A: Data Sets

Appendix A has three parts. The first, A-1, provides a description of the data used in the analysis for the United States and Canada in Part B of the study, including a presentation of the data for the various states and provinces. Appendix A-2 describes the data used in Part C for the United States and Canada, and explains the procedures for computing the regression estimate and the overtaking age estimate of the rate of return from schooling. Finally, Appendix A-3 discusses the sample data for Mexican wage earners.

## APPENDIX A-1: DATA FOR PART B: UNITED STATES AND CANADA

This appendix presents (a) regression results for all states of the United States, for whites in seventeen states, and for the provinces of Canada; (b) figures on earnings inequality, overtaking age rate of return, and schooling's "explanatory" power for the fifty-one states in the United States; and (c) means, standard deviations, and coefficients of variation for the states and provinces.

The following symbols are used:

| Variable | Symbol | Description |
|---|---|---|
| 1. Standard deviation of log of income or earnings | SD ($\ln Y$) or SD ($\ln E$) | |
| 2. Standard deviation of schooling | SD ($S$) | |
| 3. Regression estimate of rate of return from schooling (or adjusted rate of return) | $\hat{r}$ | Slope coefficient from regression of $\ln Y_i$ on $S_i$ within each region (see Chapter 3, or Appendix A-2). |
| 4. Zero schooling level of income | $\ln Y_0$ | Intercept from regression of $\ln Y_i$ on $S_i$ within each region. |
| 5. Residual income variance | Var ($U$) | Residual variance from regression of $\ln Y_i$ on $S_i$ within each region. |
| 6. Intrastate explanatory power | $\overline{R}^2$ | Adjusted coefficient of determination from regression of $\ln Y_i$ on $S_i$ within each region. |
| 7. Overtaking age rate of return | $r_m$ | See Appendix A-2. |

8. Regression equation: $\ln Y_i = (\ln Y_0) + (\hat{r})\, S_i + U_i$.

TABLE A-1

Results from Regressing the Natural Log of Income in 1959 on Schooling for Males of Twenty-five and over in the United States

| State[a] | SD(ln$Y$) (1) | SD($S$) (2) | ln$Y_{0,1}$ (3) | $\hat{r}_1$ (4) | Var($U$)$_1$ (5) | $\bar{R}_1^2$ (6) |
|---|---|---|---|---|---|---|
| *Alabama* | 1.00 | 4.21 | -.22 | .13 | 0.72 | .28 |
| | | | (.24) | (.02) | | |
| Alaska | 0.93 | 3.77 | .37 | .11 | 0.70 | .19 |
| | | | (.31) | (.03) | | |
| Arizona | 0.91 | 4.12 | .29 | .10 | 0.66 | .26 |
| | | | (.26) | (.02) | | |
| *Arkansas* | 0.99 | 4.03 | -.38 | .13 | 0.73 | .14 |
| | | | (.25) | (.03) | | |
| California | 0.84 | 3.76 | .58 | .09 | 0.61 | .15 |
| | | | (.29) | (.02) | | |
| Colorado | 0.82 | 3.67 | .41 | .09 | 0.57 | .16 |
| | | | (.29) | (.02) | | |
| Connecticut | 0.79 | 3.80 | .64 | .09 | 0.52 | .26 |
| | | | (.26) | (.02) | | |
| *Delaware* | 0.89 | 3.98 | .24 | .12 | 0.59 | .16 |
| | | | (.26) | (.02) | | |
| *D.C.* | 0.91 | 4.29 | .41 | .09 | 0.70 | .16 |
| | | | (.27) | (.02) | | |
| *Florida* | 0.91 | 4.00 | .18 | .10 | 0.68 | .18 |
| | | | (.27) | (.02) | | |
| *Georgia* | 0.97 | 4.36 | -.13 | .12 | 0.66 | .30 |
| | | | (.22) | (.02) | | |
| Hawaii | 0.79 | 4.82 | .77 | .07 | 0.51 | .17 |
| | | | (.19) | (.02) | | |
| Idaho | 0.83 | 3.30 | .23 | .10 | 0.59 | .15 |
| | | | (.32) | (.03) | | |
| Illinois | 0.85 | 3.60 | .51 | .09 | 0.61 | .14 |
| | | | (.29) | (.03) | | |
| Indiana | 0.84 | 3.39 | .37 | .10 | 0.60 | .15 |
| | | | (.30) | (.03) | | |
| Iowa | 0.91 | 3.21 | .03 | .11 | 0.71 | .14 |
| | | | (.35) | (.03) | | |
| Kansas | 0.89 | 3.38 | .13 | .11 | 0.66 | .16 |
| | | | (.33) | (.03) | | |
| *Kentucky* | 0.99 | 3.90 | -.14 | .12 | 0.76 | .23 |
| | | | (.26) | (.03) | | |
| *Louisiana* | 0.98 | 4.64 | .08 | .11 | 0.68 | .28 |
| | | | (.21) | (.02) | | |

(continued)

TABLE A-1 (continued)

| State[a] | SD(ln $Y$) (1) | SD($S$) (2) | ln $Y_{0,1}$ (3) | $\hat{r}_1$ (4) | Var($U$)$_1$ (5) | $\bar{R}_1^2$ (6) |
|---|---|---|---|---|---|---|
| Maine | 0.83 | 3.31 | .16 | .10 | 0.60 | .14 |
| | | | (.30) | (.03) | | |
| *Maryland* | 0.86 | 4.01 | .43 | .10 | 0.58 | .21 |
| | | | (.25) | (.02) | | |
| Massachusetts | 0.80 | 3.74 | .47 | .09 | 0.53 | .17 |
| | | | (.26) | (.02) | | |
| Michigan | 0.85 | 3.52 | .50 | .09 | 0.62 | .14 |
| | | | (.29) | (.03) | | |
| Minnesota | 0.91 | 3.43 | .05 | .12 | 0.67 | .18 |
| | | | (.32) | (.03) | | |
| *Mississippi* | 0.99 | 4.13 | −.61 | .14 | 0.67 | .32 |
| | | | (.22) | (.02) | | |
| Missouri | 0.95 | 3.62 | .05 | .11 | 0.75 | .18 |
| | | | (.30) | (.03) | | |
| Montana | 0.83 | 3.37 | .31 | .09 | 0.60 | .13 |
| | | | (.31) | (.03) | | |
| Nebraska | 0.87 | 3.26 | .03 | .11 | 0.64 | .16 |
| | | | (.33) | (.03) | | |
| Nevada | 0.82 | 3.37 | .58 | .09 | 0.60 | .11 |
| | | | (.32) | (.03) | | |
| New Hampshire | 0.80 | 3.41 | .39 | .09 | 0.55 | .13 |
| | | | (.29) | (.03) | | |
| New Jersey | 0.80 | 3.81 | .63 | .09 | 0.53 | .17 |
| | | | (.25) | (.02) | | |
| New Mexico | 0.91 | 4.29 | .22 | .11 | 0.61 | .26 |
| | | | (.24) | (.02) | | |
| New York | 0.84 | 3.90 | .54 | .09 | 0.59 | .16 |
| | | | (.26) | (.02) | | |
| *North Carolina* | 0.96 | 4.22 | −.10 | .11 | 0.69 | .25 |
| | | | (.23) | (.02) | | |
| North Dakota | 0.90 | 3.39 | .08 | .11 | 0.69 | .15 |
| | | | (.31) | (.03) | | |
| Ohio | 0.84 | 3.52 | .44 | .10 | 0.59 | .15 |
| | | | (.29) | (.03) | | |

(continued)

TABLE A-1 (concluded)

| State[a] | SD(ln$Y$) (1) | SD($S$) (2) | ln$Y_{0,1}$ (3) | $\hat{r}_1$ (4) | Var($U$)$_1$ (5) | $\bar{R}_1^2$ (6) |
|---|---|---|---|---|---|---|
| *Oklahoma* | 0.96 | 3.91 | -.11 | .12 | 0.71 | .24 |
| | | | (.28) | (.03) | | |
| Oregon | 0.85 | 3.38 | .35 | .09 | 0.63 | .13 |
| | | | (.32) | (.03) | | |
| Pennsylvania | 0.82 | 3.62 | .41 | .09 | 0.56 | .16 |
| | | | (.27) | (.02) | | |
| Rhode Island | 0.79 | 3.67 | .45 | .09 | 0.53 | .15 |
| | | | (.25) | (.02) | | |
| *South Carolina* | 0.96 | 4.43 | -.12 | .12 | 0.67 | .28 |
| | | | (.21) | (.02) | | |
| South Dakota | 0.93 | 3.22 | -.12 | .11 | 0.73 | .15 |
| | | | (.34) | (.03) | | |
| *Tennessee* | 1.00 | 4.07 | -.20 | .13 | 0.74 | .26 |
| | | | (.25) | (.03) | | |
| *Texas* | 0.96 | 4.32 | .11 | .11 | 0.70 | .24 |
| | | | (.25) | (.02) | | |
| Utah | 0.78 | 3.33 | .48 | .09 | 0.54 | .12 |
| | | | (.32) | (.03) | | |
| Vermont | 0.86 | 3.35 | .11 | .10 | 0.63 | .14 |
| | | | (.31) | (.03) | | |
| *Virginia* | 0.94 | 4.37 | .12 | .11 | 0.67 | .24 |
| | | | (.24) | (.02) | | |
| Washington | 0.84 | 3.43 | .33 | .10 | 0.59 | .16 |
| | | | (.31) | (.03) | | |
| *West Virginia* | 0.95 | 3.79 | .10 | .11 | 0.75 | .17 |
| | | | (.27) | (.03) | | |
| Wisconsin | 0.85 | 3.45 | .34 | .10 | 0.62 | .15 |
| | | | (.30) | (.03) | | |
| Wyoming | 0.82 | 3.38 | .45 | .09 | 0.58 | .13 |
| | | | (.31) | (.03) | | |

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

[a]Southern states are italic. The data are in 9 income and 8 schooling intervals.

**TABLE A-2**

Results from Regressing the Natural Log of Income in 1959 on Schooling for White Males of Twenty-five and over in Seventeen States

| State[a] | SD(ln$Y$) (1) | SD($S$) (2) | ln$Y_{0,1}$ (3) | $\hat{r}_1$ (4) | Var($U$)$_1$ (5) | $\bar{R}_1^2$ (6) |
|---|---|---|---|---|---|---|
| *Alabama* | 0.97 | 3.91 | -.15 | .13 | 0.70 | .26 |
| | | | (.27) | (.03) | | |
| Alaska | 0.80 | 3.00 | .94 | .07 | 0.61 | .06 |
| | | | (.38) | (.03) | | |
| *Arkansas* | 0.97 | 3.85 | -.29 | .12 | 0.72 | .23 |
| | | | (.27) | (.03) | | |
| *Delaware* | 0.85 | 3.81 | .35 | .11 | 0.55 | .23 |
| | | | (.27) | (.02) | | |
| *D.C.* | 1.02 | 4.17 | .46 | .10 | 0.89 | .14 |
| | | | (.36) | (.03) | | |
| *Florida* | 0.89 | 3.71 | .28 | .09 | 0.68 | .14 |
| | | | (.30) | (.03) | | |
| *Georgia* | 0.94 | 4.06 | -.00 | .12 | 0.66 | .25 |
| | | | (.26) | (.02) | | |
| Hawaii | 0.83 | 3.72 | .51 | .10 | 0.57 | .18 |
| | | | (.31) | (.02) | | |
| *Louisiana* | 0.91 | 4.42 | .32 | .10 | 0.65 | .22 |
| | | | (.23) | (.02) | | |
| *Maryland* | 0.83 | 3.95 | .56 | .09 | 0.55 | .19 |
| | | | (.26) | (.02) | | |
| *Mississippi* | 0.99 | 3.81 | -.40 | .13 | 0.74 | .25 |
| | | | (.30) | (.03) | | |
| New York | 0.84 | 3.94 | .57 | .09 | 0.59 | .16 |
| | | | (.26) | (.02) | | |
| *North Carolina* | 0.92 | 4.10 | .05 | .11 | 0.66 | .22 |
| | | | (.24) | (.02) | | |
| *South Carolina* | 0.88 | 4.13 | .17 | .10 | 0.62 | .21 |
| | | | (.24) | (.02) | | |
| *Tennessee* | 0.99 | 4.03 | -.17 | .13 | 0.74 | .25 |
| | | | (.26) | (.03) | | |
| *Texas* | 0.94 | 4.30 | .20 | .11 | 0.68 | .23 |
| | | | (.25) | (.02) | | |
| *Virginia* | 0.93 | 4.25 | .19 | .11 | 0.65 | .25 |
| | | | (.25) | (.02) | | |

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

[a]Southern states are italic. The data are in 9 income and 8 schooling intervals.

TABLE A-3

Results from Regressing the Natural Log of Income in 1960 on Schooling for Nonfarm Males, Twenty-five to Sixty-four, in the Provinces of Canada

| Province | SD(ln$Y$) (1) | SD($S$) (2) | AV(ln$Y$) (3) | AV($S$) (4) | ln$Y_{0,1}$ (5) | $\hat{r}_1$ (6) |
|---|---|---|---|---|---|---|
| Newfoundland | 0.87 | 3.27 | 0.84 | 6.80 | .17 | .10 |
| | | | | | (.11) | (.01) |
| Price Edward Island | 0.83 | 2.98 | 0.95 | 8.20 | .05 | .11 |
| | | | | | (.15) | (.02) |
| Nova Scotia | 0.80 | 3.00 | 1.07 | 8.39 | .21 | .10 |
| | | | | | (.13) | (.01) |
| New Brunswick | 0.79 | 3.16 | 1.03 | 7.70 | .25 | .10 |
| | | | | | (.11) | (.01) |
| Quebec | 0.77 | 3.49 | 1.25 | 7.81 | .60 | .08 |
| | | | | | (.10) | (.01) |
| Ontario | 0.72 | 3.40 | 1.40 | 9.13 | .69 | .08 |
| | | | | | (.11) | (.01) |
| Manitoba | 0.74 | 3.26 | 1.31 | 8.94 | .53 | .09 |
| | | | | | (.11) | (.01) |
| Saskatchewan | 0.90 | 3.34 | 1.18 | 8.68 | .32 | .10 |
| | | | | | (.14) | (.01) |
| Alberta | 0.78 | 3.33 | 1.36 | 9.23 | .58 | .08 |
| | | | | | (.12) | (.01) |
| British Columbia | 0.72 | 3.23 | 1.41 | 9.39 | .77 | .07 |
| | | | | | (.12) | (.01) |
| Yukon | 0.80 | 3.46 | 1.48 | 8.73 | .75 | .08 |
| | | | | | (.16) | (.02) |

*Note:* Regression equation, columns (5) to (8):

$$\ln Y_i = (\ln Y_{0,1}) + (\hat{r}_1)\, S_i + U_{1,i}.$$

Regression equation, columns (9) to (14):

$$\ln Y_i = (\ln Y_{0,3}) + (\hat{r}_{₵})\, ₵_i + (\hat{r}_{\$})\, \$_i + (\hat{r}_H)\, H_i + U_{3,i}.$$

*Source: Census of Canada: 1961*, Ottawa, Statistics Canada, Table A.11, unpublished.

$S$ = years of schooling
$₵$ = years of primary schooling
$\$$ = years of secondary schooling
$H$ = years of higher education

**TABLE A-3 (concluded)**

| $\text{Var}(U)_1$ (7) | $\bar{R}_1^2$ (8) | $\ln Y_{0,3}$ (9) | $\hat{r}_{\not\ell}$ (10) | $\hat{r}_{\not\$}$ (11) | $\hat{r}_H$ (12) | $\text{Var}(U)_3$ (13) | $\bar{R}_3^2$ (14) |
|---|---|---|---|---|---|---|---|
| 0.66 | .13 | .30 | .07 | .16 | .08 | 0.67 | .13 |
| | | (.20) | (.04) | (.07) | (.08) | | |
| 0.58 | .15 | .13 | .09 | .16 | .05 | 0.59 | .15 |
| | | (.41) | (.07) | (.07) | (.07) | | |
| 0.55 | .15 | .20 | .10 | .12 | .07 | 0.55 | .14 |
| | | (.32) | (.05) | (.06) | (.05) | | |
| 0.53 | .16 | .13 | .12 | .10 | .06 | 0.53 | .16 |
| | | (.23) | (.04) | (.06) | (.05) | | |
| 0.51 | .14 | .51 | .10 | .07 | .08 | 0.52 | .13 |
| | | (.21) | (.03) | (.04) | (.04) | | |
| 0.45 | .13 | .57 | .10 | .06 | .08 | 0.46 | .13 |
| | | (.32) | (.05) | (.04) | (.03) | | |
| 0.47 | .15 | .25 | .13 | .05 | .09 | 0.47 | .14 |
| | | (.26) | (.04) | (.05) | (.04) | | |
| 0.71 | .13 | .02 | .15 | .05 | .10 | 0.71 | .13 |
| | | (.31) | (.05) | (.06) | (.05) | | |
| 0.53 | .13 | .33 | .13 | .05 | .10 | 0.53 | .12 |
| | | (.31) | (.05) | (.05) | (.04) | | |
| 0.48 | .09 | .57 | .10 | .03 | .09 | 0.48 | .08 |
| | | (.32) | (.05) | (.04) | (.04) | | |
| 0.56 | .13 | .03 | .22 | -.06 | .06 | 0.52 | .18 |
| | | (.26) | (.04) | (.06) | (.06) | | |

TABLE A-4

Earnings Inequality, Estimated Overtaking Age Rate of Return, and Three Estimates of Schooling's Explanatory Power for the Fifty-one States

| State | SD(ln$E$) | $r_m$ | $\frac{r_m^2 \text{Var}(S)}{\text{Var}(\ln Y)}$ | $\frac{r_m^2 \text{Var}(S)}{\text{Var}(\ln E)}$ | $\frac{\hat{r}^2 \text{Var}(S)}{\text{Var}(\ln E)}$ |
|---|---|---|---|---|---|
| Alabama | .96 | .19 | .640 | .691 | .324 |
| Alaska | .98 | .17 | .464 | .415 | .178 |
| Arizona | .92 | .14 | .390 | .381 | .200 |
| Arkansas | .96 | .23 | .839 | .898 | .300 |
| California | .85 | .13 | .344 | .337 | .159 |
| Colorado | .84 | .15 | .427 | .406 | .154 |
| Connecticut | .79 | .11 | .275 | .278 | .189 |
| Delaware | .87 | .14 | .370 | .387 | .301 |
| D.C. | .89 | .12 | .331 | .345 | .188 |
| Florida | .86 | .17 | .585 | .658 | .217 |
| Georgia | .94 | .18 | .619 | .656 | .308 |
| Hawaii | .80 | .12 | .492 | .484 | .179 |
| Idaho | .88 | .15 | .351 | .311 | .140 |
| Illinois | .82 | .13 | .280 | .303 | .157 |
| Indiana | .83 | .15 | .347 | .352 | .165 |
| Iowa | .91 | .16 | .331 | .331 | .151 |
| Kansas | .89 | .16 | .360 | .357 | .173 |
| Kentucky | .96 | .19 | .531 | .567 | .239 |
| Louisiana | .93 | .17 | .610 | .679 | .302 |
| Maine | .81 | .15 | .377 | .397 | .167 |
| Maryland | .85 | .12 | .298 | .306 | .224 |
| Massachusetts | .79 | .13 | .364 | .374 | .182 |
| Michigan | .85 | .12 | .251 | .249 | .138 |
| Minnesota | .91 | .17 | .425 | .426 | .205 |
| Mississippi | 1.01 | .26 | 1.160 | 1.120 | .330 |

*Note:*

*Earnings Inequality:* Standard deviation of the natural log of earnings in 1959 for males, age fourteen and over, in the civilian labor force; *U.S. Census of Population: 1960, Vol. 1, Characteristics of the Population*, Parts 2-52, Washington, D.C., Table 124.

*Overtaking Age Rate of Return:* Estimates of rates of return to four years of high school for the earnings of males based on Mincer's overtaking age technique. The computational procedure is discussed in Appendix A-2.

*Var(ln$Y$), $\hat{r}$ and Var($S$):* Data from Table A-1.

**TABLE A-4 (concluded)**

| State | SD(ln$E$) | $r_m$ | $\frac{r_m^2 \text{Var}(S)}{\text{Var}(\ln Y)}$ | $\frac{r_m^2 \text{Var}(S)}{\text{Var}(\ln E)}$ | $\frac{\hat{r}^2 \text{Var}(S)}{\text{Var}(\ln E)}$ |
|---|---|---|---|---|---|
| Missouri | .90 | .17 | .391 | .438 | .197 |
| Montana | .87 | .14 | .332 | .300 | .121 |
| Nebraska | .89 | .17 | .406 | .386 | .162 |
| Nevada | .83 | .12 | .256 | .251 | .134 |
| New Hampshire | .79 | .12 | .279 | .290 | .153 |
| New Jersey | .78 | .11 | .285 | .303 | .196 |
| New Mexico | .91 | .15 | .493 | .499 | .272 |
| New York | .80 | .12 | .321 | .356 | .193 |
| North Carolina | .94 | .16 | .482 | .509 | .246 |
| North Dakota | .91 | .16 | .341 | .330 | .166 |
| Ohio | .82 | .13 | .306 | .324 | .186 |
| Oklahoma | .93 | .18 | .549 | .587 | .255 |
| Oregon | .84 | .14 | .292 | .302 | .132 |
| Pennsylvania | .80 | .12 | .300 | .316 | .167 |
| Rhode Island | .76 | .12 | .332 | .363 | .191 |
| South Carolina | .95 | .17 | .580 | .591 | .312 |
| South Dakota | .93 | .17 | .330 | .331 | .145 |
| Tennessee | .97 | .20 | .669 | .711 | .298 |
| Texas | .94 | .17 | .558 | .577 | .253 |
| Utah | .86 | .13 | .303 | .251 | .122 |
| Vermont | .87 | .15 | .337 | .326 | .147 |
| Virginia | .94 | .14 | .436 | .435 | .261 |
| Washington | .85 | .14 | .318 | .314 | .165 |
| West Virginia | .92 | .17 | .455 | .489 | .207 |
| Wisconsin | .85 | .13 | .287 | .284 | .163 |
| Wyoming | .86 | .14 | .333 | .305 | .140 |

TABLE A-5

Means, Standard Deviations, and Coefficients of Variation of the Parameters for the Fifty-one States and Eleven Provinces

| Parameter[a] | States | | | Provinces | | |
|---|---|---|---|---|---|---|
| | Mean (1) | SD (2) | CV (3) | Mean (4) | SD (5) | CV (6) |
| SD($S$) | 3.77 | 0.41 | 0.12 | 3.27 | 0.17 | 0.05 |
| Var($S$) | 14.38 | 3.17 | 0.22 | 10.69 | 1.08 | 0.10 |
| $\dot{r}$ | 0.10 | 0.01 | 0.14 | 0.09 | 0.01 | 0.14 |
| $\dot{r}^2$ | 0.01 | 0.00 | 0.28 | 0.01 | 0.00 | 0.27 |
| SD(ln$Y$) | 0.88 | 0.07 | 0.08 | 0.79 | 0.06 | 0.07 |
| Var(ln$Y$) | 0.79 | 0.12 | 0.15 | 0.63 | 0.09 | 0.15 |
| $\dot{r}^2$ Var($S$) | 0.16 | 0.07 | 0.42 | 0.09 | 0.02 | 0.21 |
| Var($U$) | 0.64 | 0.07 | 0.11 | 0.55 | 0.08 | 0.14 |
| $\frac{\ln Y_0}{R^2}$ | 0.22 | 0.29 | 1.30 | 0.45 | 0.26 | 0.57 |
| | 0.18 | 0.05 | 0.29 | 0.13 | 0.02 | 0.14 |
| AV($S$) | 10.28 | 0.79 | 0.08 | 8.45 | 0.78 | 0.09 |
| AV(ln$Y$) | 1.27 | 0.23 | 0.18 | 1.21 | 0.21 | 0.17 |
| SD(ln$E$) | 0.88 | 0.06 | 0.07 | | | |
| $r_m$ | 0.15 | 0.03 | 0.20 | | | |

*Source:* Same as Tables A-1, A-3, and A-4.
[a]For definition of variables, see p. 179.

## APPENDIX A-2: DATA FOR PART C: UNITED STATES AND CANADA

### The Variables

| Variable | Symbol | Source U.S.[a] | Source Canada[b] |
|---|---|---|---|
| 1. Natural log of income–mean and variance | Av(ln$Y$)<br>Var(ln$Y$) | Table 138 | [b] |
| 2. Natural log of earnings–mean and variance | Av(ln$E$)<br>Var(ln$E$) | Table 124 | |
| 3. Rate of return from schooling–regression estimate | $r_c$ | Table 138 | [b] |
| 4. Rate of return from schooling–overtaking age estimate | $r_m$ | Tables 103 and 138, and the 1/1,000 sample, 1960 Census of Population | |

| Variable | Symbol | Source U.S.[a] | Source Canada[b] |
|---|---|---|---|
| 5. Years of schooling-mean, standard deviation and variance | Av($S$)<br>SD($S$)<br>Var($S$) | Table 103 | b |
| 6. Years of age-mean, standard deviation and variance | Av($A$)<br>SD($A$)<br>Var($A$) | Table 103 | b |
| 7. Covariance of years of schooling and age | Cov($A$,$S$) | Table 103 | b |
| 8. Average years of experience (mean age minus mean schooling minus five) | Av($Exp$) | Table 103 | |
| 9. Natural log of weeks worked-mean, standard deviation and variance | Av(ln$WW$)<br>SD(ln$WW$)<br>Var(ln$WW$) | Table 118 | |
| 10. Per cent nonwhite (for the male labor force) | $p$ | Tables 97 and 138 | |
| 11. Dummy variable, $Z$ = 1, in ten states without separate race data | $Z$ | Table 138 | |
| 12. Dummy variable, $NSD$ = 1, in the sixteen Southern states (Census definition) and the District of Columbia | $NSD$ | | |

[a]References are for tables in *United States Census of Population: 1960, Volume 1, Characteristics of the Population*, Parts 2-52, Washington, D.C., Bureau of the Census. The data are from a 25 per cent sample of the population.

[b]*Census of Canada: 1961* (Ottawa: Statistics Canada), Table A.11 for the provinces, unpublished, obtained from Statistics Canada. The data are from a 20 per cent sample of private nonfarm households.

## Intervals

| Variable | United States | Canada |
|---|---|---|
| 1. Income | 9 | 13 |
| 2. Earnings | 9 | — |
| 3. Schooling | 15 | 6 |
| 4. Age | 8 | 4 |
| 5. Weeks worked | 6 | — |

United States

The fifty states and the District of Columbia are used as the units of observation. Variables for white males are defined as the value of the variable for (a) whites in the forty-one states (including the District of Columbia) for which separate white-nonwhite data are available and (b) all males in the ten remaining states.[1] Nonwhites do not constitute more than 7.6 per cent of males between the ages of twenty-five and sixty-four in any of these ten states. The analysis for white males in the subsample of thirty-nine states covers the coterminous states (i.e., excluding Alaska and Hawaii) with separate white-nonwhite data. The analysis of nonwhites is restricted to the thirty-nine coterminous states with separate race data.

The income data used to compute the mean and variance of the natural logarithm of income for all males, white males, and nonwhite males separately for each state cover the 1959 income (wage, salary, self-employment, and property income) of adult males of twenty-five and over who had an income in 1959, as reported in the 1960 Census of Population. Data on labor market income of adult males by state were not reported in the 1960 Census of Population. The mean and variance of the log of earnings is computed for 1959 earnings of males, age fourteen and over, who were in the labor force in 1960. The mean and variance of the natural log of income and earnings for each race-state were computed by using interval midpoints and the Pareto estimate for the upper open-end interval ($10,000 and over). The low value for the open-end interval reduces the impact on the income data of large nonlabor incomes.

The means, standard deviations, and variances of years of schooling and years of age, as well as the covariance of age and schooling, were computed for males between the ages of twenty-five and sixty-four. The mean and variance of the natural log of weeks worked in 1959 were computed for males in the same age group who worked in 1959. The age, schooling, and weeks worked variables were computed separately for all males, white males, and nonwhite males.

The variable "per cent nonwhite" is the percentage of males with income, age twenty-five and over, in 1959 who were nonwhite in the forty-one states with separate white-nonwhite data, and the percentage of all males between twenty-five and sixty-four who were nonwhite in the remaining ten states.

1. These ten states are Idaho, Maine, Montana, Nevada, New Hampshire, North Dakota, Rhode Island, Utah, Vermont, and Wyoming.

Two measures of the rate of return from schooling are employed in the analysis of income inequality: the regression estimate ($r_c$) and the overtaking age estimate ($r_m$).[2] A cross-classification of income (in 1959) by schooling is available for males twenty-five and over with income.[3] The regression estimate of the rate of return is the slope coefficient from a linear regression of the log of income on years of schooling, using grouped microdata within a state. It was computed for each state for all males, and for white and nonwhite males where separate race data were available.

The regression estimate is deficient as a measure of the rate of return because the income data include nonlabor market income, the population includes aged males, and labor market experience is contained in the residual. The last is the most serious problem of the three. Due to the secular increase in schooling, those with low levels of schooling tend to be older and are receiving their return on earlier investments in postschool training. A regression of the log of earnings on years of schooling in which all age groups are pooled therefore results in a downward-biased estimate of the slope coefficient, and hence of the regression estimate of the rate of return. The downward bias would not be eliminated by restricting the regressions to specific age groups. For a given age, a higher level of schooling implies fewer years of experience.

Thus, the omission of experience as an explicit explanatory variable in the regression equation biases the slope coefficient of schooling downward. The absence of a cross-classification for the states of income by schooling and age prevented the computation of an unbiased estimate. The regression estimate, however, reflects the average rate of return for all levels of schooling.

Mincer developed an alternative shortcut technique for calculating unbiased rates of return from a given level of schooling which he calls the "overtaking age rate of return." The overtaking age is the age at which the observed age-earnings profile cuts the horizontal earnings profile that would exist if there were no investment in postschool training.

The estimating procedure requires cross-classified data on in-

2. The regression estimate of the rate of return was first used in G. S. Becker and B. R. Chiswick, "Education and the Distribution of Earnings," *American Economic Review*, May 1966, pp. 358-369. Its properties are developed in greater detail in Chapter 3 of this book.

The overtaking age rate of return and the estimating procedure used here were both developed by Jacob Mincer in *Schooling, Experience, and Earnings*, NBER, 1974, Part 1. The calculations were performed by Sarah Paroush.

3. A cross-classification of earnings by schooling for the states does not exist.

come, schooling, and labor market experience (or age). There are no published classifications of income by schooling and age for the states, but a detailed age and income distribution for each schooling group does exist for males in each state. Using the overtaking age technique on these data, estimates of rates of return ($r_m$) by state were calculated for high school males.

The data permitted the calculation of the average age and the average log of income for elementary school and high school graduates in each state. It was assumed that the mean age-income point so obtained was on the age-log of income profile for that schooling level in the particular state. Using the one-in-a-thousand sample for males in the country as a whole, an age-log of income profile was generated for each level of schooling. The overtaking age for high school and for elementary school graduates was found in the aggregate data. It was assumed that each state's age-log income profile had the same shape, but not necessarily the same intercept or height, as in the aggregate data. This assumption, and the point for average age and average log of income, permitted the estimation of the log of income at the national overtaking age (or level of experience) for the two schooling levels in each state. The estimate of the overtaking age rate of return ($r_m$) was then calculated for each state from the overtaking age log of income for high school ($\ln Y_{HS}$), and for elementary school ($\ln Y_E$), and from the relation $\ln Y_{HS} = \ln Y_E + r_m S$, where $S$ is four years.

This shortcut for estimating the overtaking age rate of return has two disadvantages. First, given the indirect estimating procedure, there are probably significant errors of measurement. Second, for our purpose an estimate of the average rate of return for all schooling levels is required, whereas the overtaking age rate of return used here is for high school education alone. The overtaking age estimate was computed only for all males in the fifty-one states.

The regression and overtaking estimates are highly and significantly correlated. For males in the fifty-one states the correlation is .73. The estimates of the overtaking age rate of return, however, are consistently larger; the average $r_m$ for the states is equal to .151, while the average $r_c$ is equal to .102.

## Canada

The data for Canada come from unpublished tables of the 1961 Census for a 20 per cent sample of private nonfarm households. The data, for each of the ten provinces and the Yukon Territory, represent a cross-classification of income by schooling

and by age. The parameters of the distributions of income, schooling, and age and the regression estimate of the rate of return from schooling were computed in the same manner as for the United States. A distribution of weeks worked for nonfarm adult males exists for the provinces, but the intervals are very wide, especially at the upper end of the distribution.[4] There are four intervals: 1 to 13, 14 to 26, 27 to 39, and 40 to 52 weeks. All but 25 per cent of the observations were in the upper interval. The result was little interprovincial variation in the computed mean and variance in the log of weeks worked. By contrast, the intervals for the United States are: 1 to 13, 14 to 26, 27 to 39, 40 to 48, 48 to 49, and 50 to 52.

## APPENDIX A-3: MEXICAN WAGE EARNERS SAMPLE

The Mexican data are from a sample of 3,901 male wage earners in the cities of Monterey, Puebla, and the Federal District (Mexico) taken in the summer of 1963 by Martin Carnoy.[5] Carnoy used a stratified sample in which the number of people interviewed within each urban occupational sector was based on the proportion of wage earners in that sector given by the Mexican Census of Population of 1960. However, within an "urban occupational sector" the sample was nonrandom. The sampler depended on the cooperation of management and a relatively small number of contacts that put him in touch with firms. The workers sampled probably contain a disproportionate number from American-affiliated and local "modern" firms. Stratification reduced, but did not eliminate, the nonrandom character of the sample.

The income variable is average monthly (i.e., annual divided by twelve) earnings. The workers were specifically asked to exclude property income but to include earnings from sources other than their primary employment. The education data consist in the number of years of schooling completed. About two-thirds of all those interviewed with sixteen years and more of schooling are in five firms, and these firms do not all fall into any particular sector. The concentration of those with higher education in a few firms may reflect the nonrandomness of the sample.

4. *Census of Canada: 1961*, Vol. 3, Part 3, Table 10.

5. M. Carnoy, "The Cost and Return to Schooling in Mexico: A Case Study," Ph.D. dissertation, University of Chicago, 1964, pp. 30 and 120-133. Children, the youngest being a ten-year-old, formed a small proportion of the sample. They were interviewed on the job (selling newspapers, running errands, et cetera) in different parts of the cities.