Volume Title: Fatal Years: Child Mortality in Late Nineteenth-Century America

Volume Author/Editor: Samuel H. Preston and Michael R. Haines

Volume Publisher: Princeton University Press

Volume ISBN: 0-691-04268-3

Volume URL: http://www.nber.org/books/pres91-1

Conference Date: n/a

Publication Date: January 1991

Chapter Title: Appendix B, The State Earnings Index

Chapter Author: Samuel H. Preston, Michael R. Haines

Chapter URL: http://www.nber.org/chapters/c11548

Chapter pages in book: (p. 221 - 225)

APPENDIX B

THE STATE EARNINGS INDEX

TO ACCOUNT for spatial income variation, an index of state earnings was constructed. One obvious choice would be the 1900 state personal income estimates of Easterlin (1957). These estimates, however, are heavily influenced by the occupational structure of a state. Since we have information on an individual's occupation, we seek a measure of occupation-specific wage levels, aggregated across occupations using a set of weights that is identical for all states. For this purpose, we utilize the earnings data presented by Lebergott (1964: Tables A-23 to A-29) by state for 1899–1900 for six selected occupations or industries (farm laborers, common laborers, domestic service, cotton manufacture, woolen manufacture, and iron and steel manufacture). To obtain comparable estimates from that source, monthly wages of farm laborers were multiplied by 12; the weekly wages of domestic workers were multiplied by 52; and the daily wages of common laborers were multiplied by 5.6 days per week and 44 weeks per year. The earnings of the textile and metal workers were already on an annual basis. Thus, farm laborers and domestic service workers were assumed to have had a full year's employment, while common laborers were not. Not all occupations or industries were represented for all states. There were 48 states (data for Alaska and Hawaii were not available) and the District of Columbia. A total of 186 cells resulted. The following dummy variable regression equation was estimated:

$$Ln(Y_{ij}) = B_0 + \Sigma_{i=1}^{6} B_i I_i + \Sigma_{j=1}^{49} B_j S_j ,$$

where $Ln(Y_{ij})$ = natural log of earnings for occupation/industry i and state j;

B = coefficient;

I_i = dummy variable for occupation or industry i ($i = 1 \ldots 6$);

S_j = dummy variable for state j ($j = 1 \ldots 49$).

The regression results are presented in Table B.1. Three models are given. Model I includes only the occupational dummy variables. The omitted category is iron and steel manufacture. It had the highest average earnings, so that all the coefficients are negative deviations from iron and steel manufacturing incomes. The constant term is thus, in this case, the natural logarithm of the state average of iron and steel earnings per worker. Model II includes only dummy vari-

ables for states. The reference category is California, which had the highest average income in this sample. The constant term is therefore the logarithm of the average earnings in California, and the dummy variables measure deviations from it. Finally, Model III includes both occupational and state dummy variables.

A perusal of Table B.1 reveals that occupation is much more successful in explaining variation than is location (i.e., state of residence). The R^2 for Model I is .594 and R^2 adjusted for degrees of freedom is .583, whereas the R^2 for Model II, with only the state dummy variables, is only .292 and the adjusted R^2 only .045. The Model II equation is not even jointly significant at a 5 percent level, as measured by the F-ratio (which is only 1.180). Nonetheless, Model III, which has both occupational and state dummy variables, is *much* more successful than either of the other two in explaining variation in earnings per worker. It has an R^2 of .916 and an adjusted R^2 of .882, and it is jointly significant at a 1 percent level (as measured by the F-ratio). The conclusion that can be drawn is that, although occupation is a better predictor of earnings per worker than state of residence, both sets of variables belong in the model.

The estimated coefficients in Model III permit the estimation of predicted average state workers' annual earnings, holding constant occupational and industry composition at their averages over all states. The estimates are given in Table B.2. In that table the values of $Ln(Y)$ and Y are given and are converted to an index, dividing each value by the $Y(= 244)$ for the whole sample. This is the index used in the regressions in Chapter 4. One advantage of using the Lebergott (1964) data is that the types of earnings estimates produced are rather close in nature to those derived from the 1901 cost-of-living survey (U.S. Commissioner of Labor 1903), Douglas's study of wages (1930), and other estimates from Lebergott that were used to impute income for individual occupations. The estimated values of the state earnings index were assigned to each worker on the basis of state of residence for the regression analysis.

For purposes of comparison, Table B.2 also includes the Easterlin estimates of state-level income per capita for 1899–1900. The correlation between them is quite reasonable, with a zero-order correlation of .810 and a Spearman rank order correlation of .819.

TABLE B.1

Equations Predicting Earnings in Various Occupations and States of Residence: U.S., 1899–1900

	(1)		(2)		(3)	
	Coefficient	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*
Independent variables:						
Constant	6.1121	NC	5.9545	NC	6.5496	NC
Occupation						
Farm laborer	−0.8376	***			−0.9532	***
Common laborer	−0.3117	***			−0.2759	***
Domestic servant	−1.1511	***			−1.1516	***
Cotton manufacture	−0.5626	***			−0.4932	***
Woolen manufacture	−0.4630	***			−0.4468	***
Iron & steel mfg.	NI	NI			NI	NI
State						
Maine			−0.3562	—	−0.2872	***
New Hampshire			−0.3035	—	−0.2346	**
Vermont			−0.4135	*	−0.3445	***
Massachusetts			−0.1713	—	−0.2130	**
Rhode Island			−0.2799	—	−0.2110	*
Connecticut			−0.2226	—	−0.2644	**
New York			−0.2539	—	−0.2956	***
New Jersey			−0.2661	—	−0.3078	***
Pennsylvania			−0.2475	—	−0.2892	***
Ohio			−0.3140	—	−0.3557	***
Indiana			−0.3672	—	−0.4090	***
Illinois			−0.2238	—	−0.2535	**
Michigan			−0.3361	—	−0.3658	***
Wisconsin			−0.2960	—	−0.3257	***
Minnesota			−0.4082	—	−0.2965	***
Iowa			−0.5130	*	−0.4013	***
Missouri			−0.3376	—	−0.3672	***
North Dakota			−0.3861	—	−0.0281	—
South Dakota			−0.4532	—	−0.0952	—
Nebraska			−0.4448	*	−0.2464	**
Kansas			−0.5679	*	−0.3695	***
Delaware			−0.4351	**	−0.4556	***
Maryland			−0.5169	*	−0.5586	***
Dist. Columbia			−0.2846	—	−0.6038	***
Virginia			−0.6286	**	−0.6703	***
West Virginia			−0.3871	—	−0.4168	***
North Carolina			−0.9608	***	−0.8918	***
South Carolina			−1.0373	***	−0.9140	***
Georgia			−0.9341	***	−0.9758	***
Florida			−0.9309	***	−0.7326	***
Kentucky			−0.4875	**	−0.5292	***
Tennessee			−0.7152	***	−0.7569	***
Alabama			−0.6714	***	−0.8327	***
Mississippi			−0.7969	***	−0.8498	***
Arkansas			−0.9369	***	−0.7385	***
Louisiana			−0.9261	***	−0.7277	***
Oklahoma			−0.7954	**	−0.4374	**

TABLE B.1 (*cont.*)

	(1)		(2)		(3)	
	Coefficient	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*
Texas			−0.5158	**	−0.5455	***
Montana			−0.0012	—	0.3568	**
Idaho			−0.1314	—	0.2266	—
Wyoming			−0.0795	—	0.2785	—
Colorado			−0.3233	—	0.0347	—
New Mexico			−0.5563	—	−0.1983	—
Arizona			−0.1285	—	0.2295	—
Utah			−0.2211	—	0.1369	—
Nevada			−0.0117	—	0.3463	*
Washington			−0.2474	—	0.1106	—
Oregon			−0.3377	—	−0.0203	—
California			NI	NI	NI	NI
N	186		186		186	
R-square	0.594		0.292		0.916	
Adjusted R-square	0.583		0.045		0.882	
F-ratio	52.690	***	1.180	—	27.233	***

Source: Data from Lebergott 1964.

Note: The dependent variable is the log of annual earnings in a particular occupation/state-of-residence category. A one-tailed significance test was used in models 1 and 2. A two-tailed significance test was used in model 3. NC = not calculated; NI = not included; *** = significant at least at a 1 percent level; ** = significant at least at a 5 percent level; * = significant at least at a 10 percent level; — = not significant at least at a 10 percent level.

TABLE B.2
Estimates of Annual Earnings Levels by State: U.S., 1899–1900

	Estimates derived from Table B.1		*Easterlin's estimates*	
State	*Annual earnings ($)*	*Ratio to national average*	*Annual income ($)*	*Ratio to national average*
Alabama	163	0.6680	88	0.4356
Arizona	471	1.9303	321	1.5891
Arkansas	179	0.7336	89	0.4406
California	374	1.5328	365	1.8069
Colorado	388	1.5902	318	1.5743
Connecticut	287	1.1762	278	1.3762
Delaware	237	0.9713	220	1.0891
Dist. Columbia	205	0.8402	—	—
Florida	180	0.7377	112	0.5545
Georgia	141	0.5779	86	0.4257
Idaho	470	1.9262	221	1.0941
Illinois	291	1.1926	260	1.2871
Indiana	249	1.0205	182	0.9010

TABLE B.2 (*cont.*)

State	*Estimates derived from Table B.1*		*Easterlin's estimates*	
	Annual earnings ($)	*Ratio to national average*	*Annual income ($)*	*Ratio to national average*
Iowa	251	1.0287	202	1.0000
Kansas	259	1.0615	187	0.9257
Kentucky	221	0.9057	120	0.5941
Louisiana	181	0.7418	128	0.6337
Maine	281	1.1516	187	0.9257
Maryland	214	0.8770	204	1.0099
Massachusetts	303	1.2418	304	1.5050
Michigan	260	1.0656	185	0.9158
Minnesota	278	1.1393	207	1.0248
Mississippi	160	0.6557	84	0.4158
Missouri	259	1.0615	188	0.9307
Montana	535	2.1926	415	2.0545
Nebraska	293	1.2008	212	1.0495
Nevada	529	2.1680	395	1.9554
New Hampshire	296	1.2131	214	1.0594
New Jersey	275	1.1270	277	1.3713
New Mexico	307	1.2582	148	0.7327
New York	279	1.1434	323	1.5990
North Carolina	153	0.6270	72	0.3564
North Dakota	364	1.4918	209	1.0347
Ohio	262	1.0738	222	1.0990
Oklahoma	242	0.9918	114	0.5644
Oregon	382	1.5656	248	1.2277
Pennsylvania	280	1.1475	250	1.2376
Rhode Island	303	1.2418	293	1.4505
South Carolina	150	0.6148	74	0.3663
South Dakota	340	1.3934	183	0.9059
Tennessee	176	0.7213	101	0.5000
Texas	217	0.8893	138	0.6832
Utah	429	1.7582	183	0.9059
Vermont	265	1.0861	190	0.9406
Virginia	192	0.7869	110	0.5446
Washington	418	1.7131	296	1.4653
West Virginia	247	1.0123	117	0.5792
Wisconsin	270	1.1066	179	0.8861
Wyoming	495	2.0287	311	1.5396
Total	244	—	202	—

Source: Table B.1 and Easterlin 1957.